



JÖNKÖPING UNIVERSITY

*School of Education and
Communication*

LANGUAGE RELATED OUTCOMES OF BILINGUAL EDUCATION IN PRESCHOOL AND PRIMARY SCHOOL

A SYSTEMATIC REVIEW FROM 2000 TO 2016

LUCIO LAZZARINO

One Year Master Thesis 15 Credits

Interventions in Childhood

Spring Semester 2017

Supervisor

Håkan Nilsson

Examiner

Eva Björck-Åkesson

ABSTRACT

Author: Lucio Lazzarino

Language Related Outcomes of Bilingual Education in Preschool and Primary School

A systematic review from 2000 to 2016

Good language skills are essential to academic success. Immigrant and refugee children who enter school without previous knowledge of the societal language are more prone to failure and need of special support. The aim of this study is to describe bilingual educational program used in preschool and primary school and to examine their outcomes related to language development, both for the home language (L1) as well as the school language (L2). 17 studies were identified through a systematic literature review. Results showed a predominance of the transitional bilingual education (TBE) and two-way immersion (TWI) models in bilingual education. Language related outcomes confirmed the finding from previous literature that bilingual education doesn't inhibit L2 acquisition. Also, confirming previous literature, advantages of bilingual programs over monolingual ones are proven hard to confirm. However, several methodological issues addressed by the previous meta-analysis seem to generally persist in the most recent literature. The results of this study reiterate the need for more high-quality study in the field. Moreover, future research should also include experimentation with different languages. Finally, this argues the interest to further study and implement bilingual education programs to better accommodate the need of children with a migration background.

Pages: 44

Keywords: *Bilingual Education, Minority Children, Language Outcomes, Preschool, Primary School, Systematic Literature Review*

| Postal Address | Street address | Telephone | Fax |
|--|-----------------------|------------------|------------|
| Högskolan för lärande och kommunikation (HLK) Box 1026 551 11 JÖNKÖPING | Gjuterigatan 5 | 036-101000 | 036162585 |

TABLE OF CONTENTS

| | |
|--|-----------|
| 1. INTRODUCTION..... | 5 |
| 1.1 THEORETICAL BACKGROUND..... | 6 |
| 1.1.1 <i>Defining bilingualism.....</i> | 6 |
| 1.1.2 <i>Language proficiency at home and in school</i> | 7 |
| 1.1.3 <i>Cross-linguistic transfer</i> | 7 |
| 1.1.4 <i>Typologies of bilingual education programs</i> | 8 |
| 1.2 PREVIOUS RESEARCH | 10 |
| 1.3 RATIONALE | 11 |
| 1.4 AIM | 11 |
| 1.5 RESEARCH QUESTIONS | 11 |
| 2. METHOD | 11 |
| 2.1 SEARCH PROCEDURE | 12 |
| 2.1.1 <i>Search terms.....</i> | 12 |
| 2.2 SELECTION PROCESS..... | 12 |
| 2.2.1 <i>Inclusion/Exclusion Criteria.....</i> | 13 |
| 2.2.2 <i>Title and abstract screening.....</i> | 14 |
| 2.2.3 <i>Full-text screening</i> | 14 |
| 2.3 QUALITY ASSESSMENT..... | 14 |
| 2.4 DATA EXTRACTION | 15 |
| 2.5 METHODOLOGICAL LIMITATIONS OF THE PRESENT STUDY | 15 |
| 3. RESULTS | 17 |
| 3.1 TYPES OF PROGRAM IMPLEMENTED IN THE REVIEWED STUDIES | 18 |
| 3.1.1 <i>Preschool</i> | 18 |
| 3.1.2 <i>Primary School</i> | 19 |
| 3.2 LANGUAGE OUTCOMES OF BILINGUAL EDUCATION | 21 |
| 3.2.1 <i>Language outcomes compared to L2-only programs</i> | 23 |
| 3.2.2 <i>Language outcomes compared to other bilingual programs.....</i> | 24 |
| 4. DISCUSSION | 25 |
| 4.1 TYPES OF BILINGUAL EDUCATIONAL PROGRAMS | 25 |
| 4.2 LANGUAGE RELATED OUTCOMES OF BILINGUAL EDUCATIONAL PROGRAMS | 26 |
| 4.3 METHODOLOGICAL ISSUES OF THE STUDIES..... | 26 |
| 4.4 FUTURE RESEARCH..... | 27 |
| 5. CONCLUSION | 27 |

| | |
|--|-----------|
| REFERENCES..... | 29 |
| APPENDIX..... | 34 |
| A. TABLE DISPLAYING SEARCH WORDS USED IN EACH DATABASE DURING THE SEARCH PROCEDURE..... | 34 |
| B. QUALITY ASSESSMENT TOOL. | 35 |
| C. QUALITY ASSESSMENT SCORES OF THE REVIEWED STUDIES. | 37 |
| D. DATA EXTRACTION AND SELECTION PROTOCOL. | 38 |
| E. TABLE DISPLAYING GENERAL INFORMATION ABOUT THE REVIEWED STUDIES. | 39 |
| F. LIST OF ABBREVIATIONS. | 44 |

1. Introduction

«I prefer that a boy should begin with Greek, because Latin being in general use, will be picked up by him whether we will or not; while the fact that Latin learning is derived from Greek is a further reason for his being first instructed in the latter. I do not, however, desire that this principle should be so superstitiously observed that he should for long speak and learn only Greek, as is done in the majority of cases. Such a course gives rise to many faults of language and accent; the latter tends to acquire a foreign intonation, while the former through force of habit becomes impregnated with Greek idioms, which persist with extreme obstinacy even when we are speaking another tongue. The study of Latin ought, therefore, to follow at no great distance and in a short time proceed side by side with Greek. The result will be that, as soon as we begin to give equal attention to both languages, neither will prove a hindrance to the other. »

(Quintilian, *Institutio Oratoria*, trans. 1922)

Although one can think about bilingual education as a relatively new topic related to the recent migratory phenomena, this quote from the Latin orator Quintilian shows how bilingual education has been discussed, at least in some form, as far as 2000 years ago. What is perhaps more interesting, is that the ancient orator Quintilian recognises topics and problematics of bilingual education that are of interest even now. He identifies, for example, the difference between the societal language and languages related to other contexts; he admits the possibility to learn and speak two languages at the same time in school; and, finally, he proposes that learning two languages will not hinder either learning, if we *give equal attention to both*. These problems and questions related to language learning are very like what educational research has discussed in the recent decades, as it will be explained during this introductory chapter.

While learning Greek and Latin was the problem of the Rome's ruling elite education, today learning a new language when resettling in a new country is an extremely important aspect of refugee children's and adults' adaptation alike. The language barrier is often cited in the literature as a component of refugee children's educational special needs, both in preschool and in primary school (Dryden-Peterson, 2016; Hamilton & Moore, 2003; Strelakova & Hoot, 2008; Szente, Hoot, & Taylor, 2006).

