Spring 4-2021

The Effects of Bilingual Education on Dual Language Learners' Academic Outcomes

Lauren Thayer
University of Nebraska-Lincoln, lauren.thayer@huskers.unl.edu

THE EFFECTS OF BILINGUAL EDUCATION ON DUAL LANGUAGE LEARNERS' ACADEMIC OUTCOMES

by

Lauren Thayer

A THESIS

Presented to the Faculty of

The Graduate College at the University of Nebraska

In Partial Fulfillment of Requirement

For the Degree of Master of Arts

Major: Special Education

Under the Supervision of Professor J. Marc Goodrich

Lincoln, Nebraska

April 2021

THE EFFECTS OF BILINGUAL EDUCATION ON DUAL

LANGUAGE LEARNERS' ACADEMIC OUTCOMES

Lauren Thayer, M.A.

University of Nebraska, 2021

Advisor: J. Marc Goodrich

The population of dual language learners (DLLs) in the United States continues to grow. As the population grows, so does the need to research their unique development. Additionally, considerations for their education come to the forefront. This meta-analysis will analyze one option for DLLs – bilingual education programs. These programs offer the opportunity for DLLs to receive educational support in both of their languages. However, there has been controversy over the division of languages; specifically, concerns that children will not become proficient in the language of majority (e.g., English). This meta-analysis seeks to aggregate research on the effects of bilingual education programs on DLLs' academic outcomes and compare them to academic outcomes of DLLs' in mainstream, monolingual programs.

TABLE OF CONTENTS

TABLES AND FIGURES

The Effects of Bilingual Education on Dual Language Learners' Academic Outcomes

Dual language learners (DLLs)[1] make up about ten percent of students enrolled in the United States in public and private schools (U.S. Department of Education, 2017). This growing population's needs and special considerations bring new questions to light regarding the best way to educate and prepare DLLs for successful academic outcomes, social lives, and careers. One approach that continues to be controversial is the implementation of bilingual education. Controversy of this approach is borne from the lack of a consensus in educational research on whether bilingual education is an effective way to improve children's academic outcomes while ensuring they are proficient in English (Krashen, 1991). This meta-analysis seeks to consolidate the research on the effects of language of instruction on DLLs' academic outcomes and add to the growing body of literature to determine the most effective way to teach DLLs so that they may have the same opportunities as students who enter school already fluent in English.

DLLs' Academic Achievement

The U.S. Department of Education (2019) has reported findings based on the National Assessment of Educational Progress (NAEP) standardized testing that suggest a gap in literacy and mathematics outcomes such that some DLLs tend to perform worse than their monolingual peers. This gap is found between DLLs who are still classified as English language learners (ELLs) and their DLL peers who gained English proficiency,

[1] Although schools in the United States typically refer to DLLs as English language learners (ELL) or as having limited English proficiency (LEP), to simplify the terminology, the term DLL will be used here in place of ELL, LEP, or bilingual to describe students whose primary home language is not English. That is, students who are described as having limited English proficiency or as being English language learners in primary studies would both fall under the term DLL for the purposes of this review.

as well as native English speakers. Although these gaps are reduced when key demographics are controlled for such as SES and DLL identification rather than ELL status (Kieffer & Thompson, 2018), there is little research examining what is causing the gaps that do exist to narrow. However, researchers have recently reported evidence that these gaps are smaller in areas with more bilingual education programs than areas with fewer bilingual education programs (Goodrich, Thayer, & Leiva, 2021). Therefore, this meta-analysis examines recent evidence of the effects of bilingual education when compared to English-only education on DLLs' academic achievement.

Bilingual Education in the United States

Federal and state policies in the U.S., not to mention societal views, regarding bilingual education have changed drastically over the past century. Although the implementation of bilingual education in the United States can be traced back to the early 19th century, formal policy to regulate and guide educating English language learners (ELLs) did not emerge until the late 19th and early 20th century (see Ovando, 2003 for review). Policies enacted during this time, such as increased funding for English-only schools, were influenced by fears over increased immigration and enabled institutions such as Native American Residential Schools to be established to push for assimilation (Higham, 1992, as cited in Ovando, 2003). These schools would place immigrants and DLLs in English submersion classes. Submersion classes emphasized a sink-or-swim mentality, placing the responsibility of learning English and American culture on the student rather than the school or teacher (Ovando, 2003). These types of classes did not include non-English language supports and were originally designed to build English fluency in students as quickly as possible (Kim et al., 2015). The emphasis on replacing one's language and culture with English and American views, respectively, continued to be a prevalent mentality into the mid-1900s and was not truly challenged until the 1960s.

The first major piece of legislation to promote bilingual education in public and private schools was the National Defense Education Act (NDEA) of 1958, which was passed in response to the successful launch of Sputnik the previous year (Ovando, 2003). Federal funding for non-English language instruction was included in the bill to support teachers who taught foreign languages and provide grants to students. However, the NDEA's main goal was to encourage monolingual English students to learn a second language rather than supporting learning in two languages for students who already spoke a non-English language at home. Non-English language support in schools was not implemented until after the government passed the Bilingual Education Act (1968) under Title VII of the Elementary and Secondary Education Act (ESEA; 1965) a decade later, which led to the addition of English as a Second Language (ESL) programs in public and private elementary and secondary schools (Kim et al., 2015; Ovando, 2003). In addition to this bill, the Supreme Court ruled in the *1974 Lau v. Nichols* case that equal treatment for bilingual and monolingual students did not equate to equal opportunities for all students. That is, if a student cannot understand the instruction because of language barriers, it is the responsibility of the school to provide adequate accommodations.

Despite this ruling, a resurgence of English-only instruction occurred in the 1990s with some states passing laws restricting the use of non-English languages in the classroom. For example, California passed Proposition 227 in 1998 which stated the primary language of instruction for all students in the state should be English, leading to DLLs receiving less support in their home language. This led to a decrease in DLLs' test scores, widening the gap between them and their monolingual peers (Parish et al., 2006).

In the early 2000s, the No Child Left Behind (NCLB) Act (2001) was passed. There was a push to create equal opportunities in school, especially for children from lower SES households, those who needed special education services, and those with limited English proficiency. Because the goal of NCLB was to bring all students up to the same national standards regardless of status in a special group (e.g., DLL, low SES household, special education services), DLLs were required to complete the same high stakes tests as monolingual students. This caused concerns that students would only be "taught to the test" rather than taught for understanding because of the rigid testing standards (Darling-Hammond, 2007; Hursh, 2007). To address the shortcomings of NCLB, Every Student Succeeds Act (ESSA) was passed in 2015. This bill handed over more control to the states to design their own standards and accountability systems and encouraged schools to focus less on benchmarks and more on equitable education for all students regardless of their background (Darling-Hammond et al., 2016). That is, the ultimate goal of ESSA was to give states more power to focus on their schools' specific needs and create more meaningful measures of success. These state-specific indicators were ESSA's solution to the issue of one-size-fits-all education benchmarks set in place by NCLB.

In recent years, because of the implementation of the ESSA and the flexibility it has allowed states in developing their own curriculum, bilingual education and additional language services have become more common place. Between the 2000-2001 and 2016-2017 school years, the DLL population in the United States school system grew by about one million students (U.S. Department of Education, 2020). With this increase in DLLs, there has also been an increase in bilingual education programs. In the 2016-2017 school year, 35 states and the District of Columbia reported offering dual language programs to their students (U.S. Department of Education, 2019). Spanish was the most commonly offered partner language in these programs. This trend in bilingual education brings up new questions that must be addressed. For example, although states report offering dual language programs, these programs can differ greatly. Additionally, there is little high-quality research in this area and experts do not always agree on what constitutes the most effective program to best serve DLLs (Slavin & Cheung, 2005). Among those studies that do find a positive effect of bilingual education on student achievement, there is little overlap in how programs are defined, which can lead to unreliable conclusions (Kim et al. 2015). Additionally, there can be variability within the same program type. For example, a two-way immersion program in one study may split the language of instruction 50:50 whereas another study may split instruction across languages 90:10. These two forms of the same program type can result in differential outcomes attributable to the balance of the language of instruction in each classroom. Therefore, there is a need for clear operational definitions of bilingual education programs, including evaluations of which programs are most effective for promoting educational achievement of students who speak a language other than English at home.

