5-2022

Delivering on a Promise: A Longitudinal Cohort Study of Emergent Bilinguals' Academic Achievement in a Utah Dual Language Program

J. Eric Campbell
Utah State University

DELIVERING ON A PROMISE: A LONGITUDINAL COHORT STUDY OF EMERGENT BILINGUALS' ACADEMIC ACHIEVEMENT IN A UTAH DUAL LANGUAGE PROGRAM

by

J. Eric Campbell

A dissertation submitted in partial fulfillment
of the requirement for the degree

of

DOCTOR OF PHILOSOPHY

in

Education

Approved:

Parker Fawson, Ed.D. Major Professor	Kathleen A. J. Mohr, Ed.D. Committee Member
Alyson Lavigne, Ph.D. Committee Member	Stefinee Pinnegar, Ph.D. Committee Member
Sarah Braden, Ph.D. Committee Member	D. Richard Cutler, Ph.D. Interim Vice Provost of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2022

ABSTRACT

Delivering On a Promise: A Longitudinal Cohort Study of Emergent Bilinguals' Academic Achievement in a Utah Dual Language Program

by

J. Eric Campbell, Doctor of Philosophy

Utah State University, 2021

Major Professor: Parker Fawson, Ed.D.
Department: School of Teacher Education and Leadership

Emergent bilingual (EB) students are a growing population in the U.S. school system currently comprising over 10% of the total population. U.S. school districts have educated EB students using a myriad of practices, including Dual language immersion (DLI). Many studies have looked at the academic achievement of native-Spanish speaking EB students, yet there is a dearth of research on DLI in medium-sized school districts in rural settings. This study focused on native-Spanish speaking EB students in a DLI program in a rural Utah district. Specifically, the study compared student academic achievement in English acquisition, English language arts, mathematics, and grade point average of EB students enrolled in DLI to EB students not enrolled in a DLI program.

Student achievement data for EB students was collected from 2014-2020 ($N =$ 1,046). Using various regression methods such as ordinal logistic regression, multiple regression, and multilevel modeling (MLM), the study sought the predictive power of DLI after controlling for gender, free and reduced lunch status, and special education enrollment. Results found that on average DLI students performed as well as or better than their non-DLI peers. MLM analyses indicated that EB students enrolled in DLI had superior growth trajectories to their non-DLI peers over time.

(158 pages)

PUBLIC ABSTRACT

Delivering On a Promise: A Longitudinal Cohort Study of Emergent Bilinguals' Academic Achievement in a Utah Dual Language Program

J. Eric Campbell

Emergent bilingual (EB) students are a growing population in the U.S. school system currently comprising over 10% of the total population. U.S. school districts have educated EB students using a myriad of practices, including dual language immersion (DLI). Many studies have looked at the academic achievement of native-Spanish speaking EB students, yet there is a dearth of research on DLI in medium-sized school districts in rural settings. This study focused on native-Spanish speaking EB students in a DLI program in a rural Utah district. Specifically, the study compared student academic achievement in English acquisition, English language arts, mathematics, and grade point average of EB students enrolled in DLI to EB students not enrolled in a DLI program.

Student achievement data for EB students was collected from 2014-2020 (N = 1,046). Using various regression methods such as ordinal logistic regression, multiple regression, and multilevel modeling (MLM), the study sought the predictive power of DLI after controlling for gender, free and reduced lunch status, and special education enrollment. Results found that on average DLI students performed as well as or better than their non-DLI peers. MLM analyses indicated that EB students enrolled in DLI had superior growth trajectories to their non-DLI peers over time.

ACKNOWLEDGMENTS

I want to thank Drs. Martin Kokol, Christopher Layne, Sterling Hilton, Damon Bahr, Yvonne Freeman, and David Freeman for encouraging me to pursue a dissertation. I am deeply indebted to Paul Sweat and Linda Turner for supporting me and encouraging me to pursue this particular study on dual language immersion and to my professional colleagues across the State of Utah that have tirelessly worked to implement dual language immersion.

I want to express my appreciation for the faculty and staff at Utah State University, especially my committee members, for their continued support. I owe a special thank you to Drs. Joseph (Joe) Matthews, Courtney Stewart, and Parker Fawson, who chaired my committee at different stages along the process and to Drs. Kathleen (Kit) Mohr, Stefinee Pinnegar, and Tyson Barrett who have spent countless hours advising me throughout my dissertation journey. I need to thank my many colleagues and fellow students in the Ph.D. cohort that made the journey so much more enjoyable, especially my good friend, Kevin Thomas, that got me into this in the first place.

Finally, I cannot say enough about the support and patience that my family and friends have provided me from application through to completion. Particularly, I thank my amazing wife, Melissa, four remarkable boys, and God for their continual emotional, physical, and spiritual support.

J. Eric Campbell

CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Overview

One of the greatest challenges currently facing U.S. public schools is the education of students enrolling without English language proficiency sufficient to meet academic demands (García et al., 2008). Emergent bilingual (EB) students come to school with a native language other than English and are tasked with learning a second language while simultaneously tackling the same content as their English-only peers (García & Kleifgen, 2018). EB students now make up over 10% of the American student body and the growth rate is steadily climbing (Bialik et al., 2018). In some states, almost one in five students is not yet English proficient (Hussar et al., 2020). With the growth of this demographic in the U.S., educators, researchers, and policy makers in impacted states, schools, and districts have responded with a wide range of educational offerings. Bilingual and English-only programs—and variations in between—dot the U.S. school system landscape (Kim et al., 2015; Ovando, 2003; Tedick, 2015). Despite myriad programmatic opportunities established to assist EB students, schools and districts have continued to struggle to improve the academic outcome distance between EB students and their peers (Abedi, 2004; Hussar et al., 2020; Solano-Flores & Trumbull, 2003).

Growth of a Diverse Emergent Bilingual Population

EB students comprise the fastest growing student demographic in the U.S.

(Breiseth, 2015; Hussar et al., 2020). These students add rich diversity to classrooms. Native languages of public-school students within the U.S. number over 400 (Bialik et al., 2018) with Spanish accounting for the largest portion. Though the majority of EB students populate urban areas, up to 4.1% comprise rural student populations (Hussar et al., 2020). Regardless of location, almost eight out of 10 districts have the opportunity to educate EB students (Sugarman & Lazarín, 2020). EB students comprise a larger percentage in the younger grades with 15.9% of kindergarteners versus 4.6% of 12th graders (Hussar et al., 2020). Unfortunately, many EB students come from homes with higher poverty levels, with lower parental educational attainment, and have the highest rate of homelessness among school-aged children (National Academies of Sciences Engineering and Medicine, 2017). With the growth of this diverse population, understanding best practices for educating EB students has become a need for almost every educator in the U.S.

Educational Outcomes for Emergent Bilingual Students

Indeed, academic proficiency rates of EB students have not matched their English-only peers in Reading, Mathematics, and Science on assessments administered in English (Hussar et al., 2020; U.S. Department of Education, 2018a). Poor performance on academic assessments appears to hamper educational opportunities because, as a group, EB students have lower grade point averages (GPA), lower graduation rates, and are less likely to continue on to post-secondary education (Montoya-Ávila et al., 2018; Sheng et al., 2011; U.S. Department of Education, 2018a).

Research confirms common sense that command of academic English is a prerequisite for performing well on state and federal academic assessments administered in English (Cummins, 2008). Research suggests that even in effective educational systems, it takes EB students many years to master English proficiency at levels that allow them to have equitable access to available educational opportunities in English-only environments. Depending on the study, research suggests it takes between 3-7 years for EB students to attain English proficiency (Conger, 2008; Hakuta et al., 2000; Slama, 2014). Although not yet fully understood, the difference in times to develop academic English proficiency has been attributed to differences in native language proficiency, English exposure, school programming, instructional practices, and school context (Thompson, 2017; Umansky & Reardon, 2014). Because of the challenges in reaching English proficiency, EB students are not only at-risk of not performing as well on academic measures in English, but they are also often placed in less rigorous curricular classes and have higher school drop-out rates (Genesee et al., 2005; U.S. Department of Education, 2018b). Other studies indicate a lack of acculturation into the majority culture and lower social support as explanations of poor academic outcomes (López et al., 2002). However, a frequently identified cause of poor performance is the fact that a majority of EB students come from homes from lower socioeconomic conditions and lower levels of parental education. In sum, each of these factors has been shown to correlate with lower academic achievement of EB students (Assari et al., 2019; Selcuk, 2005).

U.S. Response to the Educational Needs of Emergent Bilingual Students

In response to the academic needs of EB students, several major pieces of legislation, including the Bilingual Education Act of 1968 and the Equal Education Opportunities Act of 1974, have increased the formal responsibility of schools to provide quality education for EB students. In addition to legislation, the federal courts have strengthened those requirements for school systems to implement and evaluate educational programs in order to provide EB students with more equitable access to grade-level content (*Castañeda v Pickard*, 1981; *Lau v Nichols*, 1974). Most recently, the reauthorizations of the Elementary and Secondary Education Act (ESEA) in 2002 and 2015, more popularly known as the No Child Left Behind Act (NCLB) and the Every Student Succeeds Act (ESSA), pushed for school systems to pay attention to, and report annually on, the achievement of all students, including the gap between EB students and their English-proficient peers.

Legislation and case law place the challenge of providing educational opportunity and evaluating those opportunities squarely on the state, district, and school-level educational leaders. Districts and schools have a great deal of control over how they can meet the requirements of educating EB students by providing equal access to high levels of academic performance. Local education leaders have a significant influence on the success of their school systems and in turn their students (Leithwood et al., 2009; Leithwood & Strauss, 2007). Consequently, the success or existence of bilingual programs depends on a principal's understanding of the essential components of such programs, carefully monitoring of their implementation, and how to evaluate available data regarding program success (Menken & Solorza, 2014, 2015).

State and Local Education Agencies (LEA) have responded differently over the years to the pressures imposed on school systems to adequately educate EB students by favoring either English-dominant or bilingual educational programs. Currently, this means the U.S. public school system has a range of educational programs across the English-dominant-bilingual spectrum. (García & Kleifgen, 2018; Tedick, 2015). In a review of the U.S. Title III programs, which serve EB students, English-dominant educational approaches such as English as a second language (ESL) were found to be the most popular forms of educating EB students (Tanenbaum et al., 2012). Less than half of all districts receiving Title III funds reported offering academic services using students' native languages or bilingual programing. Perhaps bilingual program implementation suffers because of the number of languages represented in the U.S. public school system, which complicates the process of securing well qualified teachers and appropriate resources.

Bilingual Education as a Response to Educating Emergent Bilinguals

Researchers have demonstrated the benefit of bilingual education on native-English students learning a second language (Genesee, 2004), as well as the benefits for minority-language, or EB, students (Genesee et al., 2005; Lindholm-Leary, 2017; Lindholm-Leary & Hernández, 2011). Because the largest portion of EB students in the U.S. are native-Spanish-speaking students, most research looks at Latinx student outcomes in bilingual versus traditional programs (Lindholm-Leary & Hernández, 2011).

These studies have found that Latinx EB students enrolled in bilingual programs perform as well or better than their Latinx EB peers in traditional programs (Lindholm-Leary & Howard, 2008).

In contrast to the English-only instruction, proponents of bilingual education provide evidence-based arguments that strengthening students' native and second languages yields higher academic outcomes for EB students (Cummins, 1981). In addition to the academic benefits (Genesee et al., 2005), EB students can prosper economically (Gunnarsson, 2013), cognitively (Bialystok, 2011), and culturally (Feinauer & Howard, 2014). Researchers have argued that in our current global economy and information age, bilingualism benefits communities and companies, because such economies need individuals who possess the linguistic skills to manage multiple languages (Alarcón & Heyman, 2013).

Although bilingual program components and characteristics vary, they still hold in common their primary goal of high levels of language skills in two languages. However, programs differ in their secondary goals and features. For example, transitional bilingual programs provide native-language instruction for just a few years or until EB students have enough English to assimilate them into a full English experience (Baker & Wright, 2017). In contrast, dual language bilingual programs teach students in the native language and second language for multiple years supporting a linguistically balanced philosophy with bilingualism and biliteracy proficiency as main goals. Dual language immersion (DLI) programs in the U.S. include *one-way*, where the majority of the students enrolled speak English and receive instruction in the target or partner language and *two-way* where student enrollment consists of approximately 50% native-English speakers and 50% the target language speakers (Lindholm-Leary, 2016). Instruction in the target language further differentiates programs from 90:10 to 50:50 describing the amount of time students are exposed to the target, or second language (Gomez et al., 2005; Lindholm-Leary, 2012).

Utah's Response to Educating Emergent Bilinguals

In the 2016-17 school year, more than 35 states offered dual language programs, with 30 states offering Spanish as the partner language (U.S. Department of Education, 2019). In 2017, Utah's Title III report to the federal government claimed that their dual language programs included five different languages, of which Spanish dominated their two-way models (U.S. Department of Education, 2017). A few years earlier, Utah had embarked on an arguably ambitious endeavor to open multiple dual language programs across the state with centralized support through the Utah School Board of Education staff and funding through the state legislature (Leite & Cook, 2015). The Utah initiative, the first of this magnitude with state support in the U.S., was largely focused on building economic capacity through developing bilingual capabilities among students (Leite & Cook, 2015; Tedick, 2015). The Utah model requires two teachers, a 50:50 English-to-partner-language ratio to receive state support (Watzinger-Tharp et al., 2016) rather than allowing for 90:10 or 80:20 models. Citing Utah's deference to the 50:50 model, some have argued that Utah's dual language initiative has sought to mainly benefit those of "racial privilege, wealth, and English privilege" (Valdez, Freire, et al., 2016, p. 601) rather than language-minority students for whom bilingual programs have traditionally focused. Despite concerns in the research community about the efficacy of the program model for supporting EB students, some Utah districts implemented DLI with a clear focus on improving EBs' academic opportunities (Eaton, 2016).

Notwithstanding the concerns regarding Utah's dual-language initiative's impetus and focus, it still stands to offer Latinx EB students greater academic opportunities than most states due to its rising native-Spanish speaking student population (Utah State Board of Education, 2019a), its legislative support of DLI (Utah State Board of Education, 2018), and its increasing bilingual populace (Leite & Cook, 2015). Yet, the Utah DLI model remains minimally studied. To date, only a few studies have empirically reviewed student outcomes for Utah's program (e.g., Leite & Watzinger-Tharp, 2016; Steele et al., 2019; Watzinger-Tharp et al., 2016). None of the Utah studies have looked at the outcomes of an entire school district nor how EB students have performed within a school district. Rather the studies have looked at students across multiple programs and contexts, possibly obscuring how individual programs are succeeding in their efforts to achieve the state-mandated goals.

Statement of the Problem

Similar to the whole of the U.S., Utah schools face the challenge of meeting the linguistic and academic needs of EB students. EB students in Utah, predominantly Latinx, have scored 30 points below their peers on recent English language arts and mathematics proficiency state-level assessments (Sugarman & Geary, 2018). According to research, two-way DLI Spanish programs could make a significant difference for EB students in Utah. Although some argue that the model proposed by the state of Utah did not have EB progress as the main impetus (Valdez, Freire, et al., 2016), Utah administrators and districts have looked to the model to build a more bilingual native-English populace and a more academically proficient EB population. Evidence of academic benefits for EB students acts as a counterbalance to the criticisms leveled at the Utah DLI model in benefitting mostly native-English speakers.

Significance of the Problem

Although studies of academic performance of EB students enrolled in the Utah DLI model have initially shown positive results for EB students (Steele et al., 2019; Watzinger-Tharp et al., 2016), none has looked at the impact of a single rural district on EB student academic outcomes. Closing the opportunity gap for EB students is a real concern for districts and schools across the country and in those areas where EB student populations are increasing. In Utah, a state-supported DLI program could contribute to closing the opportunity gap for EB students, yet there remain unknowns regarding its effectiveness due to program focus and the relative nascency of the program. By investing in the capacity of schools and districts to internally evaluate their DLI programs through sophisticated methods the individual LEAs and the whole state of Utah could simultaneously profit from greater understanding of the benefits of the Utah DLI model.

Purpose of the Study

This study looked at the impact of DLI on EB students in a rural community in

Utah. Also, the author sought to complement findings of earlier research on the benefits of DLI programming for EB students (Lindholm-Leary, 2017; Lindholm-Leary & Block, 2010) by determining if a DLI program had an impact on the academic outcomes of EB students within a single rural district. Given the lack of empirical research, this study was designed to provide additional results on the impact of the Utah DLI model on language and academic development of EB students enrolled in DLI programs. Regardless of the original intent of Utah's DLI model, EB students stand to benefit. In addition, this study sought to provide an additional model for local districts and administrators tasked with evaluating EB programming aimed at continuous improvement and accountability for student outcomes across multiple measures. To fulfill these purposes, this study was guided by the following questions.

Research Questions

1. How does enrollment in a DLI program predict English language development growth among EBs?
2. How does enrollment in a DLI program predict English language arts proficiency score differences among EB students in Grades 3-9?
3. How does enrollment in a DLI program predict English language arts proficiency growth among EB students in Grades 3-9?
4. How does enrollment in a DLI program predict mathematics proficiency score differences among EB students in Grades 3-9?
5. How does enrollment in a DLI program predict mathematics proficiency growth differences among EB students in Grades 3-9?
6. How does enrollment in a DLI program predict GPAs among EB students in Grades 6-10?

Limitations

This study investigated academic archival data typically collected by school districts and state education agencies each year. These data afforded rich opportunities for understanding student progress and program effectiveness. One of the strengths of this project is its potential to broaden the research on DLI programming in a mid-sized rural school district. Yet, there are also disadvantages to choosing one district such as the study may be somewhat limited in its scope of generalizability to other contexts. In addition, because parents of EB students have a choice in enrolling their students in DLI, the study design could not be experimental. Therefore, results associated with this study should not assume causation, but may add to the extant research on strong correlative associations (Bleske-Rechek et al., 2015).

Delimitations

Because of the difficulty in aggregating and securing many critical aspects of studying dual language immersion, certain factors were not addressed including teacher variables: teacher preparedness, professional development, teacher effectiveness, etc.; student variables: self-esteem, motivation, anxiety, etc.; or home variables: parent educational attainment, parental involvement, home environment, etc., all of which have been studied as having effects on students' academic achievement. Many research studies on DLI schools and districts include data on native-English speakers as well as partner-language speakers (Lindholm-Leary & Borsato, 2006; Lindholm-Leary & Genesee, 2014; Steele et al., 2013). Although the outcomes of native-English speakers in a rural district offering DLI would be important to the field, this study only focused on the impact of DLI on native-Spanish speakers. Additionally, DLI studies of minority-language students have often highlighted students' L1 and L2 proficiencies (Genesee & Lindholm-Leary, 2008; Lindholm-Leary & Block, 2010; Lindholm-Leary & Hernández, 2011). This study deviated from exploring the proficiency and growth of EB students in both languages and only examined academic English outcomes of those enrolled in DLI compared to those not enrolled in DLI.

The choice to focus on academic English results comes from the desire of the researcher to examine how DLI programming may help school administrators in complying with the challenges placed on schools to provide for EB student academic outcomes as measured by state required assessments. As a principal in one of the pilot schools in Utah (Eaton, 2016), the researcher hopes to verify if the Utah DLI model offers EB students within the school and district improved educational opportunities. Although criticisms regarding the Utah DLI model's economic and racial foci bare some credence, the motivations of the district personnel emanated from an expressed and sincere desire to provide better educational opportunities through a two-way dual language experience. From this position, this study may inform other school and district administrators in similar positions who are motivated to better understand and improve education for all students, including EBs.

Definition of Terms

Emergent bilingual (EB) can refer to any student who is in the process of learning more than one language (García et al., 2008), yet for the purposes of this study the term EB will refer to native-Spanish speakers learning English as an additional language.

Bilingual education encompasses instructional approaches that use more than one language to educate students. Bilingual education programs vary by length, intent, and structure of language use (Lindholm-Leary, 2016; Tedick, 2015; Ovando, 2003).

Dual Language Immersion (DLI) refers to a bilingual education program that uses two languages to educate students. For DLI programs in the U.S., English and a target language are used to provide instruction. DLI programs are differentiated by the amount of time that students are exposed to each language during the day. For example, 90:10 programs use the target language for 90% of the time and English for the other 10% for instruction. A 50:50 DLI program would share instructional language time equally between English and the target language (Lindholm-Leary & Howard, 2008).

One-way Dual language immersion DLI programs refer to those programs that are made up of largely native-English speakers that are receiving instruction in both English and a target language (Lindholm-Leary, 2016).

Two-Way Dual Language Immersion DLI programs refer to DLI programs that maintain balanced enrollments of native-English and target language-speakers (Lindholm-Leary, 2016). For Utah, in which this study was conducted, two-way DLI programs must have a minimum of a one-third to two-thirds mix of native-English and target language speakers. Thus, two-way programs can be comprised of one-third native-English or one-third target language speakers.

CHAPTER 2

LITERATURE REVIEW

The U.S. public education system continues to seek the best ways to educate a growing EB population while adhering to policy and legal mandates and functioning with finite resources. This quest has taken many forms, including the establishment of bilingual programs of varied configurations. Research on the efficacy of these efforts is still limited. The proposed study sought to determine any academic impact of the Utah DLI model within a rural context. The following literature review examines demographic trends and research that illustrate the types of models that have been used to educate EB students. This review also provides a summary of the research on academic and English language achievement among EB students over the last several decades. A context for understanding different pedagogical approaches to educating EB students and their philosophical underpinnings are also presented. Finally, the methods used for accounting for EB student achievement and progress afford a glimpse into what is or is not known regarding effectiveness of the pedagogical approaches used to educate EB students, followed by possible new ways of examining program efficacy.

Emergent Bilinguals as Public-School Students

EB students have been a growing demographic in the U.S. school system and, as such, have received increased attention in academic research over the years. The number of EB students in the U.S. has climbed 8.1% since 2000 (Bialik et al., 2018). By fall 2017, over 10% or 4.9 million of the U.S. student population was EB (Hussar et al., 2020). The western states have recently experienced higher rates of EB enrollment with at least 10 states identifying greater than 10% EB populations. California, Texas, and Nevada have the highest percentages of EB students at 19.2%, 18%, and 17.1%, respectively. Regardless of the state, although the majority of EB students live in urban centers (García & Kleifgen, 2018), by the 2014-15 school year 78% of districts enrolled at least one EB (Sugarman & Lazarín, 2020). The locations of where EB students have resided are in flux. Some states, like Delaware, have seen an increase of more than 7% whereas Arizona has seen a decline of about the same (Hussar et al., 2020).

EB students, commonly referred to in the U.S. school system as Limited English Proficient (LEP), English Language Learners (ELLs), and/or English Learners (ELs) are bilingual students who have acquired, or are in the process of acquiring, English at school (García et al., 2008). These labels suggest that the orientation of schooling is toward acquiring English rather than maintaining their native languages while gaining proficiency in English. Labels like LEP, ELL, and EL, have traditionally focused on how EB students' "English proficiency affects their ability to meaningfully participate and succeed in school" (Linquanti et al., 2016, p. 3) and consequently their struggles academically (Abedi, 2011; Genesee et al., 2005; Goldenberg, 2008; Kim et al., 2015; U.S. Department of Education, 2018a), rather than focusing on the potential of students to leverage their bilingualism or multilingualism as a resource (García & Kleifgen, 2018). There is concern that such labels discredit the importance of students' native language and assume that the education of EB students should be to focus on developing academic English (García et al., 2008). The U.S public system has not necessarily recognized or considered native-language maintenance as part of regular schooling (García et al., 2008).

The research literature that explores the educational development of EB students frequently uses LEP, ELL, and EL labels. However, to accurately reflect who these students are and honor their potential rather than possible deficits, the term EB will be used in this study. Accordingly, all students in bilingual programs can be considered EBs, however this study will specifically focus on the English language development (ELD) of native-Spanish speakers. Furthermore, the researcher recognizes the importance of the development and maintenance of Spanish of all students enrolled in Spanish-English bilingual programs. However, the orientation of this study is to determine the impact of the DLI program on the academic English development and academic outcomes of its native-Spanish speakers.

English Acquisition and Academic Achievement of Emergent Bilinguals

An assumption of the label EB is that these students maintain their first language to some degree while on a trajectory to attain English proficiency. Learning English is an obvious prerequisite for EB students to perform well academically when the majority of assessments in an English-dominant country are administered in English (Rumberger & Larson, 1998; Saunders & Marcelletti, 2013). Therefore, the time it takes EB students to become proficient becomes extremely important in maximizing academic opportunities. The following section first reviews the literature on the importance of English proficiency on academic outcomes and the time it takes EB students to reach proficiency. Secondly, this section reviews the academic outcomes of EB students and characteristics that affect those outcomes.

Effect of English Proficiency on Academic Achievement

Cummins (1979) pointed to the importance of English proficiency for EB students for academic achievement when proposing his *threshold hypothesis*, which

> …assumes that those aspects of bilingualism which might positively influence cognitive growth are unlikely to come into effect until the child has attained a certain minimum or threshold competence in a second language. (p. 229)

Cummins (2008) later argued that there are two thresholds in English acquisition that students need to attain. The first, *basic interpersonal communicative skills* (BICS), and the second, *cognitive academic language proficiency* (CALP). BICS refers to conversational fluency whereas CALP refers to academic language fluency. Cummins found that "educators and policy-makers frequently conflated conversational and academic dimensions of English language proficiency and that this conflation contributes significantly to the creation of academic difficulties for students who were learning English as an additional language" (p. 72).

Cummins (1979) also found that "the level of competence bilingual children achieve in their two languages acts as an intervening variable in mediating the effects of their bilingual learning experiences on cognition" (p. 229). Therefore, competence in both the students' L1 and L2 influences general cognitive processes that should correspond with better academic outcomes. In a review of multiple studies, Lindholm-Leary and Borsato (2006) found that higher levels of proficiency in both languages correlated with high academic achievement. In a study of students in Grades1-4, Lindholm and Aclan (1991) found that EB students with high levels of bilingualism reached grade-level achievement in math by Grade 3 and in English language arts by Grade 4. Rumberger and Larson (1998) found that Latinx EB students considered proficient in English had higher grades, lower rates of transiency, and were by Grade 9 more likely to graduate with the required number of credits than their non-English proficient and English-only Latinx peers. These findings substantiate that English proficiency is critical to success in U.S. public schools.

Time to L2 Proficiency

Hakuta et al. (2000) suggested that it takes EB students 3-5 years to attain English oral proficiency and 4-7 years to attain academic English proficiency. Other studies have found different results depending on a number of variables including student age, native-language proficiency, socioeconomic background, beginning English proficiency skills, type of instructional program in which the students were enrolled, and also the function of reclassification criteria (e.g., assessment, additional measures) by state (Thompson, 2017). Thus, there is not a common or nationally specified standard or expectation defining the time it takes EBs to acquire English proficiency (Hopkins et al., 2013; Robinson-Cimpian & Thompson, 2016). An EB student who does not achieve proficiency within the temporal expectations set by states or researchers is labeled as a Long-Term English Language Learner (LTELL). These students have for whatever reason did not exit the ELL designation in the schools in a timely manner (Menken et al., 2012). Thresholds for labeling EBs as LTELLs depend on operational definitions set by the individual researchers or government entities.

Age-at-Enrollment Effects on L2 Acquisition

One of the factors that affects the time to reclassification or attaining English language proficiency is the age of entrance into the public school system. In a study conducted in the New York public school system, Conger (2009) found that younger EB students gained English proficiency faster than their older counterparts with 5-year-old entrants acquiring proficiency in 1.69 median years and 10-year-old entrants reaching proficiency in 3.78 median years. Similar findings were found using scores from the World-Class Instructional Design and Assessment (WIDA) ACCESS for ELLs assessment (Cook et al. 2008). The authors found that younger students grow faster in their English proficiency than those in higher grades.

L2 Proficiency at Time of Enrollment Effects on L2 Acquisition

In addition to the age when entering the public school system, the English skills that EB students already bring to school influence the growth of English proficiency. Thompson (2017) followed a longitudinal research design to examine the time it took for EB students to attain English proficiency. The study covered nine years of data of 202,931students from the Los Angeles Unified School District (LAUSD). The author found that EB students had a 74% likelihood of becoming English proficient after nine years. EB "students who enter kindergarten with higher levels of English proficiency were approximately 13% more likely to be reclassified" as being English proficient than their peers entering with lower English skills (p. 352). Cook et al. (2008) also looked at the influence of beginning English proficiency on later English proficiency results among

EB students. Interestingly, when looking at growth rather than likelihood of attaining proficiency by a certain year, the authors found that students with lower proficiency levels grew faster. These results highlight the importance of examining proficiency by time and growth to understand better student English proficiency achievement.

Specific Language Influences on English Proficiency

Researchers have found that a student's specific L1 correlates with faster or slower rates of English acquisition. Over 400 languages and dialects are spoken by American K-12 students (Office of English Language Acquisition, 2018). Of the over 400 languages, Spanish is listed by 77% of the U.S. EB student population as their primary language (Bialik et al., 2018). The second and third most spoken native languages in the U.S. public school system are Arabic and Chinese respectively (Hussar et al., 2020). Some studies have shown that Spanish-speaking students have slower rates of English acquisition (Conger, 2010; Thompson, 2017). Conger studied English language acquisition rates among EB student in the New York City public school system. He found that Russian and Korean students were more likely to reach proficiency faster than other language groups. Of the languages represented in the sample (i.e., Spanish, Chinese, Russian, Bengali, Haitian-Creole, and Korean), Spanish-speakers were least likely to be proficient in English each year. Thompson found similar results to Conger in her study of EB students in the LAUSD. Spanish-speaking students were less likely to become proficient in English when compared to their Cantonese-, Korean-, or Filipino-speaking peers. Slama (2014) found that being Spanish-speaking correlated to a lower likelihood of acquiring English proficiency earlier when compared to other-language-speaking Massachusetts peers after controlling for low-income. The author argued that lower rates of English acquisition were the result of attending more segregated schools, thus having less exposure to English speaking peers.

Program Effects on L2 Acquisition

Considerable research in recent years has focused on differing types of instructional programming for EB students. For example, Umansky and Reardon (2014) conducted a longitudinal study of Latinx EB students in California. Of the 5,423 students sampled, the authors found that on average it took eight years for 50% of the students to reach English proficiency. Differences in instructional programing for EB students accounted for a significant portion of the variance of English acquisition among participants. Umansky and Reardon determined that

> English immersion students show an early advantage in all outcomes while students in the two-language programs catch up and, in some instances, surpass their peers in all English environments in later grades. (p. 23)

This conclusion supports Cummins (1979, 1981) findings that students who focus on L2 learning at the cost of L1 may perform well in earlier grades, but as the linguistic demand increases, competency in both L1 and L2 yields greater cognitive and academic advantages.