Refugee and immigrant children in pre-school settings are also confronted with the task to acquire a new language. However, the research pointed out the importance for children to maintain and develop their skills in the primary language, especially for children under the age of five (Loewen, 2003). This suggests that both preschool and primary school environment should accept and encourage the use of the children's first language in the classroom.

Moreover, the children from a migration background are more likely to be the object of special education measure. This leads to an overrepresentation of this population in the special school in some European countries (Wernig, Lösen & Urban, 2008; Sermier, Dessemontet & Bless, 2013).

Finally, bilingual education can be connected to children's rights. The Convention on the Rights of the Child (United Nations General Assembly, 1989) contains article pertaining education and concerning also children's cultural identity and their right to practise their culture and their own language (Smidt, 2016). The Article 29 of the CRC specifically cites the respect of the children's own cultural identity and language as one of the goals of compulsory primary education granted in the article 28 (United Nations General Assembly, 1989). UNESCO is also committed to multilingual education and emphasises the importance of proficiency of children in their mother language (Smidt, 2016).

1.1 Theoretical Background

1.1.1 Defining bilingualism

Grosjean (in Baker 2011, p. 4), defines bilingualism as «... those who use two or more languages (or dialects) in their everyday lives». Baker (2011) points out how language's use must be understood and studied in the context in which is used. An individual is a functional bilingual when he/her can use language in a variety of everyday context and events. However, a bilingual individual can have different abilities in each language and thus bilingualism is a very complex phenomenon that can have different practical outcomes. Baker (2011, p. 207) points out how the term bilingual education is «a simplistic label for a complex phenomenon». This complexity is evident also from the wide array of terms to describe children that have not adequate skill in the *Language 2* (L2) the language used at school. Baker, Basaraba, and Polanco (2016) report four different labels that, with slightly different perspectives, indicate bilingual students: *Dual Language Learner* (DLL), *Emergent Bilingual* (EB), *English Language Learner* (ELL) and *Limited English Proficient* (LEP).

In this study, we will use the terms *bilingual children/bilingual students*, to refer to those children that spoke at home a language (L1) that differs from the school and the societal language (L2).

1.1.2 Language proficiency at home and in school

Language proficiency can be assessed and divided into a variety of skills (Baker, 2011). However, there are four basic language abilities: listening, speaking, reading and writing. This four can be divided into receptive skills (listening, reading) and in productive skills (speaking, writing).

Each of these abilities can be divided into numerous sub-skills. For example, «... someone may listen with understanding in one context (e.g. shops) but not in another context (e.g. an academic lecture) » (Baker, 2011, p. 7).

| | Oracy | Literacy |
|--------------------------|--------------|-----------------|
| Receptive skills | Listening | Reading |
| Productive skills | Speaking | Writing |

This multidimensionality has been reflected in empirical research. For example, Hakuta, Butler, and Witt (2000) found that basic oral English proficiency in children takes three to five years to be fully developed, while academic language proficiency may take between four to seven years. Other results in this sense are described by Cummins (2000).

Following this type of results, Cummins (2000) introduced a distinction between *basic interpersonal communicative skills* (BICS) and *cognitive/academic language proficiency* (CALP). CALP is characterised by context reduced communication, where non-linguistic cues are virtually absent, and by cognitively demanding communication; on the other hand, BICS are characterised by a considerable level of nonlinguistic cues (e.g., body communication, context) and cognitively undemanding communication (Baker, 2011; Cummins, 2000). Cummins (Baker, 2011; 2000), suggests that BICS in a second language develop independently of the first language, while CALP develops interdependently in both languages.

According to Baker (2011, p. 172), in the case of children that appeared to be fluent in L2 and are transferred in mainstream classes «... Cummins' distinction between BICS and CALPS explains why such children tend to fail when mainstreamed».

1.1.3 Cross-linguistic transfer

The influence of the L1 is among the most cited theoretical notions in bilingual education and in second language acquisition in general (Cummins, 2001, Baker, 2011), even though a comprehensive theory of transfer is still missing (Butler, 2012). Odlin (1989, 27) defines the

notion as «the influence resulting from the similarities and differences between the target language and any other language that has been previously (and perhaps imperfectly) acquired». The transfer can result in positive outcomes (i.e., *facilitations*) or negative (i.e., *interferences*). Butler (2012) identified several factors involved in the cross-linguistic transfer, the most important being: (1) language distance, in terms of similarities between L1 and L2; (2) developmental stage; (3) age; and (4) sociolinguistic factors. However, he acknowledges that the mechanisms of these influences are still largely unknown.

On this base, Cummins (2001) proposed the *threshold* (or *interdependence*) *hypothesis*. He proposes that, for a bilingual student, positive transfer between L1 and L2 occurs only when a certain level of proficiency in L1 is reached. This should be especially true for the academic proficiency level required by the school.

The possibility of a positive transfer between L1 and L2 and the threshold hypothesis provide a strong rationale to support bilingual education programs (Baker, 2011).

1.1.4 Typologies of bilingual education programs

Terminology and distinctions between bilingual educational programs tend to be quite ambiguous and has been the object of critics (Baker, 2011; Slavin & Cheung, 2005). Cummins (2000) argues that the central issue is to know whether the program is enrichment-oriented or remedial-oriented. He subsequently divides bilingual programs into three types: (1) second language immersion programs, where children are taught in L2 with or without support from an L1 teacher; (2) developmental bilingual programs, where the language minority is taught somewhere around 50% of the time in L1; and (3) two-way immersion or dual-language programs. Both the language minority and the language majority are in the same class and are taught either 50% or 90% of the time in L2.

Table 1: Programs aimed at bilingual students, adapted from Baker et al. (2016).

| Aim | Denomination | Description |
|--|---|--|
| Monolingualism/Second language immersion | Submersion | No special service is provided to language minorities. The instruction is conducted only in L2. |
| | English as a second language (ESL) | Instruction is aimed at developing L2 skills. Two modalities are present: (1) pull-out, where the children are removed from the class and work with an L2 teacher; and (2) pull-in where the ESL teacher work in the classroom. |
| | English for speakers of other languages (ESOL) | This model contains individualized instruction for the children that are not proficient in L2. |
| Developmental/Weak Bilingual | Sheltered (or sheltered) English Submersion (SEI) | The curriculum is adapted to L2 learner's, but all the instruction (literacy included) is conducted in L2. |
| | Transitional bilingual education (TBE) | Children are taught primarily in L1, and then gradually transitioned to L2. Early-exit models conclude the transition in 2 nd or 3 rd grade, while late-exit models continue throughout elementary school. |
| Bilingualism/Biliteracy | Extended foreign language (EFL) | Instruction is 70% in L2 and 30% in L1. Language instruction time is divided by content. |
| | Two-way Immersion | Both L1 and L2 native speakers are enrolled in the program. Academic instruction is provided in both languages. |
| | Two-way bilingual education | The goal is to facilitate the development of L2 skills and the cultural integrity of students. |
| | Academic language acquisition (ALA) | Instruction is divided into language blocks. The student's goal is to become proficient in both L1 and L2. |
| | 50/50 content model | Mathematics instruction is provided in L2, while social studies and science instruction are in L1. Other subjects alternate language of instruction. |

Baker (2011), cites ten theoretically possible bilingual educational programs ranging from immersion model in L2 to a separationist model where children are only taught in L1. He discriminates between three types of program depending on the in intended aim in language outcome, that is: (1) monolingualism (e.g., L2 immersion programs); (2) weak bilingualism (e.g., transitional programs); and (3) bilingualism and biliteracy (e.g., two-way/dual language programs).