Types of Bilingual Education Program

As stated previously, studies focusing on bilingual education programs lack conformity regarding what each program entails and how it is labeled. There can be multiple names for the same program type, or a new label attached to a program by the researcher. In this section, each type of program, and the label that will be used in the current paper, is described below.

English instruction. The main goal of English instruction programs is to build English language skills, not to build or support bilingualism. Ideally, DLLs enrolled in these programs will reach English proficiency as quickly as possible so they no longer need any additional language services. The main difference between each program is the amount of DLL-specific language support they provide. There are three common English instruction programs, and each is described briefly below.

***Submersion*.** Submersion programs provide no native language support or strategies to facilitate English comprehension for DLLs. DLLs are placed in an all-English, mainstream classroom along with native English speakers. Although this type of program is no longer practiced, it historically had a “sink-or-swim” mentality such that students were expected to become proficient in English solely through exposure to the language (Kim et al., 2015).

***English as a second language (ESL)*.** ESL – or English for Speakers of Other Languages (ESOL) – programs are designed to provide additional support to DLLs who do not have sufficient English proficiency. ESL instruction can be focused exclusively on improving English oral language and comprehension, or it can also provide additional support in content areas such as reading and mathematics (Baker et al., 2016). Students are enrolled in mainstream classes but are given targeted English language instruction for English learners from a classroom aide or ESL teacher. This targeted instruction can be delivered in a few ways: ESL pull-out supports, a separate ESL class, or ESL push-in supports. In an ESL pull-out program, DLLs are removed from their class to work individually with an ESL teacher who provides additional support to the student in content areas and reading (Kim et al., 2015). ESL pull-out is most common in elementary schools (Kim et al., 2015). ESL class is a separate class period for DLLs, which is most commonly used in middle or high school (Kim et al, 2015). In addition to a mainstream content course schedule, DLLs will also have a class dedicated to English language instruction to increase English proficiency. Lastly, ESL push-in refers to instruction in which an ESL teacher is present in a mainstream classroom with a DLL student and provides support as class instruction and activities are ongoing (Kim et al., 2015; Slavin & Cheung, 2005).

Structured English immersion (SEI). SEI programs, also referred to as sheltered English or sheltered instruction observation protocol (SIOP), focus on building English proficiency through strategies to facilitate English learning. Common approaches in these classes include conducting instruction in simplified English and concentrating on developing students' English vocabulary (Slavin & Cheung, 2005). These classes can be used as a transitional period for DLLs who start school with low English proficiency because of the gradual introduction of English. Although complete English proficiency is not required, once the students in SEI can demonstrate their ability to learn in English without the use of additional supports such as simplified English, they are moved to a mainstream classroom.

Bilingual Instruction. Bilingual instruction programs differ from English instruction programs not only in the design but also the main goal. Each program described below supports and encourages bilingualism in its students, although approaches are different across programs.

Transitional bilingual education (TBE). Despite its name and the inclusion of

DLLs' home language instruction, TBE is more closely related to English instruction programs than the other bilingual instruction programs. TBE and English instruction programs' end goal is to build English proficiency in DLLs to the level where they no longer need home language instruction. However, the difference between the programs is the approach used to reach this goal. Unlike English instruction programs, TBE programs typically take place over multiple years with students starting the program in kindergarten or first grade and transferring out by fourth or fifth grade. As students progress through each year, they are slowly instructed less in their home language and more in English. For example, a Spanish-speaking DLL student who is enrolled in TBE in kindergarten may be taught with a 90:10 Spanish to English split in instruction. When they move on to first grade, this ratio shifts so that the class instruction is 70:30 Spanish to English. This gradual adjustment continues until English is the majority language of instruction. Once this occurs, the student will transfer to a mainstream classroom to be taught exclusively in English.

Two-way immersion (TWI). TWI, also known as dual language education or bilingual immersion, typically serves students who are native English speakers and students who speak a language other than English at home. There are two types of immersion education. In full immersion, students are taught in Spanish and English at a ratio of 90:10 (Kim et al., 2015). Typically, instruction switches to 50:50 Spanish to English instruction by second grade, and then the program is considered partial immersion (Kim et al., 2015; Lindholm-Leary, 2012). Because of the mix of native speakers of each language in the classroom, it is believed to be a good way for students to model their native language for their peers and encourages learning in both languages.

One-way immersion (OWI). OWI education is very similar to TWI except it is meant to only serve students from a single language group. That is, the class will only have DLLs who are learning English, or it will only have native English speakers learning a second language. For the purposes of this meta-analysis, we only included studies of OWI programs involving DLLs learning English.

Developmental Bilingual Education (DBE). DBE is also known as maintenance bilingual education and is sometimes considered a subtype of OWI programs because it serves students from one language group. It is typically only for DLL or former-ELL students so they can maintain their home language while also having the opportunity to gain proficiency in English in an environment conducive to learning (Boyle et al., 2015). Like other bilingual programs, the ratio of languages can differ depending on the goal of the program (e.g., 50:50 or 90:10), but usually stays static throughout the program.

Theoretical Background of Bilingual Education

Advocates for bilingual education programs often draw on Cummins' developmental interdependence hypothesis, which states that DLLs can apply skills learned in their L1 when acquiring their L2 (Cummins, 1979). This hypothesis is also sometimes referred to as cross-linguistic transfer. This hypothesis suggests that when DLLs learn a concept in their L1, they can apply that concept in their L2, they just need the vocabulary to express their knowledge in their L2. For example, if a child learns to count in their L1, they only need to learn the names of numbers in their L2 to express their counting knowledge in their L2, they do not need to learn to count again. In a bilingual education program, this process is utilized when DLL children learn a subject in their L1 and then use their knowledge to bolster learning in their L2. This relation has been evidenced by previous studies exploring the development of DLLs' early language and literacy skills. Melby-Lervåg and Lervåg (2011) conducted a meta-analysis to examine whether there was evidence of cross-linguistic transfer in DLL children. They found significant positive correlations between DLLs' first and second language phonological awareness. This relation was also observed for first and second language decoding skills. However, oral language skills such as vocabulary did not show evidence in support of cross- linguistic transfer. Similarly, Goodrich and Lonigan (2017) examined Spanish-English DLL preschoolers' print knowledge, phonological awareness, and oral language proficiency. Consistent with Melby-Lervåg and Lervåg's (2011) conclusions, they noted that language-independent skills, such as print knowledge and phonological awareness, did support the cross-linguistic transfer theory such that preschoolers with higher first language outcomes performed better on second language outcomes. However, language- dependent skills, such as vocabulary knowledge, did not show evidence of transfer.

Given the significant relations between DLLs' L1 and L2 language-independent skills, bilingual education programs may provide a unique opportunity to use the L1 to leverage learning in L2. The goal of most bilingual education programs is to support both of DLLs' languages. Cross-linguistic transfer may occur in bilingual education such that supporting L1 learning while providing some supports in L2 will result in all language outcomes improving over time. Based on the evidence surrounding language-dependent and language-independent skills, instruction that supports both first and second language learning, especially in early grades, may provide the support DLLs need to improve overall academic achievement, regardless of language. However, it is important to note that language proficiency may not benefit from transfer in a bilingual education context in the same way that academic skills such as reading and mathematics might. As evidenced by previous research (Goodrich & Lonigan, 2017; Melby-Lervåg & Lervåg's, 2011), language-dependent skills such as vocabulary do not show the same transfer as language-independent skills. This may hold true in a bilingual education setting as well. Indeed, because language exposure is divided between two languages in a bilingual education program, language proficiency may develop at a different rate than in a monolingual education setting.

Previous Language of Instruction Meta-Analyses

The number of previously conducted meta-analyses on the effects of language of instruction on DLLs' academic outcomes are few, and those that are available have limitations that warrant new investigation into the efficacy of bilingual education programs. One of the earliest meta-analyses included 23 studies of bilingual education programs and submersion programs (Willig, 1985). Willig reported an overall positive effect of bilingual education on English and Spanish outcome measures when looking at adjusted means. When program effects were examined more closely, students in bilingual education programs performed significantly better in reading, language, and total achievement in English than students in English submersion programs. When home language outcomes were available, Willig (1985) reported students in bilingual education programs performed better in listening comprehension, writing, total language, mathematics, and social studies in their home language than students in an English submersion program. Although these initial results were promising, the studies included in the meta-analysis have several limitations. Willig (1985) attributed the issue of the small number of studies included in the analysis to the lack of high-quality studies available at the time of publication. Variables that contributed to low study quality include missing information such as how students were assigned to programs, whether students were comparable before treatment, and lack of descriptive information about students, teachers, and curricula.