Emergent Bilingual Academic Achievement in the U.S.

Despite the criticisms of the appropriateness of high-stakes tests for accurately measuring the academic knowledge of EB students (Huang et al., 2012), the U.S. school system has continued to require academic testing of this student group to account for their levels of scholastic achievement. Since the reauthorization of the Elementary and Secondary Education Act (ESEA) in 2001, commonly known as the No Child Left Behind Act (NCLB), states have been required to report on the academic progress of students in English language arts, mathematics, science, and additionally English proficiency for EB students. Thus, for almost two decades, U.S. states, districts, and schools have measured, analyzed, and reported on the academic progress of EB students. Unfortunately, results have shown EB students to continually lag behind their peers in academic testing (Abedi, 2011; Solano-Flores & Trumbull, 2003).

The U.S. Department of Education (2018a) reported that for fourth grade, only 14% of EBs were proficient in mathematics on the 2017 National Assessment of Educational Progress (NAEP) and only 9% of EBs were proficient in reading (U.S. Department of Education, 2018a). The report also highlighted that although proficiency rates of EB students have increased slightly since 2000, recent proficiency levels have not changed much. When looking at the gaps between EBs and their peers, the reading proficiency levels on the 2019 NAEP assessment indicated a difference of 33 and 45 scaled points for grades four and eight, respectively, with EB students proficiency being lower (Hussar et al., 2020). For math, the 2019 NAEP results specified lower average scale scores of 23 points at fourth grade, and 42 points at eighth grade, for EB students compared to peers.

Factors Affecting Achievement of EB Students

Research reveals a number of variables that contribute to the difference in academic achievement scores between EB students and their peer groups. A few of the more frequently studied factors include assessment issues, socioeconomic factors, and instructional models (Genesee et al., 2005; Lindholm-Leary & Borsato, 2006; Taggart, 2018). Assessment practices have long been proven problematic for EB students due to the language complexity of content tests in English (Abedi, 2011; Hopkins et al., 2013). Abedi (2011) argued that content tests were not created with EB students in mind and that the language requirements of content tests weaken each test's validity and reliability for measuring the targeted constructs. One experiment demonstrated how EB students could perform better on a mathematics test when computerized administration was used with immediate glossary assistance (Abedi, 2009). The author found that providing a readily available computerized glossary embedded in the assessment assisted in ameliorating the effects of language complexity within the test. This finding demonstrates how assessments of content are simultaneously language tests. EB students may be unable to show what they know in content due to the language demand required of academic assessments.

Another variable, socioeconomic status, has long been established as a factor in academic achievement across students (Sirin, 2005), including EBs (Lindholm-Leary, 2001). In a study of EB Latinx students in rural communities in California, Hampton et al. (1995) found that socioeconomic status of parents of EB Latinx students was the greatest predictor of academic performance. Kieffer (2010) found in a longitudinal study that when controlling for SES "the differences in risk between [EBs] and native English speakers are substantially reduced" (p. 486). Though SES can be a great predictor in academic achievement of students, one of the major problems with using SES as a predictor is the variability in how SES is measured and also how those measures belie the construct of SES. Sirin (2005) explains that SES has three parental components including income, education, and occupation. In observational data, it could be difficult to ascertain all three indicators of SES. School districts typically only measure SES through family income as measured by eligibility for free and reduced lunch through the federal lunch program. Use of free and reduced lunch data as a factor to control is common, although admittedly limited in understanding SES in minority populations.

Accounting for socioeconomic status and using valid assessments are crucial in measuring the academic achievement of EBs. It is important to control for as many confounding factors as possible when measuring the academic outcomes of students in various programs. In addition to socioeconomic status and assessment measures, curriculum and instructional delivery have a significant impact on student achievement (Natriello, et al., 1989).

Bilingual Education in the United States

Articles and book titles written on the history of bilingual education in the United States use words and phrases like “bumpy” (Palmer et al., 2017, p. 449), “paradoxes” (Tedick, 2015, p. 1), “condemned without a trial” (Krashen, 1999), and “basements and pride to boutiques and profit” (Flores & García, 2017, p. 14). These descriptors connote a clear picture of bilingual education’s divisive and turbulent past in the U.S. Ovando (2003) describes the U.S.’s unsettled history of bilingual education using Baker and Jones’s (1998) typology of historical phases—permissive, restrictive, opportunist, and dismissive. Both Ovando and Jones argue that none of these historical periods have embraced bilingual education primarily for the sake of language minority students. Instead, the opportunistic phase promoted bilingualism as a foreign language issue while the permissive period allowed bilingual education to be implemented based on local needs. The restrictive and dismissive periods were, as they connote, times of obstruction or flat-out opposition to bilingual programs in the name of assimilation. Ovando further reasoned that bilingual education has been influenced more by political and economic forces "rather than any consistent ideology" (p. 2). Therefore, because of the lack of grounding of educational decisions in evidence-based practices, local politics, rhetoric, and special interests have dominated the public discourse and, consequently, educational policy regarding bilingual programs. This tension continues in today's political and educational arenas.

The splintered ideological and political history of bilingualism in the U.S. has at its foundations three often competing beliefs: (a) preserving the pluralistic and cultural heritage of a diverse U.S. populace, (b) development of foreign languages for economic and political strength, and (c) assimilating diverse people into one U.S. society. Ovando (2003) describes the tension through U.S. history that pit the *pluralistic* desires of those wanting to preserve the cultural heritage of different groups against the desires of *assimilationists* who wish to maintain an American society of monolinguistic and monocultural policies. Palmer et al. (2017) explain this tension as a by-product of having no common educational theory as a nation, stating

> …although linguistic and cultural diversity have always been a reality in the United States, the nation has long been ambivalent in terms of general attitudes

toward this diversity. (p. 449)

The tension between assimilationist and pluralistic discourses has created a conflicted path forward for bilingual education in the U.S. for EB students.

Although the overall American bilingual history appears disjointed, many local communities and movements have produced examples of successful bilingual programming for EB students as well as for native-English speakers. Many early bilingual programs mirrored the cultural identity of the local area, such as German-English efforts in the mid-west and Pennsylvania, French-English in Louisiana, and Spanish-English in New Mexico (Gándara & Escamilla, 2017; Ovando, 2003; Palmer et al., 2017). In the 1960s and 70s, two Spanish dual language programs started in the U.S. (Ovando, 2003). One program began in Miami, FL among Cuban refugee families serving both native-English and native-Spanish students (Gándara & Escamilla, 2017) and the other program began in Culver City, California, which served only native-English students for the first several years (de Jong, 2016).

Today, the number of dual language programs has reached over 3,000 (Lam & Richards, 2020). This number does not include the other types of bilingual programs such as transitional bilingual and developmental bilingual that will be discussed in a later section. Although bilingual programs can serve both native-English and non-native-English speakers, the majority serves students coming to school without being English proficient (Gándara & Escamilla, 2017).

Government Influence on Bilingual Education for Emergent Bilinguals

It would be hard to understand the U.S. school system's approach to educating EB students without addressing how federal legislation has played a significant role in the education of EB students. One of the major pieces of legislation that put an end to the "sink or swim" (Kim et al., 2015) de facto EB educational programing within the U.S. was the Civil Rights Act of 1964. Ovando (2003) argued that the Civil Rights Act of 1964, with the creation of the Office of Civil Rights, ushered in the beginning of changes towards linguistic diversity for the country. The Act stated, "No person on the United States shall, on the grounds of race, color, or national origin, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity receiving Federal financial assistance". A couple of years later the Bilingual Education Act of 1968, the "first official federal recognition of the needs" of EB students, encouraged instruction in languages other than English (Stewner-Manzanares, 1988, p. 1). The act eventually provided funds in the form of grants to offer bilingual education for language-minority students, but the legislation did not mandate schools to address the instructional practices for EB students.

A landmark Supreme Court case, *Lau v. Nichols,* pushed the school system to actively address the needs of EB students. This case from California authorized increased educational attention and opportunities for language-minority students. The U.S. Supreme Court ruled in favor of native-Chinese speaking students and their families stating, "There is no equality of treatment merely by providing students with the same facilities, textbooks, teachers, and curriculum; for students who do not understand

English are effectively foreclosed from any meaningful education" (*Lau v Nichols*, 1974). The ruling strengthened the application of the Civil Rights Act of 1964 and led to an increased advocacy of bilingual education and accountability for providing improved opportunities for EB students (National Academies of Sciences Engineering and Medicine, 2017).

In that same year, the Equal Education Opportunities Act of 1974 strengthened the ruling of *Lau v. Nichols* to address the needs of EB students not just in schools receiving federal funding, but all schools (Ovando, 2003). The next major lawsuit that would affect the education of EB students was brought before the U.S. Court of Appeals for the Fifth Circuit in 1981. *Castañeda v. Pickard* established a three-pronged test for schools and districts to measure their instructional programming for EB students. The ruling required that instructional programming for EB students must be (a) based on "sound" educational theory, (b) have adequate resources to be implemented well, and finally must be (c) evaluated for effectiveness after a reasonable time (*Castañeda v Pickard*, 1981). These major pieces of legislation and litigation delineate the three points upon which many instructional programs for EB students should be measured.

In contrast to the Supreme Court cases and federal legislation providing opportunities for bilingual education, some state governments passed laws which prevented bilingual programs in schools. Although English-only policies were introduced in some parts of the U.S. as early as the 18th and 19th centuries, the *English-Only Movement* describes the late 1990s and early 2000s legislation passed in California, Arizona, and Massachusetts (Pac, 2012). To date, only Arizona has retained its laws, although it has begun to loosen its restrictions by allowing more opportunities for EB students to receive instruction in their native language (Mitchell, 2019). Although the legislation has been appealed in most cases, English-only and monolingual sentiments still remain. Some have argued that educational assessments, practice, and curriculum themselves have subtly perpetuated English-only policies (Gándara & Escamilla, 2017; Pac, 2012).

Educational Programs for Emergent Bilingual Students

Because educational programs for EB students are adjudicated at the state and local level, they reflect the dominating political and ideological sentiments of the local region and time. The three competing beliefs of preserving pluralism, furthering economic progress, and assimilating all into one society continue to influence the spectrum of educational offerings for EB students. Pluralism and assimilation trends clearly affect educational program goals and philosophies. Yet, economic interests have arguably been used to justify many of the different programs (Kelly, 2018).

Language Instruction Educational Programs (LIEP) contain certain characteristics of these beliefs which help delineate the differences in educational approaches and goals (National Academies of Sciences Engineering and Medicine, 2017; U.S. Department of Education et al., 2012). Valdés et al. (2017) use the terms *transitional* to describe those programs that seek to move an EB from a primary language to a secondary one without concern for the first language and *compensatory* for those programs which seek to assist an English learner in attaining proficiency in English while maintaining their first language. Likewise, Tedick (2015) refers to those programs that provide English instruction without assisting in maintaining their primary language as *subtractive* and those programs that maintain their first language while providing for English acquisition as *additive*. The terms compensatory and additive embody a similar understanding which is that EB students learning English are gaining an additional skill and not subtracting or transitioning from one to another. The educational language programs commonly offered to EB students fall into one of these two camps.

A recent review of types of LIEPs by the U.S. Department of Education et al. (2012), categorized one half of the programs or models as having an ESL approach and the others having a bilingual focus. The authors of this report defined the ESL approach as an instructional framework that "focuses on instruction [in English] as the primary means to help ELs acquire the language and ultimately meet high academic standards" (p. ix). This type of program falls under the subtractive models. In contrast, the U.S. Department of Education et al.'s bilingual approach refers to a framework "based on a commitment to the understanding that instruction in students' L1 [first language] will help them to meet the goals of and attaining English proficiency and meeting high academic achievement standards" (p. ix). The ESL approaches use the second language (English) or L2 as the primary language for instruction, while the bilingual programs recognize the importance of EL student's L1 as a key to L2 improvement.

In addition to categorizing LIEPs as either ESL or bilingual, the list of instructional models is distinguished by their goals. As shown in Table 2.1, the goals become progressively broad as the programs move along a continuum from the single

Table 2.1

Language Instruction Education Programs

Model	Description	Approach	Goals	Features
English as a second language (ESL) instruction; English language development (ELD)	Explicit language instruction focuses on development of proficiency in the English language, including grammar, vocabulary, and communication skills.	ESL	Proficiency in English	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 ESL	Language instruction uses content as a medium for building language skills.	ESL	Proficiency in English; Preparation to meet academic achievement standards	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.
Sheltered Instruction (SI); Specially designed academic instruction in English (SDAIE); The Sheltered Instruction Observation Protocol (SIOP)	Instruction that simultaneously introduces both language and content, using specialized techniques to accommodate ELs linguistic needs. Instruction focuses on the teaching of content rather that the English language itself, even though acquisition of English may be one of its instructional goals.	ESL	Proficiency in English; Preparation to meet academic achievement standards	SI may be used for EL-only classrooms or for mixed classrooms with ELs and non-ELs. Instruction is likely to be delivered by a general education teacher but may be delivered by an ESL certified teacher.
Transitional bilingual education (TBE); Early-exit bilingual	Students begin in grade K or 1 by receiving instruction all or mostly in their L1, and transition incrementally over to English. Typically, transition to all English is complete by mid- to late elementary grades. L1 is used to leverage L2 acquisition, but L1 proficiency is not a program goal.	Bilingual	Proficiency in English; Preparation to meet academic achievement standards	Some TBE programs begin with L1 exclusively, others begin with a majority of L1 and use some L2. Typically, students complete their transition by around grade 3, but may exit as early as grade 2, or as late as grade 5.
Developmental bilingual education (DBE); Late-exit bilingual; Maintenance bilingual	Students begin in grade K or 1 by receiving instruction all or mostly in their L1, and transition incrementally over to English. Regardless of when or whether students attain English, the program is designed to keep them enrolled through its completion (typically, the end of elementary school), using a 50-50 language balance through the end.	Bilingual	Proficiency in English; Preparation to meet academic achievement standards; Bilingualism and biliteracy	Programs follow either a 50-50 model or a 90-10 model (which ultimately transitions to a 50-50). Programs may balance languages by dividing instructional time according to content area, class periods, instructor, week, unit, or semester. Teachers may be bilingual, or teachers who teach in English may use sheltered instruction techniques to make their instruction accessible for ELs.
Two-way immersion (TWI); Dual immersion (DI); Dual language immersion (DLI)	ELs and non-ELs receive instruction in English and a non-English language.	Bilingual	Proficiency in English; Preparation to meet academic achievement standards; Bilingualism and biliteracy; Biculturalism	Programs follow either a 50-50 model or a 90-10 model (which ultimately transitions to a 50-50). Programs may balance languages by dividing instructional time according to content area, class periods, instructor, week, unit or semester. Teachers may be bilingual, or teachers who teach in English may use sheltered instruction techniques to make their instruction accessible for ELs.

Adapted from (National Academies of Sciences Engineering and Medicine, 2017; U.S. Department of Education et al., 2012) goal of English proficiency to, at the other end of the continuum, a program including goals in English proficiency, high academic achievement, bilingualism, biliteracy, and biculturalism. The first model, ESL or ELD, focuses solely on the goals of helping EL students attain proficiency in English, rather than maintaining the students' L1. The next goal, preparation to meet academic achievement standards, as added to the subsequent models brings a renewed focus on supporting students to perform academically but does not yet constitute the compensatory or additive component.

In contrast to ESL programs, the bilingual approaches add the aim of bilingualism and biliteracy to the models, which value a student's L1 proficiency while adding English proficiency. Bilingualism in this context is not merely having experiences with two languages or even "a person who *is two monolinguals in one person"* (Valdés, 2015, p. 38), but rather speaks to a holistic approach where bilingualism is "grounded in the idea that what is known and understood in one language contributes to what is known and understood in the other" (Hopewell & Escamilla, 2015, p. 39). Thus, bilingualism as a language acquisition approach, leverages L1 to the attainment of L2 through academic content.

Tedick's (2015) description of *subtractive* and *additive* programs assists in clarifying the differences between approaches and goals of LIEPs models. Examples of subtractive programs include ESL, or ELD programs, and Structured English immersion where students are immersed in intensive English instruction so that they can be moved quickly into mainstreamed English content classes. Other subtractive models include content-based ESL and sheltered instruction (SI), which both focus on using content as the main medium for language instruction. One bilingual model, Transitional Bilingual Education (TBE)—where students' native language is used for a very short period of time while students learn English in order to be mainstreamed into dominant language classes (Tedick, 2015)—also fits a subtractive structure because the goal of this model is to exit students as quickly as possible out of L1 instruction into English mainstream classes.

Additive bilingual programs such as Developmental Bilingual Education (DBE) and Dual Language Immersion (DLI) maintain both the native and the second language (National Academies of Sciences Engineering and Medicine, 2017; Tedick, 2015). DBE focuses on providing minority-language students instruction in both languages until the end of elementary school. Both TBE and DBE models leverage the first language of the minority-language students to acquire the second language, yet DBE uses both languages over several years thus facilitating ELs' opportunities to maintain their first language, while acquiring English and increasing content knowledge (National Academies of Sciences Engineering and Medicine, 2017; U.S. Department of Education et al., 2012). Dual language immersion (DLI) adds biculturalism as a goal (National Academies of Sciences Engineering and Medicine, 2017; U.S. Department of Education et al., 2012).

Dual Language Immersion Programs

DLI programs distinguish themselves by "[adhering] to the principles of additive bilingualism and biliteracy and cultural pluralism" (Tedick et al., 2011, p. 1) more so than the other bilingual programs. This is accomplished by honoring the value of both English and the target language as well as the cultures that use those languages. Umansky et al. (2015) published an article entitled The Promise of Bilingual and Dual Immersion Education. The authors, citing the cognitive, cultural, and economic advantages of bilingualism along with the positive academic outcomes, make the argument to increase investment into high-quality DLI programs. The following section reviews the elements of different program types that make up DLI, the growth of such programs in the U.S. and specifically, the Utah model of DLI. In addition, the literature related to the *promise* of DLI programs to provide EB students with a high-quality education is reviewed along with the potential issues and criticisms that DLI programs have drawn.

Dual Language Immersion Characteristics and Academic Outcomes

DLI programs are differentiated both by the amount of time dedicated to the partner language and by the percent of students who are native-English and native speakers of the partner language. *One-way* models are made up of second language learners, while *two-way* models maintain more even numbers of students from both language groups. In 90:10 models, sometimes called full immersion, teachers deliver instruction 90% of the time in the target language and 10% in English for the first few years of school (Lindholm-Leary, 2016). As students advance in years so does the amount of instruction in English. Therefore, for a 90:10 Spanish DLI program, EB students would receive the majority of their instruction in Spanish for their first several years and then would gradually receive more of their content instruction in English. For native-English speakers, they would be almost totally immersed in the second language—

Spanish—for the first few years. In contrast, 50:50 models maintain an equal amount of instruction in Spanish and English for all students for the duration of the program.

Investigating DLI Academic Achievement of Emergent Bilingual Students

The research literature regarding DLI programs has shown that both 50:50 and 90:10 programs can provide EB students with superior academic outcomes compared to traditional English mainstream education (Kim et al., 2015; Lindholm-Leary & Borsato, 2006; Lindholm-Leary & Genesee, 2010). The 50:50 programs have been shown to provide better academic outcomes in language arts and mathematics assessments over transitional or developmental bilingual and ESL programs and are only slightly less effective than 90:10 programs in both English and target language outcomes (Acosta et al., 2019; Collier & Thomas, 2009; Morita-Mullaney et al., 2020). Students in 90:10 programs had better Spanish fluency than those in 50:50 programs (Lindholm-Leary & Howard, 2008), thus supporting DLI's more broad goals of building bilingualism.

Although the efficacy of DLI programs as a whole has been established in the literature, the methods in studying DLI programs have evolved. As sophistication of statistics have progressed in analyzing academic results of students enrolled in DLI programs, so too have the opportunities for improvements in individual programs. A vast majority of studies include large samples from highly concentrated language minority students in urban regions (Lindholm-Leary & Block, 2010; Thomas & Collier, 2002). In earlier studies of DLI outcomes, researchers compared mean scores of students' academic and linguistic outcomes by program type (e.g., Marian et al., 2013). Other studies used

ANOVA to study the difference between groups (e.g., Cobb et al., 2006) and regression to control for a number of confounding variables (e.g., Tran et al., 2015). More recent studies have accounted for the interdependency of student outcomes influenced by common variables clustered within schools or regions, thus accounting for between-location variances (e.g., Steele et al., 2013; Watzinger-Tharp et al., 2016). The literature review below discusses a few of these sample studies to demonstrate the advantages of the types of studies that have been conducted.

In a seminal study, Thomas and Collier (2002) conducted a large national study of 210,054 EB students. They compared the academic outcomes of EB students across multiple programs including traditional English mainstream, ESL, transitional bilingual, developmental bilingual, and one- and two-way DLI programs. The longitudinal study found that EB students enrolled in DLI programs achieved higher scores in English language arts, mathematics, science, social studies, and writing than their EB peers in English-only and traditional English-as-a-second-language programming by late-elementary and middle-school grades. The study used percentile scores to compare results, which showed how students in DLI and bilingual programs improved over time.

In another comparative study of 659 students from predominantly low-income schools in California, Lindholm-Leary and Block (2010) found that English proficient Hispanic and Hispanic-EB students in 90:10 programs outperformed their EB and English proficient Hispanic peers in English language arts and mathematics in Grades 4-6. In almost all of the grades compared, the DLI students scored above the state average in both content areas compared to their non-DLI peers. In both the Thomas and Collier

(2002) and the Lindholm-Leary and Block comparison studies, longitudinal analyses included comparing annual outcome measures across groups of students enrolled in different EB educational programs which provided a way of comparing groups for program efficacy.

Other studies have employed ANOVA or regression analyses. Cobb et al. (2006) used a 2 x 2 x 2 repeated measures ANOVA to compare the means of academic outcomes of native- Spanish-speaking EB students in DLI to non-DLI peers. By employing a repeated measure ANOVA, the authors found significant differences in outcomes of native-Spanish speaking students based on program. Consistent with other findings, those native-Spanish-speaking students enrolled in DLI performed better than their peers in mathematics and writing, though smaller effects were found in reading. The Cobb et al. study used sixth- and seventh-grade measures in their design. Although repeated measures design has many advantages over comparing means across multiple years through separate analyses as used in the above studies, this design only includes those students with scores in all of the time points under study. This restricted the number of students in the sample, which lessened the design's ability to account for actual variance. Another weakness of repeated measures ANOVA noted by the authors is that this design assumes that "there are no significant (unaccounted for) individual differences in systematic changes over time (such as linear slopes) and there are equal correlations among all possible pairs of *Y* values measured at different times" (Cohen et al., 2003, p. 578). The assumption that there are no significant, unaccounted for, individual differences is impractical with the observational data collected for that study.

Additionally, repeated measures ANOVA do not account for the shared variance that is present in clustered data, such as dependent variance of students' scores clustered within students and students clustered within schools.

In a more recent study, Valentino and Reardon (2015), measured the differential effect between programs for EBs including DLI. Their study of 13,750 EB students used a multilevel model (MLM) analysis which accounted for the clustered nature of student scores within students and within schools. By employing MLM, the authors' statistical model accounted for a truer model of the clustered factors effecting outcomes. The authors found that the growth trajectory of EB students in DLI programs increased at a much steeper rate than traditional English immersion services, transitional bilingual, and developmental bilingual programs for English language arts. In mathematics, the authors found that although all students' standardized math scores declined from second to seventh grades, the DLI students declined more slowly than did students in the other programs studied. MLM analyses allowed for the measuring of the trajectory of growth while accounting for the variance explained by clustered contexts.

MLM has increased in use in educational efficacy research (Gustafsson, 2010) and, in specific, the study of DLI (e.g., Steele et al., 2013; Valentino & Reardon, 2015; Watzinger-Tharp et al., 2016). Some of the advantages of MLM overcome the weaknesses of repeated measures ANOVA in that all student data can be used even with missing data (Peugh, 2010), which better models the variances associated in clustered data. The hope would be that MLM would become more widely used in effectiveness research for new and well established DLI programs across the U.S.

Dual Language Immersion Growth in the U.S.

DLI programs have grown from one program in Miami, FL in 1963, among the Cuban refugees, to over 2,000 programs nationwide (Gross, 2016; Ramirez, 2016). Much of the growth in DLI can be attributed to a number of states such as California, Delaware, Georgia, Louisiana, Minnesota, Ohio, Oregon, Texas, and Utah endorsing DLI instruction (American Councils for International Education, n.d.). Currently, the majority of DLI programs use Spanish as the target language, though use of many other languages has increased including, "Arabic, Armenian, Cantonese, Filipino, French, German, Greek, Haitian Creole, Hebrew, Hmong, Italian, Japanese, Khmer, Korean, Mandarin Chinese, Polish, Portuguese, Russian, Ukrainian, Urdu, and Vietnamese" and 14 Native American languages (Thomas & Collier, 2019, p. 104).

Although DLI programs have been growing in number and with more languages offered, there are many difficulties in implementing such educational programs, not the least of which is the difficulty in finding and developing qualified teachers (De La Garza et al., 2015). Some states have attempted to remedy this issue with hiring native-speaking guest teachers from other countries. Although this seems like a perfect solution to teacher shortages, there are a number of obstacles to consider such as pedagogical differences between other countries and the U.S. system, the lack of pedagogical training and experience, and cultural differences between teachers from other countries and the minority students in the U.S. whom they teach (Freeman et al., 2018). These potential staffing challenges can vary based on the social circumstances of hiring schools and available resources.

Dual Language Immersion Growth in Utah

DLI started in Utah as early as 1979 (Leite & Cook, 2015). Yet, Utah's push to increase DLI programs as a state started in 2008 with Senate Bill 41, which sought to enhance second language offerings in the state by providing access and support for pilot programs. The first programs started in elementary schools as one-way DLI for French, Chinese, or Spanish and two-way DLI for Spanish. The USBE classified two-way programs as those having, ideally, a 1:1 ratio of students with and without the target language but would allow two-way programs with a 2:1 ratio (Utah State Board of Education, 2018). DLI grew from 25 programs servicing approximately 1,400 students in 2008 to over 118 programs providing DLI for approximately 25,000 students by 2015 (Leite & Cook, 2015).

Utah's Dual Language Immersion Model

The Utah State Board of Education (USBE) supports and assists new DLI programs through required teacher trainings and on-site visits. The USBE requires certain assurances be provided to maintain state support (Utah State Board of Education, 2019c). Assurances required for implementing school districts cover the (a) essential elements of the instructional model, (b) the open enrollment policy for students, and (c) requirements for professional development and assessment of students.

The Utah instructional model currently requires a 50:50, two-teacher model. The target language and English are to be distributed equally across the instructional day. Content and language requirements are set to ensure a balance of language exposure and practice with one teacher for the target-language content and one for English content areas. Figures 2.1 and 2.2 provide a visual breakdown of content areas by language of instruction. Mathematics is covered entirely in the target language in the early elementary

Figure 2.1

Dual Language Immersion Instructional Time: Grades 1-3

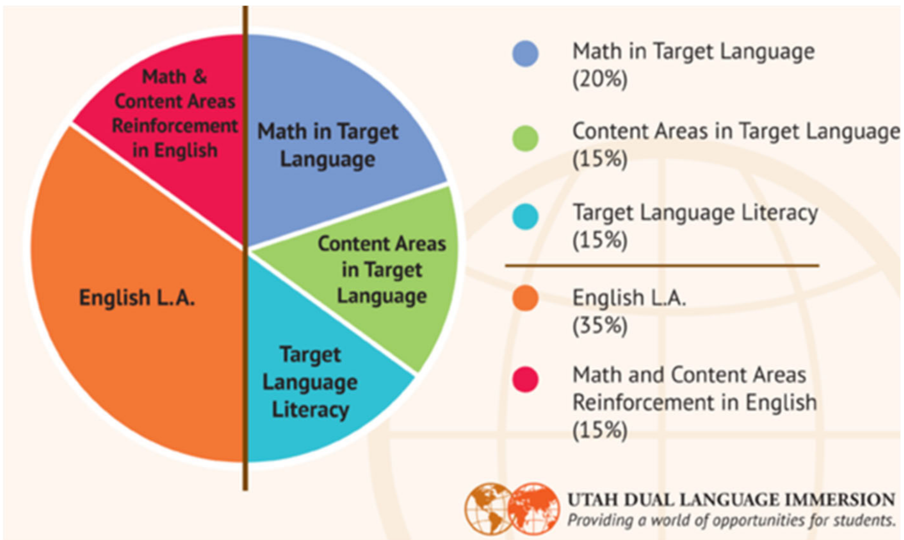
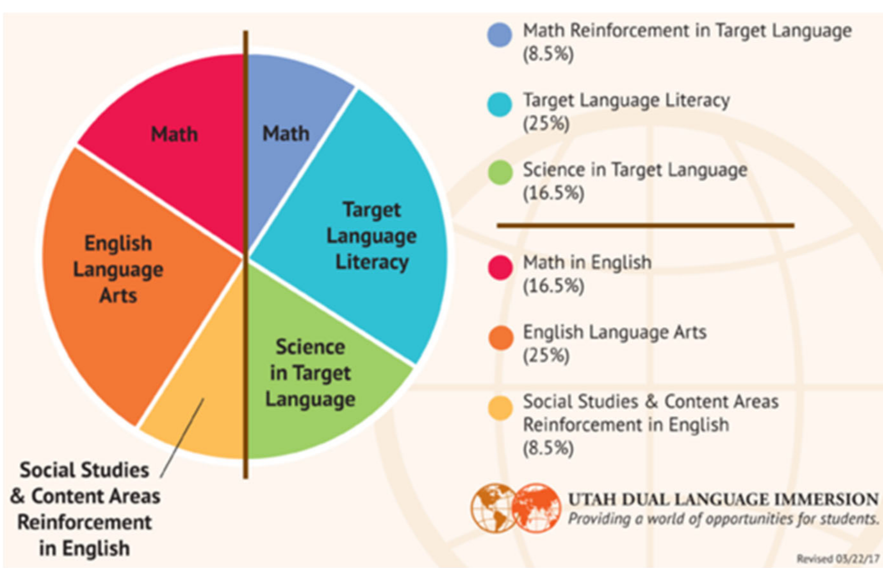


Figure 2.2

Dual Language Immersion Instructional Time: Grades 4-5



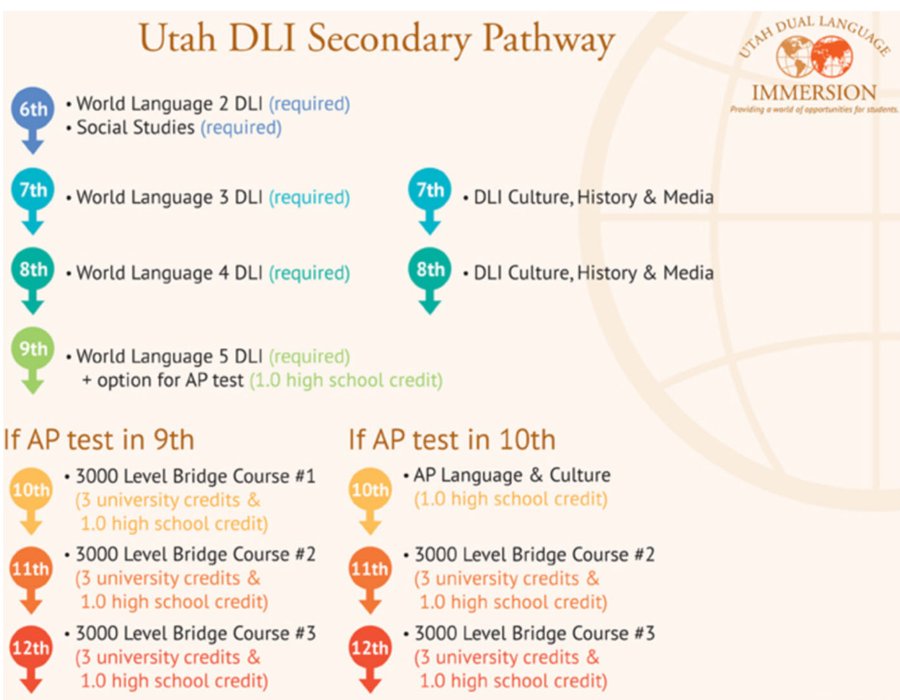
grades with some reinforcement of concepts in English. In fourth grade, the instruction of mathematics shifts from almost entirely in the target language to more instruction in English with reinforcing of concepts in target language. Throughout all elementary years, English language arts receives roughly the same number of minutes as non-DLI peers.