Baker et al. (2016), in a review of bilingual education, found several different typologies of bilingual education, which are summarized in Table 1 and connected with the distinction made by Baker (2011) and Cummins (2000). The programs' denominations and descriptions were summarized and reorganized from Baker et al. (2016) and combined with Cummins (2000) and Baker (2011) programs' categorizations.

1.2 Previous Research

Bilingual education has been widely studied during the last decades, especially in the United States. A previous meta-analysis (Greene, 1997; Rossell & Baker, 1996; Slavin & Cheung, 2005; Willig, 1985) who tried to establish the efficacy of bilingual educational program lead to contradicting results. According to Cummins (2000, p. 194), «... few educational issues in North America have become as volatile or as ideologically loaded as the debate on the merits or otherwise of bilingual education».

The first meta-analysis that tried to summarize results in this area has begun to appear during the eighties, such as the study from Willig (1985), in response to a review of 28 studies by Baker and De Kanter (1981). The debate about the effectiveness of bilingual educational problem continued in the following years and revolved mostly around methodological issues in the research field and the criteria to apply in systematic reviews (see, for example, Willig, 1987).

In 1997, Greene (1997) reviewed a systematic analysis by Rossell and Baker (1996), which included 75 studies and concluded unfavourably in the regard of bilingual education programs in the U.S.A. In his revision, Greene found that only 11 of the studies were conducted by acceptable methodological standards. His conclusion was opposite to the Rossell and Baker's review, arguing that methodological valid studies supported a moderate positive effect of bilingual educational programs (Greene, 1997).

Slavin and Cheung (2005, p. 249), in their review, point out how «... quantitative research on the outcomes of bilingual education has diminished in recent years, but policy and practice are still being influenced by conflicting interpretations of research in this topic». Their main conclusion is that the number of high-quality studies of the issue is still little, and that also conflicting researchers agree on the necessity of improving the methodological quality of this kind of studies (Slavin & Cheung, 2005).

To improve the quality of this studies some reviewers agreed (Greene, 1997; Rossell & Baker, 1996; Slavin & Cheung, 2005; Willig, 1985) for the need of randomized, longitudinal studies, to control effects of bilingual education across more years and to avoid selection bias of participants, as well as meaningful comparison groups. Moreover, Slavin and Cheung (2005) pointed out the low significance of studying programs that lasted less than one year, due to the relative known long time needed by these programs to work.

1.3 Rationale

The recent affluence of refugees and immigrant children in the European countries poses the problem to their inclusion in the scholastic system. As acquiring adequate language skill in each dimension is one of the main goals of primary education it is important to examine what strategies are more effective to help the language to develop. This systematic review examines therefore what programs of bilingual education are already in use and their effectiveness. Moreover, some of the previous reviews of the literature on bilingual education pointed out the need for more high-quality studies of the topic.

1.4 Aim

The aim of this thesis is to describe the bilingual programs that are used in primary and preschool education and to examine their outcomes related to the children's language skill acquisition in both their L1 and L2.

1.5 Research Questions

1. What type of bilingual education programs are implemented in preschool and primary education?
2. What are the language related outcomes of bilingual education in preschool and primary education regarding the development of both L1 and L2?

2. Method

This study reports a systematic literature review. The method comprises a systematic search of all relevant studies concerning a specific topic, a selection process by previously established inclusion and exclusion criteria and a quality assessment. The studies identified by this means are successfully analysed and then summarised (Jesson, Matheson, & Lacey, 2011). Figure 2 provides a flowchart depiction of the process.

2.1 Search Procedure

Searches for this systematic literature review were performed in March 2017 using four different databases: ERIC, PsycINFO (PI), Scopus (SC) and Web of Science (WS). The first two are thesaurus search engine focused in the fields of psychology and education, the other two are generic free-text search engines. The rationale for the databases' choice is the following: (1) ERIC and PI cover the areas of educational research and educational psychology, covering the main fields interested by our study; (2) however, bilingual education interest also other disciplines (e.g., applied linguistics). To assure the maximal coverage possible of this other fields and overcome the terminology's ambiguity in the field (see introduction), SC and WS free-text search engine were also used. A flowchart depicting the search procedure is presented in figure 2. Inclusion and exclusion criteria were established to respect the research aim and questions.

2.1.1 Search terms

Search terms in ERIC and PsycINFO were selected through the databases' thesauri. Free-text search was used for SC and WS. A table showing search terms for each database during the search procedure is presented in Appendix A. Search terms were connected with AND and OR operators to create search strings.

2.2 Selection Process

A total of 1540 articles were found. The references and abstracts were saved in the Zotero reference manager software for the selection process. First, 324 duplicates were found and excluded. A selection process was then carried out with the other 1216 articles left. The articles were screened by title and abstract and subsequently, a full-text screening was realised.

2.2.1 Inclusion/Exclusion Criteria

Inclusion and exclusion criteria for the present literature review are presented in Table 2. Only studies focusing on preschool and primary education were included, consequently to the research aim and questions.

Studies focusing on non-societal languages and/or heritage programs were also excluded from this review. This is because of the elective character of such programs. Moreover, the focus of the present study is the typical situation when the school/societal language is different from the language spoken at home (Baker, 2011).

The review also excluded studies prior to 2000. The rationale for this criterion is two-fold: (1) Studies conducted prior to 2000 are widely reported in previous systematic review and meta-analysis (Greene, 1997; Rossell & Baker, 1996; Slavin & Cheung, 2005; Willig, 1985); and (2) the practices in bilingual education have evolved and new methodological standards for the research in the field have been agreed upon (Slavin & Cheung, 2005).

Moreover, studies that researched bilingual interventions lasting less than one academic year have been excluded, consistently with previous meta-analysis (Greene, 1997; Slavin & Cheung, 2005).

Table 2: Inclusion and exclusion criteria.

| | Inclusion | Exclusion |
|---------------------|--|--|
| <i>Participants</i> | Preschool and primary school children Children coming from a linguistic minority, including immigrant and refugee children. | Preschool and primary school children with speech impairments and/or disabilities. |
| <i>Intervention</i> | Any type of intervention described as bilingual. | Intervention/programs lasting less than a scholastic year Foreign language education Heritage programs |
| <i>Comparison</i> | Studies comparing bilingual programs/interventions vs. monolingual programs/interventions | |
| <i>Outcomes</i> | Language skills related outcomes | Other academic outcomes |
| <i>Study Design</i> | Quantitative, Qualitative, Mixed Methods | Case studies |

*Other*Paid articles, articles not
published as full text.