Greene (1998) also conducted a meta-analysis to compare students in different language programs and reported similar results to Willig (1985). This meta-analysis focused on studies included in Rossell and Baker's (1996) review of the bilingual education literature. Rossell and Baker collected studies they determined to be methodologically sound and tallied the number of results in which bilingual education programs led to better outcomes than English-only instruction, and the number in which the opposite occurred (Greene, 1998; Rossell & Baker, 1996). Greene (1998) extended these findings by including statistical analyses to account for differences across studies, such as varying sample sizes. After removing studies that did not meet inclusion criteria, there were 11 studies included in the analysis, with only five of them having randomized samples. Greene (1998) concluded that students in bilingual education performed better overall in reading and mathematics when tested in English (g = .18) or in Spanish (g = .74) than students in English-only programs. These results remained when only including studies that utilized random assignment. However, these results should be viewed in the context of the several limitations. First, this analysis included only 11 studies, limiting the potential generalizability of the results. Additionally, Greene (1998) did not report descriptive variables such as program types compared, ages or grades of the participants, first and second languages of the participants, and specific tests that measured outcomes.

Lastly, Greene's (1998) meta-analysis included studies that compared DLLs to non-DLLs, creating a comparison in the results that is not consistent with the primary question involved in evaluating bilingual education programs (i.e., are bilingual education programs effective at improving academic outcomes for DLLs?).

More recently, Rolstad, Mahoney, and Glass (2005) conducted a meta-analysis as an update to Willig's (1985) original paper by only including studies that were published after 1985. They found 17 studies that fit the inclusion criteria with the majority of studies focused on students in first through eighth grades, although there were a few that focused on secondary grades. Overall, there was a small, positive effect of bilingual education (g = 0.08). When separating outcomes by academic subject, there was a negative effect of bilingual education on English reading outcomes (g = -0.06) and a positive effect on English mathematics outcomes (g = 0.08). However, these results include the full sample and do not differentiate between the two different comparison groups (i.e., DLL vs. monolingual and DLL vs. DLL). When comparing DLL students in bilingual education to those in a mainstream program, there was a larger positive effect of enrollment in a bilingual education program overall (g = 0.23). Limitations of Rolstad and colleagues' (2005) meta-analysis are similar to those that came before it. A small number of studies included in the analysis and a lack of descriptive information contributes to a lack of generalizable results.

Slavin and Cheung (2005) conducted a review (n = 17) on how language of instruction affects reading instruction for DLLs. This review did not aggregate effect size values to get an overall value comparing bilingual programs to monolingual programs, but they did acknowledge that 12 out of their 17 studies reported better outcomes for bilingual children in bilingual programs. There were no significant differences between student outcomes in the remaining five studies. Limitations for this analysis include a small number of studies and the inclusion of studies that did not report the appropriate statistical information to compute effect sizes, meaning no statistical differences between programs could be established.

Current Meta-Analysis

Given the limitations of prior meta-analyses examining the effectiveness of bilingual education programs, this meta-analysis will extend previous findings by only including studies that compare two DLL groups in different education programs.

Quantitative data will be used to examine outcomes including reading, language, writing, mathematics, and science in English and in participants' home language, when available. Additionally, the current meta-analysis will use program type as a moderator in analyses to identify any differences between all programs. Lastly, there are a plethora of additional studies, unpublished dissertations and theses, and other gray literature to include in the current meta-analysis that were unexamined in previous meta-analyses either due to the literature not meeting inclusion criteria or because of the time of publication. The specific research questions of the current meta-analysis are as follows:

RQ 1: Are there differences in English and home language academic outcomes between DLLs in a bilingual education program compared to those in an all-English, mainstream programs?

RQ 1 Hypothesis: DLLs in a bilingual education program will not have significantly different L2 academic outcomes from their DLL peers in a L2-only program. However, DLLs in a bilingual education will perform significantly better on L1 outcomes than DLLs in a mainstream, L2-only program (Farver, Lonigan, & Eppe, 2009).

RQ 2: Does program effectiveness differ by academic outcome (e.g., reading, mathematics)?

RQ 2 Hypothesis: For skills that are language-specific (e.g., vocabulary knowledge), DLLs will do better in the language they are instructed in (e.g., Spanish vocabulary benefits from Spanish instruction, English vocabulary benefits from English instruction). In contrast, for skills that are not highly dependent on language and represent more abstract cognitive abilities (e.g., word reading, mathematics calculations), DLLs in bilingual programs will perform better in both languages.

RQ 3: Does program effectiveness differ across different types of bilingual education programs (e.g., dual language, transitional bilingual education)?

RQ 3 Hypothesis: Programs that focus on supporting both languages equally across time will produce better outcomes regardless of language than programs that focus on students achieving English proficiency.

Method

The current meta-analysis used EBSCOhost, ProQuest Dissertations & Theses A&I, and Education Resources Information Center (ERIC) to search for published articles, book chapters, dissertations, and theses. Within EBSCOhost, the databases included in the search were Academic Search Premier, Primary Search, PsycARTICLES, Psychology and Behavioral Sciences Collection, and PsycINFO. The following search terms were used for all databases:

Bilingual* OR English Language Learner* OR Dual Language Learner* OR multilingual OR language minority OR language-minority OR “ELL” OR “DLL” OR “ESL” OR “E2L” OR “ESOL” OR “EAL”

AND Dual language program* OR two-way bilingual immersion program* OR “TWBI” OR two-way dual language program* OR “TWI” OR one-way immersion program* OR one-way dual language program* OR language immersion OR foreign language immersion OR developmental bilingual language program* OR dual language education OR bilingual education OR developmental bilingual education program* OR native language program OR world language immersion program* OR heritage language immersion program*

AND literacy OR read* OR writ* OR math* OR “academic achievement”.

In addition to these search terms, ProQuest includes a feature used to exclude certain subjects that do not pertain to the goal of the search. Irrelevant articles may be included in the initial results because of words that are used in the search that are common across disciplines. This feature was used to exclude articles tagged with topics including, but not limited to film studies, engineering, and home economics. Please see Appendix A for the complete list of excluded topics.

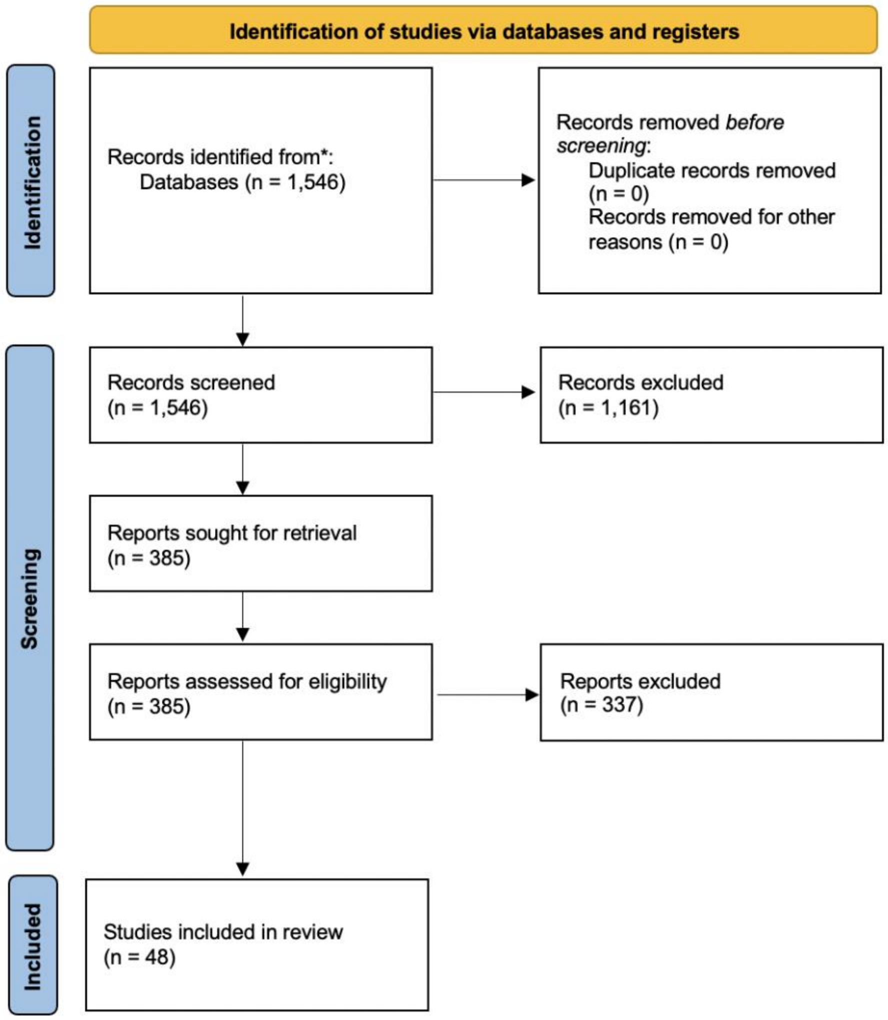
Inclusion Criteria

There were several criteria all studies must meet to be included in the current meta-analysis. All studies must have been available through the online databases no later than January 15, 2021 and written in English. The current meta-analysis is focused on