The Utah model not only requires the separation of content by language but requires the separation of languages by teachers. Teachers are asked to only speak to students in the language they are assigned to teach both in and out of the classrooms. Additionally, DLI target-language teachers are not to teach students not enrolled in DLI.

The structure of DLI looks different in the secondary grades (see Figure 2.3). Students no longer receive instruction in the target language for 50% of the time. The

Figure 2.3

Dual Language Immersion Secondary Level Class Progression



Utah model currently requires two classes in the target language each year in middle school. The secondary classes focus on preparing students to take the AP Spanish exam in the student's 9th- or 10th-grade year. Once the AP Spanish test is passed, students have the opportunity to take 300-level college classes in Spanish while still being enrolled in high school through concurrent enrollment.

In addition to instructional model assurances, Utah schools must maintain an open enrollment policy for all students, regardless of ability. Enrollment in DLI is typically started in Kindergarten or Grade 1. For schools having more entries than spots, a lottery has been instituted to provide a fair enrollment process.

Students enrolled in DLI are required to take the American Council on the Teaching of Foreign Languages' (ACTFL) Assessment of Performance toward Proficiency of Languages (AAPPL) which measures target-language acquisition. In addition to the AAPPL assessment, which is taken in the target language, students take all other state mandated assessments in English. To help all students at varying proficiency levels achieve at high levels on these assessments, DLI teachers are required to attend state trainings (Utah State Board of Education, 2019b).

Benefits and Criticisms of Utah's Dual Language Immersion Model

Utah's required model provides some ancillary benefits to DLI programs and their schools but also has come under some criticism. As stated, one of the requirements for the Utah model is the separation of languages. Thomas and Collier (2019) list separation of languages, cost-effectiveness, and leveraging academic instruction by teacher language as the benefits of separating the languages of instruction. These authors suggest that separation of language needs “to be handled with sensitivity to the nuances of this issue” (Thomas & Collier, 2019, p. 101). Separation of languages was touted as an important aspect of DLI programming but has been challenged through recent research naming translanguaging as an effective pedagogical strategy for EB students (Freeman et al., 2018). Freeman et al. explain that translanguaging is the process of bilinguals or multilinguals using one, unified, complex language system rather than two or more separate language systems. Thus, the theory follows that bilinguals do not learn nor produce each language separately, suggesting that teaching them separately fails to recognize bilinguals’ authentic language experience. The one complex language system advanced in translanguaging theory is discounted with the separation of languages required in the Utah model.

One requirement of the Utah’s DLI program, the two-teacher model, offers a cost potential by allowing for classrooms to maintain full loads of students rather than hiring an additional teacher to teach Spanish as a supplement to individual classes as offered in Foreign Language for Elementary Students or FLES programs. Thus, bilingual teachers teaching in the target language are not an additional cost but rather fill a full-time equivalent (FTE) position. FTEs are the same for a school before and after implementing DLI. Additionally, a two-teacher model affords teacher focus on one language allowing native-target-language teachers to teach content in their language of strength and for near-native-second-language teachers hired to teach the target language the ability to focus on one language of instruction.

Some DLI model criticisms are based in the ideological tension between building human capital and promoting equity. In a review of print media about implementation of Utah's DLI model, Valdez, Delavan, et al. (2016) argue that Utah's model detracts from the "equity/heritage" framework, that has traditionally surrounded bilingual programs (p. 851. Flores (2016) argues that language-as-resource, or DLI programs,

> …inadvertently reproduce White supremacist, imperialist, and capitalist relations of power both through the expectation that all people should master the idealized language practices of hegemonic Whiteness and through the otherizing of people who are unable or unwilling to fit this ideal. (p. 33)

In a similar vein of thought, Valdez, Freire, et al. (2016) critique the Utah DLI model as furthering *gentrification*. Gentrification, in this context, describes the increased opportunities presented to the majority culture while decreasing the opportunities for the minority language population for whom bilingual programs such as DLI were intended. Good student outcomes in DLI programs may mean the White middle- and upper-class students, now bilingual, may take opportunities from bilingual EB students (Flores, 2016).Valdez, Freire, et al. contest that by providing more access to bilingual programs for majority students, opportunities for minority language children to pursue "routes to economic empowerment" are diminished due to increased competition (p. 613).

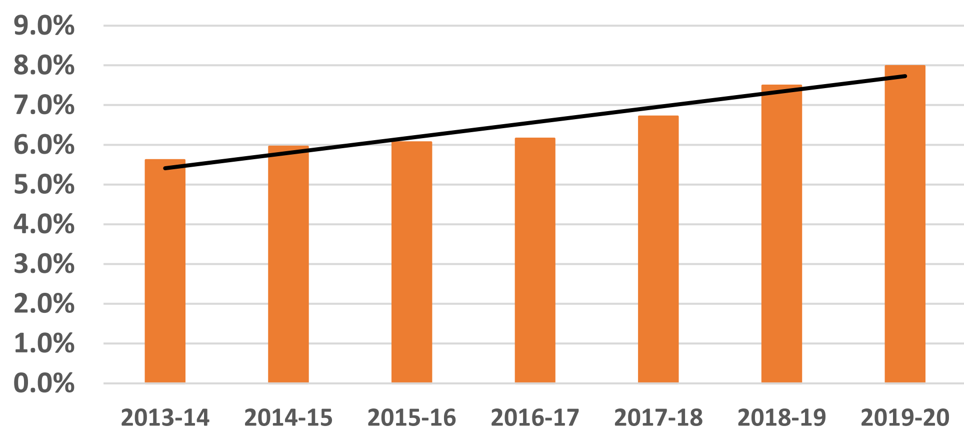
The criticisms of DLI in Utah likely merit reflection by Utah leaders, policymakers, and educators. Although providing DLI programs for native-English speakers could further upset the power balance between language-minority students and their native-English speaking peers, attention to academic performance and outcomes is an important step towards empowerment (Ladson-Billings, 1995). Gay (2018) argues that disempowerment of minority populations comes from a lack of opportunities for high academic achievement. Ladson-Billings advocates for not just the empowerment of individual minority students, but the collective community through adhering to the tenets of *culturally relevant pedagogy*—(a) academic success, (b) cultural competence, and (c) competence to challenge the status quo of the current social order. Thus, there is a need to elevate the cultural status of EB students while simultaneously raising their academic standing and capacity to direct their future through sufficient educational progress.

Utah Emergent Bilingual Demographics

Because of the population growth of native-Spanish speaking students, Utah has an opportunity to increase the number of two-way DLI programs. Though smaller than the national average (9.6% EB students), Utah's EB population continues to increase (McFarland et al., 2019). The Utah State Board of Education recently reported the total population of EB students for the state at 53,234 (Utah State Board of Education, 2019a). Figure 2.4 depicts the growth of EB students as a percentage of the total K-12 population from the 2013-14 school year to the 2019-20 school year. Since the 2013-14 school year, Utah's EB population has grown 54.7% from a total population of 34,394 to 53,234, representing a change from 5.6% to 8% of Utah's total student enrollment. In the 2014-15 school year, 15 of Utah's 41 school district had at least 5% of their enrollment comprised of EB students (U.S. Department of Education, 2018b). Native Spanish-speaking students comprise the largest percentage of EB students in Utah at 77% with Navajo a distant second at 2% (Sugarman & Geary, 2018). The growth of Spanish-speaking EB students in Utah makes implementing two-way programs including adequate numbers of native-target-language speakers more plausible.

Figure 2.4

Emergent Bilingual Growth Trend for Utah K-12 Population



Adapted from Utah State Board of Education. (2019). *Fall Enrollment by Demographics and Grade Levels*. Retrieved from https://www.schools.utah.gov/data/reports?mid=1424&tid=4

Accounting for Dual Language Immersion Outcomes

As mentioned earlier, government intervention in the form of legislation and litigation has increased the attention given to EB students at the national, state, district, school, and classroom levels. From the influence of the Civil Rights Act of 1964 to the current ESSA Act and from *Lau v. Nichols* (1974) to *Castañeda v. Pickard* (1981), school systems have adopted increased scrutiny in their efforts to provide EB students with researched-based programs. Where *Lau v. Nichols* paved the way for bilingual programs, *Castañeda v. Pickard* established a standard of implementation requiring a scientific base of programming, adequate program support, and evaluation of program effectiveness. Currently, the ESSA maintains the requirement that states report the academic progress of English language acquisition and academic achievement of EB students.

Schools and districts have a significant influence on the implementation and success of DLI programs. Menken and Solorza (2014) determined that under the pressure of high stakes testing and accountability, principals in New York City dismantled many bilingual programs even though there had been a history of bilingual support in the district. The authors found that, principals, feeling tremendous pressure for their schools to perform well on state assessments, had a “myopic focus on English as the overriding instructional goal for emergent bilinguals” (p. 106) and often blamed bilingual programs when EB students did not perform well.

In stark contrast to the actions of the principals studied by Menken and Solorza, Souto-Manning et al. (2016) quote another New York City principal of an elementary school offering DLI saying:

> You can’t lead a school if all you are doing is reacting to mandates and compliance issues. You have to have a vision, a plan and then assess how the mandates and compliance issues fit within your vision, your mission. (p. 58)

This idea of assessing how compliance and mandates fit within a plan to educate students illustrates the concept of *coherence* between *internal accountability* and *external accountability* (Fullan & Quinn, 2016). Internal accountability is when individuals take on the work of continued improvement towards student success whereas, external accountability is how the organization monitors the improvement to ensure congruence of performance with “societal expectations and requirements” (p. 111). The coherence comes when internal accountability of schools precedes external accountability, which empowers schools to take on the work of continual improvement and not relying on or waiting for periodic external sources for validation of results (Elmore, 2004).

In the case of DLI programs, schools and districts must engage in the work of continual improvement through consistent program evaluation, which can include many aspects that allow a system to improve. Stufflebeam (1968) highlights four areas that schools and districts could and should engage in when evaluating their DLI programs: context, input, process, and product. *Context* looks at a program's ability to meet specified goals, *input* concerns evaluating alternative programs, *process* addresses implementation, and *product* highlights the outcomes of implemented programs.

For this study of an established program in Utah, using the product lens for evaluating DLI makes the most sense. Because the program has been in existence for a number of years, students participating in DLI from the early grades have now reached the secondary-school level. Strong research studies of DLI programs must be longitudinal, have a control group with whom to compare DLI participants, follow students at least through middle school, and must employ rigorous statistical methods (Salazar, 1998). Using these standards will assist schools in making the types of decisions necessary to strengthen internal accountability. Thus, this review of relevant literature provides both the background to inform extended research and calls for the utilization of extensive data and more sophisticated analyses to meet coherent accountability standards associated with public-school, language-focused programs.

CHAPTER 3

METHODS

This study aimed to evaluate how enrollment in a DLI program predicted the academic outcomes of native-Spanish-speaking EB students in a rural district in Utah. Researchers have claimed that DLI enrollment provides EB students with greater opportunities for academic achievement and progress (Collier & Thomas, 2017; Lindholm-Leary, 2012). However, while DLI programs have been increasing in popularity across the county (Arteagoitia & Yen, 2020), the majority of EB students do not have access to DLI (Tanenbaum et al., 2012). Utah has embraced DLI programs, but the impetus stemmed from an economic focus of increasing bilingualism among its mainstream student population, not necessarily to provide bilingual education for its minority populations, whom such programs have typically served (Valdez, Freire, et al., 2016). With both praise and criticisms leveled at two-way DLI programs in Utah and the challenge that school administrators have in implementing and evaluating quality programs for EB students, this study of a DLI program in a Utah school district helps fill a need in the research literature. Its design highlights Utah's specific context, but also contributes to the national challenge, as other school systems grapple with implementing programs to benefit EB students. A common goal of education is to increase achievement in key subjects (e.g., language arts and mathematics). A parallel goal for DLI programs is to support development in both the target and mainstream languages. This study measured achievement in these areas using the following research questions.

Research Questions

1. How does enrollment in a DLI program predict English language development growth among EBs?
2. How does enrollment in a DLI program predict English language arts proficiency score differences among EB students in Grades 3-9?
3. How does enrollment in a DLI program predict English language arts proficiency growth among EB students in Grades 3-9?
4. How does enrollment in a DLI program predict mathematics proficiency score differences among EB students in Grades 3-9?
5. How does enrollment in a DLI program predict mathematics proficiency growth among EB students in Grades 3-9?
6. How does enrollment in a DLI program predict GPAs among EB students in Grades 6-10?

Research Design

To answer the proposed questions, this study applied different regression analyses to explore the influence of DLI participation on native-Spanish speaking EB student outcomes as identified in the research questions above. Multiple regression techniques are often used in educational policy studies that seek to compare the effects of treatments, in this case DLI instruction versus more traditional ESL services, when controlling for multiple background variables that influence outcomes (Cohen et al., 2003). Another benefit of multiple regression analysis is its ability to describe the shape of the relationships between the dependent versus the independent variables (Cohen et al., 2003). In determining the shape of a relationship, a researcher can determine if the relationship is linear or curvilinear and, thus, better predict the influence that a independent variable has on the dependent variable. This study explored the magnitude and direction of the influence of participation in a DLI program on academic achievement measures through ordinary least squares (OLS) multiple regression, ordinal logistic regression (OLR), and multilevel modeling (MLM) or (i.e., hierarchical linear modeling, HLM; Raudenbush & Bryk, 2002).

Participants

For this study, a rural Utah school district that has implemented DLI Spanish in all of its elementary schools provided student data for analysis. Given its extensive implementation of DLI programming, the rural Utah district provided an excellent case to explore impacts of the Utah model on EB student outcomes as well as adding to the larger body of research on bilingual education. The EB and former EB-student population for the selected district exceeded 10% of the total 7,300 students enrolled, Prekindergarten through Grade 12. Participants for this study were EB students enrolled in a DLI program and EB students who received traditional EL services. These two categories of EB students afforded a comparison of outcomes based on DLI enrollment in order to understand whether the DLI program benefitted EB students in comparison to their peers not participating in DLI programming.

As with all Utah DLI programs, parents in this district were given a choice to enroll in the DLI program. In the selected school district, all parents needed to declare their desire to enroll their students into the DLI program at the time of initial enrollment. A lottery was instituted whenever requests exceeded available slots. To include sufficient students of the target language, native-Spanish-speaking students were given preference for DLI placement in kindergarten and first grade. Therefore, all native-Spanish speaking students who initially requested a spot were enrolled in DLI. After first grade, native-Spanish speaking students were placed in DLI classrooms if slots were available, and the students were deemed similarly proficient in speaking and reading of Spanish through an informal assessment conducted by the Spanish-speaking DLI teacher.

The native-Spanish speaking EB students not enrolled in DLI received traditional ESL pullout services. These services included receiving one-on-one and small-group instruction in a separate setting from the students' main class. Non-DLI EB students received, on average, 20 minutes of ELD instruction daily. ELD included explicit instruction focused on English grammar, vocabulary, and communication skills based on the students' needs. Because students were not randomly assigned to the DLI program or traditional ESL program, the sample served as a convenience sample.

Measurements

Schools annually administer a number of assessments including summative, formative, and interim measures. Summative assessments, like end-of-level state administered tests, serve schools, districts, and policymakers in evaluating the performance of programs (Perie et al., 2007). Therefore, this study made use of such summative assessments. Extant summative data of academic achievement and growth collected annually by the LEA was obtained from the selected district to answer the proposed research questions.

In addition to summative assessments, another common school metric used to measure academic success of students, individual GPAs, were utilized to indicate general academic achievement. The following section will review the types of assessments and metrics used within the state of Utah to measure English acquisition, English language arts and mathematics achievement, and academic success.

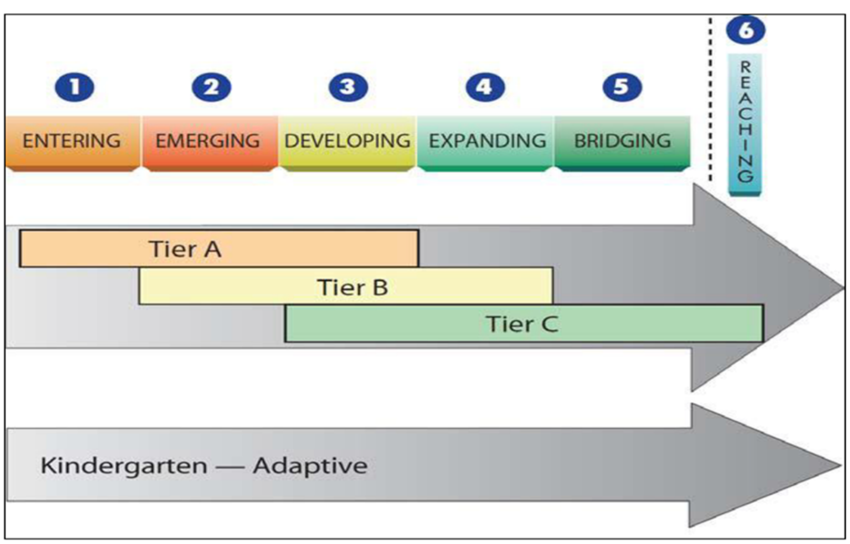
Assessing Comprehension and Communication in English State to State for English Language Learners

The Assessing Comprehension and Communication in English State to State (ACCESS) for ELLs 2.0, developed by the World-Class Instructional Design and Assessment (WIDA) Consortium, assesses proficiency of ELD standards (WIDA, 2019). EB students are assessed in four domains: listening, reading, writing, and speaking. These domains are assessed along five English language proficiency standards: Social and Instructional Language, Language of Language Arts, Language of Math, Language of Science, and Language of Social Science (Kenyon et al., 2011). The test results indicate six levels of proficiency: (1) Entering, (2) Emerging, (3) Developing, (4) Expanding, (5) Bridging, and (6) Reaching (WIDA, 2019). Proficiency levels are reported using a two-digit decimal number, where the first digit represents the overall level, from one to six, and the decimal represents the “proportion of the range between cut scores that the student’s scale score represents” (WIDA, 2019, p. 20). For example, a score of 3.6 means that an EL student is at level 3 or “Developing” and is more than half of the way from the scaled cut score of three towards the scaled cut score for a level 4—"Expanding.” In addition to proficiency levels, results are reported as scale scores from 100 to 600 for each domain and an overall composite score. Scale scores "form an interval scale and are continuous across grades from Kindergarten to Grade 12" (WIDA, 2019, p. 20).

Starting in 2014, Utah began administering the ACCESS for ELLs screener assessment to identify students needing EL services and the annual assessment to measure EB student progress. To ensure that students receive test items appropriate for their level of ELD, WIDA has established three tiers (A, B, or C) or forms of the WIDA ACCESS available for EB students. Figure 3.1 depicts how the tiers A, B, and C overlap and correspond to the ELD levels. Thus, EB students receive either Form A, B, or C, based on their previous years' score. For students who have not received a previous ACCESS score, a short screener is administered to assist the test administrator in determining the most appropriate form for the student.

Figure 3.1

Tier Structure for ACCESS for ELLs



Note. A graphic depicting the tier structure for ACCESS for ELLs (WIDA, 2019).

WIDA (2019) reported high reliability of the overall composite score of their ACCESS measure across grade-level clusters using a stratified Cronbach alpha with Kindergarten, $\alpha = .973$; Grade 1, $\alpha = .934$; Grade 2, $\alpha =.944$; Grade 3, $\alpha = .932$; Grades 4-5, $\alpha = .940$; Grades 6-8, $\alpha = .944$; and Grades 9-12, $\alpha = .949$ (p. iv). Reliability of the overall composite score is important when determining EL student proficiency from year to year. EB students have to achieve a proficiency level of five, or Bridging, before they are deemed to have reached a level of proficiency to no longer be eligible to receive English language services.

To ensure validity of assessment for items that require a human scorer, scoring directors and/or team leaders established high inter-rater reliability by recalibrating scoring sets frequently during the first week of scoring and then intermittently with all raters that score the writing and speaking portions of the assessment. The Rasch fit statistic was also used to measure construct validity of the WIDA items. The Rasch model accounts for the variability in the difficulty of items when measuring particular constructs (Boone, 2016). Table 3.1 shows a summary of the infit and outfit mean square scores, according to the Rasch model, for each domain of the ACCESS assessment. Infit mean square scores determine the validity of those items that correspond to a test taker's general level of ability by assessing variability in responses for items intended for respondents at their expected levels. In contrast, the outfit mean squares statistic is more influenced by outlier observations where students at lower ELD levels for some reason score proficient on a more difficult item (WIDA, 2019). By reviewing both infit and outfit scores, the test's items can be deemed appropriate as the mean square more closely approaches a score of 1. Linacre (2002) defines a productive means square score to be between .5 and 1.5. Linacre also highlights that infit deviations are more crucial to a measurement's validity than outfit. As such, Table 3 shows how all domains across all grade-level clusters fall within the .5 to 1.5 acceptable range. Thus, the WIDA ACCESS for ELLs provides a valid instrument for measuring the acquisition of English.

Table 3.1

WIDA Task Analysis (Rasch Fit Statistic)

	Listening		Reading		Writing A		Writing B/C		Speaking	
Grade(s)	Average infit mean square	Average outfit mean square	Average infit mean square	Average outfit mean square	Average infit mean square	Average outfit mean square	Average infit mean square	Average outfit mean square	Average infit mean square	Average outfit mean square
1	0.98	0.98	0.99	0.98	0.69	0.86	0.57	0.57	0.75	0.64
2-3	0.98	1.03	0.97	0.97	0.41	0.40	0.54	0.53	0.73	0.62
4-5	0.98	1.02	0.99	0.99	0.39	0.41	0.59	0.59	0.66	0.52
6-8	0.98	0.98	0.99	0.99	0.46	0.49	0.54	0.52	0.57	0.41
9-12	0.97	0.96	0.98	0.98	0.58	0.61	0.76	0.74	0.57	0.43

Note. Adapted from (WIDA, 2019).

Measurements of English Language Arts and Mathematics Achievement

Analyses targeting ELA achievement as an outcome measure utilized scores from Utah's end-of-level Student Assessment of Growth and Excellence (SAGE) test for school years 2015 through 2018 and the Readiness Improvement Success Empowerment (RISE) assessment for school year 2019. The SAGE assessment was first administered in 2014. The assessment was computer-adaptive and aligned with the Utah Core Standards, which were influenced by the Common Core State Standards (Jacobsen, 2016; Utah Education Association, n.d.). In 2018, the Utah State Board of Education changed the vendor that delivered Utah's end-of-level assessments for grades 3-8 (Knowles, 2018). Where SAGE assessed students in Grades 3-11, RISE only assessed students in Grades 3-8 though it maintained a similar format in that it was computer adaptive. Data used for this study included SAGE and RISE results from 2015-2019. Analyses were conducted for English language arts and mathematics outcomes for Grades 3-9, as the first cohort of DLI students had Grade 9 SAGE assessment results.

Knowles (2018) pointed out that both SAGE and RISE assessments have maintained item bank questions and used the "same performance level descriptors and similar scale scores" (para. 6) for students in English Language Arts (ELA), Mathematics (MA), and Science (SC). Because the assessment changed in name and delivery, but little in content and format, they will be referred to here as SAGE/RISE. Across both assessments, test questions target the Utah Core Standards in content area via multiple item types such as drag and drop, drop-down, point and click, and student-generated responses. All test items went through a review to confirm appropriateness, coverage of content standards, clarity, freedom from bias, and proper difficulty (Questar, 2019).

Table 3.2 offers evidence of the reliability and validity of the SAGE/RISE assessment for ELA and MA through marginal reliability and correlation statistics. A marginal reliability score measures the overall reliability of the assessment by accounting for ability ranges of students. Pearson correlations are used to measure the internal validity of items within a test by measuring the relationship of the subtests that make up the overall ELA and MA assessments. Table 3.2 lists the lowest and highest correlation for each test by grade level to provide a general view of the internal validity of

SAGE/RISE. The subtests for ELA measured Reading Literature, Reading Informational Text, Listening Comprehension, and Language. The subtests for MA measured for Grades 3-5: Operations and Algebraic Thinking, Numbers and Operations of Base Ten, Number and Operations-Fractions, and Measurement and Data and Geometry; Grade 6-7, Ratios and Proportional Relationships, The Number System, Expressions and Equations, Geometry and Statistics and Probability; and Grade 8, Expressions and Equations, Functions, Geometry and The Number System, and Statistics and Probability.

Table 3.2

2019 SAGE/RISE Reliability and Validity for English Language Arts and Mathematics

	Marginal Reliability		Pearson Correlation Ranges			
	ELA	MA	ELA		MA	
Grade			Low	High	Low	High
3	0.90	0.91	0.55042	0.63794	0.5756	0.66102
4	0.90	0.93	0.50864	0.70154	0.60542	0.73465
5	0.90	0.93	0.58133	0.69767	0.64438	0.72198
6	0.92	0.94	0.58804	0.72977	0.66952	0.80803
7	0.91	0.95	0.56825	0.73174	0.75171	0.81538
8	0.90	0.94	0.58395	0.6844	0.78331	0.83277

Note. Adapted from (Questar, 2019)

SAGE/RISE results include proficiency scores and scale scores. Scale scores provide a way to compare scores across versions of the test (Tan & Michel, 2011). Additionally, Utah vertically aligned scale scores for both the English language arts and math assessments (Utah Education Association, n.d.). Therefore, SAGE/RISE scale scores provide a continuous variable across grades to strengthen statistical analysis.

Grade Point Average

In this study, GPA was used as a measure of academic achievement, similar to related studies (Lindholm-Leary & Borsato, 2006) and a proxy for academic success (Boutakidis et al., 2014; Dickinson & Adelson, 2016). Although not a perfect measure of academic content knowledge or performance, GPA assumes additional academic success elements such as motivational effects, levels of students' organizational skills, and teachers' subjective judgements of student achievement (Dickinson & Adelson, 2016). In the Utah DLI model, students progressing in the DLI program at the secondary level take, at most, two classes each year in Spanish. Therefore, GPA scores reflect a general overall academic performance metric, rather than being overly influenced by participation in DLI. In the selected district, GPA is figured on a four-point scale (0-4) and is calculated for students in Grades 6-12. The GPA measurement used for this study is a composite end-of-year GPA.

Data Collection

The selected district requires all research conducted within the school district to be approved by the Director of Research and Evaluation. As outlined in the district expectations, all personally identifiable information was omitted from the data file provided by the district. A randomly generated student ID was requested to mark an individual student's academic data for logistical purposes. The key for the student IDs was maintained by the school district and not provided to the researcher. Additionally, data results are reported in the aggregate so as to ensure individual student anonymity.

EB student data for WIDA ACCESS for ELLs, SAGE/RISE summary data, and composite GPA score data were requested for the 2014-2020 school years. EB student data included all EB students who were eligible to receive ELL services, as outlined by the district, and attending school during the years identified. Data were provided upon Utah State University IRB and district approval.

The original student file received from the district identified EB students that at some time during their school career in the school district had received EL services either through DLI or from traditional EL services (N = 1,064). Some issues with the district's enrollment system made identifying EB students from Spanish-speaking homes difficult. The registration system defaults to English as the home language if a parent does not specifically choose a home language. Some EB students that appear to have received EL services or had been identified EB at one time by taking the WIDA ACCESS test had English labeled as the home language, which would normally not trigger a language assessment. Thus, some students listed with a home language of English, but with a Hispanic ethnic code have been included in the study. Some EB students identified in the sample came from households where languages other than Spanish were spoken (n = 15). These and three EB students from Asian descent with a home language listed as English were removed from the sample. Thus, 18 EB students were excluded from any analyses to better analyze the 1,046 EB students who identified as EB students with Spanish-language backgrounds.

Of the 1,064 identified EB students, 787 enrolled in DLI. Of the enrolled DLI students, 52 had sporadic or incomplete enrollments. These 52 participants were included if they were enrolled in a DLI classroom for at least four years and began the program in kindergarten through second grade. Consequently, 17 of the 52 students identified with incomplete enrollment were included in the analyses. The total number of students with available data and the identified number of years of DLI enrollment was 752. The outcomes for this group of DLI EB students were compared with those of the 312 EB students not enrolled in the DLI program.

Data Analysis

An MLM statistical technique was employed to answer Research Questions 1, 3, and 5. Each question sought to understand the influence of DLI versus traditional English-language services on the growth trajectory of English acquisition, English language arts achievement, and math achievement respectively. MLM offers several advantages when looking at growth of individual students across two different groups. MLM allows analysis of within-person and between-person changes simultaneously (Singer & Willett, 2003). This is accomplished by layering different models over one another to account for variables nested within one another. Specifically, this investigation sought any interaction between the Level-1 dependent variable (WIDA Score) and the Level-2 independent variable (DLI status). A list of variables used in this study with their different levels is presented in Table 3.3. Gender, socioeconomic status, and special education status were all used as control variables in addition to DLI as the main differentiating factor of focus in this study. Essentially, the “level-1 submodel...describes how each person changed over time and, a level-2 model...described how these changes differ across people" (Singer & Willett, 2003, p. 45). Adding schools for a Level-3 model was considered for the analysis but due to the low number of schools there was a danger of overfitting the model.