2.2.2 Title and abstract screening

A title and abstract screening were realised with the 1216 articles left after duplicate exclusion. After title and abstract screening, 40 articles were selected for full-text screening. Most of the articles were excluded to the inclusion/exclusion criteria. During title and abstract screening, some more articles were identified as duplicates and excluded. The articles that didn't contain enough information to determine exclusion were included for the full-screening review.

2.2.3 Full-text screening

Three articles were excluded because the full text was not freely accessible. The rest of the articles were excluded because they studied elective bilingual programs that included a language not used either by the children's family or in the society. One article was excluded because it focused on an additional intervention other than the bilingual program. That left 19 articles for data extraction and analysis. One study (López & Tashakkori, 2006) reported a mixed method study, but given that only the quantitative part was related to the research question of the present study, only the latter was extracted and analysed. The remaining 19 articles reported quantitative studies.

2.3 Quality Assessment

Study quality has been assessed with a tool adapted from the Quantitative Research Assessment Tool (CCERC, 2013). A complete version of the tool used can be found in the appendix B. The tool has been chosen because it accommodates different research design and puts emphasis on randomised selection of participants, sample sizes and quality of statistical analysis, which are an important issue in the study of bilingual education.

The tool assesses studies on 11 points with a three-point scale (-1, 0, 1). Possible score lies therefore in a ± 11 range. For this study, we considered scores below zero as insufficient and rationale to exclusion (LQ). Studies lying in the 0-5 range were considered medium quality (MQ), while studies having a score in the range 6-11 were rated as high quality (HQ) (see figure 1). The complete quality evaluation for each study is available in Appendix C. Only six

studies were evaluated as high quality, while the remaining were rated as medium quality. However, no study was excluded because of quality concerns.

2.4 Data extraction

The studies were analysed using a data extraction protocol (see Appendix D). The protocol included information about the bilingual program type and its description, population characteristics, measure tools, methods, statistical procedures, results, and conclusion. Bilingual program descriptions were transcribed and, if available, L1:L2 ratio was included. Detailed information about the comparison group(s) was also included. Moreover, the extraction protocol included information about the language dimensions assessed by each study.

The protocol was filled in an excel sheet. The complete Excel file containing all the information is available from the author upon request.

2.5 Methodological limitations of the present study

The literature search was conducted on multiple databases to ensure the best possible results. However, the terminology in this subject is quite extended and somewhat undisciplined (Baker et al., 2016). To overcome this, other reviewers, like Baker et al. (2016), hand-searched relevant journals to assure the greatest possible coverage. This was not done in the present review due to time constraints.

Moreover, the reliability of systematic literature reviews is assured by the presence of more reviewers. In this case, only one researcher did the present literature review, decreasing the reliability of this study.

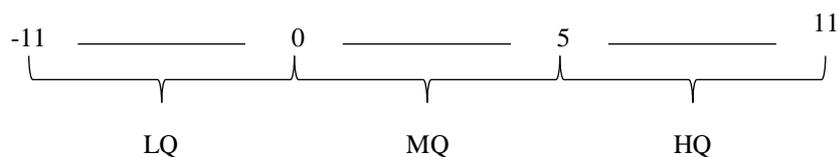


Figure 1: Quality Assessment Scale

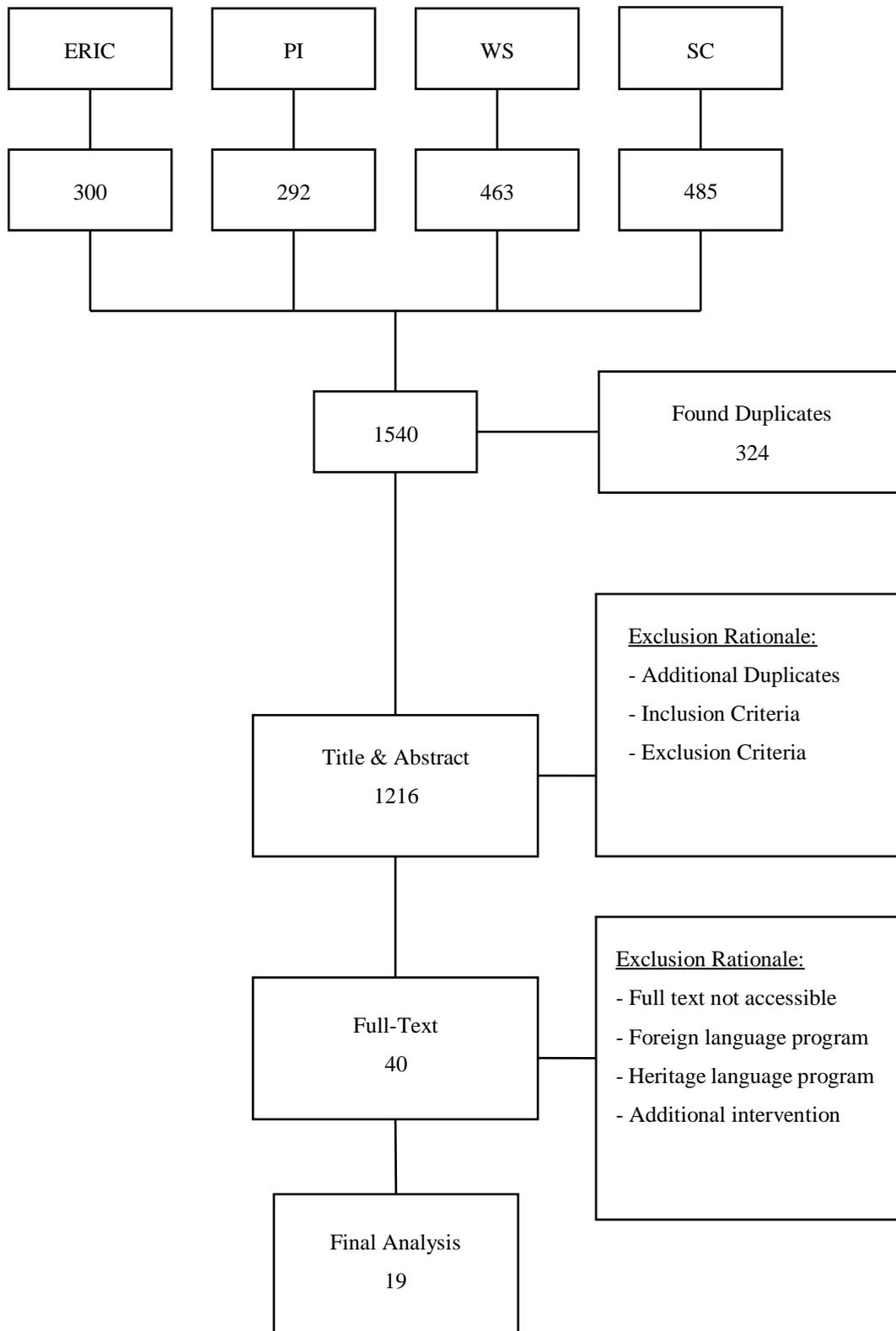


Figure 2: Flowchart showing search procedure.