DLL children's academic achievement in non-traditional language programs. Therefore, to be included in the review, a study must have included DLLs, a bilingual education program, as well as an academic outcome. The studies needed to include typically developing DLLs. That is, studies that only focused on DLLs with learning disabilities or a physical diagnosis such as being visually impaired were not included. Studies that were conducted as experimental, quasi-experimental, or ex-post facto group design studies that measured quantitative outcome variables were included. This eliminated any study that was non-experimental, case studies, and qualitative studies. Lastly, studies needed to report effect sizes or data appropriate for calculating effect sizes. Finally, participants in the studies needed to be language-minority children. That is, the participants' first language must not be an official language of their country, and their second language must be the language of majority. For example, a study about Spanish speakers in the United States would meet this inclusion criteria because Spanish is not the majority language of the United States. On the other hand, a study that recruited French speakers from Quebec, Canada would not meet this inclusion criteria because French and English are both official languages in Canada and they can be used interchangeably in social situations. There were 47 studies that met all inclusion criteria and are included in the final analyses. See Figure 1.

Figure 1

PRIMSA Flowchart



Coded Variables

Study and participant descriptive characteristics were coded for this analysis. Study descriptive characteristics included study author(s), year of publication, experimental design, country the study took place in, and publication status. Additionally, characteristics of the experiment such as outcome variable, language of outcome measure, treatment group program, control group program, and languages of programs were also coded. For participants, grade level, L1 and L2, age of acquisition, and percentage of participants who were eligible for free and reduced-price lunch were coded.

Statistical Analyses

All analyses were conducted using the statistical software R (R Core Team, 2020). Within R, the statistical package robumeta (Fisher, Tipton, & Zhipeng, 2017) was used to conduct all main and moderator analyses. Robumeta was used to analyze the data because there were multiple effect sizes per study. Every study coded had the information needed to calculate the effect size or reported the effect sizes in the results. Hedge's g was used as the current measure of effect size because it is unbiased and accounts for unequal groups in the study (Hedges, 1981). The author also conducted multiple moderator analyses to explore the effects of academic subject, type of bilingual program, and grade level on the effect of treatment.

Interrater Reliability

All studies were coded on multiple variables including descriptive characteristics (e.g., study characteristics, participant characteristics, measure characteristics) and study results (e.g., means and standard deviations). All studies were coded by the first author. An undergraduate student coded 25% of the studies to establish interrater reliability. Prior to reliability coding, the author and undergraduate student met over the course of three weeks to review the codebook and practice coding one study to clarify any questions. There was an average of 98.75% agreement between the author and a trained coder across the studies. Any discrepancies in coding were resolved by the author reviewing the original study and determining which code was correct.

Study Quality

All studies were coded for eight quality indicator variables based on a modified version of the National Institute of Health's Study Quality Assessment Tools for the quality assessment of controlled intervention studies. The quality indicators coded in this meta-analysis were (1) study design, (2), blindness of assessors to participant group assignment, (3) reporting that groups were not statistically different from each other at pre-test, (4) less than 20% attrition in the treatment group, (5) no statistical difference in the rate of attrition between the two groups, (6) fidelity to the intervention, (7) measure reliability, and (8) sufficient power to detect effects.

As with the other coded variables such as study and participant characteristics, quality indicator variables were double coded by the author and a trained undergraduate. Twenty-five percent of the studies were double coded. There were no discrepancies between the author and the second coder's codes.

Publication Bias

Publication bias analyses were conducted on this dataset. Scientific research papers are less likely to be published if the findings were nonsignificant and because of this, gray literature, or literature such as theses and dissertations, are less likely to be included in meta-analyses (Duval & Tweedie, 2000). This can cause a spurious positive outcome in meta-analysis results. To address this and detect the possibility of publication bias, this meta-analysis will use a funnel plot, Egger's test, and a trim and fill funnel plot analysis.

Results

Descriptive Statistics

This meta-analysis included 45 studies (k = 540) from 1972 to 2020 with an average of 12 effect sizes per study (range: 1 – 84). Most studies (n = 41) focused on Spanish-English bilingual programs, but there were also bilingual programs that had Russian-Hebrew (n = 1), Chinese-English (n = 1), and various languages (n = 2). The majority of the studies took place in the United States (n = 43) and two from outside the United States (Israel, n = 1; Canada, n = 1). Twelve of the included studies were published and 33 were not. Please see Table 1 for additional descriptive information. The average effect of a bilingual education program on DLLs' academic outcomes when compared to DLLs in mainstream English programs was $g = 0.028$ ($SE = 0.085$, $p = .742$, CI 95% [-0.144, 0.201], Prediction Interval 95% [-1.403,1.459]). This indicated there was no significant overall effect of enrollment in a bilingual education program on DLLs' academic outcomes. However, the prediction interval indicated substantial heterogeneity in effect sizes across studies, warranting exploration of potential moderators.

Table 1

Summary of Descriptive, Participant, and Treatment Characteristic

Characteristic	*n(k)*	%	Characteristic	*n(k)*	%
Publication Year			*Study Type*		
1970s	2(26)	4.4	RCT	4(116)	19.8
1980s	5(48)	8.1	Quasi-experiment	21(294)	50.1
1990s	1(18)	3.1	Archival/Ex Post Facto	23(177)	30.2
2000s	11(109)	18.5			
2010 - present	29(387)	65.9	*Country of Study*		
Publication Status			United States	46(576)	98.1
Published	13(284)	48.4	Canada	1(5)	.9
Non-published	35(303)	51.6	Israel	2(6)	1
Academic Subject					
Reading	39(281)	47.9	*Grade*		
Language	28(246)	41.9	Early Elementary (PK-3)	37(471)	80.2
Writing	9(25)	4.3	Late Elementary (4-6)	20(116)	19.8
Mathematics	13(35)	6			
Treatment Instructional Program			*Control Instructional Program*		
DL	30(263)	44.8	SEI	35(275)	46.9
TBE	26(314)	53.5	Mainstream	14(312)	53.1
SI	1(10)	1.7			

Note. n = Number of studies, RCT = Randomized Control Trial, DL = Dual Language, TBE = Transitional Bilingual Education, SI = Spanish Immersion, SEI = Structured English Immersion.
Some n sums are larger than the total number of studies in the analysis (n = 48) because some studies included multiple types of variables.

Moderator Analyses

All moderation models were run using meta-regression analyses. Therefore, moderation estimates in the tables should be interpreted as coefficients in the regression model, rather than effect sizes for a specific outcome.

Academic Subject. Academic subject was included as a moderator to further investigate the effects of bilingual education on specific academic domains (Table 2). Reading was entered into the meta-regression model as the intercept; however, it was not significantly different from zero (*p* = .758). Language and writing were not significantly different from reading (*p* = .221 and *p* = .439, respectively). However, mathematics did differ significantly from reading (*p* = .013), suggesting DLLs in a bilingual program performed significantly better on mathematics outcomes than their peers in a mainstream program.