Table 3.3

Dependent and Independent Variables for Each Level of the Multi-Level Modeling Model for English Language Acquisition

Hierarchical Level	Example of Hierarchical Level	Example Variables	Variable Type
Level-2	Student Level	Gender	Dichotomous
		Socioeconomic status	Dichotomous
		Special Education status	Dichotomous
		DLI status	Dichotomous
Level-1	Time Level	Composite WIDA scale score[a]	Continuous
		Time/Event(s)	Continuous

Note. All variables listed are independent unless otherwise specified.
[a] Dependent variable.

Importantly, student WIDA ACCESS for ELL scores are nested within individual students while MLM controls for student factors such as gender, SES, and EL status. Each of the control variables were coded as Y for yes, N for No and for gender F for female and M for male. In conducting the regression analyses, the categories of N were used and M for gender as the reference category. Because the focus of the analyses was on growth trajectory, WIDA ACCESS scale scores were used in the analyses without centering as results for growth did not need centering for interpretation.

Similar to Research Question 1, the other research questions that used MLM for analysis (Research Questions 3 and 5) nested English language arts scores and math scores, respectively, within students while controlling for student variables as outlined in Tables 3.4 and 3.5. One of the major advantages to using MLM is its ability to explore growth trajectory including the handling unbalanced data sets (Barkaoui, 2014). Even

Table 3.4

Dependent and Independent Variables for Each Level of the Multi-Level Modeling Model for English Language Arts Proficiency

Hierarchical Level	Example of Hierarchical Level	Example Variables	Variable Type
Level-2	Student Level	Gender	Dichotomous
		Socioeconomic status	Dichotomous
		Special Education status	Dichotomous
		DLI status	Dichotomous
Level-1	Time Level	SAGE/RISE English language arts scale score[a]	Continuous
		Time/Event(s)	Continuous

Note. All variables listed are independent unless otherwise specified.
[a] Dependent variable.

Table 3.5

Dependent and Independent Variables for Each Level of the Multi-Level Modeling Model for Mathematics Proficiency

Hierarchical Level	Example of Hierarchical Level	Example Variables	Variable Type
Level-2	Student Level	Gender	Dichotomous
		Socioeconomic status	Dichotomous
		Special Education status	Dichotomous
		DLI status	Dichotomous
Level-1	Time Level	SAGE/RISE Mathematics scale score[a]	Continuous
		Time/Event(s)	Continuous

Note. All variables listed are independent unless otherwise specified.
[a] Dependent variable.

though the number of EB students differs by program status, MLM allows for the inclusion of all data. Thus, MLM allows for more accurate predictions of actual growth. Likewise, MLM allows for missing data points that arise because of mobility factors, clerical errors, etc. Students with even one data point were included in the data set, which provided for better estimates (Peugh, 2010).

For Research Questions 1, 3, and 5, this study explored the interaction of the Level-1 composite scores changing over time with enrollment in DLI at Level-2. Often an effect size is used to describe the magnitude of difference among groups. However, with MLM this statistic can be problematic (Nezlek, 2012). In many cases, the statistic to use to better explain the effect of the difference is the pseudo R^2, similar to R^2, which provides an explanation of the amount of variance attributed to the variable (Aguinis et al., 2013). The type and strength of interaction highlighted how DLI status formed the shape and trajectory of academic growth by enrollment group.

Research Questions 2 and 4 sought to identify the probability of EB students scoring highly proficient on their state administered, end-of-level assessment in English language arts and mathematics. The ordinal logistic regression analyses applied to the data determined if the odds of students achieving scores deemed highly proficient for ELA and MA on the SAGE and RISE assessment were statistically different, based on participation in DLI. Like MLM, the analyses controlled for student variables, including gender, SES, and special education status. Unlike MLM, students without sufficient data were excluded from these analyses. Because logistic regression does not look at growth of student outcomes over time, a comparison was made year to year by cohort, providing a general picture of student achievement comparing DLI participants and their non-participating EB peers.

To answer Research Question 6, using GPA as a measure for student academic success, OLS regression was applied to determine the influence of DLI participation on EB GPA scores while controlling for student-level variables similar to previously discussed analyses. Comparing GPA scores of EBs provided insight into how participation in DLI covaried with scholastic engagement. Like regression analyses other than MLM, students with missing data were not included. The regression analysis of GPAs explored each student's data from Grade 6-10 for the 2019 school year.

CHAPTER 4

RESULTS

Carefully planned analyses were employed to ascertain the predictive effect of DLI enrollment on student achievement in English language acquisition, language arts, mathematics, and grade point average. Each of these areas of student achievement is a snapshot in the academic career of a student in Kindergarten through Grade 12. As noted in the methods chapter, the regression analyses were conducted using data collected from a Utah school district's EB students over a period of 7 years (2014 to 2020). Some analyses involved data from all 7 years while some data represent fewer years based on the sample collected from the school district. To quantify the predictive effect of DLI on these measures of student academic achievement, all analyses conducted were forms of regression analyses including multilevel, ordinal logistic, and multiple linear regression.

The analyses presented in this chapter follow the numerical progression of the questions proposed in this study, which mirrors the progression of academic achievement outcomes that EB students typically experience throughout their educational careers, from early elementary to high school. EB students start with an English language acquisition assessment as they enter school for the first time, whether in kindergarten or their first year within the U.S. Question 1 of this study was addressed using Utah's assessment of English language acquisition, the WIDA ACCESS, as the outcome measure to identify differences in English-acquisition growth based on enrollment in the district's DLI program.

In Utah, students take their first end-of-level assessments for English language arts and mathematics in Grade 3. Question 2 addresses the differences, by DLI enrollment, in academic achievement levels based on a 4-point scale outcome for English language arts on Utah's SAGE and RISE assessments. The analyses for Question 2 used an ordinal logistic regression analysis of individual Grades 3-9 to determine if DLI enrollment predicted a difference in odds of achievement. Question 2 considered grade-level odds of achieving a 4 on the SAGE and RISE assessments in English language arts. Question 3 addressed the difference in academic growth of students separated by enrollment in DLI across time as measured by scale scores across grade levels. Questions 4 and 5 mirror Questions 2 and 3 but focus on mathematics achievement outcomes. Questions 4 concerns the odds of scoring a 4 on the SAGE and RISE mathematics assessment, and Question 5 asks if DLI enrollment influences mathematics growth across time using scale scores to measure student mathematics achievement growth from Grades 3-9. Finally, Question 7 targeted grade point average (GPA) as a measure of student achievement. As a proxy for school academic engagement, GPA was analyzed using a regression analysis to determine how DLI influences GPA outcomes in the Grade 6-10.

Longitudinal Multilevel Analysis of English Acquisition

The first question of this study asked how enrollment in DLI predicted English language acquisition outcomes. WIDA ACCESS scores for each EB were compiled by grade level for the analysis. Tables 4.1 and 4.2 display the number of observations by variable for each grade level, as well as the mean and standard deviation of the WIDA ACCESS scale scores separated by DLI enrollment.

Table 4.1

Emergent Bilingual Student Demographic and WIDA ACCESS Scale Score Distribution for Grades K-5

Variable	Kindergarten				Grade 1				Grade 2				Grade 3				Grade 4				Grade 5			
	Y		N		Y		N		Y		N		Y		N		Y		N		Y		N	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
IEP																								
Y	32	8	0	0	41	8.4	1	10	44	9.1	7	20.6	40	9.4	15	22.1	39	11.2	21	28.4	29	11.6	29	35.8
N	368	92	5	100	446	91.6	9	90	441	90.9	27	79.4	385	90.6	53	77.9	309	88.8	53	71.6	220	88.4	52	64.2
SES																								
Y	358	89.5	2	40	439	90.1	4	40	441	90.9	24	70.6	392	92.2	53	77.9	320	92	59	79.7	227	91.2	68	84
N	42	10.5	3	60	48	9.9	6	60	44	9.1	10	29.4	33	7.8	15	22.1	28	8	15	20.3	22	8.8	13	16
Gender																								
F	205	51.2	4	80	239	49.1	6	60	232	47.8	16	47.1	201	47.3	32	47.1	151	43.4	31	41.9	107	43	36	44.4
M	195	48.8	1	20	248	50.9	4	40	253	52.2	18	52.9	224	52.7	36	52.9	197	56.6	43	58.1	142	57	45	55.6
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
WIDA ACCESS Scale Score	211.3	50.7	219	69.6	272.7	22.7	270.1	34.5	296.5	25.9	306	24.7	318.7	30.4	335	26.3	345.9	30.2	345.7	24.6	356.7	30.2	357	25.4

Table 4.2

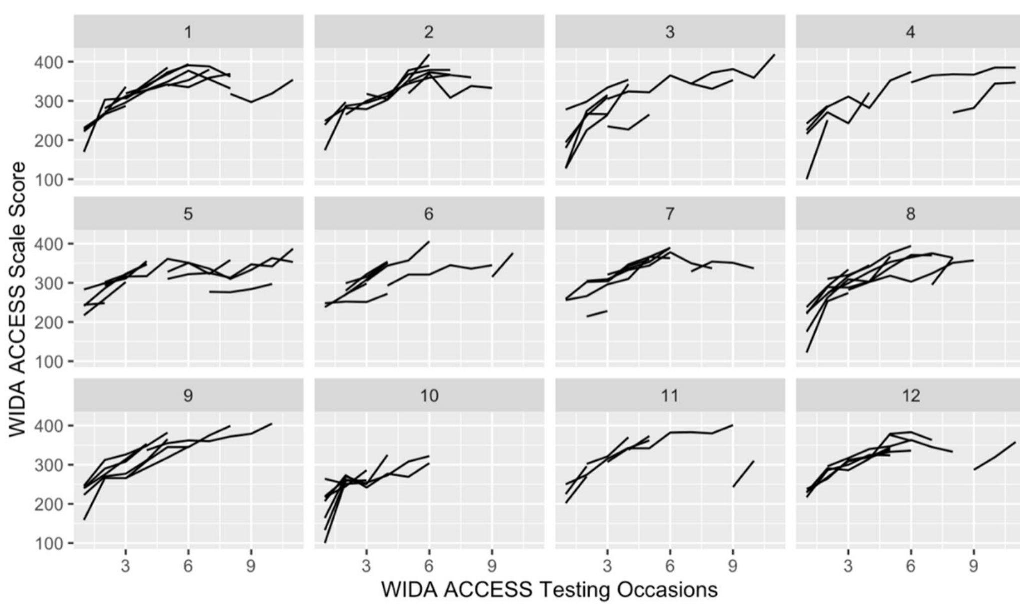
Emergent Bilingual Student Demographic and WIDA ACCESS Scale Score Distribution for Grades 6-10

Variable	Grade 6				Grade 7				Grade 8				Grade 9				Grade 10			
	Y		N		Y		N		Y		N		Y		N		Y		N	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
IEP																				
Y	24	13.9	29	42	16	15	30	42.9	9	18	26	37.1	5	13.5	11	22.4	6	25	9	25
N	149	86.1	40	58	91	85	40	57.1	41	82	44	62.9	32	86.5	38	77.6	18	75	27	75
SES																				
Y	158	91.3	59	85.5	101	94.4	60	85.7	49	98	64	91.4	36	97.3	44	89.8	24	100	33	91.7
N	15	8.7	10	14.5	6	5.6	10	14.3	1	2	6	8.6	1	2.7	5	10.2	0	0	3	8.3
Gender																				
F	68	39.3	25	36.2	37	34.6	31	44.3	18	36	28	40	15	40.5	20	40.8	10	41.7	16	44.4
M	105	60.7	44	63.8	70	65.4	39	55.7	32	64	42	60	22	59.5	29	59.2	14	58.3	20	55.6
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
WIDA ACCESS Scale Score	344.4	30.4	344	28.9	345.6	34.8	340.4	39.4	349.9	32.9	343	40.2	372.9	24.2	358	39.2	378.3	32.8	351.8	43

To explore this research question, this study used a longitudinal multilevel model. Using an MLM approach allowed for the exploration of English acquisition growth over time using scaled scores across grade levels. (As mentioned previously, WIDA ACCESS scale scores allow for growth comparisons across grades.) Individual student growth trajectories vary considerably in intercepts and slopes. Figure 4.1 provides a visual representation of the varying growth curves present within the sample. Each square in Figure 4.1 provides a random sampling of students from Kindergarten–Grade 12 represented along the x-axis and the WIDA ACCESS scale scores along the y-axis. The figure shows that the majority of students experienced a positive increase in growth across the grade levels, though the slopes differed individually, Additionally, some of the students had only a few years of data due to transiency or exiting from EL services that triggers cessation of WIDA ACCESS testing. Figure 4.1 illustrates the advantages of

Figure 4.1

WIDA ACCESS Scale Score Growth Sample



using MLM analyses due to the variability of individual growth trajectories of EB student WIDA ACCESS outcomes over time and the issue of missing or inconsistent data points. Accounting for missing data points is one of the advantages of running a MLM analysis.

Multilevel Model Analysis of WIDA ACCESS Scale Scores

To better understand the EB students' ELD, regression models were built using the lme4 package (Bates et al., 2015) in R. Each model displayed in Table 4.1 shows that the analysis proceeded in an iterative process of adding predictor variables to each successive model starting with the unconditional mean model (M0). Fixed factors included DLI status (dli), grade level (wida_grade), receiving special education services (iep), eligibility for free and reduced lunch (ses), and gender (genderc). Wida_grade was added as a quadratic factor and as a random variable to allow for varying slopes over grade levels. The M0 provided the basis upon which to judge the explanatory power of each model. The M0 is written as an equation below:

$$\mathrm{Y}_{ti} = \beta_{0i} + \beta_{1i}T + \varepsilon_{ti}$$

explains where Y equals the predicted WIDA ACCESS scale score for student *i* at testing occasion *t*. β_{0i} is the grand mean of all WIDA ACCESS scale scores. β_{1i} is the coefficient for student *i* at *T*, which is the repeated measure—grade level, with ε_{ti} being the unexplained error.

Each model was fitted using a sample of 878 EB students across Kindergarten-Grade 10 from 2014 to 2020. The unconditional model had an intercept of Y = 305.18, $p < .001$, which represents the mean score of all scale scores. With DLI (dli) enrollment as the factor of interest, the first model (M1) had dli as a fixed variable and ID or student as the level-2 factor. Variables were added to M1 to build successive models. After adding dli to M0, the time variable, wida_grade was added as a fixed variable to create the second model which accounted for the level-1 repeated measure. Table 4.3 displays how each succeeding model was created with fixed and random factors. An interaction term was introduced for models 4 (M4) and 5 (M5).

Table 4.3

Model Comparison of WIDA ACCESS Scale Scores

Models	
wida_M0	wida_scale_score ~ 1 + (1 | ID)
wida_M1	wida_scale_score ~ dli + (1 | ID)
wida_M2	wida_scale_score ~ dli + wida_grade + (1 | ID)
wida_M3	wida_scale_score ~ dli + wida_grade + iep + ses + genderc + (1 + wida_grade | ID)
wida_M4	wida_scale_score ~ wida_grade * dli + I(wida_grade^2) * dli + iep + ses + genderc + (1 | ID)
wida_M5	wida_scale_score ~ 1 + wida_grade * dli + I(wida_grade^2) * dli + iep + ses + genderc + (I(wida_grade) | ID)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
wida_M0	3	36089	36107	-18041	36083			
wida_M1	4	35962	35987	-17977	35954	128.45	1	< 2.2e-16 ***
wida_M2	5	33555	33586	-16773	33545	2408.83	1	< 2.2e-16 ***
wida_M3	10	32644	32705	-16312	32624	921.79	5	< 2.2e-16 ***
wida_M4	11	32163	32230	-16071	32141	482.41	1	< 2.2e-16 ***
wida_M5	13	32005	32084	-15989	31979	162.53	2	< 2.2e-16 ***

Note. ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

M5 indicated the best fit with a log likelihood of -15989 and $X^2(2, N = 878) = 408.95, p < .001$. The model was significantly improved over M4 when grade level was allowed to vary by individual in the second level explaining the amount of variance as a random effect. M5 had an ICC = .76 and a conditional $R^2 = .89$. Thus, M5 was able to explain 89% of the variance in WIDA ACCESS scale score outcomes with its defined fixed and random effects. Table 4.4 depicts the significance of all variables associated

Table 4.4

Model Comparison for Repeated Measures Multilevel Analysis of WIDA-ACCESS Outcomes

	Model 0			Model 1			Model 2		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	305.18	302.26 – 308.10	<0.001*	297.76	294.77 – 300.76	<0.001*	227.79	224.29 – 231.29	<0.001*
dli [N]				41.85	34.86 – 48.84	<0.001*	-11.38	-17.86 – -4.89	0.001*
wida_grade							18.63	18.04 – 19.22	<0.001*
Random Effects									
σ^2	2053.32			2054.36			835.9		
τ_{00}	1290.44 $_{ID}$			1022.91 $_{ID}$			1099.64 $_{ID}$		
τ_{11}									
ρ_{01}									
ICC	0.39			0.33			0.57		
Marginal R^2 / Conditional R^2	0.000 / 0.386			0.074 / 0.382			0.517 / 0.791		

	Model 3			Model 4			Model 5		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	197.38	190.30 – 204.46	<0.001*	151.52	144.18 – 158.86	<0.001*	148.9	141.42 – 156.38	<0.001*
dli [N]	10.65	5.46 – 15.84	<0.001*	59.6	41.71 – 77.50	<0.001*	78.4	58.06 – 98.74	<0.001*
wida_grade	23.47	22.33 – 24.61	<0.001*	45.62	44.07 – 47.16	<0.001*	46.92	45.09 – 48.76	<0.001*
iep [N]	27.2	21.63 – 32.76	<0.001*	28.09	21.80 – 34.38	<0.001*	29.02	23.21 – 34.84	<0.001*
ses [N]	5.42	-0.35 – 11.19	0.066	8.11	1.72 – 14.51	0.013*	6.93	0.89 – 12.98	0.025*
genderc [M]	0.89	-2.59 – 4.37	0.616	0.6	-3.33 – 4.53	0.765	0.02	-3.67 – 3.71	0.991
wida_grade^2				-2.7	-2.86 – -2.54	<0.001*	-2.8	-3.00 – -2.59	<0.001*
wida_grade * dli [N]				-18.86	-24.39 – -13.33	<0.001*	-23.88	-30.09 – -17.67	<0.001*
dli [N] * wida_grade^2				1.26	0.84 – 1.67	<0.001*	1.58	1.11 – 2.06	<0.001*
Random Effects									
σ^2	470.68			559.79			463.21		
τ_{00}	3078.47 $_{ID}$			674.63 $_{ID}$			1464.87 $_{ID}$		
τ_{11}	205.30 $_{ID.wida\_grade}$						40.82 $_{ID.I(wida\_grade)}$		
ρ_{01}	-0.94 $_{ID}$						-0.82 $_{ID}$		
ICC	0.79			0.55			0.76		
Marginal R^2 / Conditional R^2	0.617 / 0.920			0.644 / 0.839			0.547 / 0.891		

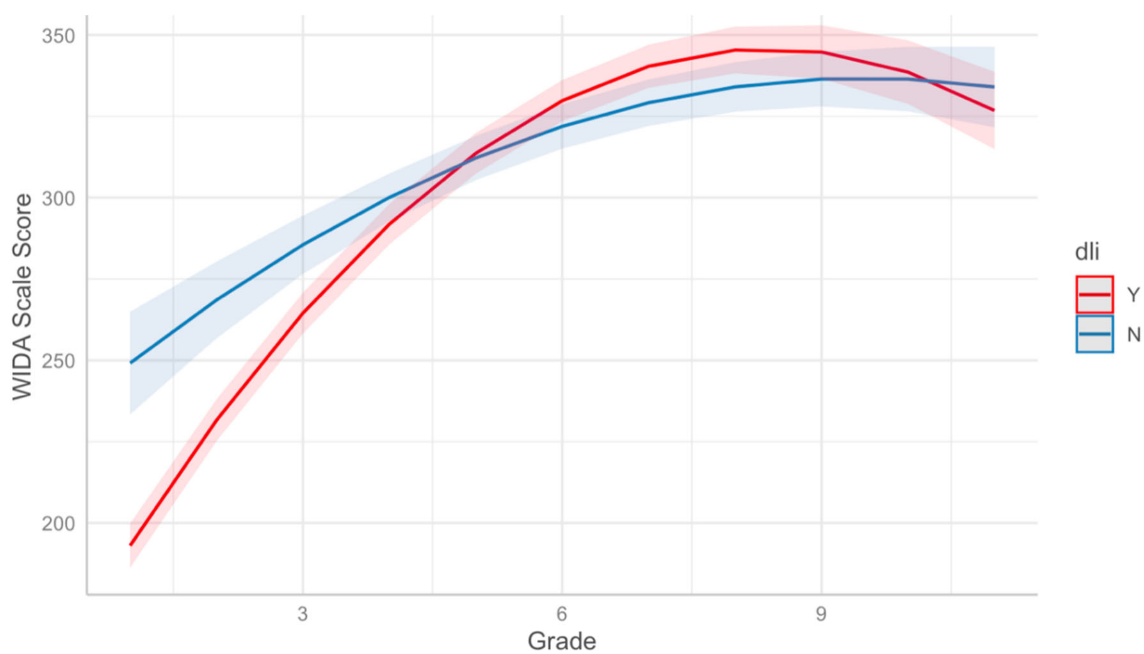
Note. N = 878. CI = confidence interval ; dli [N] = students not enrolled in DLI; iep [N] = students not enrolled in special education; ses[N] = students not receiveing free or reduced lunch; genderc[M] = male.

with this model. All variables in M5 were significant except gender with dli[N]=78.4, $p< .001$; wida_grade= 46.92, $p< .001$; iep[N]= 29.02, $p< .001$; ses[N]= 6.93, $p< .001$, wida_grade^2= -2.8, $p< .001$; wida_grade*dli= -23.88, $p< .001$; and dli[N]*wid_grade^2= 1.58, $p< .001$. There was a significant interaction between DLI status and grade level, indicating that DLI enrollment moderates the growth curve of English acquisition over time.

The interaction between DLI status and time or grade levels explains how DLI status moderated the scale score value over time. The interaction depicted in Figure 4.2 demonstrates how students enrolled in DLI had lower WIDA ACCESS scale scores up until Grade 4, then surpassed their non-DLI peers until Grade 10 when the mean scores converged. On average, EB Kindergarten students enrolled in DLI were 50 scale-score

Figure 4.2

WIDA ACCESS Predicted Scale Score



points lower than their non-DLI peers, yet by Grade 6 DLI students surpassed their peers in English language acquisition. Growth trajectories of DLI students were steeper than their non DLI peers and the leveling of scale scores was delayed compared to their non-DLI peers.

Summary of Multilevel Model Analysis of WIDA ACCESS Outcomes

The results of the longitudinal multilevel analysis of WIDA ACCESS outcomes suggested that DLI enrollment is associated with a steeper growth curve in English acquisition outcomes over time delineated by grade levels. When accounting for SES, special education enrollment, and gender, DLI enrollment had a significant positive effect on the growth curve of EB students. EB students enrolled in DLI experienced a benefit in English language acquisition compared to their non-DLI peers over time. On average, although students enrolled in DLI initially tested below their peers on English-language assessment, by Grade 6 EB students enrolled in DLI surpassed their non-DLI peers in their English language proficiency development. As Figure 4.2 illustrated, DLI student scale scores exceeded their non-DLI peers' scale scores during the intermediate grades.

Ordinal Logistic Regression Analysis of Language Arts Outcomes

To answer Research Question 2 about whether enrollment in DLI predicts language arts achievement, the study used ordinal logistic regression analyses to ascertain the likelihood of EB students attaining a certain proficiency based on enrollment in DLI. Student SAGE/RISE scores for Grades 3-9 were collected from 2014-2019. Each grade-level analysis contains the scores of EB students across multiple cohorts of students. Ordinal logistic regression assumes that the dependent variable is ordinal and that the independent variables are ordinal, categorical, or continuous. As noted previously, SAGE/RISE proficiency data for language arts are reported on a 1-4 scale with 1 indicating below proficient, 2 indicating approaching proficient, 3 indicating proficient, and 4 highly proficient. The dichotomous independent variables used for the regression analysis included DLI enrollment (dli), socio-economic status based on eligibility for free and reduced lunch (ses), enrollment in special education (iep), and gender (genderc). Salient findings for each control variable as well as the DLI enrollment are shared below.

Each grade level (Grades 3-9) was analyzed using the MASS package (Venables & Ripley, 2002) in R. Included in the analyses were the Variance Inflation Factor (VIF) to test for both multicollinearity and the Wald statistic with the Brant test (Brant, 1990) for proportional odds (also called the parallel odds assumption). Any VIF score above 5.0 signals a possible problem with multicollinearity. Additionally, for the Brant test, any p-value less than .05 could indicate a violation of the parallel odds assumption within that model or any of the variables tested (Lee, 2019). When running the Brant test for the available grade level data, a warning indicated that some issues were present with the calculation of the parallel odds, possibly caused by a skewed distribution of scores across proficiency levels for all grade levels within the sample. Thus, caution should be taken in interpretating the results of the ordinal logistic regression presented below.

EB enrollment showed that more students were enrolled in DLI than not. Table 4.5 depicts the number of students by elementary grade level and by independent variable

Table 4.5

Elementary Language Arts Proficiency Score Counts by Grade and Demographics (2014-2019)

Proficiency	DLI (Y)												DLI (N)											
	Gender				IEP				SES				Gender				IEP				SES			
	F		M		Y		N		Y		N		F		M		Y		N		Y		N	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Grade 3																								
1	98	29.5	121	36.4	33	9.9	186	56.0	205	61.7	14	4.2	10	20.8	20	41.7	8	16.7	22	45.8	21	43.8	9	18.8
2	30	9.0	44	13.3	2	0.6	72	21.7	65	19.6	9	2.7	5	10.4	5	10.4	0	-	10	20.8	9	18.8	1	2.1
3	20	6.0	16	4.8	0	-	36	10.8	30	9.0	6	1.8	7	14.6	1	2.1	0	-	8	16.7	7	14.6	1	2.1
4	2	0.6	1	0.3	0	-	3	0.9	3	0.9	0	-	0	0.0	0	-	0	-	0	-	0	-	0	-
Grade 4																								
1	91	31.3	115	39.5	22	7.6	184	63.2	190	65.3	16	5.5	24	25.8	34	36.6	20	21.5	38	40.9	44	47.3	14	15.1
2	30	10.3	35	12.0	2	0.7	63	21.6	57	19.6	8	2.7	14	15.1	10	10.8	0	-	24	25.8	18	19.4	6	6.5
3	5	1.7	11	3.8	0	-	16	5.5	11	3.8	5	1.7	6	6.5	3	3.2	0	-	9	9.7	9	9.7	0	-
4	2	0.7	2	0.7	0	-	4	1.4	3	1.0	1	0.3	2	2.2	0	-	0	-	2	2.2	0	-	2	2.2
Grade 5																								
1	73	31.3	99	42.5	19	8.2	153	65.7	158	67.8	14	6.0	33	24.6	53	39.6	21	15.7	65	48.5	67	50.0	19	14.2
2	18	7.7	25	10.7	0	-	43	18.5	35	15.0	8	3.4	14	10.4	14	10.4	1	0.7	27	20.1	24	17.9	4	3.0
3	5	2.1	7	3.0	0	-	12	5.2	10	4.3	2	0.9	14	10.4	5	3.7	0	-	19	14.2	13	9.7	6	4.5
4	2	0.9	4	1.7	0	-	6	2.6	4	1.7	2	0.9	1	0.7	0	-	0	-	1	0.7	1	0.7	0	-

Note. Grade 3 N = 380. Grade 4 N= 384. Grade 5 N= 367.

group that scored below proficient (1), approaching proficient (2), proficient (3), or highly proficient (4) ratings on the language arts assessment. Subgroup numbers and percentages for control variables of IEP, SES, and gender are listed in Tables 4.5 and 4.6 by DLI enrollment. Additionally, Tables 4.5 and 4.6 illustrate how percentages of EB students scoring at level 3 or 4 on the language arts assessment indicate that the district struggled in assisting both DLI and non-DLI students in attaining proficiency across grade levels.

Given the data, many of the EB elementary and secondary students fell in the below and approaching proficient categories for English language arts outcomes. Another way to visually depict the distribution of the scores is in the histogram in Figure 4.3. Each grade level is skewed right with most proficiency scores falling in the below and approaching proficient categories. In general, it appeared that the school district had not been successful in having the majority of EB students reach proficiency on the language arts SAGE and RISE assessments regardless of DLI enrollment.

Tables 4.5 and 4.6 and Figure 4.3 illustrate the issue with a skewed sample distribution of scores which may have contributed to an error warning when running the ordinal logistic regression analysis for each grade. As noted earlier, due to the warning and skewed distribution, great care should be taken in the interpretation of the results of this regression analysis. Conducting an ordinal logistic regression with the achievement data received may not have been the most effective in answering the research question.