3. Results

A total of 19 articles reporting 17 different studies were found through research in the ERIC, PsychInfo, Web of Science and Scopus databases. All of them were published between 2000 and 2016 in peer-reviewed journals related to education and applied linguistics. Every study was conducted in the United States and concerned mainly students that spoke Spanish (L1) at home and English (L2) in school. This is due to the predominance of the Spanish-speaking minority in the U.S. (Baker et al., 2016). Only three studies (2, 6 and 14) focused on the preschool environment (see Table 3).

Table 3: Summary of the reviewed studies.

| SN | Article | School | Program | L1 | L2 |
|----|------------------------------------|-------------------|---------|---------|---------|
| 1 | Bae (2007) | Primary School | KETWIP | Korean | English |
| 2 | Barnett et al. (2007) | Preschool | TWI | Spanish | English |
| 3 | Berens et al. (2013) | Primary School | TWI | Spanish | English |
| 4 | Carlisle and Beeman (2000) | Primary School | SI | Spanish | English |
| 5 | Carlo et al. (2014) | Primary School | TBE | Spanish | English |
| 6 | Duran et al. (2013; 2015; 2010) | Preschool | TBE | Spanish | English |
| 7 | Hofstetter (2004) | Primary School | TBE | Spanish | English |
| 8 | Lindholm-Leary (2014) | Primary School | TBE | Spanish | English |
| 9 | Lopéz and Tashakkori (2004) | Primary School | TWI | Spanish | English |
| 10 | Lopéz and Tashakkori (2006) | Primary School | TWBE | Spanish | English |
| 11 | Marian et al. (2013) | Primary School | TWI | Spanish | English |
| 12 | Murphy (2014) | Primary School | TBE | Spanish | English |
| 13 | Nakamoto et al. (2012) | Primary School | TBE | Spanish | English |

| | | | | | |
|----|-------------------------|-------------------|-------------------------|---------|---------|
| 14 | Ryan (2007) | Preschool | Bilingual Preschool | Spanish | English |
| 15 | Slavin et al. (2011) | Primary School | TBE (Early- Exit) | Spanish | English |
| 16 | Soltéro-González (2016) | Primary School | Paired Literacy | Spanish | English |
| 17 | Tong et al. (2008) | Primary School | TBE | Spanish | English |

Note. KETWIP = Korean/English Two-Way Program, TWI = Two-Way Program, TBE = Transitional Bilingual Education, SI = Spanish Immersion, DL = Dual Language.

3.1 Types of program implemented in the reviewed studies

The first research question of this study asked what types of bilingual education programs have been implemented between 2000 and 2016. As shown in Table 3, 8 studies investigated transitional bilingual programs (studies 3, 5, 6, 7, 8, 12, 13, 15, 17); 6 studies examined two-way immersion programs (studies 1, 2, 9, 10, 11); study 4 focused on a Spanish immersion program; study 14 described the investigated program as ‘bilingual preschool’; finally study 16 examined the effects of paired vs. sequential literacy instruction, which can be found in several different bilingual educational programs. Not all the studies included detailed descriptions of the various programs, this will be pointed out case by case. Generally, the most described feature was the proportional amount of instruction in either L1 or L2, and/or in the following, we summarise program descriptions in preschool and primary school environment.

3.1.1 *Preschool*

Only three studies (2, 6, and 14) focused on preschool. Reportedly, there is a lack of research concerning bilingual education in this educational environment (Barnett, Yarosz, Thomas, Jung, & Blanco, 2007; Durán, Roseth, Hoffmann, & Robertshaw, 2013; Durán, Roseth, & Hoffman, 2010; Durán, Roseth, & Hoffman, 2015; Ryan, 2007).

Study 2 reported results from a preschool adopting a two-way immersion program aimed at both Spanish and English speakers (57% Spanish speakers, 37% English speakers and 6% had another home language). Students alternated every week between Spanish and English classes.

Study 6 reported a three-year longitudinal study of a preschool transitional bilingual education program targeting Spanish-speaking students. During Year 1 and the first half of year 2 classes used exclusively Spanish. From January of the Year 2, English has gradually introduced a ratio of 70% Spanish and 30% English was reached (Durán et al., 2013). Interestingly, this is the only study reviewed where the researchers were active in the creation of the program.

Study 14 reports a retrospective comparison between the same program that changed from bilingual to monolingual in two different years, namely 2003-04 and 2004-05 (Ryan, 2007). However, the study doesn't give any other significant information about the characteristics of the bilingual program implemented during the year 2003-04.

3.1.2 Primary School

Fourteen studies (3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17) reported studies focused on primary school. Transitional bilingual education (TBE) programs were the most studied, with 7 studies (5, 7, 8, 12, 13, 15, 17), followed by two-way immersion programs reported in 5 studies (3, 9, 10, 11).

Transitional Bilingual Education (TBE). Transitional Bilingual Education programs were generally described as aimed to transition bilingual students to the L2. The second language was introduced gradually with different ratios during the years in which the programs were implemented, with the final goal set to transition to a full-time L2 instruction. However, not every study discussing TBE reported the exact ratio between L1:L2 across the different grades. Study 5 didn't describe the TBE program at all. Table 4 shows the available information about L1:L2 ratios for the study focusing on transitional bilingual education for each grade.

Table 4: L1:L2 ratios in TBE programs.

| SN | L1:L2 ratio | | | | | |
|----|-------------|-------|-------|-------|-------|-------|
| 5 | NA | | | | | |
| 7 | K-1 | | 5 | | | |
| | 70:30 | | 15:85 | | | |
| 8 | NA | | | | | |
| 12 | NA | | | | | |
| 13 | K | 1 | 2 | 3 | 4 | 5 |
| | 75:25 | 70:30 | 40:60 | 35:65 | 25:75 | 20:80 |
| 15 | NA | | | | | |
| 17 | 70:30 | | | | | |

Note. NA: Not Available; K, 1, 2, 3, 4, 5: Primary school grades..

Another important feature that distinguishes TBE is the time when the transition to L2 is considered complete: early-exit models described in the reviewed studies transitioned around grade 2-3. On the other hand, late-exit TBE programs can continue to offer part of the instruction in the L1 all along primary education.

Two-way Immersion (TWI) programs and Dual Language (DL) programs. Two-way Immersion programs are generally aimed at both bilingual and monolingual students. Again, different ratios between the two languages are present, but for different motivations than for Transitional Bilingual Education programs.