Table 2

Moderation of Academic Subject

					95% Confidence Interval	
	Estimate	Std. Err.	df	p-value	Lower	Upper
Intercept	0.034	0.110	27.7	.758	-0.192	0.260
Language	-0.207	0.167	37.7	.221	-0.545	0.130
Writing	0.239	0.281	4.41	.439	-0.513	0.992
Mathematics	0.471	0.167	14.8	.013	0.115	0.828

Next, language of the outcome measure was included in the moderator analysis to determine if academic performance was affected by whether outcomes were measured in the child's L1 or L2 (Table 3). Outcome language was significantly different from zero (*p* < .001); there was a significant, negative effect of language of outcome measure such that second language outcomes were worse than first language outcomes. Reading outcomes for first language measures was the intercept for this model and was significantly different from zero (*p* < .001). That is, there was a significant, positive effect of bilingual education programs on DLLs' L1 reading skills (*g* = 0.898), Effects for L1 language skills (*g* = 0.573) were significantly smaller than were effects for L1 reading skills, and effects for L1 mathematics skills (*g* = 1.449) were significantly larger than were effects for L1 reading skills. Effects on reading and writing (*g* = .989) skills were not significantly different from each other. Specifically, across all outcomes, effects of bilingual education programs on first language outcomes were approximately one standard deviation larger than effects on second language outcomes. Effects on academic skills in the second language were -0.064 for reading, -0.389 for language, 0.487 for mathematics, and 0.027 for writing outcomes.

Table 3

Moderation of Academic Subject by Language of Outcome Measures

					95% Confidence Interval	
	Estimate	Std. Err.	df	p-value	Lower	Upper
Intercept	0.898	0.180	16.7	<.001	0.517	1.278
Outcome Language	-0.962	0.187	16.9	<.001	-1.358	-0.567
Language	-0.325	0.153	36.5	.040	-0.635	-0.015
Writing	0.091	0.204	4.4	.679	-0.457	0.638
Mathematics	0.551	0.17	14.9	.005	0.195	0.908

Note. Outcome language was dummy coded (0 = first/home language, 1 = second language)

Grade. Grade was analyzed to determine whether it moderated the relation between enrollment in a bilingual program and overall academic performance (Table 4). Because of study constraints (i.e., studies including aggregated grade outcomes), grade was coded as a dichotomous variable for this analysis. Group one included preschool to third grade and group two included fourth to sixth grade. Group one was included as the intercept of the model and was not significantly different from zero. Similarly, group two was not significantly different from group one. Overall, grade did not significantly moderate the relation between enrollment in a bilingual education program and academic outcomes.

Table 4

Moderation of Grade

	Estimate	Std. Err.	df	p-value	95% Confidence Interval	
					Lower	Upper
Intercept	-0.018	0.114	31.6	0.875	-0.251	0.214
Late Elementary (4-6)	0.152	0.137	26.3	0.277	-0.129	0.433

Program Type. Program type was included as a moderator to determine if one type of bilingual program produced significantly different academic outcomes than another bilingual program (Table 5). Dual language was included as the intercept of the model and it was significantly different from zero, indicating a positive overall effect of dual language programs on DLLs' academic outcomes. Transitional bilingual programs were significantly lower than dual language programs (g = -0.229), suggesting DLL students in transitional bilingual programs perform worse than their peers in dual language programs.

Table 5

Moderation of Program Type

	Estimate	Std. Err.	df	p-value	95% Confidence Interval	
					Lower	Upper
Intercept	0.244	0.105	25.4	.029	0.027	0.460
Transitional Bilingual	-0.473	0.159	41.6	.004	-0.795	-0.152

To further examine the effect of program type on academic outcomes, type of bilingual program was separated by the language of the outcome measures. There was a significant effect of language of outcome measure ($p < .001$) such that DLLs' L2 outcomes were significantly worse than their L1 outcomes. L1 outcomes in dual language was the intercept in this model and it was significantly greater than zero (g = 0.910), indicating that dual language programs had large, positive effects on DLLs' academic outcomes in their home language. However, effects of dual language programs on DLL's academic outcomes in their second language were significantly smaller (*g* = 0.115). Outcomes for transitional bilingual programs were significantly different from dual language programs. For home language outcomes, the effect of transitional bilingual education programs was positive (*g* = 0.464). In contrast, the effect of transitional bilingual education programs on second language outcomes was negative (*g* = -0.331).

Table 6

Moderation of Program Type by Language of Outcome Measure

					95% Confidence Interval	
	Estimate	Std. Err.	df	p-value	Lower	Upper
Intercept	0.910	0.188	15.2	<.001	0.508	1.313
Outcome Language	-0.795	0.190	17.3	<.001	-1.197	-0.393
Transitional Bilingual	-0.446	0.153	42.7	.005	-0.755	-0.137

Finally, grade was added to the program type moderation model to determine if effects of program type varied by grade. Like before, dual language was the intercept in the model, however it was not significantly different from zero when grade was added. The effect of grade was not statistically significant, indicating that effects of different dual language programs did not vary significantly across early and late elementary school.

Table 7

Moderation of Program Type by Grade

					95% Confidence Interval	
	Estimate	Std. Err.	df	p-value	Lower	Upper
Intercept	0.200	0.133	23.4	.145	-0.075	0.475
Grade	0.136	0.140	26.1	.339	-0.152	0.424
Transitional Bilingual	-0.468	0.159	41.2	.005	-0.788	-0.148

Study Quality

Study quality indicator averages are reported in Table 8.

Table 8

Study Quality

Quality Indicator	RCT	Quasi-experimental	Archival
1. Study design	19.8%	50.1%	30.2%

Quality Indicator	Yes (%)	No (%)	Partial (5)	NR (%)
2. Blind to group assignment	0	18.1	-	81.8
3. Equal groups (pre-test)	21.9	7.3	-	70.7
4. < 20% attrition in treatment group	17.7	15.5	-	66.6
5. Equal attrition	11.3	15.9	-	72.7
6. Fidelity	11.3	6.8	-	81.8
7. Measure reliability	47.7	-	4.5	47.7
8. Power analysis	11.3	13.6	-	75
Average	17.3	12.8	4.5	70.9

Sensitivity Analysis. Because there were low quality studies included in the main analysis, and one of the main limitations of prior meta-analyses was the lack of high-quality studies included, I conducted a sensitivity analysis where only studies rated as high-quality were included. To be marked as high-quality, a study had to use a randomized control trial design or use a quasi-experimental design and report participants in the treatment and control groups as not statistically different from each other. After screening, there were 12 studies that met the criteria to be considered high quality (k = 163). The average effect size of these studies was g = -0.252 (SE = 0.180, p = .190). When compared to the main analysis, only including high quality studies did not change the main outcome. That is, DLLs enrolled in a bilingual program still did not differ from

DLLs in a mainstream, monolingual program.

To further explore the effects of only including high quality studies, moderator analyses were conducted. The moderator analyses in this section were identical to the ones conducted previously except for the studies included. In the moderation analysis of academic subject, reading was used as the intercept and was not different from zero (g = 0.015, p = .758). This matches the previous finding. Mathematics was no longer significantly different from reading in this analysis (g = 0.316, p = .171). However, effects on language outcomes were significantly smaller than effects on reading outcomes (g = -0.737, p = .003). When academic subject and language of outcome measure were included in the model simultaneously as moderators, from the pattern of results was the same as in the main analysis.

A moderation of grade was also explored. However, like in the main analysis, early elementary scores were not significantly different from zero, and late elementary scores were not significantly different from early elementary scores.