Table 4.6

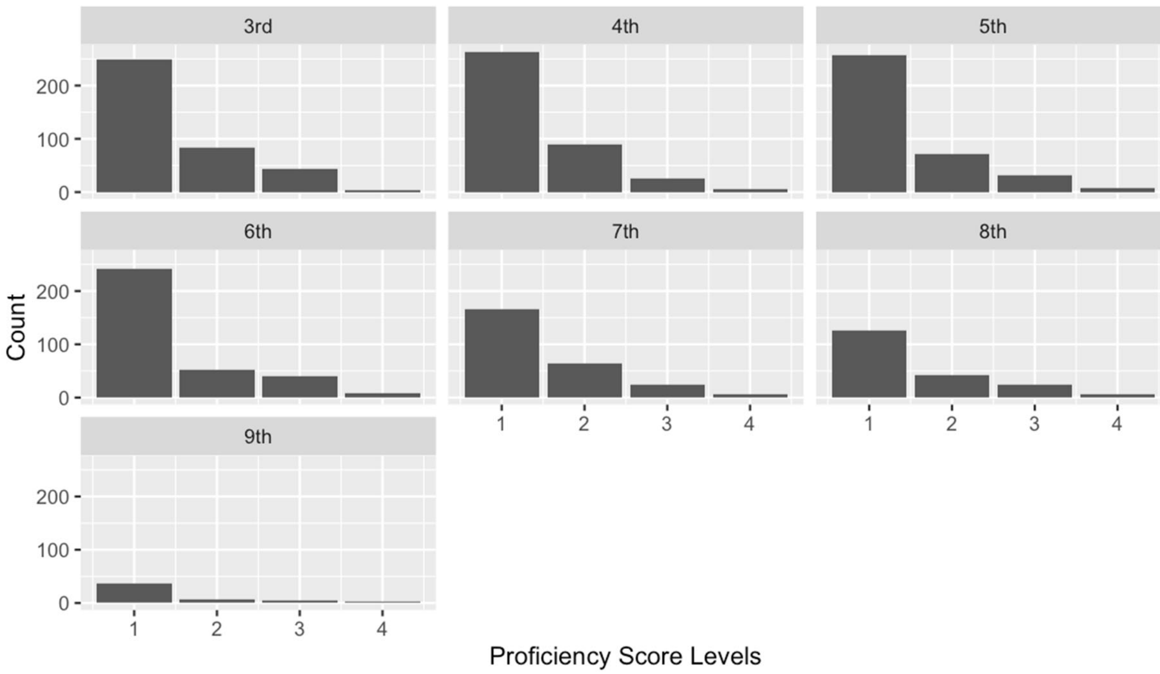
Secondary Language Arts Proficiency Score Counts by Grade and Demographics (2014-2019)

Proficiency	DLI (Y)												DLI (N)											
	Gender				IEP				SES				Gender				IEP				SES			
	F		M		Y		N		Y		N		F		M		Y		N		Y		N	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Grade 6																								
1	55	30.2	79	43.4	16	8.8	118	64.8	125	68.7	9	4.9	38	23.6	69	42.9	28	17.4	79	49.1	84	52.2	23	14.3
2	19	10.4	9	4.9	0	-	28	15.4	23	12.6	5	2.7	14	8.7	10	6.2	0	-	24	14.9	19	11.8	5	3.1
3	6	3.3	9	4.9	0	-	15	8.2	12	6.6	3	1.6	20	12.4	6	3.7	0	-	26	16.1	21	13.0	5	3.1
4	1	0.5	4	2.2	0	-	5	2.7	4	2.2	1	0.5	2	1.2	2	1.2	0	-	4	2.5	4	2.5	0	-
Grade 7																								
1	26	25.7	36	35.6	8	7.9	54	53.5	57	56.4	5	5.0	39	24.2	64	39.8	28	17.4	75	46.6	84	52.2	19	11.8
2	13	12.9	16	15.8	0	-	29	28.7	26	25.7	3	3.0	20	12.4	16	9.9	1	0.6	35	21.7	28	17.4	8	5.0
3	2	2.0	4	4.0	0	-	6	5.9	5	5.0	1	1.0	15	9.3	4	2.5	0	-	19	11.8	14	8.7	5	3.1
4	1	1.0	3	3.0	0	-	4	4.0	1	1.0	3	3.0	2	1.2	1	0.6	0	-	3	1.9	3	1.9	0	-
Grade 8																								
1	23	29.1	28	35.4	6	7.6	45	57.0	48	60.8	3	3.8	30	24.8	44	36.4	16	13.2	58	47.9	60	49.6	14	11.6
2	7	8.9	11	13.9	0	-	18	22.8	15	19.0	3	3.8	15	12.4	10	8.3	0	0.0	25	20.7	14	11.6	11	9.1
3	5	6.3	4	5.1	0	-	9	11.4	7	8.9	2	2.5	11	9.1	5	4.1	1	0.8	15	12.4	15	12.4	1	0.8
4	1	1.3	0	-	0	-	1	1.3	0	-	1	1.3	4	3.3	2	1.7	0	-	6	5.0	3	2.5	3	2.5
Grade 9																								
1	5	29.4	6	35.3	0	-	11	64.7	10	58.8	1	5.9	14	41.2	12	35.3	5	14.7	21	61.8	21	61.8	5	14.7
2	4	23.5	1	5.9	0	-	5	29.4	3	17.6	2	11.8	2	5.9	0	-	0	-	2	5.9	2	5.9	0	-
3	0	-	1	5.9	0	-	1	5.9	1	5.9	0	-	2	5.9	2	5.9	0	-	4	11.8	4	11.8	0	-
4	0	-	0	-	0	-	0	-	0	-	0	-	1	2.9	1	2.9	0	-	2	5.9	1	2.9	1	2.9

Note. Grade N = 343. Grade 7 N= 262. Grade 8 N= 200. Grade 9 N= 51

Figure 4.3

Language Arts Proficiency by Grade Level



Language Arts Proficiency for Special Education Emergent Bilingual Students

The results of the analyses across all grades indicated that one of the controlled variables, eligibility for special education services (IEP), held statistical significance across all grades ($p < .05$). However, this predictor variable also proved problematic for Grade 3 in violating the Proportional Odd (PO) assumption required for model fit with a Brant test indicating $p < .05$. This was not an issue for the subsequent grade levels, although Grade 6 and Grade 9 data indicated some issues with calculating the estimates and CI. The IEP variable for students not participating in special education had significant estimates for Grade 4, *OR* = 12.79, 95% CI [3.76, 80.3], $p < .01$; Grade 5, *OR* = 21.3, [4.47, 382.32], $p < .01$; Grade 7, *OR* = 24.58, [5.10, 441.94], $p < .01$; and Grade

8, *OR* = 14.22, [2.81, 259.66], $p = .012$. As expected, IEP status proved to be a significant factor in predicting language arts outcomes. Thus, EB students receiving special education services demonstrated a more difficult time than their peers in reaching higher levels of achievement on the SAGE and RISE assessments in language arts, regardless of DLI enrollment.

Language Arts Proficiency by Gender for Emergent Bilingual Students

As a control variable, gender proved significant at Grade 6, *OR*= .48, 95% CI [.29, .77], $p < .01$; Grade 7, *OR* = .57, [.34, .95], $p = .03$; and Grade 8, *OR* = .52, [.29, .92], $p = .03$. The *OR* or odds ratio indicated that males had 52% lower odds in Grade 6, 43% lower odds in Grade 7, and 48% lower odds than their female peers of scoring a level-4 achievement score for all EB students. The findings that female EBs performed better than their male peers confirm previous research results of gender differences among Spanish-speaking EB students (see Lapayese et al., 2014)

Language Arts Proficiency by Dual Language Immersion Enrollment

The predictor variable of focus, DLI status, was only significant in Grade 4, *OR* = 1.75, [1.04, 2.93], $p = .02$; Grade 5, *OR* = 1.79, [1.12, 2.87], $p = .02$; and Grade 6, *OR* = 1.72, [1.06, 2.8], $p = .03$. These results indicate that non-DLI students had 75% higher odds of achieving a level 4 score in Grade 4 than their DLI peers and 79% higher odds in Grade 5 and 72% higher odds in Grade 6 of achieving a level 4 score. Across grade levels, DLI status did not remain a significant predictor of language arts achievement as measured by SAGE and RISE assessments. Additionally, when estimates were significant, the odds ratio showed that non-DLI students had higher odds of attaining a highly proficient score than their peers. DLI students not performing as well as their peers in the lower grades seems to follow the growth analysis of WIDA ACCESS, in that DLI students did not surpass their peers until late elementary. If there is a possible link to WIDA ACCESS and student achievement it would follow logically that WIDA ACCESS success would precede SAGE and RISE language arts achievement.

Results for each ordinal logistic regression model are displayed in Table 4.7 for elementary grades and Table 4.8 for secondary grades. Pseudo R^2 values for the Grade 3-9 models range from .083 to .147 suggesting a low to moderate explanatory power of the variance in the English language outcome explained by the predictor variables (Cohen, et.al., 2003). Grade 9 had the lowest pseudo $R^2 = .08$ and Grade 7 the highest at pseudo $R^2 = .15$. Although pseudo R^2 values trended low for most of the models, each model was markedly improved by including all the variables as opposed to just including DLI as the variable of interest.

Summary of Ordinal Logistic Regression Analysis of Language Arts Outcomes

The results do not demonstrate a consistent trend of the influence of DLI status on English language arts achievement outcomes. Significant estimates for DLI enrollment in Grades 4-6 suggest that non-DLI participants were more likely than their DLI peers to receive a level 4 score for language arts achievement. As proposed earlier, lower odds of high achievement on language arts assessment may be correlated with lower results in

Table 4.7

Elementary Grades Language Arts Achievement Ordinal Logistic Regression Analysis

	Grade 3				Grade 4				Grade 5			
Predictors	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*
genderc [M]	0.8	0.52 – 1.23	0.309	1.019821	0.91	0.59 – 1.42	0.675	1.028463	0.78	0.50 – 1.24	0.292	1.017885
iep [N]	12.58	3.76 – 78.22	0.001*	1.023706	12.79	3.76 – 80.30	0.001*	1.062255	21.3	4.47 – 382.32	0.003*	1.042113
ses [N]	1.39	0.70 – 2.67	0.335	1.030848	1.61	0.86 – 2.96	0.134	1.041103	1.51	0.82 – 2.72	0.182	1.029854
dli [N]	1.29	0.67 – 2.43	0.434	1.029895	1.75	1.04 – 2.93	0.035*	1.074519	1.79	1.12 – 2.87	0.016*	1.043801
1|2	19.22				27.8				49.74			
2|3	75.23				153.68				195.19			
3|4	1367.5				867.65				1198.82			
Observations		380				384				367		
R^2Nagelkerke		0.085				0.094				0.106		

Note. *$p < .05$

Table 4.8

Secondary Grades Language Arts Achievement Ordinal Logistic Regression Analysis

	Grade 6				Grade 7				Grade 8				Grade 9			
Predictors	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*
genderc [M]	0.48	0.29 – 0.77	0.003*	1.015982	0.57	0.34 – 0.95	0.034*	1.017567	0.52	0.29 – 0.92	0.028*	1.002941	0.77	0.20 – 2.76	0.691	1.021714
iep [N]	25070910.93		<0.001*	1.044954	24.58	5.10 – 441.94	0.002*	1.065398	14.22	2.81 – 259.66	0.012*	1.050396	72111954.11		<0.001*	1.093587
ses [N]	1.13	0.59 – 2.10	0.705	1.034986	1.37	0.70 – 2.64	0.35	1.042986	1.69	0.84 – 3.37	0.138	1.065054	1.08	0.20 – 4.81	0.924	1.034435
dli [N]	1.72	1.06 – 2.80	0.028*	1.045234	1.05	0.62 – 1.78	0.861	1.043348	1.21	0.67 – 2.23	0.528	1.044501	0.92	0.26 – 3.35	0.903	1.061501
1|2	44475334.98				26.83				18				144276202			
2|3	120048461.8				125.24				61.52				352464353.8			
3|4	800773002.2				656.25				335.98				1380918874			
Observations	343				262				200				51			
R^2Nagelkerke	0.159				0.147				0.129				0.083			

Note. *p< .05

growth in English acquisition in the early elementary grades. Significant effects of DLI were not found in Grade 3 or Grades 7-9. Importantly, skewed results likely contributed to error warnings when conducting the possible violation of the parallel odds assumption placing the interpretation of the overall results in jeopardy. It is important to note that the explanatory power of each model was low and at best moderate, with a range of R^2 values ranging from .083 to .147 (Cohen et al., 2003). Thus, the models created did not account for much of the variance in language arts outcomes among EB students.

Longitudinal Multilevel Analysis of Language Arts

In contrast to the OLR analysis conducted, Question 3 targeted how DLI enrollment predicted English language arts growth among EB students. Observations for the analysis included assessment data from 686 EB participants across Grades 3-9 with the largest number of observations amounting to 2,660 for Grade 3 and the smallest number of responses, 357, for Grade 9. Tables 4.9 and 4.10 show the number of observations used in the analysis by variable, grade level, and DLI enrollment. Additionally, both tables indicate the mean and distribution of language arts scale scores on the SAGE and RISE assessment by grade level. The data indicate that mean scale scores are similar for many of the grades.

Though mean scores were comparable, students' individual slopes and intercepts can vary; thus, a longitudinal mixed model approach was used to identify those factors that may have influenced students' language arts outcomes. Figure 4.4 depicts individual

Table 4.9

Emergent Bilingual Student Demographic and Language Arts Scale Score Distribution for Grades 3-5

Variable	Grade 3				Grade 4				Grade 5			
DLI	Y		N		Y		N		Y		N	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
IEP												
Y	245	10.5	56	16.7	168	8.2	140	21.5	133	8.2	154	16.4
N	2079	89.5	280	83.3	1869	91.8	511	78.5	1498	91.8	784	83.6
SES												
Y	2121	91.3	259	77.1	1827	89.7	497	76.3	1449	88.8	735	78.4
N	203	8.7	77	22.9	210	10.3	154	23.7	182	11.2	203	21.6
Gender												
F	1050	45.2	154	45.8	896	44	322	49.5	686	42.1	434	46.3
M	1274	54.8	182	54.2	1141	56	329	50.5	945	57.9	504	53.7
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
SAGE/RISE LA Scale Score	268.6	57.6	268.2	61.2	290.5	61.1	298	66.7	325.1	64.2	339	64.4

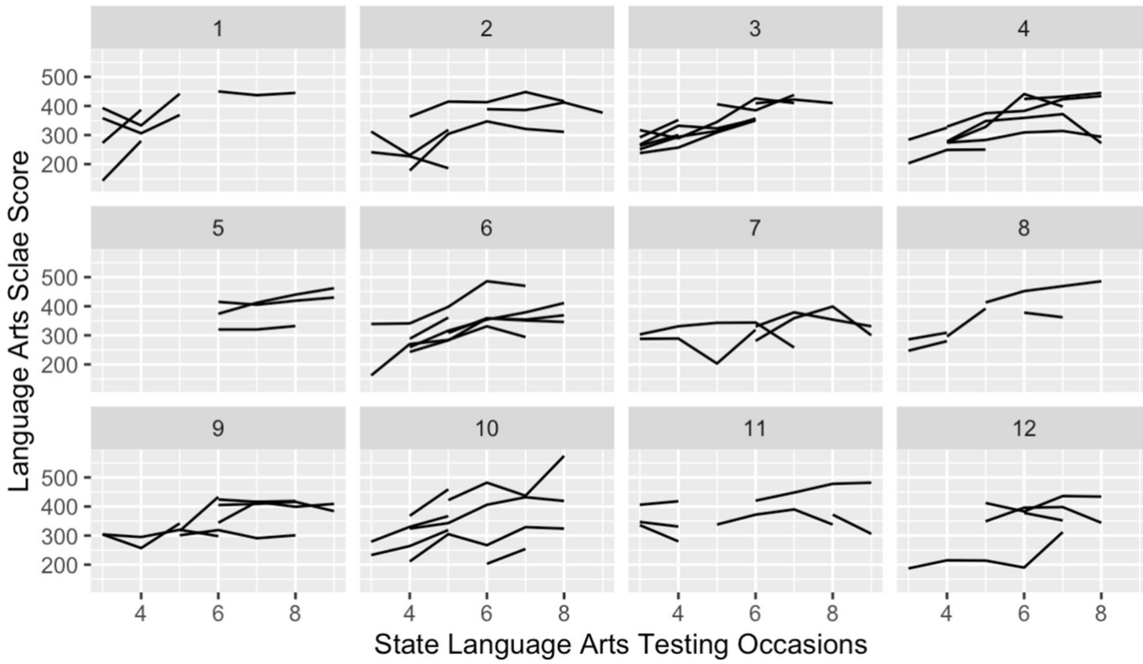
Table 4.10

Emergent Bilingual Student Demographic and Language Arts Scale Score Distribution for Grades 6-9

Variable	Grade 6				Grade 7				Grade 8				Grade 9			
DLI	Y		N		Y		N		Y		N		Y		N	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
IEP																
Y	112	8.8	196	17.4	56	7.9	203	18	42	7.6	119	14	0	0	35	14.7
N	1162	91.2	931	82.6	651	92.1	924	82	511	92.4	728	86	119	100	203	85.3
SES																
Y	1148	90.1	896	79.5	623	88.1	903	80.1	490	88.6	644	76	98	82.4	196	82.4
N	126	9.9	231	20.5	84	11.9	224	19.9	63	11.4	203	24	21	17.6	42	17.6
Gender																
F	567	44.5	518	46	294	41.6	532	47.2	252	45.6	420	49.6	63	52.9	133	55.9
M	707	55.5	609	54	413	58.4	595	52.8	301	54.4	427	50.4	56	47.1	105	44.1
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
SAGE/RISE LA Scale Score	351.4	67.7	370.2	63.8	375.2	68.5	375.3	69.9	390.7	65.8	399.4	76.4	412.9	55.8	376.1	94.5

Figure 4.4

Language Arts Scale Score Growth Sample



language arts growth curves for samplings of EB students. For clarity, a random sample of the 686 participants' data has been displayed across 12 charts displaying SAGE and RISE scales scores from Grade 3 through Grade 9. Grades are represented on the x-axis, and SAGE and RISE scale scores fall on the y-axis. Evident in Figure 4.4 are the varying intercepts and slopes embodied in this reasonably large sample. The sample growth curves also illustrate the need for a mixed method approach, as some of the samples have missing observations across Grades 3-9. As mentioned previously, a mixed method approach accounts for such missing observations (Cohen et al., 2003).

The lme4 package within R (Bates et al., 2015) was used to assess patterns regarding performance and DLI status. Table 4.11 shows a comparison of the models fitted for the RISE/SAGE Language Arts outcomes as a scale score for EB students

Table 4.11

Model Comparison for Repeated Measures Multilevel Analysis of Language Arts

	Model 0			Model 1			Model 2		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	321.99	316.74 – 327.25	<0.001	308.04	301.90 – 314.18	<0.001	197.92	189.67 – 206.17	<0.001
dli [N]				43.17	32.46 – 53.88	<0.001	9.77	-0.09 – 19.63	0.052
grade							24.19	22.62 – 25.76	<0.001
Random Effects									
σ^2		2312.91			2311.05			1157.47	
τ_{00}		3923.28 $_{ID}$			3519.90 $_{ID}$			3013.31 $_{ID}$	
τ_{11}								77.66 $_{ID.grade}$	
ρ_{01}								-0.40 $_{ID}$	
ICC		0.63			0.6			0.74	
Marginal R^2 / Conditional R^2		0.000 / 0.629			0.070 / 0.631			0.292 / 0.819	

	Model 3			Model 4			Model 5		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	141.77	126.53 – 157.01	<0.001	81.96	56.33 – 107.59	<0.001	81.26	55.72 – 106.81	<0.001
dli [N]	12.75	3.46 – 22.04	0.007	-61.2	-103.05 – -19.34	0.004	-66.62	-109.17 – -24.08	0.002
grade	23.99	22.43 – 25.55	<0.001	48.34	39.64 – 57.04	<0.001	48.66	39.65 – 57.67	<0.001
iep [N]	62.87	49.78 – 75.95	<0.001	61.28	47.83 – 74.73	<0.001	62.06	49.00 – 75.11	<0.001
ses [N]	23.37	11.04 – 35.70	<0.001	25.34	12.81 – 37.86	<0.001	23.41	11.11 – 35.72	<0.001
genderc [M]	-4.61	-12.85 – 3.64	0.274	-5.16	-13.63 – 3.32	0.233	-4.83	-13.06 – 3.40	0.25
grade^2				-2.12	-2.93 – -1.30	<0.001	-2.17	-3.05 – -1.29	<0.001
grade * dli [N]				26.43	11.58 – 41.27	<0.001	28.79	13.49 – 44.08	<0.001
dli [N] * grade^2				-2.39	-3.67 – -1.10	<0.001	-2.6	-3.96 – -1.24	<0.001
Random Effects									
σ^2		1161.27			1190.11			1056.1	
τ_{00}		2506.20 $_{ID}$			2625.22 $_{ID}$			2070.60 $_{ID}$	
τ_{11}		78.32 $_{ID.grade}$						0.61 $_{ID.I(grade\char`^2)}$	
ρ_{01}		-0.44 $_{ID}$						0.01 $_{ID}$	
ICC		0.71			0.69			0.73	
Marginal R^2 / Conditional R^2		0.378 / 0.820			0.390 / 0.810			0.378 / 0.833	

Note. N = 686. CI = confidence interval ; dli [N] = students not enrolled in DLI; iep [N] = students not enrolled in special education; ses[N] = students not receiveing free or reduced lunch; genderc[M] = male.

taking the end-of-level assessment between 2015 and 2019. Each model includes Grades 3–9. The unconditional means model (M0; the "null" model) can be written as:

$$Y_{ti} = \beta_{0i} + \beta_{1i}T + \varepsilon_{ti}$$

where Y is the student outcome of language arts scale score for individual i at time measurement t. β_{0i} is the expected estimate of language arts scale score at the first measurement time for the student i, β_{1i} is the estimate of repeated measure T at the occasion of testing, whereas ε_{ti} is the measurement of residual error.

Multilevel Regression Models for Language Arts Outcomes

The unconditional means model conducted for the language arts scale scores included data on 686 student participants, with 1,987 observations from Grades 3-9. The model indicates a log likelihood of -11094.7, which was used to compare the fit of the subsequent models and to verify if the subsequent models better explained the variance. In addition, the unconditional model confirmed the need to conduct a multilevel analysis with an intraclass correlation (ICC) of .63, indicating that, as expected, the clustering of students' individual scores explains 63% of the variance and therefore merited this type of analyses. The random effects results of the unconditional means model, or Model 0, also suggested that there was significant variability among participants at an estimate of 62.64 with a lower limit of 58.62 and an upper limit of 66.92. Table 4.11 shows that model M0 had an estimated grand mean of 321.99 for all students measured as a scale score with a significance level of $p < .001$. Although multilevel models can take a "top-down" approach, where all possible fixed explanatory variables are placed in the first model and then each successive model removes non-significant effects, this study took the more common "bottom-up" approach, where each successive model adds additional explanatory variables to determine the best model fit (Hox et al., 2018, p. 43). As shown in Table 4.11, model five (M5) which includes the random and fixed effects explains 83% of the variance in the scaled scores of SAGE and RISE Language arts outcomes.

M5 is significantly improved over the other models listed in Table 4.12, based on likelihood ratio tests, over the preceding models, X^2 (2, N = 686) = 57.49, p < .001, with the lowest deviance score of 20,959 and the lowest log likelihood score of -10,479. After controlling for the intra-individual parameter of enrollment in DLI and demographic variables IEP, SES, and gender, M5 adds the random effect of grade in its quadratic form.

Table 4.12

Model Comparison of Language Arts Scale Scores

Models	
M0	la_scale_score ~ 1 + (1 | ID)
M1	la_scale_score ~ dli + (1 | ID)
M2	la_scale_score ~ dli + grade + (1 + grade | ID)
M3	la_scale_score ~ dli + grade + iep + ses + genderc + (1 + grade | ID)
M4	la_scale_score ~ 1 + grade * dli + I(grade^2) * dli + iep + ses + genderc + (1 | ID)
M5	la_scale_score ~ 1 + grade * dli + I(grade^2) * dli + iep + ses + genderc + (I(grade^2) | ID)

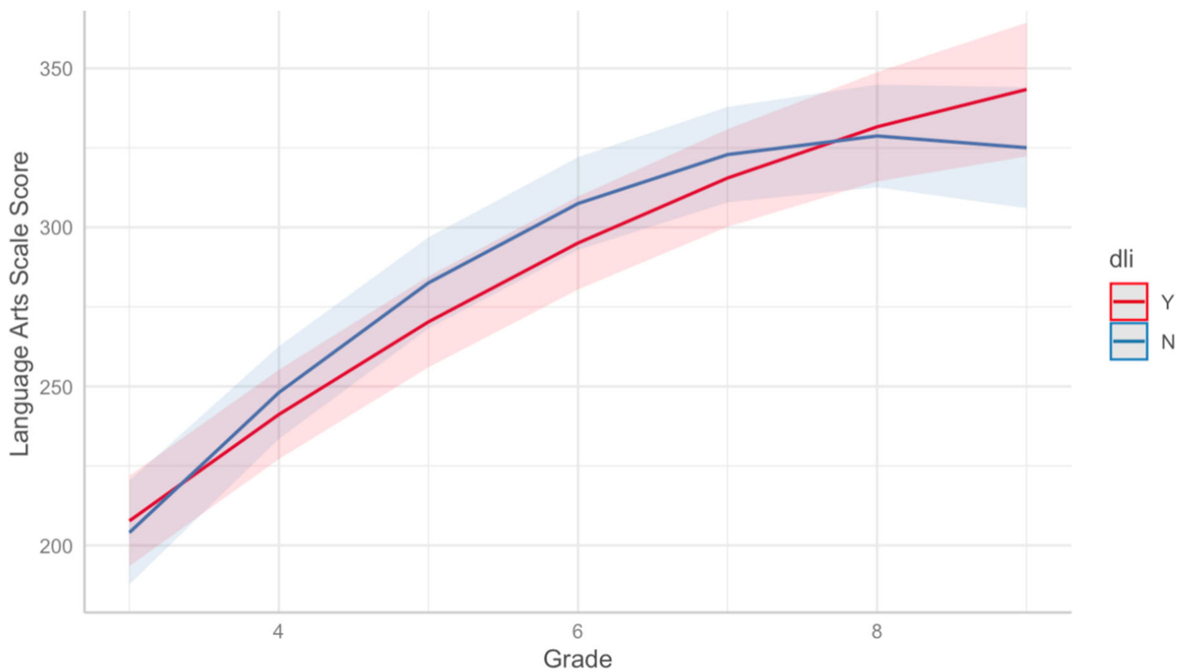
	npar	AIC	BIC	loglik	deviance	Chisq	Df	Pr(Chisq)
M0	3	22196	22212	-11095	22190			
M1	4	22138	22160	-11065	22130	59.774	1	1.06e-14 ***
M2	7	21198	21237	-10592	21184	946.014	3	< 2.2e-16 ***
M3	10	21098	21154	-10539	21078	105.334	3	< 2.2e-16 ***
M4	11	21038	21100	-10508	21016	62.276	1	2.985e-15 ***
M5	13	20985	21057	-10479	20959	57.497	2	3.272e-13 ***

Note. Repeated measures multilevel modelling was used for this analysis.
p ≤ .10. *p< .05. **p< .01. ***p<.001

To represent the interaction of DLI and the curvilinear growth of students in language arts outcomes, Figure 4.5 demonstrates visually that while DLI and non-DLI EB students scored similarly on the language arts measure in Grade 3, DLI students then lagged behind their peers in scale score outcomes until Grade 8 when they surpassed their peers in language arts achievement growth. Concomitantly, growth among non-DLI students in this sampling appeared to level off starting in Grade 7. The interaction demonstrated in Figure 4.5 is a graphical depiction of how enrollment in DLI significantly ($p < .001$) moderated the growth trajectories of students' language arts scale scores.

Figure 4.5

Language Arts Predicted Scale Score



Summary of Multilevel Model Analysis of Language Arts Outcomes

The results of the MLM analysis of language arts outcomes demonstrated that DLI enrollment was, indeed, a predictor of language arts outcomes. The final model explained 83% of the variance and included significant factors such as eligibility for free and reduced lunch, enrollment in special education services, and enrollment in DLI. Gender was not a significant predictor of language arts outcomes for EBs in this analysis.

The time factor was revealed a significant interaction between DLI enrollment and grade level and illustrated how EBs enrolled in DLI lagged behind their peers in the lower grades but eventually surpassed their peers in language arts growth. Interestingly, growth curve slopes for EB DLI students remained consistent without a leveling off like those of their non-DLI peers.

Ordinal Logistic Regression Analysis for Mathematics Outcomes

Question 4 asked if DLI enrollment predicts proficiency in mathematics outcomes as measured by end-of-level assessments in Utah schools. Like the language arts analysis reviewed previously, an OLR analysis was chosen to highlight the possible predictive effects of DLI enrollment on mathematics achievement outcomes while controlling for other predictive factors such as enrollment in special education (IEP), gender (genderc), and eligibility for free and reduced lunch (SES). Outcome measures were SAGE and RISE mathematics proficiency levels measured on a 4-point scale from below proficient to highly proficient. Tables 4.13 and 4.14 provide distribution counts for each of the predictor variables by enrollment in DLI.

Similar to the analyses of proficiency scores for language art outcomes, the majority of EB students in this district achieved a proficiency level of 1 or 2. Tables 4.13 and 4.14 display the number and percentage of students scoring at each proficiency score for SAGE and RISE mathematics assessment by DLI enrollment. Unlike the language arts data, the mathematics proficiency increased in the secondary grades for DLI

Table 4.13

Elementary Mathematics Proficiency Score Counts by Grade and Demographics (2014-2019)

Proficiency	DLI (Y)												DLI (N)											
	Gender				IEP				SES				Gender				IEP				SES			
	F		M		Y		N		Y		N		F		M		Y		N		Y		N	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Grade 3																								
1	89	26.6	97	29.0	34	10.2	152	45.5	172	51.5	14	4.2	13	26.5	21	42.9	9	18.4	25	51.0	27	55.1	7	14.3
2	35	10.5	42	12.6	0	-	77	23.1	69	20.7	8	2.4	3	6.1	5	10.2	0	-	8	16.3	5	10.2	3	6.1
3	19	5.7	30	9.0	1	0.3	48	14.4	44	13.2	5	1.5	4	8.2	0	-	0	-	4	8.2	4	8.2	0	0.0
4	9	2.7	13	3.9	0	-	22	6.6	20	6.0	2	0.6	3	6.1	0	-	0	-	3	6.1	2	4.1	1	2.0
Grade 4																								
1	91	30.8	103	34.9	19	6.4	175	59.3	179	60.7	15	5.1	29	31.5	31	33.7	20	21.7	40	43.5	47	51.1	13	14.1
2	26	8.8	26	8.8	4	1.4	48	16.3	46	15.6	6	2.0	7	7.6	8	8.7	0	-	15	16.3	11	12.0	4	4.3
3	8	2.7	30	10.2	1	0.3	37	12.5	31	10.5	7	2.4	5	5.4	4	4.3	0	-	9	9.8	6	6.5	3	3.3
4	5	1.7	6	2.0	0	-	11	3.7	8	2.7	3	1.0	4	4.3	4	4.3	0	-	8	8.7	6	6.5	2	2.2
Grade 5																								
1	80	33.5	104	43.5	19	7.9	165	69.0	167	69.9	17	7.1	41	30.6	46	34.3	22	16.4	65	48.5	70	52.2	17	12.7
2	13	5.4	17	7.1	0	-	30	12.6	25	10.5	5	2.1	11	8.2	19	14.2	0	-	30	22.4	22	16.4	8	6.0
3	6	2.5	13	5.4	0	-	19	7.9	14	5.9	5	2.1	8	6.0	6	4.5	0	-	14	10.4	11	8.2	3	2.2
4	1	0.4	5	2.1	0	-	6	2.5	5	2.1	1	0.4	2	1.5	1	0.7	0	-	3	2.2	2	1.5	1	0.7

Note. Grade 3, N = 383. Grade 4, N= 387. Grade 5, N= 373.