Study 9 (López & Tashakkori, 2004) omitted a thorough description of the TWI program they studied.

Study 3 (Berens, Kovelman, & Petitto, 2013), reports a comparison between two different approaches to TWI. The first one applies a ratio L1:L2 equals to 50:50 and the second one a ratio of 90:10. This difference is explained by two different assumptions that can be made in this type of program. A 50:50 ratio is motivated by the children can learn simultaneously in two different domains and that benefits the global acquisition by constructing better representations of the two languages. Conversely, a 90:10 assumes that children speaking only the minority language can profit best by first learn to read in their L1 and subsequently transfer these skills to the L2. Study 10 reported of another variation, where a 40:60 ratio was adopted. Another approach in TWI program design is to divide the language according to the subject taught. Study 11 described a program that included sequential reading instruction (L1->L2) and different L1:L2 ratios across the years. Table 5 replicates Table 4 by showing the available information on L1:L2 ratios for two-way immersion programs (TWI).

Other. In study 4 (Carlisle & Beeman, 2000) a Spanish Immersion program was the study focus. The program was described as an 80:20 maintenance bilingual program like a TBE one, where the main difference consisted in that the first literacy training was done exclusively in L1.

Study 16 (Soltero-González, Sparrow, Butvilofsky, Escamilla, & Hopewell, 2016), draw a similar distinction and focused on a comparison between sequential literacy, where literacy instruction was carried out in L1 and only subsequently in L2, and paired literacy, where the instruction was carried simultaneously in both languages.

Table 5: L1:L2 ratios in TWI programs.

| SN | L1:L2 ratio | | | | | |
|----|---------------|-------|--------|-------|-------|-------|
| 3 | 50:50 / 90:10 | | | | | |
| 9 | NA | | | | | |
| 10 | 60:40 | | | | | |
| 11 | K | 1 | 2 | 3 | 4 | 5 |
| | 100:0 | 100:0 | 100:00 | 40:60 | 60:40 | 60:40 |

Note. NA: Not Available; K, 1, 2, 3, 4, 5: Primary school grades.

3.2 Language outcomes of bilingual education

The second research questions of the present study concerned the language related outcomes of bilingual educational programs. Table 6, shows which of the four main language abilities (see section 1.1 Theoretical Background) were assessed for each study. Most the studies focused on the reading and listening ability. However, five studies (1, 2, 7, 8, 9) didn't measure any outcome related to L1. Only two studies (6 and 12) measured all dimension in both L1 and L2.

Table 6: Language dimensions assessed in each study.

| SN | L1 | | | | L2 | | | |
|----|----|---|---|---|----|---|---|---|
| | R* | L | W | S | R* | L | W | S |
| 1 | | | | | | | X | |
| 2 | | | | | X | | | |
| 3 | X | | | X | X | | | X |
| 4 | X | X | X | | X | X | X | |
| 5 | X | | | | X | | | X |
| 6 | X | X | X | X | X | X | X | X |
| 7 | | | | | X | X | X | X |
| 8 | | | | | X | X | X | X |
| 9 | | | | | X | | | |
| 10 | X | X | | X | X | X | | X |
| 11 | X | | | | X | | | |
| 12 | X | X | X | X | X | X | X | X |
| 13 | X | X | | | X | X | | |
| 14 | X | | | | X | | | |
| 15 | X | | | | X | | | |

| | | | | | | | |
|----|---|---|---|--|---|---|---|
| 16 | X | | X | | X | | X |
| 17 | X | X | | | X | X | |

**Measures assessing precursors of reading are included here.*

Note: Initials refer to the main linguistic abilities dimensions: R(eading), L(istening), W(riting) and S(peaking). Every study in the present review compared outcomes of a bilingual education program to a control group. The problem in comparing the results from the studies resides in the control group characteristics: some of the studies compared bilingual programs to monolingual programs, while other compared different typologies of bilingual programs with each other. Table 7 shows the characteristics of target program and comparison program for each study. Eleven studies (1, 2, 4, 5, 6, 7, 8, 9, 13, 14, 15) compared a bilingual program with a variation of a monolingual program (e.g., English-Only, Structured English Immersion, etc.), while six studies compared a bilingual education program with another bilingual program. Some studies contained both types of comparison.

Table 7: Target and comparison group(s) for each study.

| SN | Target Program(s) | Comparison Program |
|----|---------------------|---------------------|
| 1 | KETWIP | EO |
| 2 | TWI | SEI |
| 3 | TBE | DL, EI |
| 4 | SI | EI |
| 5 | TBE | EI |
| 6 | TBE | EO |
| 7 | TBE | Immersion |
| 8 | TBE | Immersion |
| 9 | TWI | Immersion |
| 10 | TWBE | TBE |
| 11 | TWI | TPI |
| 12 | TBE | DL |
| 13 | TBE | EI, DL |
| 14 | Bilingual Preschool | EO |
| 15 | TBE (Early-Exit) | SEI |
| 16 | Paired Literacy | Sequential Literacy |
| 17 | TBE (70:30*) | TBE (80:20*) |

Note. *L1:L2 ratio. EO = English Only, SEI = Structured English Immersion, EI = English Immersion, DL = Dual Language,

Some studies contained both types of comparison. Language outcomes description will consider this difference in the following presentations of the language related outcomes of

bilingual education programs. Appendix E presents a summary of each study and the language-related outcomes.

3.2.1 Language outcomes compared to L2-only programs

None of the examined studies reported an advantage of monolingual programs compared to different typologies of bilingual education. However, several studies reported none or only marginal differences between the bilingual group and the comparison group (studies 2, 7, 9, 14 and 15).

Study 1 (Bae, 2007) compared English writing proficiency of children in a two-way Korean English program with a peer in monolingual classrooms. The results showed no difference in L2 writing proficiency. This is the only study that focused on a different L1-L2 pair than Spanish-English.

Study 3 (Berens et al., 2013), reported advantages in reading performance of children in the bilingual program compared to monolingual peers in a single-language context.

Study 4 (Carlisle & Beeman, 2000) compared Spanish Immersion (SI) and English Immersion classes in both Spanish and English skills. The results showed that the SI class' reading and writing skills were as strong as those of the EI class. However, EI class showed an advantage in listening skills.

Study 6 (Duran et al., 2010; 2013; 2015) reported in three articles a three-year longitudinal, randomized evaluation of a Transitional Bilingual Education (TBE) preschool program. Results from the first year showed that TBE was associated with enhanced Spanish language and literacy acquisition without cost to the English development. Year 2 results showed strong gains in Spanish receptive vocabulary for the TBE group, without similar gains in English for the control group. Overall, the TBE group showed stronger gains in both languages. Year 3 results confirmed this trend and the overall validity of native-language support and, again, no evidence of inhibition of English acquisition for the TBE group.