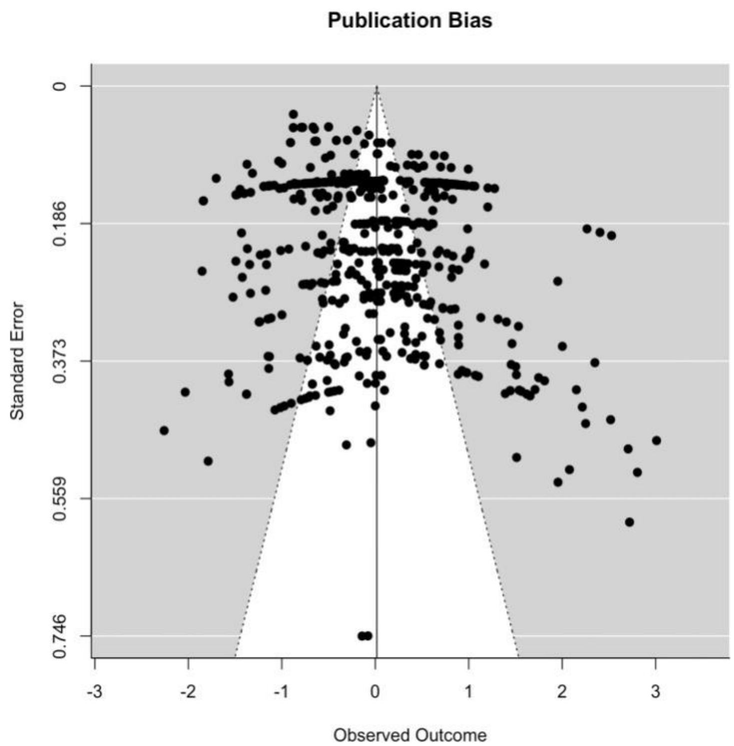
Lastly, program type was used as a moderator. Dual language programs were used as the intercept. Unlike in the main analysis, dual language program scores were not significantly different from zero, and transitional bilingual program scores were not significantly different from dual language program scores. This indicates there was the effects of dual language and transitional bilingual programs did not differ. Language of outcome measure was added to the program type moderation model with dual language L1 scores as the intercept. The results of this analysis matched the main analysis results. That is, there was an overall significant effect of outcome language such that L1 outcomes were significantly better than L2 outcomes regardless of program (L1 outcomes: reading g = 0.933, language g = 0.273, writing g = 1.119, mathematics g = 1.653; L2 outcomes: reading g = -0.269, language g = -0.929, writing g = -0.083, mathematics g = 0.451). Finally, program type and grade were used as moderators. However, these results did not differ from the main analysis – there was no significant moderation.

Publication Bias

I expected there to be no publication bias given the overall effect of bilingual education programs was not statistically different from zero. A moderation model was run with publication status included to determine if there was a significant difference between outcomes of published and non-published studies. Non-published status was included as the intercept and did not differ from zero (p = .920). Published articles did not differ from non-published articles in this model (p = .451). That is, there was no significant difference in outcomes between published and unpublished articles included in this meta-analysis. To determine if there was publication bias, first, a funnel plot was created. Visually, there is slight asymmetry toward positive effect sizes. To test this further, Egger's test was used by adding standard error into a meta-regression model as a moderator to account for effect sizes being nested within studies. According to this model, there was no significant asymmetry in the funnel plot (p = .156). Finally, a trim and fill funnel plot was created; however, there were no effect sizes that were trimmed and filled, suggesting limited evidence of publication bias. Overall, these three analytic approaches converged to suggest minimal evidence of publication bias.

Figure 2

Publication Bias Funnel Plot



Discussion

Although the DLL population continues to increase, there is still no consensus on whether bilingual education programs are an appropriate medium to deliver academic instruction. This hesitation is mainly caused by fears that DLL students will not become proficient in English if they are exposed to two languages during school – thus splitting their exposure to English in half (Krashen, 1991). This meta-analysis sought to consolidate research on the effects of language of instruction on DLLs' academic outcomes and answer three questions. First, are there differences in English and home language academic outcomes between DLLs in a bilingual education program compared to those in an all-English, mainstream programs? Second, does program effectiveness differ by academic outcome? And third, does program effectiveness differ across different types of bilingual education programs?

Effects of Bilingual Education

Overall, there was no significant difference between DLL students in bilingual education programs and DLL students in a mainstream, monolingual programs. Although there were no significant differences in L2 outcomes, DLLs in bilingual education programs did perform significantly better in L1 outcomes than their peers in monolingual programs. These results provide evidence that there are no detrimental effects of supporting DLLs' L1 and L2 in school. Indeed, these results support Cummins' developmental interdependence hypothesis (Cummins, 1979) such that even with less exposure to each language during school, there were no negative effects on overall academic outcomes in their L2. Students in bilingual programs were able to build knowledge and skills in both languages without negative consequences to L2 outcomes. That is, students may be able to leverage their learning in either their L1 or L2 to support learning in their other language, regardless of the divided exposure.

To further explore this relation between language of outcome measures and bilingual education, I examined whether there were differences in outcomes between specific academic subjects. When looking at whether there were differences between specific academic outcomes, only mathematics was significantly different from zero such that DLLs in a bilingual education program performed better on mathematics outcomes than their peers in a mainstream program. Prior researchers have provided evidence of a discrepancy in mathematics scores between DLL and monolingual students that can be attributed to their limited English proficiency in conjunction with the complexity of mathematics-specific language and vocabulary (Linquanti & Cook, 2013; Powell et al., 2020). If students are unable to become proficient in mathematics-specific language while foundational knowledge is being built, the compounding effects of this disparity of knowledge can become more pronounced as mathematics concepts become more advanced (Driver & Powell, 2016). However, if students are taught mathematics concepts and procedures in their first language – therefore removing one obstacle to their ability to learn – DLLs can perform at similar levels as their monolingual peers (Turner & Celedón-Pattichis, 2011). Indeed, results from this meta-analysis support this conclusion DLL students in a bilingual program performed significantly better on mathematics outcomes than DLL students in mainstream education programs.

Further, DLL students in bilingual education programs performed similarly to their peers in mainstream programs in reading, writing, and language outcomes. This suggests there is no negative effect of dual language exposure on DLLs academic outcomes. That is, dividing instruction between two languages does not hinder DLL students' ability to learn. In fact, in the case of mathematics specifically, it may support understanding and improve outcomes.

Effects of Different Program Types

Type of bilingual program had a significant effect on student outcomes. Dual language program outcomes were significantly greater than zero and Spanish immersion outcomes were significantly better than dual language outcomes. However, transitional bilingual program outcomes were, overall, significantly lower than dual language program outcomes. To parse these differences out further, language of outcome measure was examined and was found to have a significant effect on outcomes. Across all programs, L2 outcomes were significantly lower than L1 outcomes. Interestingly, however, transitional bilingual education was the only program out of the three to have a negative effect on English outcomes. This may be due to how transitional bilingual education programs are designed. In transitional bilingual education programs, students begin with a high amount of exposure to their first language (e.g., Spanish) and a lower amount of exposure to their second language (e.g., English) in early elementary grades. Around third or fourth grade, the ratio becomes an even division of first and second language exposure and each year that follows will have a higher ratio of L2 to L1 exposure. Because high exposure to the L2 is delayed, DLL students may not build L2 proficiency as quickly as students in other types of bilingual education, thus causing lower L2 outcomes overall. Additionally, because of the nature of transitional bilingual education programs, grade was investigated as a contributing factor to bilingual program outcomes. There was no overall difference between outcomes in early elementary school (PK – grade 3) and late elementary school (Grade 4 – Grade 6). However, the small difference between outcomes in early and late elementary school indicate that negative effects of transitional bilingual programs on outcomes in late elementary school are very small (g = -0.31 in early elementary grades and g = -0.151 in later elementary grades). Future research should explore the long-term effects of transitional bilingual education programs on students' L2 academic outcome beyond the transition to L2 instruction. Although transitional programs show a negative effect on L2 outcomes in elementary school, it is possible that because they are focused on providing a foundation in students' L1 first, there may be a long-term, positive effect on L2 academic outcomes because of the initial support in L1.

Limitations

Results of this meta-analysis should be viewed in the context of its limitations. One limitation is the design of the studies included in this analysis. The majority of studies were quasi-experimental (n = 21) or used archival data in an ex post facto design (n = 23) while only four studies used a randomized control trial (RCT) design. These designs are useful when students are already enrolled in a bilingual program. However, it does not ensure there are no confounding differences between participants in each group. Even if students in each group (e.g., bilingual and monolingual programs) do not score significantly differently in their pre-test scores, there can be additional factors that may influence students' performance. For example, parents who enroll their child in a bilingual education program may place a greater value of bilingualism in their child than parents who do not enroll their child in a bilingual education program. A parent who places great value on bilingualism may support both of their child's languages outside of school. Although researchers cannot control for every factor such as these, RCT designs can mitigate the risk that these confounding factors will have an effect on final results. To address this limitation, I conducted a sensitivity analysis with high-quality studies to directly test the effect of study quality. Overall, the average effect of bilingual education was not significant in the sensitivity analysis, which is consistent with the results of the main analysis. There were differences in the moderation analyses. However, because of how few high-quality studies there were (n = 12), these differences should be considered in context and carefully as there was limited power to detect effects.