Table 4.14

Secondary Mathematics Proficiency Score Counts by Grade and Demographics (2014-2019)

Proficiency	DLI (Y)												DLI (N)											
	Gender				IEP				SES				Gender				IEP				SES			
	F		M		Y		N		Y		N		F		M		Y		N		Y		N	
	n	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%	*n*	%
Grade 6																								
1	67	35.8	79	42.2	16	8.6	130	69.5	134	71.7	12	6.4	55	33.7	67	41.1	28	17.2	94	57.7	99	60.7	23	14.1
2	12	6.4	17	9.1	0	-	29	15.5	25	13.4	4	2.1	13	8.0	14	8.6	1	0.6	26	16.0	19	11.7	8	4.9
3	4	2.1	5	2.7	0	-	9	4.8	8	4.3	1	0.5	5	3.1	5	3.1	0	-	10	6.1	9	5.5	1	0.6
4	0	-	3	1.6	0	-	3	1.6	2	1.1	1	0.5	2	1.2	2	1.2	0	-	4	2.5	3	1.8	1	0.6
Grade 7																								
1	28	27.5	35	34.3	8	7.8	55	53.9	58	56.9	5	4.9	47	29.2	62	38.5	27	16.8	82	50.9	89	55.3	20	12.4
2	11	10.8	14	13.7	0	-	25	24.5	20	19.6	5	4.9	17	10.6	16	9.9	1	0.6	32	19.9	27	16.8	6	3.7
3	3	2.9	7	6.9	0	-	10	9.8	10	9.8	0	0.0	10	6.2	7	4.3	0	-	17	10.6	12	7.5	5	3.1
4	1	1.0	3	2.9	0	-	4	3.9	2	2.0	2	2.0	2	1.2	0	-	0	-	2	1.2	2	1.2	0	-
Grade 8																								
1	24	29.6	28	34.6	8	9.9	44	54.3	51	63.0	1	1.2	35	27.6	46	36.2	16	12.6	65	51.2	66	52.0	15	11.8
2	7	8.6	9	11.1	0	-	16	19.8	11	13.6	5	6.2	18	14.2	15	11.8	2	1.6	31	24.4	24	18.9	9	7.1
3	5	6.2	6	7.4	0	-	11	13.6	9	11.1	2	2.5	4	3.1	5	3.9	0	-	9	7.1	6	4.7	3	2.4
4	0	-	2	2.5	0	-	2	2.5	1	1.2	1	1.2	3	2.4	1	0.8	0	-	4	3.1	3	2.4	1	0.8
Grade 9																								
1	3	16.7	4	22.2	0	-	7	38.9	6	33.3	1	5.6	13	39.4	8	24.2	3	9.1	18	54.5	17	51.5	4	12.1
2	5	27.8	4	22.2	0	-	9	50.0	7	38.9	2	11.1	3	9.1	7	21.2	2	6.1	8	24.2	8	24.2	2	6.1
3	1	5.6	1	5.6	0	-	2	11.1	1	5.6	1	5.6	1	3.0	0	-	0	-	1	3.0	1	3.0	0	-
4	0	-	0	0.0	0	-	0	-	0	-	0	-	1	3.0	0	-	0	-	1	3.0	1	3.0	0	-

Note. Grade 6, N = 350. Grade 7, N= 263. Grade 8, N= 208. Grade 9, N= 51.

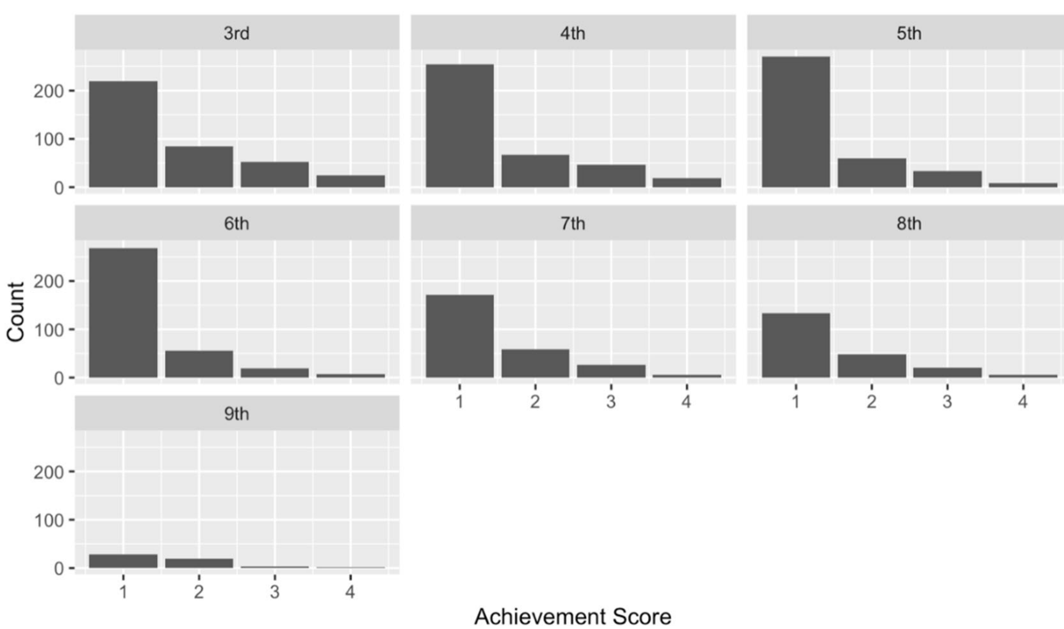
participants. The percentage of DLI EB students performing at a level 3 or 4 exceeded that of the non-DLI students consistently starting in Grade 7.

One of the concerns with running the OLR analysis by grade was the lower sample size, especially in Grade 9, which affects the ability of the OLR analysis to have sufficient data to understand how DLI EB status may affect proficiency. Figure 4.6 describes the overall distribution of achievement levels by grade for all EB students. As with the ELA outcomes, achievement levels for mathematics were skewed toward the lower levels. Thus, with the distribution of scores across achievement levels by independent variable at some grade levels it was problematic to fit all models.

Each model created in the OLR analysis met the assumptions of OLR by having a 4-level ordinal outcome measure of SAGE/RISE proficiency. The assumptions of

Figure 4.6

Mathematics Achievement Historgram by Grade Level



collinearity and parallel odds were tested for each grade-level model using the Variance Inflation Factor (VIF) and the Brant test respectively, as mentioned earlier. The OLR analysis was conducted using the polr function of the MASS package in R. Analysis of each grade level data set included all predictor variables (i.e., dli, ses, iep, gender). Models including all predictors maintained a better fit when comparing R2 statistics; therefore, each grade level analysis kept all predictor variables.

Mathematics Proficiency for Special Education Emergent Bilingual Students

Similar to the OLR analysis of language arts outcomes, IEP as a control held statistical significance ($p < .05$) in most grades. Grade 9 for the mathematics analysis was the one exception ($p = .94$). The model for Grade 3 indicated that IEP violated the proportional odds assumption since the *p* value for IEP Grade 3 was less than the chosen alpha level of $p < .05$ (Lee, 2019). For Grade 3 through Grade 8, IEP was a significant predictor of the SAGE and RISE mathematics outcome except for Grade 5 which had an unstable estimate. With estimates for Grade 4, *OR*= 5.37, 95% CI [2.21,16.11], $p < .01$; Grade 6, *OR*= 17.46, CI [3.66, 313.36], $p < .01$; Grade 7, *OR* = 21.55, CI [4.49, 388.49], $p < .01$; and Grade 8, *OR* = 6.83, CI [1.92, 43.56], $p = .012$, the data evidenced that receiving special education services was correlated with lower odds of attaining a highly proficient score on the SAGE and RISE mathematics assessment regardless of enrollment in DLI.

The only other control variable to show significance was SES in Grade 8, *OR* = 2.16, CI [1.08, 4.29], $p = .03$. Gender as a variable did not return any significant effect for any of the grade levels.

Mathematics Proficiency by Dual Language Immersion Enrollment

Tables 4.15 and 4.16 display the OLR model for each grade level including all control variables and DLI. DLI was only significant in the Grade 5 model, *OR* = 1.91, CI [1.18, 3.09], $p < .01$. The Odds Ratio estimate in the Grade 5 model indicated that non-DLI students had 91% higher odds of scoring as highly proficient than their DLI peers. These results suggest that DLI enrollment did not play a significant role in predicting mathematics student achievement, and for Grade 5 it indicated that those not enrolled in DLI had better odds of achieving at the highest level. These results could be due to the scope and sequence of mathematics instruction within the DLI program as mathematics is taught primarily in Spanish in the lower elementary grades and then is switched to being primarily taught in English from Grade 4 on. However, testing for SAGE and RISE mathematics assessments remains in English across all elementary grades. Perhaps this language-for-content-instruction transition may affect the overall mathematics outcomes seen in this study.

Also noted, the pseudo R^2 values for the grade-level models ranged from .067 to .127, suggesting that all models had low explanatory power. The two strongest grade-level models Grade 3 and Grade 5 reached a pseudo R^2 of .127. Although ultimately reaching low power, the models were improved by retaining all control variables, SES, IEP, and gender with dli as the independent variable of interest.

Table 4.15

Elementary Grades Mathematics Achievement Ordinal Logistic Regression Analysis

	Grade 3				Grade 4				Grade 5			
Predictors	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*
genderc [M]	1.2	0.80 – 1.80	0.381	1.019	1.47	0.96 – 2.26	0.077	1.026	1.39	0.87 – 2.24	0.173	1.018
iep [N]	38.6	8.23 – 689.20	<0.001*	1.025	5.37	2.21 – 16.11	0.001*	1.063	16434045.4		<0.001*	1.043
ses [N]	1.15	0.60 – 2.17	0.671	1.031	1.77	0.98 – 3.14	0.056	1.039	1.6	0.88 – 2.88	0.119	1.028
dli [N]	0.62	0.31 – 1.20	0.166	1.032	1.2	0.71 – 1.99	0.493	1.075	1.91	1.18 – 3.09	0.008*	1.043
1|2	44.83	29.89 – 67.24	<0.001		12.4	8.09 – 19.01	<0.001		61892041.1	3359.45 – 993590	<0.001	
2|3	138.96	18.74 – 1030.29	<0.001		32.9	12.38 – 87.41	<0.001		196087111	5486.60 – 242548	<0.001	
3|4	521.45	274.28 – 991.34	<0.001		135.35	75.51 – 242.60	<0.001		1028423994	5473.60 – 1863689	<0.001	
Observations		383				387				373		
R^2 Nagelkerke		0.127				0.067				0.127		

Note. *$p < .05$

Table 4.16

Secondary Grades Mathematics Achievement Ordinal Logistic Regression Analysis

	Grade 6				Grade 7				Grade 8				Grade 9			
Predictors	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*	*Odds Ratios*	*CI*	*p*	*VIF*
genderc [M]	1.23	0.75 – 2.05	0.414	1.017	0.9	0.54 – 1.51	0.697	1.016	0.89	0.50 – 1.57	0.682	1	1.25	0.41 – 3.81	0.695	1.015
iep [N]	17.46	3.66 – 313.36	0.006*	1.049	21.55	4.49 – 388.49	0.003*	1.059	6.83	1.92 – 43.56	0.012*	1.044	0.92	0.14 – 7.90	0.936	1.098
ses [N]	1.3	0.66 – 2.49	0.437	1.036	1.23	0.63 – 2.37	0.536	1.04	2.16	1.08 – 4.29	0.03*	1.057	1.3	0.31 – 5.16	0.709	1.031
dli [N]	1.36	0.81 – 2.27	0.245	1.049	0.87	0.51 – 1.47	0.603	1.039	0.9	0.50 – 1.63	0.726	1.03	0.38	0.11 – 1.23	0.119	1.068
1|2	64.86	39.09 – 107.63	<0.001		29.73	17.71 – 49.92	0.001		10.59	5.97 – 18.78	0.003		0.73	0.23 – 2.26	0.78	
2|3	256.33	34.19 – 1921.90	<0.001		116.19	15.42 – 875.60	<0.001		44.96	10.12 – 199.75	<0.001		7.79	1.07 – 56.76	0.093	
3|4	1018.85	524.02 – 1980.96	<0.001		728.57	374.27 – 1418.28	<0.001		221.33	110.84 – 441.97	<0.001		33.42	8.09 – 138.00	0.022	
Observations		350				263				208				51		
R^2Nagelkerke		0.078				0.112				0.098				0.067		

Note. $^*p < .05$

Summary of Ordinal Logistic Regression Analysis of Mathematics Outcomes

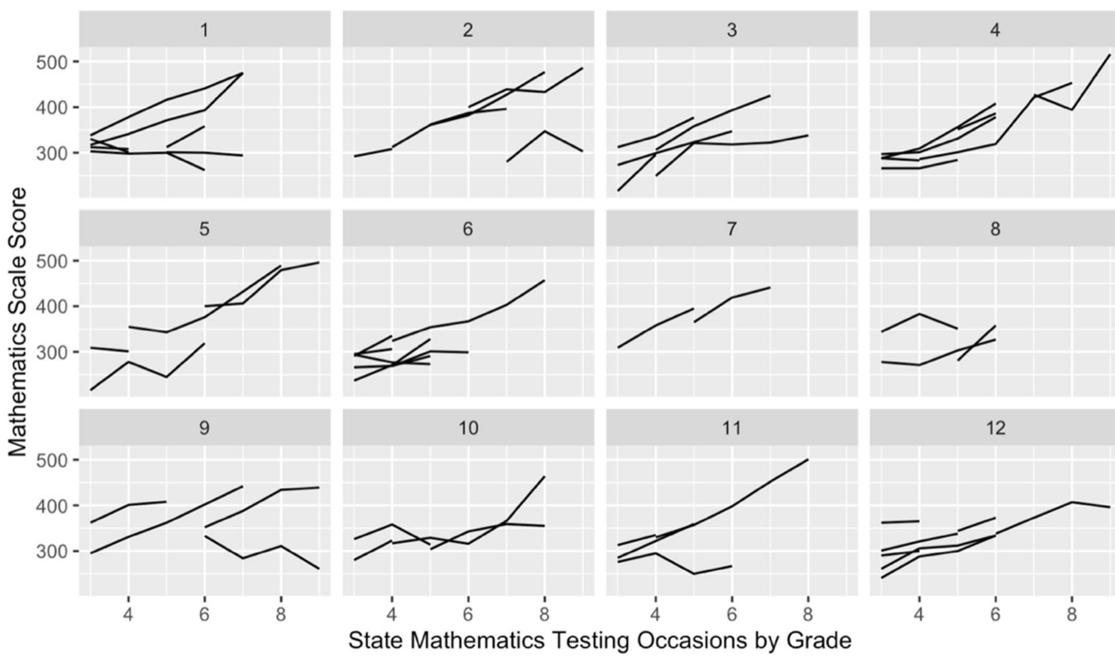
The results for the OLR analysis of mathematics outcomes suggests that DLI is not a consistent predictor of mathematics achievement as measured by Utah's SAGE and RISE summative assessments. Though mathematics-proficiency percentages for DLI-enrolled EB students seemed to improve over their non-DLI peers in the secondary grades, no statistically significant effect was found for DLI enrollment in the OLR analysis. A skewed sample of proficiency scores across all EB students regardless of DLI enrollment may contribute to the lack of statistically significant results. Additionally, low power in the analysis as measured by R^2 was largely reflective of the lack of consistent data across grades and achievement levels for EB students.

Longitudinal Multilevel Analysis of Mathematics Achievement

In contrast to Question 4, which concerns assessing achievement level and DLI for mathematics, Question 5 addresses individual growth trajectories of EB students' mathematics achievement and DLI status over time. Using scale scores rather than achievement measured on a 4-point scale potentiated a longitudinal growth curve analysis, Figure 4.7 illustrates how growth intercepts and slopes vary across individuals. To improve clarity, a random sampling of students' growth trajectories represented in 12 separate graphs—to minimize crowding of observations—have been included in Figure 4.7 with the x-axis representing grade level progression and the y-axis indicating scale scores. As expected, slopes and intercepts varied by individual. Also of note is the presence of missing values for some students across grades. Using a longitudinal mixed model approach accounts for missing data and maintains model specification integrity (Peugh, 2010).

Figure 4.7

Mathematics Scale Score Growth Sample



Multilevel Regression Grade Level Models

For this longitudinal mixed methods analysis, the lme4 package in R was employed to assess patterns of mathematics performance and DLI status. Like the language arts longitudinal analysis discussed earlier, several models progressively added individual predictors to the unconditional means model (M0) to identify the best fit model. The unconditional means model written below:

$$Y_{ti} = \beta_{0i} + \beta_{1i}T + \varepsilon_{ti}$$

includes *Y* as the mathematics scale score for individual *i* at time measurement *t*. β_{0i}

represents the grand mean of all mathematics scales scores or expected measurement for student *i* at the first measurement time. β_{1i} is the coefficient of repeated measure T at the testing occasion with ε_{ti} as the measure of error.

The unconditional model and subsequent models included data on 697 students with 14,105 observations from 2015 to 2019 in Grades 3-9. Table 4.17 displays the log likelihood of M0 as -11095, which was used to compare the fit of subsequent models. Table 4.17 also exhibits the progression of models that first added DLI as a factor in M1 and then time in M2. The final model, M5, demonstrated the best model fit with the log likelihood of -10479 a significant $X^2 = (2, N = 697) = 408.95, p < .001$. The final model includes time in its quadratic form, grade^2, and grade as a random effect allowing for students' scores over time to vary. The full model, displayed as previously presented in Table 4.16, also shows the interaction variable of dli and grade (grade * dli) included in

Table 4.17

Model Comparison of Mathematics Scale Scores

Models	
M0	ma_scale_score ~ 1 + (1 | ID)
M1	ma_scale_score ~ dli + (1 | ID)
M2	ma_scale_score ~ dli + grade + (grade | ID)
M3	ma_scale_score ~ dli + grade + iep + ses + genderc + (1 + grade | ID)
M4	ma_scale_score ~ 1 + grade * dli + I(grade^2) * dli + iep + ses + genderc + (1 | ID)
M5	ma_scale_score ~ 1 + grade * dli + I(grade^2) * dli + iep + ses + genderc + (I(grade) | ID)

	npar	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
M0	3	21818	21834	-10905.8	21812			
M1	4	21735	21758	-10863.6	21727	84.468	1	<2e-16 ***
M2	7	19859	19898	-9922.4	19845	1882.328	3	<2e-16 ***
M3	10	19762	19818	-9871	19742	102.933	3	<2e-16 ***
M4	11	20133	20194	-10055.4	20111	0	1	1
M5	13	19728	19801	-9850.9	19702	408.951	2	<2e-16 ***

both M4 and M5. Of important note, solely adding the quadratic form of the time variable did not improve the model as shown in M4. Only when the time variable was allowed to vary across individuals at the second level was the model improved. This finding points to the importance of running a repeated measure multilevel model to better identify the effect of the predictor variables, namely DLI, on the mathematics outcome.

When reviewing M5 for fit, Table 4.18 shows the R^2 statistics for M5 had a R^2 = .98, suggesting that the model explains 98% of the total variance in the mathematics outcome of both random and fixed factors with 22% of the fixed effects explained, R^2 = .22. Table 4.18 lists the progression of models used for the analysis and illustrates how all of the predictive factors have a significant effect except SES: dli[N]= -53.58, $p < .001$; Grade = 3.47, $p = .04$; IEP[*N*] = 39.19, $p < .001$, ses[*N*]= 5.93, $p = .177$; gender[M] = 6.75, $p = .02$. Additionally, the model shows that the interaction of dli and the quadratic form of time is significant with dli[N]*grade^2 = -2.57, $p < .001$, suggesting that dli moderates the relationship of growth of mathematics scale score as students progress from elementary to secondary schools.

One way to illustrate the effect of the interaction of DLI and grade level is to chart the slopes of the DLI status across grade levels. Figure 4.8 provides a graph of the curvilinear slopes of EBs enrolled in DLI and those not enrolled. The figure shows that in Grade 3 DLI and non-DLI EB students start relatively similarly. Growth curves tend to lag for DLI-enrolled EBs, but then they experience a steeper growth curve in the middle grades. Towards later middle school, Grades 7-8, DLI students surpass their non-DLI peers in math achievement as measured by scale scores.

Table 4.18

Model Comparison for Repeated Measures Multilevel Analysis of Mathematics

	Model 0			Model 1			Model 2		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	336.18	332.03 – 340.34	<0.001*	322.74	318.01 – 327.46	<0.001*	218.71	211.77 – 225.65	<0.001*
dli [N]				41.95	33.62 – 50.28	<0.001*	8.17	1.63 – 14.71	0.014*
grade							22.89	21.40 – 24.38	<0.001*
Random Effects									
σ^2		1192.19			1192.13			192.97	
τ_{00}		3049.30 $_{ID}$			2668.37 $_{ID}$			6917.42 $_{ID}$	
τ_{11}								321.28 $_{ID.grade}$	
ρ_{01}								-0.88 $_{ID}$	
ICC		0.72			0.69			0.94	
Marginal R^2 / Conditional R^2		0.000 / 0.719			0.097 / 0.721			0.348 / 0.959	

	Model 3			Model 4			Model 5		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	179.41	167.98 – 190.85	<0.001*	211.99	200.68 – 223.30	<0.001*	224.29	210.63 – 237.95	<0.001*
dli [N]	10.94	4.62 – 17.25	0.001*	-58.82	-70.65 – -46.99	<0.001*	-53.58	-72.24 – -34.93	<0.001*
grade	22.79	21.31 – 24.26	<0.001*	4.77	2.67 – 6.87	<0.001*	3.47	0.24 – 6.71	0.035*
iep [N]	39.09	29.89 – 48.29	<0.001*	49.58	40.15 – 59.02	<0.001*	39.19	29.97 – 48.41	<0.001*
ses [N]	5.14	-3.45 – 13.72	0.241	14.89	5.96 – 23.82	0.001*	5.93	-2.68 – 14.54	0.177
genderc [M]	6.63	0.84 – 12.42	0.025*	3.68	-2.28 – 9.65	0.226	6.75	0.94 – 12.55	0.023*
grade^2				1.85	1.65 – 2.04	<0.001*	2.05	1.77 – 2.33	<0.001*
grade * dli [N]				24.91	21.32 – 28.51	<0.001*	25.72	20.41 – 31.03	<0.001*
dli [N] * grade^2				-2.25	-2.56 – -1.94	<0.001*	-2.57	-2.96 – -2.17	<0.001*
Random Effects									
σ^2		193.05			470.83			189.91	
τ_{00}		6837.34 $_{ID}$			1555.31 $_{ID}$			6688.60 $_{ID}$	
τ_{11}		316.80 $_{ID.grade}$						314.87 $_{ID.I(grade)}$	
ρ_{01}		-0.90 $_{ID}$						-0.89 $_{ID}$	
ICC		0.93			0.77			0.97	
Marginal R^2 / Conditional R^2		0.390 / 0.959			0.513 / 0.887			0.222 / 0.979	

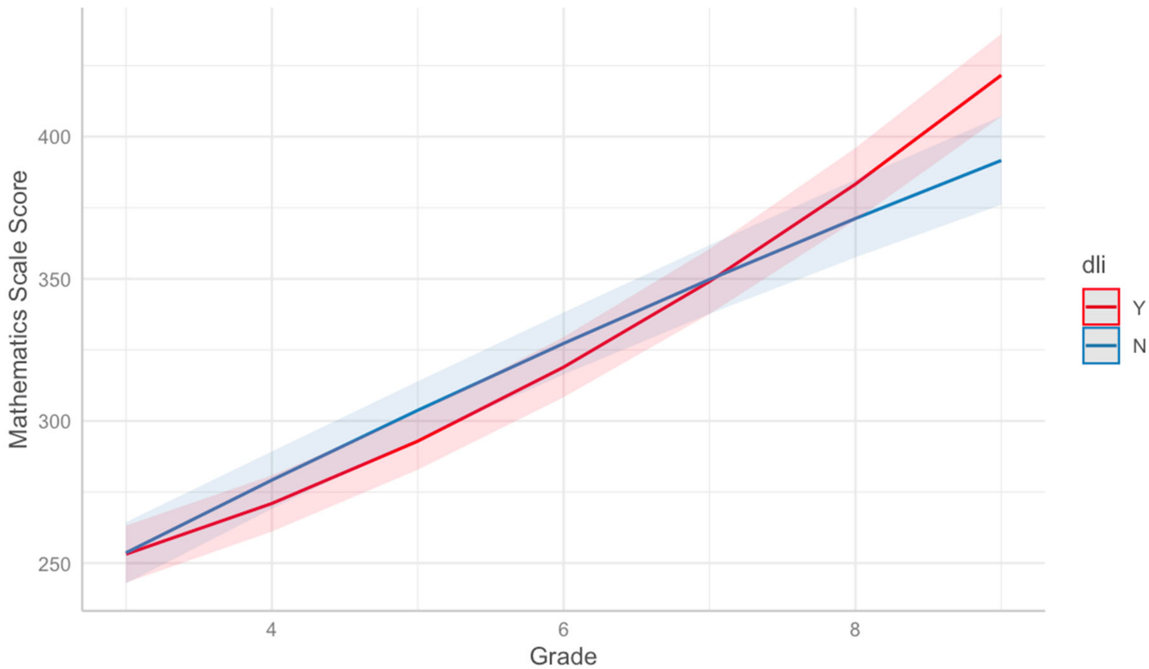
Note. *$p < .05$.

Summary of Multilevel Model Analysis of Mathematics Outcomes

The results of the longitudinal multilevel or mixed effects analysis on mathematics achievement across grade levels suggest that DLI enrollment does have a curvilinear relationship with mathematics achievement. By adding the quadratic form of grade level and introducing an interaction term of DLI and grade to the equation, results show that DLI enrollment has a significant predictive effect on mathematics outcomes after accounting for SES, gender, and IEP status. Interestingly, this finding is similar to

Figure 4.8

Mathematics Predicted Scale Score



that of the language arts achievement conducted in this study. In the sample data, enrollment in DLI predicted better mathematics growth for EBs across their school career from Grade 3-Grade 9.

Grade Point Average Analysis

Question 6 targeted if and how DLI predicted a GPA score for EB students. In the district studied, only secondary grade-level students accrue a GPA. Table 4.19 shows the demographics and GPA distribution across grades. Across all EB students and grades, the grand mean of GPAs of 2.59 is higher than means found in Grade 9 and Grade 10 within this district. Inversely, the middle schools report higher means than the total school mean

Table 4.19

Descriptive Statistics of Students in GPA Analysis by Grade

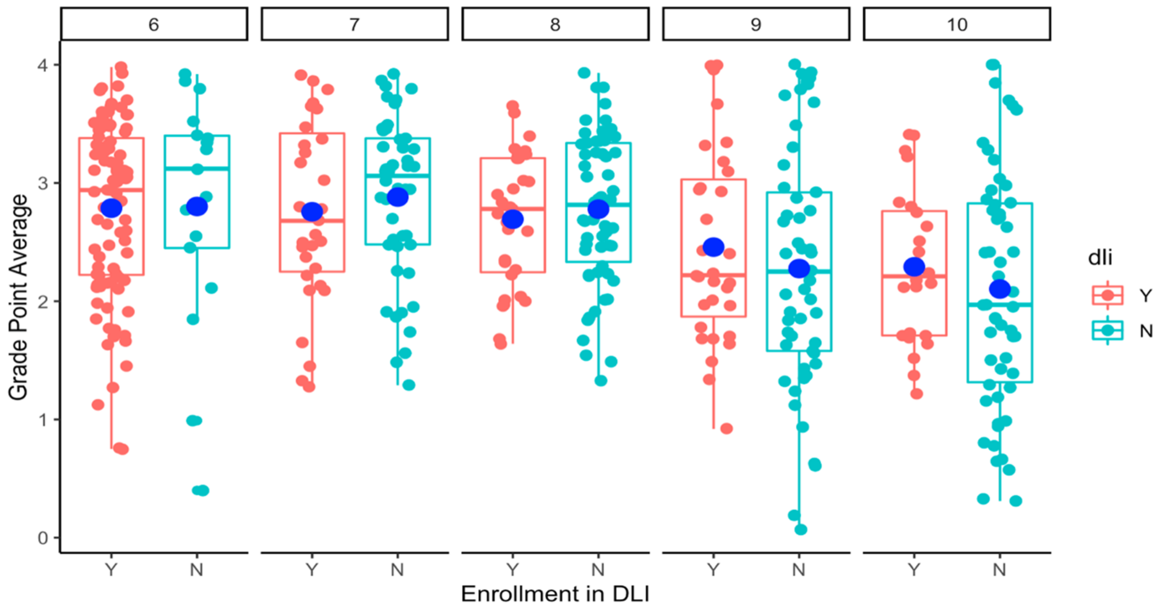
	6 (N=115)	7 (N=77)	8 (N=89)	9 (N=87)	10 (N=78)	Overall (N=446)
Gender						
F	56 (48.7%)	33 (42.9%)	38 (42.7%)	36 (41.4%)	40 (51.3%)	203 (45.5%)
M	59 (51.3%)	44 (57.1%)	51 (57.3%)	51 (58.6%)	38 (48.7%)	243 (54.5%)
Special Education						
Y	12 (10.4%)	14 (18.2%)	12 (13.5%)	9 (10.3%)	12 (15.4%)	59 (13.2%)
N	103 (89.6%)	63 (81.8%)	77 (86.5%)	78 (89.7%)	66 (84.6%)	387 (86.8%)
Free & REduced Lunch						
Y	103 (89.6%)	67 (87.0%)	71 (79.8%)	72 (82.8%)	63 (80.8%)	376 (84.3%)
N	12 (10.4%)	10 (13.0%)	18 (20.2%)	15 (17.2%)	15 (19.2%)	70 (15.7%)
Dual Language Immersion						
Y	96 (83.5%)	31 (40.3%)	31 (34.8%)	32 (36.8%)	24 (30.8%)	214 (48.0%)
N	19 (16.5%)	46 (59.7%)	58 (65.2%)	55 (63.2%)	54 (69.2%)	232 (52.0%)
Grade Point Average						
Mean (SD)	2.79 (0.779)	2.83 (0.725)	2.75 (0.613)	2.34 (0.949)	2.16 (0.911)	2.59 (0.840)
Median [Min, Max]	2.94 [0.400, 3.98]	2.95 [1.28, 3.92]	2.79 [1.33, 3.93]	2.23 [0.0700, 4.00]	2.14 [0.310, 4.00]	2.68 [0.0700, 4.00]
Missing	4 (3.5%)	0 (0%)	0 (0%)	3 (3.4%)	0 (0%)	7 (1.6%)

for Grade 6, Grade 7, and Grade 8 at 2.79, 2.83, and 2.75, respectively. Hence, it appears that on average EB GPAs are generally lower in high school that in the middle school.

Looking at the distribution of the GPAs across grade levels and DLI enrollment, it appears that although means dropped for all EBs, enrollment in DLI may have moderated the rate of decline in GPAs. To further explore this possibility, Figure 4.9 shows the distribution of GPAs by DLI enrollment. GPAs seem to be lower for DLI-enrolled students for Grade 6 and Grade 7. At Grade 8, mean grades become similar among DLI enrollment groups and then begin to differ for Grades 9 and 10 with DLI-enrolled students attaining better GPAs on average than their non-enrolled peers. GPAs for this dataset represent grade-level cohort groups. Although the data represent GPAs for the 2019 school year and not a longitudinal dataset, the distribution across grades as a measure of time informed the regression model that was built through a multiple regression analysis.