Study 8 (Lindolm-Leary, 2014) showed that students in the bilingual group had more growth in language skills than other peers in immersion programs.

Study 11 (Marian, Shook, & Schroeder, 2013) showed significant progress in reading for the TWI-S group compared to monolingual classroom peers.

Study 13 (Nakamoto, Lindsey, & Manis, 2012) compared Transitional Bilingual Education (TBE) and Dual Language (DL) with English immersion students. The results showed that

TBE and DL students performed better on tests. Moreover, phonological and decoding skills in Spanish were a significant predictor to English reading comprehension.

3.2.2 Language outcomes compared to other bilingual programs

Study 3, 10, 12, 16 and 17 compared different typologies of bilingual programs with each other. Study 3 (Berens et al., 2013) differences in reading performances according to the L1:L2 ratio. Students in the 50:50 Two-Way Immersion (TWI) program reported better performances in various reading tasks compared to the TBE 90:10 program students. The authors argued that the 50:50 model, in combination with phonological training in the early childhood, could constitute an optimal model of bilingual education.

Study 10 (López & Tashakkori, 2006) compared a TBE program with a Two-Way Bilingual Education (TWBE) program. The type of program was significant in the determining the level of Spanish reading skills (which direction check), but no significant difference was found for English performances. Moreover, TWBE didn't increase the amount of time needed to be considered proficient in English and TWBE students acquired oral English at a faster rate than TBE students.

Study 12 (Murphy, 2014) showed only some advantage in both L1 and L2 oral proficiency for the Dual Language (DL) group, but no other difference between the groups.

Study 16 (Soltéro-González et al., 2016) focused on the literacy acquisition in bilingual programs comparing paired and sequential literacy acquisition. Paired literacy acquisition is typical of TWI bilingual programs that divide the time equally between L1:L2: in this model literacy in both language is acquired simultaneously. Sequential literacy, which is the acquisition of literacy first in the L1 and subsequently in the L2, is typical of TBE programs: in this study, the TBE program had a ratio of L1:L2 equal to 90:10. Students in the paired language program outperformed the sequential literacy students in writing, reading in both L1 and L2.

Study 17 (Tong, Irby, Lara-Alecio, & Mathes, 2008) compared two typologies of TBE programs. The treatment group was composed of students enrolled in a 70:30 TBE program, while the control group was composed by students of an 80:20 TBE program. The results showed advantages in reading and oral proficiency in both languages of students enrolled in the treatment group (70:30 TBE).

4. Discussion

The results of the data coming from the systematic review will be discussed in this section to answer to each research question. Additionally, we will discuss methodological issues of the study that had been reviewed for the present study as well as future research.

4.1 Types of bilingual educational programs

Most of the study reviewed focused on either Two-Way Immersion (TWI) and/or Transitional Bilingual Education (TBE) programs. This is consistent with some previous literature that indicated in those two approaches the most used ones (Slavin & Cheung, 2005). The distinction in three main categories of bilingual programs proposed by Cummins (2000) and Baker (2011) summarizes well the situation of bilingual education programs, which can be easily divided into the three main categories, depending on the intended outcome, namely: (1) program that are taught exclusively in L2 or with little support for L1, with an aim of monolingualism; (2) limited bilingual programs, that aim to a transition to a fully L2 instruction, such as TBE programs; and (3) programs that aim at full biliteracy and bilingualism for the students, such as TWI programs. The present review shows how transitional bilingual education (TBE) and two-way immersion bilingual education (TWI) are the most common programs implemented in bilingual education.

However, confusion in the label applied to the different programs, cited by Baker et al. (2016) is still problematic. Some studies reviewed revealed some confusions in the naming and categorizing of these programs. Often, the L1:L2 ratio seemed the only rationale to apply one label or the other, without further consideration. This ended up in some programs labeled as TWI, while the remainder of the program's description suggested aim and modality closer to a TBE program. Another example of this issue is the 'Dual Language' denomination. Often used as a synonym to TWI, it is often not clear what differences between the two are present if any. Another issue that emerged from the previous literature (Baker, 2011; Baker et al., 2016), namely the lack of description of these programs, is consistent with the present findings. Only a handful of studies provided detailed descriptions of the content of the program, while most of them limited the description to the ratio of L1:L2 time of instruction. While L1:L2 ratios are certainly a valid metric to compare and describe bilingual programs, a more detailed discussion

on the content and methods of these programs will certainly help to determine what works and what doesn't in a more detailed way.

Finally, we confirmed the existence of a research gap concerning bilingual education in preschool (Barnett et al., 2007; Durán et al., 2010; Ryan, 2007).

4.2 Language related outcomes of bilingual educational programs

Consistent with some of the previous meta-analysis (Baker et al., 2016; Greene, 1997; Slavin & Cheung, 2005), the results of this study seem to confirm that bilingual education doesn't inhibit the learning of a second language. This is an important result because most of the critics aimed at bilingual education programs in the past were that maintaining L1 in a student's curriculum would have blocked his advancements in the L2. None of the reviewed studies offered evidence in that sense.

However, and consistently with some previous results (Greene, 1997), evidence of the superiority of bilingual education vs. monolingual education is not conclusive. A considerable number of studies in the present review found little or no evidence of advantages in language outcomes for students enrolled in a bilingual program, rendering difficult to address the hypothesis of threshold, cross-transfer, and interdependence advanced by Cummins (2000). This can be due in part to inherently methodological challenges of this research area (see 4.3). Finally, a simple survey of the language dimension assessed by each study revealed a focus of the research on L2 outcomes over L1. Only a few studies attempted to measure outcomes for each dimension for both languages. This reveals a bias in the present research that seems to put more importance on the L2 while omitting to value L1 outcomes and bilingualism in general (Wright, 2012).

4.3 Methodological issues of the studies

A previous meta-analysis (Greene, 1997; Slavin & Cheung, 2005) trying to synthesise the research on bilingual education efficacy revealed major problems in the quality of this type of studies. They revealed the necessity for multi-year, longitudinal, randomized studies with a consistent number of participants.

The present review, similarly to last recent meta-analysis of Baker et al. (2016), revealed that those issues are still present and still needs to be addressed. Only a small fraction of the study

reviewed presented a randomized control study and most the study spanned only a year of pre-post measures. Moreover, the few RCT studies included in this review featured a very low number of participants, resulting in a very low power of generalisation. The same study (Durán et al., 2013, 2010, 2015) also revealed the probable inadequacy of some measurement tool and the need to develop more sensitive and specialized instruments to study bilingual students.

4.4 Future Research

The present study revealed some prominent issues that should be considered in future research. First, this study identified an important gap in the actual research, namely the lack of research on bilingual education in preschool children. Given how some literature (Loewen, 2003) points out the importance of language development in the early years (e.g., literacy precursors), bilingual education in preschool seem an important area to further investigate and develop.

Second, this study revealed how some studies in this review didn't assessed language skills in the L1. Evaluation and comparison of L1 proficiency between bilingual students in a bilingual program and in a monolingual one is an important outcome and should be more investigated.