Conclusion

This meta-analysis examined the effects of bilingual education on DLLs' academic outcomes when compared to their peers in mainstream education and added to existing literature by extending previous meta-analyses on this topic. The results from this meta-analysis provide evidence that bilingual education is a viable option for DLLs to not only gain English proficiency but also have continued support of their first language. In particular, programs that divided language exposure evenly and consistently over grade levels, like dual language programs, tended to be the most beneficial for both L1 and L2 outcomes. This may provide additional context for Cummins' developmental interdependence hypothesis (Cummins, 1979) such that effects of transfer are most prominent and effective when both languages are supported in a structured learning environment such as a bilingual education program. Researchers should continue to examine the benefits of bilingual education while ensuring high-quality research techniques - such as RCT group designs – are used.

References

*References marked with an asterisk indicate studies included in the meta-analysis

*Acosta, S. T. (2010). *High-stakes reading assessment and English oral language development: A study of third grade English language learners in a Texas school district* [Unpublished doctoral dissertation]. Texas A&M University.

Baker, D., Basaraba, D. L., & Polanco, P. (2016). Connecting the present to the past: Furthering the research on bilingual education and bilingualism. *Review of Research in Education*, *40*, 821-883. https://doi.org/10.3102/0091732X16660691

Boyle, A., August, D., Tabaku, L., Cole, S., Simpson-Baird, A. (2015). Dual language education programs: Current state policies and practices. *Office of English Language Acquisition (OELA) and Office of State Support (OSS)*.

*Carlisle, J. F. & Beeman, M. M. (2000). The effects of language of instruction on the reading and writing achievement of first-grade Hispanic children. *Scientific*

Studies of Reading, *4*(4), 331-353. https://doi.org/10.1207/S1532799XSSR0404_5

*Calero, F. R. (2012). *Fifth-grade English language learner academic self-concept, student-teacher relationships, self-regulated learning, parental academic support, native language support, interest, usage, proficiency and academic achievement.* [Unpublished doctoral dissertation]. Dowling College.

*Coletti, S. L. (2012). *The effects of transitional bilingual education versus structured English immersion instructional models on English language development as measured by the California English Language Development Test* [Unpublished doctoral dissertation]. University of California, Santa Barbara.

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

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of ‘No Child Left Behind’. *Race Ethnicity and Education*, *10*(3), 245-260. https://doi.org/10.1080/13613320701503207

Darling-Hammond, L., Bae, S., Cook-Harvey, C. M., Lam, L., Mercer, C., Podolsky, A., Stosich, E. L. (2016). *Pathways to new accountability through the Every Student Succeeds Act*. Learning Policy Institute.

*Davis, J. M. (2008). *The effectiveness of a late-exit/transitional bilingual program related to the reading achievement of Hispanic limited English proficiency elementary school students* [Unpublished doctoral dissertation]. Texas A&M University-Kingsville.

*DeCamps, M. (2016). *A comparison of English language learner programs: A pilot* *study*. [Unpublished doctoral dissertation]. Fairleigh Dickinson University.

*de la Garza, J. V. (1984). *An evaluation study of a transitional bilingual education program* [Unpublished doctoral dissertation]. University of Arizona.

*Diemer, R. (2017). *Reading achievement in a dual language setting: An examination of the characteristics representative of proficient English language learners*. [Unpublished doctoral dissertation]. University of South Dakota.

*Dow, P. A. (2008). *Dual-language education: A longitudinal study of students' achievement in an El Paso County, Texas school district* [Unpublished doctoral dissertation]. University of Texas at El Paso.

Driver, M. K. & Powell, S. R. (2016). Culturally and linguistically responsive schema intervention: Improving word problem solving for English language learners with mathematics difficulty. *Learning Disability Quarterly*, *40*(1), 41-53. https://doi.org/10.1177/0731948716646730

*Durán, L. K., Roseth, C. J., & Hoffman, P. (2009). An experimental study comparing English-only and transitional bilingual education on Spanish-speaking preschoolers' early literacy development. *Early Childhood Research Quarterly*, *25*, 207-217. https://doi.org/10.1016/j.ecresq.2009.10.002

*Durán, L., Roseth, C. J., & Hoffman, P. (2015). Effects of transitional bilingual education on Spanish-speaking preschoolers' literacy and language: Year 2 results. *Applied Psycholinguistics*, *36*, 921-951. https://doi.org/10.1017/S0142716413000568

Duval, S. & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455-

463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Elementary and Secondary Education Act, Public Law No. 89-10, (1965). https://www.govinfo.gov/content/pkg/COMPS-748/pdf/COMPS-748.pdf

*Escatel, G. C. (2018). *A comparative analysis of third-grade, Hispanic students reading scores on the ACCESS* [Unpublished doctoral dissertation]. Capella University.

Every Student Succeeds Act, Public Law No. 114-95, (2015). https://www.congress.gov/bill/114th-congress/senate-bill/1177

Farver, J. M, Lonigan C. J., & Eppe, S. (2009). Effective early literacy skill development for young Spanish-speaking English language learners: An experimental study of two methods. *Child Development*, *80*(3), 703-719. https://doi.org/10.1111/j.1467-8624.2009.01292.x

*Figueroa, L. R. (2007). *The development of pre-reading and reading knowledge in English and Spanish in a dual language education context* [Unpublished doctoral dissertation]. University of California Santa Barbara.

Fisher, Tipton, & Zhipeng (2017). Robust variance meta-regression. R package. http://cran.uni-muenster.de/web/packages/robumeta/robumeta.pdf

*Franquiz, M. I. (1998). *The effects of bilingual education on academic achievement, language development, and self-esteem of Hispanic children* [Unpublished doctoral dissertation]. Florida State University.

*Gallegos-Jaramillo, L. (1985). *Factors influencing reading ability and self esteem: A study of bilingually education and non-bilingually educated students* [Unpublished doctoral dissertation]. University of Illinois at Urbana-Champaign.

*Galvan-Luis, S. (2010). *The impact of bilingual education on academic achievement* *and language development of third grade English language learners in Texas* [Unpublished doctoral dissertation]. Texas A&M University Corpus Christi.

*Garza, E. (1982). *The effects of transfer from Spanish reading to English reading* [Unpublished doctoral dissertation]. Brigham Young University.

Greene, J. P. (1998). A meta-analysis of the effectiveness of bilingual education. *The Tomas Rivera Policy Institute*.

Goodrich, J. M. & Lonigan, C. J. (2017). Language-independent and language-specific aspects of early literacy: An evaluation of the common underlying proficiency model. *Journal of Educational Psychology*, *109*(6), 782-793. https://psycnet.apa.org/doi/10.1037/edu0000179

Goodrich, J. M., Thayer, L., Leiva, S. (2021). Evaluating achievement gaps between monolingual and multilingual students. *Educational Researcher*. https://doi.org/10.3102/0013189X21999043

*Haubrich, M. K. (2010). *An examination of the relationship between language-based instructional strategies and academic achievement* [Unpublished doctoral dissertation]. Tarleton State University.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational and Behavioral Statistics, 6*(2), 107-128. https://doi.org/10.3102/10769986006002107

*Herrera, M. K. (2020). *Dual language education, English as a new language support, transitional bilingual education and monolingual classroom education: A comparison of academic achievement among third graders* [Unpublished doctoral dissertation]. Fairleigh Dickinson University.

Higham (1992). Crusade for Americanization. In J. Crawford (Ed.), *Language loyalties:*

A source book on the official English controversy (pp. 72-85). The University of Chicago Press.

*Hipfner-Boucher, K., Lam, K., & Chen, X. (2014). The effects of bilingual education on the English language and literacy outcomes of Chinese-speaking children. *Written Language and Literacy*, *17*(1), 116-138. https://doi.org/10.1075/wll.17.1.06hip

Hursh, D. (2007). Exacerbating inequality: The failed promise of the No Child Left Behind Act. *Race Ethnicity and Education*, *10*(3), 295-308. https://doi.org/10.1080/13613320701503264

*Jaime-Mileham, M. G. (2020). *Strategies to support native Spanish language acquisition and English development among preschoolers* [Unpublished doctoral dissertation]. California State University, Fresno.