Figure 4.9

Grade Point Average Distribution by Grade



Note. Blue dots indicate median for each boxplot.

Multiple Regression Analysis of Grade Point Average

To test the question of how DLI enrollment predicts GPA as an outcome, the analysis used multiple regression to control for SES, gender, and enrollment in special education.

Table 4.20 provides several models each using DLI enrollment as a predictor as well as demographic predictors in each subsequent model. DLI did not significantly predict GPA at the $p < .05$ level other than when it was coupled with SES. Model 6 (M6) combines each of the demographic variables and adds grade as an additional predictor. M6 illustrates that each of the variables tested other than DLI status had significant estimates at $p < .05$.

Table 4.20

Model Comparison for Multiple Regression Analysis of GPA and Predictors

	M1			M2			M3			M4		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	2.66	2.55 – 2.78	<0.001*	2.64	2.53 – 2.75	<0.001*	2.79	2.65 – 2.94	<0.001*	2.41	2.17 – 2.65	<0.001*
Dual Language Immersion: N	-0.14	-0.30 – 0.02	0.081	-0.17	-0.33 – -0.02	0.032*	-0.15	-0.30 – 0.01	0.067	-0.12	-0.28 – 0.04	0.131
Free&Reduced Lunch: N				0.28	0.06 – 0.51	0.013*						
gendercM							-0.23	-0.39 – -0.08	0.004*			
Special Education: N										0.28	0.05 – 0.51	0.017*
R^2 / R^2 adjusted		0.007 / 0.005			0.021 / 0.016			0.026 / 0.022			0.020 / 0.015	

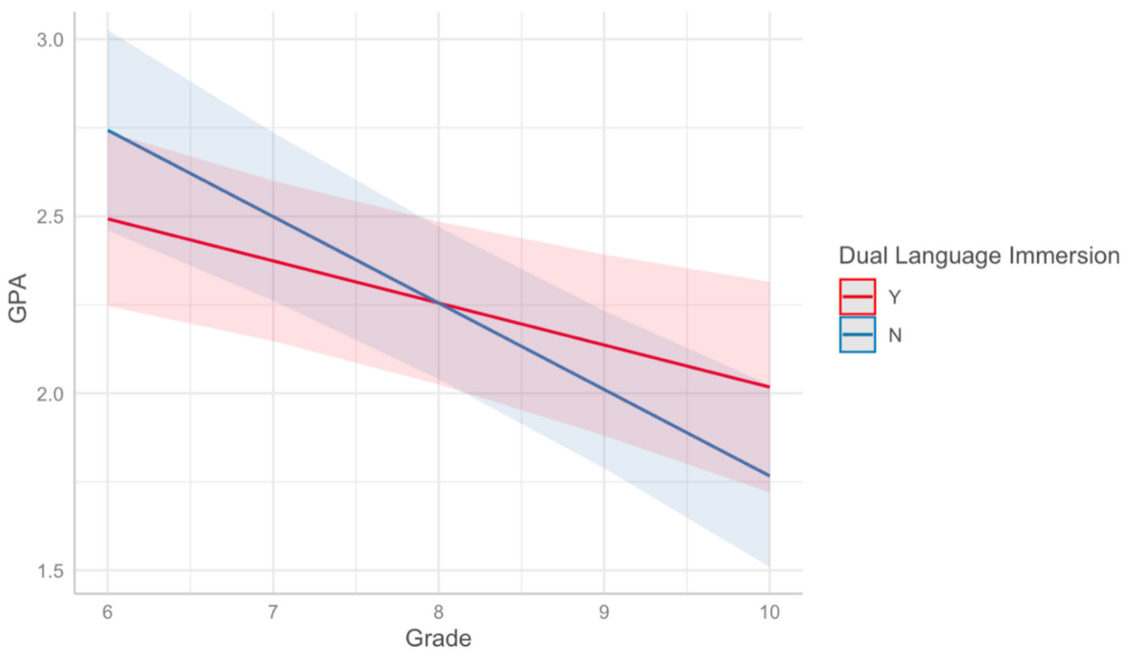
	M5			M6			M7		
Predictors	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*	*Estimates*	*CI*	*p*
(Intercept)	3.91	3.48 – 4.33	<0.001*	3.84	3.36 – 4.31	<0.001*	3.42	2.82 – 4.02	<0.001*
Dual Language Immersion: N	0.03	-0.13 – 0.19	0.692	0.01	-0.15 – 0.17	0.896	1	0.11 – 1.89	0.027*
Free&Reduced Lunch: N				0.31	0.10 – 0.52	0.005*	0.3	0.09 – 0.51	0.005*
gendercM				-0.21	-0.36 – -0.06	0.006*	-0.21	-0.36 – -0.06	0.006*
Special Education: N				0.22	0.00 – 0.45	0.048*	0.23	0.01 – 0.45	0.038*
graden	-0.17	-0.23 – -0.11	<0.001*	-0.17	-0.23 – -0.12	<0.001*	-0.12	-0.19 – -0.05	0.002*
dliN:graden							-0.13	-0.24 – -0.02	0.026*
R^2 / R^2 adjusted		0.081 / 0.077			0.126 / 0.116			0.136 / 0.124	

Note. *$p<.05$.

When an interaction variable of DLI and grade were introduced to the last model (M7), the R^2 value improved and had an adjusted R^2 value of .124, which has been suggested as a low effect (a score of .13 being the threshold for a moderate effect; Cohen et al., 2003). The interaction variable dli*grade was significant at p = .03. Thus, DLI status moderated the effect of GPA score across grade levels as depicted in Figure 4.10.

Figure 4.10

Grade Point Average Predicted Score 2019



Summary of Regression Analysis of Emergent Bilingual Students' Grade Point Averages

The results of the multiple regression model on GPA outcomes suggest that, although a small effect, DLI does statistically predict GPA. When controlling for SES, gender, and IEP status, DLI status was a significant predictor (p = .03) of GPA outcome. This effect was significantly moderated by grade p = .03 when an interaction was introduced into the model. The interaction suggests that GPA levels did not decrease as markedly across grades for EB students enrolled in DLI as compared to their non-DLI peers.

Results Summary

The above analyses looked at the predictive role of DLI enrollment on English language acquisition, English language arts and mathematics achievement, and grade point average as measures of academic achievement. A series of regression analyses, including multilevel modeling, ordinal logistic regression, and multiple regression, was conducted to identify the effect of DLI enrollment on academic achievement of EB students. Results from the analyses reflected varied success in determining DLI's influence. The OLR analyses proved to have limited power in determining the predictive effect of DLI on student achievement. In contrast, the MLM analyses yielded stronger results concluding that over time DLI enrollment predicted better achievement results for EB students in Grades 7-9 when compared to their non-DLI peers.

MLM model results demonstrated that EB students enrolled in DLI had a steeper growth curve than their non-DLI enrolled peers in English acquisition until the beginning of high school as measured by WIDA ACCESS scale scores. Both mathematics and English language arts outcomes had similar results to WIDA outcomes, showing a steeper positive trend for DLI enrolled students. Although scale scores for both mathematics and language arts showed similar starting points (on average) in Grade 3 for both DLI and non-DLI EBs, results demonstrated that in the early elementary years DLI students' scores lagged behind their peers but eventually surpassed their peers in scale scores in the middle-school grades.

Using the OLR analyses to determine DLI influence on mathematics and language arts outcomes proved inconsistent. DLI, as a factor, did not afford a consistent significant estimate when looking at results across individual grade levels. Models provided a significant DLI estimate in only three of seven grades analyzed for language arts and in one of seven grades for mathematics. When DLI status was associated with a significant estimate, the result indicated that non-DLI students had better odds of receiving a highly proficient score. In general, the OLR models demonstrated low predictive power for academic outcomes as measured on the 4-point proficiency scale for SAGE and RISE assessments.

The multiple regression analysis of grade-point average among secondary students indicated that GPA means for EBs decreased from Grade 6 through Grade 9, although DLI-enrolled students demonstrated a shallower decline. Thus, enrollment in DLI was associated with better GPAs on average over time. Using GPA as a surrogate measure of academic engagement suggested that DLI students maintain academic success better than their non-DLI peers.

In general, regression models that controlled for time by factoring in grade level yielded better explanatory power. Additionally, modeling the curvilinear relationship of results over time better predicted the nuanced relationship of DLI influence on academic outcomes. The current study found that DLI outcomes were better measured through longitudinal designs which took grade level growth into account.

CHAPTER 5

DISCUSSION AND IMPLICATIONS

One of the greatest challenges of U.S. school districts has been to provide access to grade-level curriculum for all students. This has been especially difficult for students who are simultaneously learning English as an additional language in school. As identified herein, EB students must learn academic content while simultaneously becoming proficient in academic English. EB students have a lot that they need to accomplish in their educational careers. Students not only need to learn sufficient academic language associated with language arts, mathematics, science, social science, and other curricular areas to graduate from an American educational system, but they must do so at a pace determined for English-dominant students. Clearly, the language and academic expectations on EB students are tremendous.

The American educational system has clamored to find better ways to assist EB students in taking on the challenge of learning academics and an additional language simultaneously, but with varying degrees of success. School districts have implemented many educational programs to assist EB students in meeting myriad academic and language proficiency goals. Programs such as pull-out English language services, sheltered instruction, developmental bilingual education, and dual language immersion have all achieved varied outcomes. Bilingual programs, like dual language immersion (DLI) have gained in popularity. Utah’s DLI model has received much attention in recent years and has been modeled in other states currently increasing their own DLI offerings. Although touted as progressive in its scope, one of the criticisms of Utah’s DLI program has been a focus on the benefit of such programs for the native-English speaking majority student population over that of the EB students enrolled in such programs (Valdez, et al., 2016). Research on the potential outcomes for both subpopulations of DLI participants can inform such concerns.

Study Goals

This study focused on exploring how DLI enrollment predicted important academic outcomes of EB students. The main objective of this study was to use sophisticated analyses to observe if enrollment in DLI classrooms predicted better academic outcomes for EB students than their non-native-speaking peers not enrolled in DLI classrooms in a rural district adhering to the Utah Dual Language Immersion model. Specifically, this study used regression analyses to determine how DLI enrollment predicted academic outcomes in English acquisition, English language arts, and grade point average. Multilevel modeling analyses were employed to explore longitudinal outcome data for English acquisition, English language arts, and mathematics. Ordinal logistic regression techniques were used to determine the odds of scoring highly proficient on English language arts and mathematics assessments, and multiple regression assisted in determining DLI enrollment influence on grade point average outcomes.

Salient Findings

This study not only confirmed the advantages of DLI instruction for EB students but offers additional evidence that using multilevel modeling to analyze academic outcomes can afford a nuanced understanding of student growth over time. Longitudinal analyses are often neglected in program evaluation in the field of education. The multilevel modeling and multiple regression modeling used here proved useful in providing analyses with better fitted models yielding salient findings. Results of this study offer some evidence that there is indeed a greater academic benefit for EB students enrolled in DLI as opposed to enrollment in more traditional EL programs. EB students performed as well or better over time than their non-DLI peers. Although benefits appear only after multiple years of participation, the longitudinal analyses demonstrated that academic outcomes improved at a greater rate in English language acquisition, English language arts, and mathematics for DLI students than for their non-DLI peers. Additionally, the study discerned a later benefit for DLI students with GPA as an outcome at the secondary level.

English Language Acquisition of Emergent Bilingual Students

Acquiring high levels of English proficiency is a requisite for overall high academic performance in the U.S. All high-stakes assessments in Utah are conducted in English. For EB students, scoring at high levels on the WIDA ACCESS assessment indicates that a student is drawing closer to mastering the academic English needed to perform well on other scholastic tasks. This focus on academic English is inherent in the design of the WIDA test as it intentionally targets more than conversational language skills. Within the current study, using a MLM analysis to account for the complexities of time and curvilinear growth proved beneficial in understanding the predictive effect of

DLI enrollment on English language acquisition. Looking at only the mean scores of the grade-level groups would not have delineated the true story of English language growth within the district. The interaction of time and DLI enrollment indicated that participation in DLI over time produces positive English language results over time. The study showed how EB students enrolled in the selected district's DLI program, on average, were able to surpass their peers' performance on the WIDA ACCESS by Grade 6.

English language performance is crucial to other academic achievement, especially on assessments (Abedi, 2011). As no surprise, performance trends on the WIDA ACCESS preceded similar performance on both the English language arts and mathematics assessments, further supporting the fact that English language acquisition is a prerequisite to performance on academic assessments provided in English. However, this requisite language development can be accomplished within a dual-language program. Learning in two languages appears to achieve the expected levels, if given sufficient time.

Language Arts Achievement Among Emergent Bilingual Students

This study found that for English language arts outcomes, as measured by Utah's SAGE and RISE summative assessments, EB students enrolled in DLI performed the same or below their peers until Grade 8, at which time DLI students surpassed their non-DLI peers. The OLR analysis did not yield consistently significant results across grade levels for DLI enrollment. Some results suggested that DLI enrollment was less advantageous for students in attaining language arts proficiency. However, issues with the size of the sample at some grade levels and the skewed sample across all four levels of proficiency by grade may have contributed to these weaker results. Thus, the OLR analyses proved less valuable in this study.

In contrast, the MLM analysis indicated that DLI was advantageous for language arts growth when a time and dli status interaction term was introduced into the model. This finding indicated that students enrolled in DLI received an increased benefit in language arts growth over time and on average surpassed their peers by Grade 8. The MLM analysis accounted for the complex and nuanced nature of language arts achievement growth over time by modeling the curvilinear growth and interaction term.

Yet, growth did not necessarily equate to language arts proficiency. The OLR model and data showed that the majority of EB students did not score proficient on the SAGE and RISE language arts assessment across all grades regardless of DLI enrollment. One factor that may have affected the differing outcomes of the two different analyses is the composition of the SAGE/RISE assessments. As Abedi (2011) suggests, academic assessments are not made with EB students in mind. Scale score growth, while significant, may not rise to the thresholds needed to reach proficiency levels needed, especially with the language demand that such assessments contain.

Mathematics Achievement Among Emergent Bilingual Students

Mathematics achievement in the study paralleled the results found in the language arts analyses in some ways. The OLR analysis in mathematics did not show that DLI enrollment was a significant factor in the predicting mathematics proficiency outcomes.

Descriptive mathematics SAGE and RISE data, like those for language arts, showed that the majority of DLI and non-DLI EB students in this district scored at proficiency levels 1 and 2. Descriptive statistics indicated that a greater percentage of DLI students scored proficient or above proficiency than their non-DLI peers in the secondary grades, although no statistically significant difference was found in the OLR analysis. In addition to constraints with the analyses, these incongruities could reflect aspects of the assessments, their scoring protocols, or the static nature of the proficiency levels.

Similar to language arts, the MLM analysis yielded stronger fit models of growth using DLI as a factor. The results were detected by using scale score measurements that accounted for nuanced growth. Scale score measurements are better suited to compare results over different tests and over different years, therefore better for measuring growth over time (Utah State Board of Education, 2019c). An interaction term between time, measured by grade level and DLI enrollment indicated that DLI students performed comparably or exceeded their non-DLI peers in mathematics growth over time.

Grade Point Average for Emergent Bilingual Students

The GPA analysis found that EB students enrolled in DLI achieved higher GPAs over time. Although GPAs for EB students declined from Grade 6-10, DLI students had a flatter decline compared to their non-DLI peers. An interaction between enrollment and grade level was used to improve the model fit. Using GPA as a proxy for academic engagement, results showed that DLI moderated the engagement across grade levels in positive ways.

Salient Findings Summary

Overall, this study demonstrated that EB students in DLI classes performed as well as or better than their non-DLI peers over time. Students who stayed with DLI saw better academic results in English language acquisition, language arts, mathematics, and grade point average than their peers. This study finds that although some have criticized Utah's DLI model's intent—in that it privileges the native-English-speaking students (Valdéz et al., 2016)—native-Spanish speaking EBs are served well in the Utah model. The positive findings for EB students enrolled in DLI programs suggest that DLI instruction can booster academic success of EB students. Importantly, as evident in this research, ascertaining the impact of DLI programs over time requires more sophisticated statistical methods that can be applied to available longitudinal data.

Implications for Research

Research targeting DLI programs has investigated academic outcomes of native-Spanish-speaking students via different statistical means. One statistical analysis, multilevel modeling (MLM), has recently seen a significant increase in usage. The current study sought to add to the previous research on native-Spanish-speakers enrolled in DLI programs by looking at longitudinal academic data using MLM. Employing MLM in the context of a rural school in Utah, this study adds a unique contribution to the extant research on DLI outcomes for EB students.

Confirmation of Previous Dual Language Immersion Research

This study confirms much of the previously conducted research on DLI. Like the seminal work of Thomas and Collier (2002), this study found that EB students enrolled in DLI performed higher than their non-DLI peers in English language arts and mathematics with gains improving over time. Valentino and Reardon (2015) found that EB students enrolled in bilingual programs experienced greater rates of growth in English language arts as did this current study using a longitudinal perspective. Likewise, when measuring English acquisition, Umansky and Reardon (2014) found that DLI students were slower to demonstrate English proficiency, but by high school they had superior rates of English proficiency compared with their non-DLI peers. This study found that growth was greater when measured by the WIDA ACCESS assessment for DLI students, in that EB students enrolled in DLI exceeded the performance of their non-DLI peers by late elementary school.

Addition to Dual Language Immersion Research

Most of the studies investigating the effects of DLI programming have been conducted in major metropolitan areas (see Steele et al., 2015; Umansky & Reardon, 2014). Even those studies conducted on the Utah model have only looked at students across multiple districts throughout the state and not at any individual district (see Steele et al., 2019; Watzinger-Tharp et al., 2018). This study is unique in that it investigated EB student achievement in a DLI program in a rural Utah context. Considering the results of DLI in rural school districts as well as urban settings better considers the overall impact of DLI across the state. Certainly, district and school contexts can play a significant factor in evaluating the successes and challenges associated with DLI implementation.

This current study also adds to the emerging body of research that employs MLM analyses in exploring the effects of DLI programs. The advantages of MLM, including allowing for missing data over time, adjusting for individual trajectories in growth, and analysis over time provides researchers and program evaluators powerful tools for analyzing quasi-experimental research designs. By applying a MLM analysis to the longitudinal data of this district, differences in program influence were detected. MLM growth studies strengthen the parameter estimates by accounting for time and for individual student growth and intercepts compared to comparison studies (Gustafsson, 2010). Comparing differences in outcomes by year, such as in the ORL analyses conducted in this study, did not allow for exploring the nuanced differences in program influence as did the MLM analyses. For example, when evaluating programs, most districts compare mean scores for grade levels, yet this practice ignores individual growth trajectories such as those viewed in Figures 4.1, 4.4, and 4.8. MLM analyses accounted for the variance in individual slopes allowing for a more robust analysis of program effects.

Study Limitations

One of the limitations of this study is that it was unable to explore the influence of individual schools as a Level 3 factor and take advantage of one of the benefits of using MLM. The number of schools was too small to conduct a meaningful analysis of school-level effects. Similar research in larger districts would be able to consider this Level 3 analysis.

Another limitation was that of a smaller sample size in the upper grades, due to only having the smaller initial cohort to draw from. The smaller sample size in high school created larger standard errors in the Grade 9 analyses. Having larger numbers would allow for better analyses in the future. This district's middle grades contained a large DLI sample so that when this cohort reaches the high-school grades in the future, the data could allow for a more robust analysis of the effects of DLI throughout the high school.

Additionally, by conducting this research in only one school district, generalizability of the effects of the Utah model on EB students enrolled in DLI is weakened. Focusing on EB students across the state would provide a more robust and powerful look at how Utah's DLI model benefits its students. Such an analysis would provide a much better program evaluation of the Utah model than the current study.

Additionally, generalizability of the results of this study across DLI programs nationwide is difficult as this study focused on schools implementing the Utah model. The Utah model has many strict requirements for curriculum, teacher qualifications, training, and instructional delivery (Utah State Board of Education, 2019b). As such, the results of this particular study serve to add to the literature regarding Utah's DLI program more directly and general DLI programs more distally.

When studying the effects of programming on academic achievement, the researcher is dependent upon the scope and reliability of the measurements used. By using Utah's SAGE and RISE assessments as the measures of language arts and mathematics achievement, generalizing findings to other states becomes more difficult. The only measurement used on a national scale was the WIDA ACCESS. Using commonly used nationally normed assessments, such as NAEP, Acadience, etc. could benefit the strength of this type of study.

One of the major goals of this study was to evaluate the effectiveness of a rural school district's DLI program for EB students. The current study only investigated the program through outcome data of native-Spanish speaking EB students. Understanding the effects of DLI on native-English speakers would add to understanding the program effectiveness in this rural school district. Additionally, to conduct a more thorough assessment of the DLI program within the district, subsequent studies should focus on the other areas of program evaluation suggested by Stufflebeam (1968) such as context, input, and process. To better understand the DLI program in this rural district, analyses of teacher practices, stakeholder satisfaction, and implementation effectiveness, etc. could provide the needed information to answer the broader question of whether this program is truly effective and how it benefits its participants.

Implications for Practice

The current study offers additional evidence that, indeed, DLI can make a positive difference in the academic success of EB students. The analyses and results of the study have the potential to advance teacher confidence in their ability to affect meaningful academic change in the lives of EB students; provide additional evidence for administrators when communicating DLI program goals with educators and parents and the time needed to achieve them; and offer insight for district personnel in designing program evaluation efforts.

Having knowledge of the efficacy of the DLI program on EB academic outcomes has a tremendous power to boost the confidence of teachers who are performing the influential work with students and that they can affect meaningful change in the academic trajectory of EB students. Bandura (1993) reminds us that by comparing positive results with that of others we can increase our level of self-efficacy. For professionals the collective belief that teachers within a school can affect meaningful learning and growth in all children is highly influential in the academic progress of students (Goddard et al., 2000; Hattie & Zierer, 2017). Thus, by studying the benefits of DLI on EB academic outcomes, teachers' confidence in their abilities to serve students could increase which, in turn, might increase opportunities to affect meaningful change in the academic lives of EB students.

The effects of principal leadership on student achievement have been well established (Dufour & Marzano, 2011; Grissom et al., 2021). School principals affect student outcomes indirectly by improving teacher skills and abilities (Cotton, 2003). Specifically, school administrators support teachers by building a positive and supportive school culture that challenges negative assumptions, beliefs, and expectations; celebrating teacher effort and success; and focusing on clear goals while monitoring school processes and teacher effectiveness towards attaining those goals (Marzano et al, 2001). This study supports principals in pushing for clear goals for EB student achievement and growth by demonstrating the effectiveness in increasing EB academic achievement. This study offers administrators additional evidence when communicating the benefits of DLI to school staff and the wider school community.

District administration, like school administration, affects student achievement in positive ways when clear goals for schools are set and progress toward those goals are monitored (Marzano & Waters, 2009). One of the ways in which district administration accomplishes this goal is to have school administration present their student achievement outcomes longitudinally. The analyses used in this study offer a more robust type of longitudinal analysis for school districts to employ to discern the benefits of educational programming across multiple years. MLM assisted the study in identifying growth patterns of students rather than depending on comparisons of aggregate data using separate years of data. In this study, the longitudinal analyses provided stronger statistical models and offered greater insight into the nature of DLI participation on outcomes. Likewise, future program evaluations would benefit from the more robust MLM analysis technique rather than techniques that only look at specific years or grades.

The robustness of an analysis is always dependent on the method used and the quality of the input. Evaluating programs or educational changes at the school or district level would benefit from measures that are sensitive enough to identify changes and that can be measured appropriately over time. Understanding the effects of DLI on student outcomes was facilitated by accounting for time as a major factor. As expected, the results demonstrated how learning is both individual and curvilinear. Analyses that use proficiency scales of 1-4, such as SAGE and RISE, or even 1-6, such as WIDA ACCESS, do not have the same explanatory power of understanding growth across time as do scale scores. By using larger or more finite scales, which allow for identifying nuanced changes in outcome measurement, researchers, administrators, and stakeholders can better realize the changes associated with programmatic changes and choices.

Suggestions for Future Studies

This current quantitative study looked at the results for EB students in a Utah-based DLI program in a rural setting, adding to the already rich literature of DLI outcomes in U.S. schools. Many studies have focused on metropolitan areas and have not explored the influence of DLI on students in rural settings like the current study. Often a lack of resources in rural areas plays an important role for students with limited incomes or other stress factors (Miller et al., 2019). Little research has been conducted to understand the influence that a rural setting has on EB students in a DLI program.

This study also added to the nascent literature on Utah's DLI model. Although some studies have explored academic results of the Utah DLI model (see Leite & Watzinger-Tharp, 2016; Steele et al., 2019; Watzinger-Tharp et al., 2016), none have focused on the academic outcomes of EB students in the Utah model. Additional quantitative studies are needed that explore the influence of DLI on EB academic achievement. Since the Utah model has only recently graduated its first cohort of students, a number of studies are needed and expected to investigate the effects of DLI on academic programming and outcomes for native-English as well as EB students.

For example, this study found that DLI had a moderating effect on GPA. More studies are needed to understand that moderating effect and other possible contributing factors. One such factor is that the DLI model changes from an all-day experience in elementary school to a one to two class experience throughout the school day. How does the change in structure effect EB student outcomes for DLI versus non-DLI students?

Another unexplored factor in the Utah model includes how peer interactions effect student outcomes. In the case of this rural district, DLI students remained with the same cohort of students for their academic instruction in elementary school. Knowing how the change in structure when DLI students are exposed to a different set of academic peers in the secondary grade levels affect student outcomes would be beneficial for researchers and educators to understand when looking to improve the Utah DLI model.

Additional studies are needed that broaden the scope of this current study. Conducting studies that include more Utah schools serving EB students would allow for greater scrutiny of the influence of DLI on EB outcomes. Furthermore, by capitalizing on the multilevel model analyses, research should look at the influence of the school by increasing the sample size of schools as a Level-3 factor. Though five elementary schools were too few to carry statistical power in this study, broadening the scope of the study to include all Utah elementary schools serving EB students would certainly add power to the results and increase generalizability. Moreover, studies targeting or identifying school-level influence would allow for better evaluation of how individual DLI programs are progressing toward state-mandated goals for EB students.

As academically high-achieving, high-growth schools are identified across the state, follow-up studies identifying and highlighting high-leverage teaching practices should be conducted allowing all programs to learn from those best-practices. Qualitative and mixed method studies aimed at identifying those components that make up high-quality DLI programs within the Utah model would benefit leadership, program implementation, teacher practices, and especially students and their families as districts engage in their perpetual improvement efforts.

Conclusion

One of the main purposes of this study was to examine if the implementation of a Utah DLI program fulfilled the promise of improved academic achievement for EB students in a medium-sized, rural, school district. The results of the study demonstrated that, on average, EB students who remained in the DLI program for over 5 years performed as well or better than their non-DLI peers. The results showed that DLI students performed better in English language acquisition, language arts, mathematics, and maintained better grade point averages after controlling for variables such as time, lower family incomes, special education, and gender. Assisting EB students in attaining academic achievement continues to be a major focus of all U.S. schools. Though reaching proficiency on measures of academic achievement such as language arts and mathematics assessments remain a challenge, results from this study on academic growth trajectories of EB students enrolled in DLI holds promise for improvements in the educational programming for these students.

REFERENCES

Abedi, J. (2004). The no child left behind act and english language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14. http://www.aera.net/publications/?id=332

Abedi, J. (2009). Computer testing as a form of accommodation for english language learners. *Educational Assessment, 14*, 195-211.

Abedi, J. (2011). Assessing english language learners: Critical issues. In M. del Rosario Basterra, E. Trumbull, & G. Solano-Flores (Eds.), *Cultural validity in assessment: Addressing linguistic and cultural diversity* (pp. 65-87). Routledge.

Acosta, J., Williams, J., III, & Hunt, B. (2019). Dual language program models and english learners: An analysis of the literacy results from a 50/50 and a 90/10 model in two California schools. *Journal of Educational Issues, 5*(2), 1-12. https://files.eric.ed.gov/fulltext/EJ1222762.pdf

Aguinis, H., Gottfredson, R. K., & Culpepper, S. A. (2013). Best-practice recommendations for estimating cross-level interaction effects using multilevel modeling. *Journal of Management, 39*(6), 1490-1528.

Alarcón, A., & Heyman, J. M. (2013). Bilingual call centers at the us-mexico border: Location and linguistic markers of exploitability. *Language in Society, 42*(1), 1. https://doi.org/10.1017/S0047404512000875

American Councils for International Education. (n.d.). *The national dual language immersion research alliance: Where practitioners and researchers align*. https://www.americancouncils.org/sites/default/files/DLI-portfolio.pdf

Arteagoitia, I., & Yen, S. J. (2020). Equity in representing literacy growth in dual language bilingual education for emerging bilingual students. *TESOL Quarterly: A Journal for Teachers of English to Speakers of Other Languages and of Standard English as a Second Dialect, 54*(3), 719-742. http://dx.doi.org/10.1002/tesq.588

Assari, S., Bazargan, M., & Caldwell, C. H. (2019). Association between parental educational attainment and youth outcomes and role of race/ethnicity. *JAMA Network Open, 2*(11), e1916018. https://doi.org/10.1001/jamanetworkopen.2019.16018

Baker, C., & Jones, S. P. (1998). *Encyclopedia of bilingualism and bilingual education*. http://dist.lib.usu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED463660&site=eds-live

Baker, C., & Wright, W. E. (2017). *Foundations of bilingual education and bilingualism* (6th ed.). Multilingual Matters.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117-148. https://doi.org/10.1207/s15326985ep2802_3

Barkaoui, K. (2014). Quantitative approaches for analyzing longitudinal data in second language research. *Annual Review of Applied Linguistics, 34*, 65-101. https://doi.org/10.1017/S0267190514000105

Bates, D. Mächler, M., Bolker, B. & Walker, S. (2015). *Lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.2014.

Bialik, K., Scheller, A., & Walker, K. (2018). 6 facts about English language learners in U.S. Public schools. *FACT TANK: News in the Numbers*. 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/Revue canadienne de psychologie expérimentale, 65*(4), 229-235. https://doi.org/10.1037/a0025406

Bleske-Rechek, A., Morrison, K. M., & Heidtke, K. (2015). Causal inference from descriptions of expereimental and non-experimental research: Public undersatnding of correlation-versus-causation. *The Journal of General Psychology, 142*(1), 48-70.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE - Life Sciences Education, 15*(4). http://dx.doi.org/10.1187/cbe.16-04-0148

Boutakidis, I. P., Rodríduez, J. L., & Knutson Miller, K. (2014). Academic engagement and achievement among latina/o and non-latina/0 adolescents. *Journal of Latinos and Education, 13*, 4-13.

Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, *46*(4), 1171–1178. https://doi.org/10.2307/2532457

Breiseth, L. (2015). *What you need to know about ELLs: Fast facts*. Colorín Colorado. https://www.colorincolorado.org/article/what-you-need-know-about-ells-fast-facts

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

Cobb, B., Vega, D., & Kronauge, C. (2006). Effects of an elementary dual language immersion school program on junior high achievement. *Middle Grades Research Journal, 1*(1), 27-47. http://www.infoagepub.com/products/middle-grades-research-journal-1-1

Cohen, J., Cohen, P., G., W. S., & Aiken, L. S. (2003). *Applied regression/correlation analysis for te behavioral sciences* (3rd ed.). Routledge.

Collier, V. P., & Thomas, W. P. (2009). *Educating English learners for a transformed world*. Fuente Press.

Collier, V. P., & Thomas, W. P. (2017). *Validating the power of bilingual schooling: Thirty-two years of large-scale, longitudinal research* (Vol. 37). Cambridge University Press. https://doi.org/10.1017/S0267190517000034

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.

Conger, D. (2010). Does bilingual education interfere with English-language acquisition? *Social Science Quarterly, 91*(4), 1103-1122.

Cook, H. G., Wilmes, C., Boals, T., & Santos, M. (2008). *Issues in the development of annual measurable achievement objectives for wida consortium states.* (WCER working paper no. 2008-2). Madison: University of Wisconsin-Madison, Wisconsin Center for Educational Research. https://wcer.wisc.edu/docs/working-papers/Working_Paper_No_2008_02.pdf

Cotton, K. (2003). *Principals and student achievement: What the research says.* Association for Supervision and Curriculum Development.

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research, 49*(2), 222-251.

Cummins, J. (1981). *The role of primary language development in promoting educational success for language minority students.* In California State Department of Education (Ed.), *Schooling and language minority students: A theoretical framework* (pp. 3-49). Califonia State University.

Cummins, J. (2008). BICS and CALP: Empirical and theoretical status of the distinction. In B. Street & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (2nd ed., Vol. 2, pp. 71-83). Springer Science & Business Media LLC.

de Jong, E. J. (2016). Two-way immersion for the next generation: Models, policies, and principles. *International Multilingual Research Journal, 10*(1), 6-16.

De La Garza, T., Mackinney, E., & Lavigne, A. L. (2015). Dual language instruction and achievement: A need and a void in the midwest. *Mid-Western Educational Researcher, 27*(4), 363-382.

Dickinson, E. R., & Adelson, J. L. (2016). Choosing among multiple achievement measures: Applying multitrait-multimethod confirmatory factor analysis to stedent assessment, act, and student gpa data. *Journal of Advanced Academics, 27*(1), 4-22.

DuFour, R., & Marzano, R. J. (2011). *Leaders of learning: How district, school, and classroom leaders improve student achievement*. Solution Tree Press.

Eaton, S. E. (2016). *Integration nation: Immigrants, refugees, and america at its best*. The New Press.

Elmore, R. (2004). *School reform from the inside out: Policy. practice, and performance*. Harvard Education Press.

Feinauer, E., & Howard, E. R. (2014). Attending to the third goal cross-cultural competence and identity development in two-way immersion programs. *Journal of Immersion & Content-Based Language Education, 2*(2), 257-272.

Flores, N. (2016). A tale of two visions: Hegemonic whiteness and bilingual education. *Educational Policy, 30*(1), 13-38.

Flores, N., & García, O. (2017). A critical review of bilingual education in the united states: From basements and pride to boutiques and profit. *Annual Review of Applied Linguistics, 37*, 14-29.

Freeman, Y., Freeman, D., & Mercuri, S. (2018). *Dual language essentials for teachers and administrators*. Heinemann.

Fullan, M., & Quinn, J. (2016). *Coherence: The right drivers in action for schools, districts, and systems*. Corwin.

Gándara, P., & Escamilla, K. (2017). Bilingual education in the united states. In O. Garcia, A. M. Y. Lin, & S. May (Eds.), *Bilingual and multilingual education* (pp. 1-14). Springer International.

García, O., & Kleifgen, J. A. (2018). *Educating emergent bilinguals: Policies, programs, and practices for english learners* (2^{nd} ed.). Teachers College Press.

García, O., Kleifgen, J. A., & Falchi, L. (2008). From English language learners to emergent bilinguals Equity Matters, Research review no. 1. *Campaign for Educational Equity.* Teachers College, Columbia University.

Gay, G. (2018). *Culturally responsive teaching* (3rd ed.). Teachers College Press.

Genesee, F. (2004). What do we know about bilingual education for majority language students? In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism and multilingualism* (pp. 547-576). Blackwell.

Genesee, F., & Lindholm-Leary, K. (2008). Dual language education in Canada and the USA. In J. Cummins & N. H. Hornberger (Eds.), *Encyclopedia of language and education* (Vol. 5, pp. 253-263). Springer.

Genesee, F., Lindholm-Leary, K., Saunders, W., & Christian, D. (2005). English language learners in U.S. Schools: An overview of research findings. *Journal of Education for Students Placed At Risk, 10*(4), 363-385. http://dx.doi.org/10.1207/s15327671espr1004_2

Goldenberg, C. (2008). Teaching english langauge learners: What the research does-and does not-say. *American Educator* (Summer 2008), 3-44. https://www.aft.org/sites/default/files/periodicals/goldenberg.pdf

Gomez, L., Freeman, D., & Freeman, Y. (2005). Dual language education: A promising 50-50 model. *Bilingual Research Journal, 29*(1), 145-164.

Goddard, R. D., Hoy, W. K., & Hoy, A. W. (2000). Collective teacher efficacy: Its meaning, measure, and impact on student achievement. *American Educational Research Journal*, 37(2), 479-507. https://doi.org/10.3102/00028312037002479

Grissom, J. A., Egalite, A. J., & Lindsay, C. A. (2021). *How principals affect students and schools: A systematic synthesis of two decades of research*. http://www.wallacefoundation.org/principalsynthesis

Gross, N. (2016, April 11). Dual-language programs on the rise across the U.S. *Latino ED Beat*.

Gunnarsson, B.-L. (2013). Multilingualism in the workplace. *Annual Review of Applied Linguistics, 33*, 162-189.

Gustafsson, J.-E. (2010). Longitudinal designs. In B. P. M. Creemers, L. Kyriakides, & P. Sammons (Eds.), *Methodologial advances in educational effectiveness research* (pp. 77-101). Routledge.

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 Language Minority Research Institute. https://web.stanford.edu/~hakuta/Publications/(2000)%20-%20HOW%20LONG%20DOES%20IT%20TAKE%20ENGLISH%20LEARNERS%20TO%20ATTAIN%20PR.pdf

Hampton, S., Ekboir, J. M., & Rochin, R. I. (1995). The performance of latinos in rural public schools: A comparative analysis of test scores in grades 3, 6, and 12. *Hispanic Journal of Behavioral Sciences, 17*(4), 480-498. https://doi.org/10.1177/07399863950174005

Hattie, J., & Zierer, K. (2017). *10 Mindframes for visible learning: Teaching for success*. Routledge.

Hopewell, S., & Escamilla, K. (2015). How does a holistic perspective on (bi/multi)literacy help educators address the demands of the common core state standards for english language learners/emergent bilinguals? In G. Valdés, K. Menken, & M. Castro (Eds.), *Common core, bilingual and english language learners: A resource for educators* (pp. 39-40). Caslon Publishing.

Hopkins, M., Thompson, K. D., Linquanti, R., Hakuta, K., & August, D. (2013). Fully accounting for english learner performance: A key issue in esea reauthorization. *Educational Researcher, 42*(2), 101-108. http://dx.doi.org/10.3102/0013189X12471426 Hox, J. J., Moerbeek, M., & Van der Schoot, R. (2018). *Multilevel analysis: Techniques and applications* (3rd ed.). Routledge.

Huang, J., Han, T., & Schnapp, K. (2012). Do high-stakes test really address english language learners' learning needs? A discussion of issues, concerns, and implications. *International Journal of Learning & Development, 2*(1), 499-507.

Hussar, B., Zhang, J., Hein, S., Roberts, A., Cui, J., Smith, M., Bullock Mann, F., Barmer, A., & Dilig, R. (2020). *The condition of education 2020*. (NCES 2020-144) U.S. Department of Education:National Center for Education Statistics. https://nces.ed.gov/pubs2020/2020144.pdf

Jacobsen, M. (2016, May 5, 2016). Gov. Gary Herbert's "big change" over common core puzzling for some. *Deseret News*. https://www.deseret.com/2016/5/5/20587963/gov-gary-herbert-s-big-change-over-common-core-puzzling-for-some

Kelly, L. B. (2018). Interst convergence and hegemony in dual language: Bilingual education, but for whom and why? *Language Policy, 17*(1), 1-21.

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

Kieffer, M. J. (2010). Socioeconomic status, English proficiency, and late-emerging reading difficulties. *Educational Researcher, 39*(6), 484-486. http://dx.doi.org/10.3102/0013189X10378400

Kim, Y. K., Hutchinson, L., & Winsler, A. (2015). Bilingual education in the United States: An historical overview and examination of two-way immersion. *Educational Review, 67*(2), 236-252.

Knowles, H. (2018, June 7). *Utah education board ends sage testing*. https://kutv.com/news/local/utah-education-board-ends-sage-testing

Krashen, S. D. (1999). *Condemned without a trial: Bogus arguments against bilingual education*. Heinemann (Portsmouth).

Ladson-Billings, G. (1995). But that's just good teaching! The case for culturally relevant pedagogy. *Theory Into Practice, 34*, 159-165. https://doi.org/10.1080/00405849509543675

Lam, K., & Richards, E. (2020, May 23). *More us schools teach in eglish and spanish, but not enough to help latino kids*. USA Today. https://www.usatoday.com/in-depth/news/education/2020/01/06/english-language-learners-benefit-from-dual-language-immersion-bilingual-education/4058632002/

Lapayese, Y., Hutching, K., & Grimalt, O. (2014). Gender and bilingual education: An exploratory study of the academic achievment of latina and latino English learners. *Journal of Latinos and Edcuation, 13,* 152-160.

Lau v Nichols, 414 U.S. 563 (1974).

Lee, E. (2019, May 25). Ordinal logistic regression and its assumptions-brant test. *Data Analysis in R*. https://medium.com/evangelinelee/brant-test-for-proportional-odds-in-r-b0b373a93aa2

Leite, J., & Cook, R. (2015). Utah: Making immersion mainstream. In P. Mehisto & F. Genesee (Eds.), *Building bilingual education systems: Forces, mechanisms and counterweights* (pp. 83-96. Cambridge University Press.

Leite, J., & Watzinger-Tharp, J. (2016, October 22. *English learner achievement in ela in utah's dual language immersion programs* [Paper presentation]. Sixth International Conference on Immersion and Dual Language Education, Minneapolis, MN, United States.

Leithwood, K., Louis, K., Wahlstrom, K., Anderson, S., Mascall, B., & Gordon, M. (2009). *How successful leadership influences student learning: The second installment of a longer story*. https://doi.org/10.1007/978-90-481-2660-6_35

Leithwood, K., & Strauss, T. (2007). District contributions to school leaders' sense of efficacy: A qualitative analysis. *Journal of School Leadership, 17*, 735-770.

Lindholm, K., & Aclan, Z. (1991). Bilingual proficiency as a bridge to academic achievement: Results from bilingual/immersion programs. *The Journal of Education, 173*(2), 99-113.

Lindholm-Leary, K. (2001). *Dual language education*. Multilingual Matters.

Lindholm-Leary, K. (2012). Success and challenges in dual language education. *Theory Into Practice, 51*(4), 256-262. https://doi.org/10.1080/00405841.2012.726053

Lindholm-Leary, K. (2016). Bilingualism and academic achievement in children in dual language programs. In E. Nicoladis & S. Montanari (Eds.), *Bilingualism across the lifespan: Factors moderating language proficiency* (pp. 203-223). American Psychological Association. https://doi.org/10.1037/14939-000

Lindholm-Leary, K. (2017, April 30). *Educational trajectories of Latino English language learner students in dual-language programs* [Paper presentation]. American Educational Research Association, San Antonio, TX, United States. https://works.bepress.com/kathryn_lindholm-leary/S2/

Lindholm-Leary, K., & Block, N. (2010). Achievement in predominantly low SES/ Hispanic dual language schools. *International Journal of Bilingual Education and Bilingualism, 13*(1), 43-60. http://www.informaworld.com/openurl ?genre=article&id=doi:10.1080/13670050902777546

Lindholm-Leary, K., & Borsato, G. (2006). Academic achievement. In F. Genesee, K. Lindholm-Leary, W. M. Saunders, & D. Christian (Eds.), *Educating English language learners: A synthesis of research evidence* (pp. 176-222). Cambridge University Press.

Lindholm-Leary, K., & Genesee, F. (2010). Alternative educational programs for english language learners. In C. D. O. Education (Ed.), *Research on english language learners* (pp. 323-382). California Department of Education.

Lindholm-Leary, K., & Genesee, F. (2014). Student outcomes in one-way, two-way, and indigenous language immersion education. *Journal of Immersion and Content-Based Language Education*, *2*(2), 165-180.

Lindholm-Leary, K., & Hernández, A. (2011). Achievement and language proficiency of latino students in dual language programmes: Native English speakers, fluent English/previous ELLs, and current ELLs. *Journal of Multilingual & Multicultural Development, 32*(6), 531-545. https://doi.org/10.1080/01434632.2011.611596

Lindholm-Leary, K., & Howard, E. E. (2008). Language development and academic achievement in two-way immersion programs. In T. W. Fortune & D. J. Tedick (Eds.), *Pathways to multilingualism: Evolving perspectives on immersion education* (pp. 177-200). Multilingual Matters Ltd.

Linquanti, R., Cook, H. G., Bailey, A., & MacDonald, R. (2016). *Moving toward a more common definition of english learner: Collected guidance for states and multi-state assessment consortia*. http://www.ccsso.org/sites/default/files/2017-10/MoreCommonDefinition-Final_0.pdf

López, E. J., Ehly, S., & García-Vázquez. (2002). Acculturation, social support and academic achievement of Mexican and Mexican American high school students: An exploratory study. *Psychology in Schools, 39*(3), 245-257.

Marian, V., Shook, A., & Schroeder, S. R. (2013). Bilingual two-way immersion programs benefit academic achievement. *Bilingual Research Journal, 36*(2), 167-186. http://dx.doi.org/10.1080/15235882.2013.818075

Marzano, R. J., & Waters, T. (2009). *District leadership that works: Striking the right balance*. Solution Tree Press.

Marzano, R. J., Waters, T., & McNulty, B. A. (2001). *School leadership that works: From research to results*. ASCD

McFarland, J., Hussar, B., Zhang, J., Wang, X., Hein, S., D'iliberti, M., Forrest Cataldi, E., Bullock Mann, F., & Barmer, A. (2019). *The condition of education 2019*. https://nces.ed.gov/pubs2019/2019144.pdf

Menken, K., Kleyn, T., & Chae, N. (2012). Spotlight on "long-term English langauge learners": Characteristicd and prior schooling experiences of an invisible population. *International Multilingual Research Journal, 6*(2), 121-142.

Menken, K., & Solorza, C. (2014). *No child left bilingual: Accountability and the elimination of bilingual education programs in new york city schools* (0895-9048). http://dx.doi.org/10.1177/0895904812468228

Menken, K., & Solorza, C. (2015). Principals as linchpins in bilingual education: The need for prepared school leaders. *International Journal of Bilingual Education and Bilingualism, 18*(6), 676-697.

Mitchell, C. (2019). *English-only laws in education on verge of extinction*. Education Week.

Miller, P., Votruba-Drzal, E., & Coley, R. L. (2019). Poverty and academic achievement across the urban to rural landscape: Associations with community resources and stressors. *The Russell Sage Foundation Journal of the Social Sciences*, *5*(2), 106-122. https://doi.org/10.7758/RSF.2019.5.2.06

Montoya-Ávila, A., Ghebreab, N., & Galindo, C. (2018). Toward improving the educational opportunities for black and latinx young children: Strengthening family-school partnerships. In S. Sonnenschein & B. E. Sawyer (Eds.), *Academic socialization of young black and latino children: Building on family strengths* (pp. 209-231). Springer.

Morita-Mullaney, T., Renn, J., & Chiu, M. M. (2020). Contesting math as the universal language: A longitudinal study of dual language bilingual education language allocation. *International Multilingual Research Journal*. https://doi.org/10.1080/19313152.2020.1753930

National Academies of Sciences Engineering and Medicine. (2017). *Promoting the educational success of children and youth learning english: Promising futures*. National Academies Press. http://dx.doi.org/10.17226/24677

Natriello, G., Pallas, A. M., & Alexander, K. (1989). On the right track? Curriculum and academic achievement. *Sociology of Education, 62*(April), 109-118.

Nezlek, J. B. (2012). Multilevel modeling for psychologists. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. Sher (Eds.), *Apa handbook of research methods in psychology* (Vol. 3, pp. 219-241). American Psychological Association.

Office of English Language Acquisition. (2018). *Languages spoken by English learners*. Author. https://ncela.ed.gov/files/fast_facts/FastFacts-Languages-Spoken-by-ELs-2018.pdf

Ovando, C. J. (2003). Bilingual education in the United States: Historical development and current issues. *Bilingual Research Journal, 27*(1), 1-24.

Pac, T. (2012). The English-only movement in the us and world in the twenty-first century. *Perspective on Global Development and Technology, 11*(1), 192-210.

Palmer, D., Zuñiga, C. E., & Henderson, K. (2017). A dual language revolution in the united states?: On the bumpy road from compensatory to enrichment education for bilingual chidren in texas. In W. E. Wright, S. Boun, & O. Garcia (Eds.), *The handbook of bilingual and multilingual education* (pp. 447-460). John Wiley and Sons.

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system: A policy brief*. Aspen Institute. https://files.eric.ed.gov/fulltext/ED551318.pdf

Peugh, J. L. (2010). A practical guide to multilevel modeling. *Journal of School Psychology, 48*(1), 85-112. https://doi.org/10.1016/j.jsp.2009.09.002

Questar. (2019). *Utah readiness improvement success empowerment (rise) assessment technical manual*. Utah State Board of Education.

Ramirez, M. (2016). English one day, Español the next: Dual-language learning programs on the rise. *eddigest*. https://hechingerreport.org/english-one-day-espanol-the-next-dual-language-learning-expands-with-a-south-bronx-school-as-a-model/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed., Vol. 1). Sage Publications.

Robinson-Cimpian, J. P., & Thompson, K. D. (2016). The effects of changing test-based policies for reclassifying english learners. *Journal of Policy Analysis and Management, 35*(2), 279-305.

Rumberger, R. W., & Larson, K. A. (1998). Toward explaining differences in educational achievement among mexican american language-minority students. *Sociology of Education,* 68-92.

Salazar, J. J. (1998). A longitudinal model for interpreting thirty years of bilingual education research. *Bilingual Research Journal, 22*(1), 19-30. https://doi.org/10.1080/15235882.1998.10668671

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

Selcuk, R. S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417.

Sheng, Z., Sheng, Y., & Anderson, C. J. (2011). Dropping out of school among ELL students: Implications to schools and teacher education. *The Clearning House, 84*, 98-103.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurence*. Oxford University Press.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453. https://doi.org/10.3102/00346543075003417

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. http://dx.doi.org/10.3102/0002831214528277

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of english-language learners. *Educational Researcher, 32*(2), 3-13. https://doi.org/10.3102/0013189X 032002003

Souto-Manning, M., Madrigal, R., Malik, K., & Martell, J. (2016). Bridging languages, cultures, and worlds through culturally relevant leadership. In S. Long, M. Souto-Manning, & V. M. Vasquez (Eds.), *Courageous leadership in early chidhood education: Taking a stand for social justice* (pp. 57-68). Teachers College.

Steele, J. L., Slater, R., Li, J., Zamarro, G., & Miller, T. (2013). *The effect of dual-language immersion on student achievement in math, science, and english language arts*. Society for Research on Educational Effectiveness. https://eric.ed.gov/?id=ED564065

Steele, J. L., Slater, R., Li, J., Zamarro, G., Miller, T., Li, J., Burkhauser, S., & Bacon, M. (2015). *Effects of dual language immersion on students' academic performance.* EDRE Working Paper No. 2015-09. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2693337

Steele, J. L., Watzinger-Tharp, J., Slater, R., Roberts, G., & Bowman, K. (2019). Student performance under dual language immersion scale-up in Utah. *Utah State Board of Education. See https://www.schools.utah.gov/file/10d44ed-5b6e-40*,71-a7

Stewner-Manzanares, G. (1988). *The bilingual education act twenty years later*. National Clearinghouse for Bilingual Education.

Stufflebeam, D. L. (1968). *Evaluation as enlightenment for decision-making*. Evaluation Center, Ohio State University.

Sugarman, J., & Geary, C. (2018). *English learners in Utah: Demographics, outcomes, and state accountability policies* (Fact Sheet). Migration Policy Institute. https://www.migrationpolicy.org/sites/default/files/publications/EL-factsheet2018-Utah_Final.pdf

Sugarman, J., & Lazarín, M. (2020). *Educating english learners during the covid-19 pandemic: Policy ideas for states and school districts*. Migration Policy Institute. https://www.migrationpolicy.org/sites/default/files/publications/mpi-english-learners-covid-19-final.pdf

Taggart, A. (2018). Latina/o students in k-12 schools: A synthesis of empirical research on factors influencing academic achievement. *Hispanic Journal of Behavioral Sciences, 40*(4), 448-471.

Tan, X., & Michel, R. (2011). *Why do standardized testing programs report scaled scores?: Why not just report the raw or percent-correct score?* ETS R & D Connections, 16, 1-6. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.365.2863&rep=rep1&type=pdf

Tanenbaum, C., Boyle, A., Soga, K., Le Floch, K. C., Golden, L., Prtroccia, M., & Toplitz, M. (2012). *National evaluation of title III implementation-report on state and local implementation*.Office of Planning, Evaluation and Policy, US Department of Education. https://files.eric.ed.gov/fulltext/ED531982.pdf

Tedick, D. J. (2015). United States of America: The paradoxes and possibilities of bilingual educaiton. In P. Mehisto & F. Genesee (Eds.), *Building bilingual education systems: Forces, mechanisms and counterweights* (pp. 1-22). Cambridge University Press.

Tedick, D. J., Christian, D., & Fortune, T. W. (2011). The future of immersion education: An invitation to “dwell in possibility.” In D. J. Tedick, D. Christian, & T. W. Fortune (Eds.), *Immersion education: Practices, policies, possibilities* (pp. 1-10). Multilingual Matters.

Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement*. Center for Research on Education.

Thomas, W. P., & Collier, V. P. (2019). Dual language education for all. In D. E. DeMatthews & E. Izquierdo (Eds.), *Dual language education: Teaching and leading in two languages* (Vol. 18, pp. 91-105). Springer.

Thompson, K. D. (2017). English learners’ time to recassification: An analysis. *Educational Policy, 31*(3), 330-363.

Tran, N. A., Behseta, S., Ellis, M., Martinez-Cruz, A., & Contreras, J. (2015). The effects of spanish english dual language immersion on student achievement in science and mathematics. *EJEP: EJournal of Education Policy*. https://eric.ed.gov/?id=EJ1158168

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., Valentino, R. A., & Reardon, S. F. (2015). *The promise of bilingual and dual immersion education* (CEPA Working Paper No. 15-11). Stanford Center for Education Policy Analysis. https://eric.ed.gov/?id=ED580372

U.S. Department of Education. (2017). *Consolidated state performace report: Parts I and II*. https://www.schools.utah.gov/file/e803c7a4-3c13-459c-97a6-da92b4579c52

U.S. Department of Education. (2018a). *Academic performance and outcomes for english learners: Performance on national assessments and on-time graduation rates*. U.S. Department of Education. https://www2.ed.gov/datastory/el-outcomes/index.html#datanotes

U.S. Department of Education. (2018b). *Our nation's English learners: What are their characteristics?* U.S. Department of Education. https://www2.ed.gov/datastory/el-characteristics/index.html#intro

U.S. Department of Education. (2019). *Dual language learning programs and English learners*. Office of English Language Acquisition. https://ncela.ed.gov/files/fast_facts/19-0389_Del4.4_DualLanguagePrograms_122319_508.pdf

U.S. Department of Education, Office of Planning Evaluation and Policy Development, & Policy and Program Studies Service. (2012). *Language instruction educational programs (LIEPs): A review of the foundational literature*. https://www2.ed.gov/rschstat/eval/title-iii/language-instruction-ed-programs-report.pdf

Utah Education Association. (n.d.). *Utah sage testing*. Utah Education Association. https://www.myuea.org/issues_action/position_statements_information/sage_testing.aspx

Utah State Board of Education. (2018). *Why immersion?* Utah School Board of Education. http://www.utahdli.org/whyimmersion.html

Utah State Board of Education. (2019a). *Fall enrollment by demographics and grade levels*. https://www.schools.utah.gov/data/reports?mid=1424&tid=4

Utah State Board of Education. (2019b). *Utah DLI program fidelity assurances grades 1-5/6.* http://utahdli.org/images/Updated%20DLIAssurances%20Grades%201-5_6.pdf

Utah State Board of Education. (2019c). *Utah RISE report implementarion guide.* https://www.graniteschools.org/wp-content/uploads/2019/09/Utah-RISE-ISR-Interpretation-Guide.pdf

Valdés, G. (2015). What is bilingualism/multilingualism? In G. Valdés, K. Menken, & M. Castro (Eds.), *Common core, bilingual and english language learners* (pp. 38-39). Caslon Publishing.

Valdés, G., Poza, L., & Brooks, M. D. (2017). Language acquisition in bilingual education. In W. E. Wright, S. Boun, & O. García (Eds.), *The handbook of bilingual and multilingual education* (pp. 56-74). John Wiley & Sons.

Valdez, V. E., Delavan, G., & Freire, J. A. (2016). The marketing of dual language education policy in utah print media. *Educational Policy, 30*(6), 849-883. http://dx.doi.org/10.1177/0895904814556750

Valdez, V. E., Freire, J. A., & Delavan, M. G. (2016). The gentrification of dual language education. *Urban Review: Issues and Ideas in Public Education, 48*(4), 601-627. http://dx.doi.org/10.1007/s11256-016-0370-0

Valentino, R. A., & Reardon, S. F. (2015). Effectiveness of four instructional programs designed to serve English learners: Variation by ethnicity and initial English proficiency. *Educational Evaluation and Policy Analysis, 37*(4), 612-637. https://doi.org/10.3102/0162373715573310

Venables, W. N., & Ripley, B. D. (2002). Random and mixed effects. In *Modern applied statistics with S* (pp. 271-300). Springer, New York, NY.

Watzinger-Tharp, J., Swenson, K., & Mayne, Z. (2016). Academic achievement of students in dual language immersion. *International Journal of Bilingual Education & Bilingualism, 21*(8)*,* 913-928.

Watzinger-Tharp, J., Swenson, K., & Mayne, Z. (2018). Academic achievement of students in dual language immersion. *International Journal of Bilingual Education and Bilingualism*, *21*(8), 913-928.

World-Class Instructional Design and Assessment. (2019). *Annual technical report for acess for ELLs 2.0.* https://www.ride.ri.gov/Portals/0/Uploads/Documents/Instruction-and-Assessment-World-Class-Standards/Assessment/ACCESS/ACCESS-ATR-20172018-Online.pdf

CURRICULUM VITAE

J. ERIC CAMPBELL

Director, Equity & Diversity
Wasatch County School District
101 East 200 North
Heber City, UT 84032
435-654-0280
eric.campbell@wasatch.edu

EDUCATION

Ph.D., Education, emphasis Instructional Leadership, 2021
Utah State University, Logan Utah
Dissertation: "Delivering on a Promise: A Longitudinal Cohort Study of Emergent Bilinguals' Academic Achievement in a Utah Dual Language Program"
Advisor: Dr. Parker Fawson

M.Ed., Educational Leadership, 2004
Brigham Young University, Provo, Utah

B.A., Major: History Teaching, Minor: Spanish Teaching, 1998
Brigham Young University, Provo, Utah

RESEARCH INTERESTS

Educational Leadership
Dual Language Immersion
Program Evaluation
Multilevel Analysis
Bilingual Education
Psychological Safety

PROFESSIONAL EXPERIENCE

July 2021 – Present, Wasatch County School District, Heber City, UT
Director of Equity and Diversity

- Direct all activities related to equity and diversity within the district
- Manage funds and activities related to federal funds such as Title I, Title II, Title III, and Title VI

July 2011 – Present, Wasatch County School District, Heber City, UT
Director of Elementary Education

- Supervise and mentor elementary principals
- Facilitate vetting and purchasing of elementary curriculum

July 2004-2011, Wasatch County School District, Heber City, UT
Principal of Heber Valley Elementary School

- Directed all instructional programs for Pre-Kindergarten—Fifth Grades
- Managed facilities and staff
- Implemented Two-way Dual Language Immersion Program

August 2000-2004, Wasatch County School District, Heber City, UT
Junior High Alternative Education Director

- Directed and managed the alternative placement center (PASS/AMPS) for 8th-9th grade
- Assisted with managing the discipline in the school
- Taught social skills
- Taught 9th grade Geography for Life
- Advised 25 8th grade students

PUBLICATIONS

Cox, J., Davies, D. R., Burlingame, G. M., Campbell, J. E., Layne, C. M., & Katzenbach, R. J. (2007). Effectiveness of a trauma/grief–focused group intervention: A qualitative study with war–exposed Bosnian adolescents. *International journal of group psychotherapy*, *57*(3), 319-345.

Nadelson, L. S., Throndsen, J., Campbell, J. E., Arp, M., Durfee, M., Dupree, K., Poll, T. & Schoepf, S. (2016). Are they using the data? Teacher perceptions of, practices with, and preparation to use assessment data. *International Journal of Education*, *8*(3), 50-70.

CONFERENCE PRESENTATIONS

Gallagher, E., Rocha, M., & Campbell, J. E. (2018, March) Facilitating deep mathematical thinking through comprehensive mathematical instruction (CMI) in dual language immersion programs. National Association for Bilingual Education (NABE) annual conference, Albuquerque, NM.

Campbell J. E. (2021, November) Tackling our toughest conversations like a hostage negotiator. Learning Edge Conference, Provo, UT.

COMMITTEES

Brigham Young University Elementary Education Partnership Advisory Committee (EEPAC), 2011—Present

Utah Valley University Professional Development Coordinating Council (PDCC), 2011—Present

SKILLS

Languages: English (native), Spanish (conversational)