Kieffer, M. J. & Thompson, K. D. (2018). Hidden progress of multilingual students on NAEP. *Educational Researcher*, *47*(6), 391-398. https://doi.org/10.3102%2F0013189X18777740

Kim, Y. K., Hutchinson, L. A., & Winsler, A. (2015). Bilingual education in the United States: An historical overview and examination of two-way immersion. *Educational Review*, *67*(2), 236-252. https://doi.org/10.1080/00131911.2013.865593

Krashen, S. D. (1991). Bilingual education: A focus on current research. *National Clearinghouse for bilingual Education*.

*Leafstedt, J. M. & Gerber, M. M. (2005). Crossover of phonological processing skills: A study of Spanish-speaking students in two instructional settings. *Remedial and Special Education*, *26*(4), 226-235.

https://doi.org/10.1177/07419325050260040501

*Lewis, G. D. (2012). *The effects of bilingual Spanish language response to intervention processes in comparison to an English response to intervention process on English and Spanish literacy in grades* [Unpublished doctoral dissertation]. National-Louis University.

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

Linquanti, R. & Cook, H. G. (2013). Toward a “common definition of English learner”: Guidance for states and state assessment consortia. *Council of Chief State School Officers.*

*Lopez, M. E. (2000). *A comparative study on the role of phonological awareness on Spanish and English reading acquisition for Spanish speaking first-graders* [Unpublished doctoral dissertation]. University of Oregon.

Melby-Lervåg, M. & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*(1), 114-135. https://doi.org/10.1111/j.1467-9817.2010.01477.x

*Merlos, R. L. (1978). *Effects of bilingual education on the cognitive characteristics of the Spanish speaking children in Chicago public schools* [Unpublished doctoral dissertation]. Loyola University Chicago.

*Meyer, T. R. (2017). *The effectiveness of dual language and sheltered English immersion ESOL programs: A comparative study* [Unpublished doctoral dissertation]. Liberty University.

*Murillo, H. A. (1986). *A comparison of the effects of dual language and intensive English instruction on kindergarten students in the Laredo independent school district* [Unpublished doctoral dissertation]. University of Nebraska-Lincoln.

*Nakamoto, J., Lindsey, K. A., & Manis, F. R. (2012). Development of reading skills from K-3 in Spanish-speaking English language learners following three programs of instruction. *Reading and Writing*, *25*, 537-567. https://doi.org/10.1007/s11145-010-9285-4

National Defense Education Act, Public Law No. 85-864 (1958). https://www.govinfo.gov/content/pkg/STATUTE-72/pdf/STATUTE-72-Pg1580.pdf

No Child Left Behind Act (2001), Public Law No. 107-110.

*Ortega, R. E. (2013). *Achievement, engagement, and English language acquisition of fourth-grade ELL students following five years of instruction in an English as a second language program or Instruction in a dual language program in the same urban elementary school* [Unpublished doctoral dissertation]. University of Nebraska.

Ovando, C. J. (2003). Bilingual education in the United States: Historical development and current issues. *Historical and Current Issues*, *27*(1), 1-24. https://doi.org/10.1080/15235882.2003.10162589

Powell, S. R., Berry, K. A., Tran, L. M. (2020). Performance differences on a measure of mathematics vocabulary for English learners and non-English learners with and without mathematics difficulty. *Reading & Writing Quarterly*, *32*(2), 124-141. https://doi.org/10.1080/10573569.2019.1677538

R Core Team (2020). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*. https://www.R-project.org

*Reese, L., Goldenberg, C., & Saunders, W. (2006). Variations in reading achievement among Spanish-speaking children in different language programs: Explanations and confounds. *The Elementary School Journal*, *106*(4), 363-385. https://doi.org/10.1086/503637

Rolstad, K., Mahoney, K. & Glass, G. V. (2005). The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy*, *19*(4), 572-594. https://doi.org/10.1177/0895904805278067

Rossell, C. H. & Baker, K. (1996). The educational effectiveness of bilingual education. *Research in the Teaching of English*, *30*(1), 7-74. https://www.jstor.org/stable/40171543

*Saldate, M., Mishra, S. P., & Medina, M. (1985). Bilingual instruction and academic achievement: A longitudinal study. *Journal of Instructional Psychology*, *12*(1), 24-30.

*Sanders, A. N. (2010). *The effectiveness of two-way bilingual immersion programs in closing the achievement gap for minority students* [Unpublished doctoral dissertation]. Texas Southern University.

*Schwartz, M. (2014). The impact of the *First Language First* model on vocabulary development among preschool bilingual children. *Reading and Writing*, *27*, 709-732. https://doi.org/10.1007/s11145-013-9463-2

*Slavin, R. E., Madden, N., Calderón, M., Chamberlain, A., & Hennessy, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis*, *33*(1), 47-58. https://www.jstor.org/stable/41238537

*Slavin, R. E., Madden, N., Calderón, M., Chamberlain, A., & Hennessy, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis*, *33*(1), 47-58. https://doi.org/10.3102%2F0162373711398127

Slavin, R. E. & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Education Research*, *75*(2), 247-284. https://doi.org/10.3102/00346543075002247

*Tazi, Z. (2011). *The effects of bilingual instruction on the English emergent literacy skills of Spanish-speaking preschool children* [Unpublished doctoral dissertation]. University of New York.

*Tong, F. (2006). *Oral English development and its impact on emergent reading achievement: A comparative study of transitional bilingual and structured English immersion models* [Unpublished doctoral dissertation]. Texas A&M University.

*Tong, F., Lara-Alecio, R., Irby, B., Mathes, P., Kwok, O. (2008). Accelerating early academic oral English development in transitional bilingual and structured English immersion programs. *American Educational Research Journal*, *45*(4), 1011-1044. https://www.jstor.org/stable/27667161

Turner, E. E. & Celedón-Pattichis, S. (2011). Mathematical problem solving among Latia/o kindergarteners: An analysis of opportunities to learn. *Journal of Latinos and Education*, *10*(2), 146-169. https://doi.org/10.1080/15348431.2011.556524

U.S. Department of Education, National Center for Educational Progress (2017). NAEP

Reading report card for the nation and states.

U.S. Department of Education, Office of English Language Acquisition (2019). Dual language education programs: current state policies and practices. https://www.air.org/sites/default/files/downloads/report/Dual-Language-Education-Programs-Current-State-Policies-Feb-2017-rev.pdf

U.S. Department of Education, National Center for Education Statistics, Common Core of Data (2020). Local education agency universe survey, 2000-01 through 2016-2017. https://nces.ed.gov/programs/digest/d18/tables/dt18_204.20.asp?current=yes

*Valdez, A. (2012). *The impact of writing achievement of two bilingual education models for English language learners* [Unpublished doctoral dissertation]. Walden University.

*Vega, L. D. (2014). *Effects of an elementary two way bilingual Spanish-English immersion school program on junior high and high school student achievement* [Unpublished doctoral dissertation]. Colorado State University.

*Vela, A., Jones, D., Mundy, M., & Isaacson, C. (2017). Determining the effectiveness of bilingual programs on third grade state exam scores. *Research in Higher Education Journal*, *33*.

*Wages, M. M. (2013). *A comparison of two bilingual programs on student reading achievement in a public elementary school in Texas* [Unpublished doctoral dissertation]. Capella University.

*Walichowski, M. F. (2009). *An alternative oral proficiency and expressive vocabulary assessment of kindergarten English language learners* [Unpublished doctoral dissertation]. Texas A&M University.

*Williams, J. L. (2011). *A difference in ELL: A study of 50/50 dual-immersion bilingual literacy program for second grade English language learners* [Unpublished doctoral dissertation]. Capella University.

Willig, A. C. (1985). A meta-analysis of selected studies on the effectiveness of bilingual education. *Review of Educational Research*, *55*(3), 269-317. https://doi.org/10.3102/00346543055003269

*Zirkel, P. A. (1972). *An evaluation of the effectiveness of selected experimental bilingual education programs in Connecticut* [Unpublished doctoral dissertation]. University of Connecticut.

*Zito-Nash, J. (2017). *Impact of sheltered instruction observation protocol (SIOP) on reading achievement of English language learners in the primary grades* [Unpublished doctoral dissertation]. University of St. Francis College.