Third, as mentioned by Wright (2012), most of the research concerning the efficacy has been realised in the United States and is not easily generalizable to other contexts. This implies that further research with other L1:L2 languages and in other context is needed to provide a more conclusive answer on the efficacy of such programs.

Finally, methodological issues pointed out by the previous meta-analysis seem to persist in the studies contained in this review. Further research should try to assess this problematic and implement higher quality research designs.

5. Conclusion

The aim of this study was to describe the bilingual programs implemented in preschool and primary school and to examine the outcomes related to language proficiency both in L1 and L2. While language outcomes don't constitute the only and one rationale to support the implementation of such programs, removing the language barrier for a student that come from a household where the school language is a different one remains these programs' primary goal. The present study consisted of a systematic literature review realised by just one reviewer. This constitutes the main limitation of the present study.

The results emerged from this systematic review of the literature showed a predominance of two approaches, namely transitional and two-way immersion programs. A lack of detail in the description concerning the implementation of this program emerged also from the present study. Moreover, a certain confusion and ambiguity in the denominations used to describe bilingual problems have been found.

Regarding the outcome related to language proficiency, the results from this review suggest that following a bilingual program doesn't inhibit the acquisition of the second language. However, the advantages of these programs are more difficult to show, and numerous studies found no significant statistical differences between bilingual and monolingual programs. This could be imputed to methodological limitations in this area of research that still needs to be addressed, as well as to the tendency of some study to concentrate mostly on the outcomes related to the L2. Sustained by other authors, the study claim that this fact constitutes a bias against bilingual education and should, therefore, be addressed in the future.

Future research should focus on covering the gap of research in preschool as well as design more high-quality research to address the problematic. Moreover, research on bilingual programs outside the U.S. is warranted to be able to generalise the results. Additionally, other L1/L2 combinations should be addressed, especially those who involve more linguistic distance between the two languages. Only one of the studies included in this review (Bae, 2007) focused on such a combination (Korean/English).

A possible way to address this kind of issues would be to adopt a paradigm of action research where the researchers would be included also in the program design and implementation. This could be an effective way to design the needed randomized, longitudinal studies in this field. In the present study, only the longitudinal study by Durán et al. (2010; 2013; 2015) adopted such a perspective.

Two thousand years after the discussion of bilingual education by Quintilian, it is, therefore, possible to confirm some of his intuitions, most importantly that learning in two languages is possible without either of the two hindering the other or the general curriculum. Perhaps now the question that is more urgent to address is to understand what the optimal implementations are and whether it would be possible to expand bilingual education to a wider audience.

References

- Bae, J. (2007). Two-way immersion shows promising results: Findings from a new study. *Language Learning*, 57(2), 299–332. doi: 10.1111/j.1467-9922.2007.00410.x.
- Baker, C. (2011). *Foundations of bilingual education and bilingualism* (5th ed). Bristol, UK ; Tonawanda, NY: Multilingual Matters.
- Baker, D. L., Basaraba, D. L., & Polanco, P. (2016). Connecting the present to the past: Furthering the research in bilingual education and bilingualism. *Review of Educational Research*, 40, 821–883. doi: 10.3102/0091732X16660691.
- Baker, K. A., & De Kanter, A. A. (1981). Effectiveness of bilingual education: A review of the literature. Final draft report. Retrieved from <http://eric.ed.gov/?id=ED215010>
- Barnett, W. S., Yarosz, D. J., Thomas, J., Jung, K., & Blanco, D. (2007). Two-way and monolingual English immersion in preschool education: An experimental comparison. *Early Childhood Research Quarterly*, 22(3), 277–293. doi: 10.1016/j.ecresq.2007.03.003.
- Berens, M. S., Kovelman, I., & Petitto, L.-A. (2013). Should bilingual children learn reading in two languages at the same time or in sequence? *Bilingual Research Journal*, 36(1), 35–60. doi: 10.1080/15235882.2013.779618.
- Butler, Y. G. (2012) Bilingualism/multilingualism and second-language acquisition, in *The handbook of bilingualism and multilingualism*, second edition (eds T. K. Bhatia and W. C. Ritchie), John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9781118332382.ch5.
- 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. doi: 10.1207/S1532799XSSR0404_5.
- Carlo, M. S., Barr, C. D., August, D., Calderón, M., & Artzi, L. (2014). Language of instruction as a moderator for transfer of reading comprehension skills among Spanish-speaking English language learners. *Bilingual Research Journal*, 37(3), 287–310. doi: 10.1080/15235882.2014.963739.
- CCEERC. (2013). Quantitative research assessment tool. Retrieved from <http://www.researchconnections.org/content/childcare/understand/research-quality.html>

- Cummins, J. (2000). *Language, power, and pedagogy: bilingual children in the crossfire*. Clevedon [England]; Buffalo [N.Y.]: Multilingual Matters.
- Dryden-Peterson, S. (2016). Refugee education in countries of first asylum: Breaking open the black box of pre-resettlement experiences. *Theory and Research in Education*, 14(2), 131–148. doi: 10.1177/1477878515622703.
- Durán, L. K., Roseth, C., Hoffman, P., & Robertshaw, M. B. (2013). Spanish-speaking preschoolers' early literacy development: A longitudinal experimental comparison of predominantly English and transitional bilingual education. *Bilingual Research Journal*, 36(1), 6–34. doi: 10.1080/15235882.2012.735213.
- Durán, L. K., Roseth, C. J., & Hoffman, P. (2010). An Experimental study comparing English-only and transitional bilingual education on Spanish-speaking prepre-schoolers' early literacy development. *Early Childhood Research Quarterly*, 25(2), 207–217. doi: 10.1016/j.ecresq.2009.10.002
- Durán, L. K., Roseth, C. J., & Hoffman, P. (2015). Effects of transitional bilingual education on Spanish-speaking prepre-schoolers' literacy and language development: Year 2 results. *Applied Psycholinguistics*, 36(4), 921–951. doi: 10.1017/S0142716413000568.
- Greene, J. P. (1997). A Meta-Analysis of the Rossell and Baker review of bilingual education research. *Bilingual Research Journal*, 21(2). Retrieved from <http://proxy.library.ju.se/login?url=http://search.proquest.com/docview/62383161?accountid=11754>.
- Hakuta, K., Butler, Y. G., & Witt, D. (2000). How long does it take English learners to attain proficiency? In *University of California Linguistic Minority Research Institute, Policy Report 2000-2001*. On WWW at: <http://www.stanford.edu/hakuta/Docs/HowLong.pdf> (pp. 36–53). UNESCO Publishing.
- Hamilton, R. J., & Moore, D. (2003). *Educational interventions for refugee children: theoretical perspectives and implementing best practice*. London; New York: RoutledgeFalmer.
- Hofstetter, C. H. (2004). Effects of a transitional bilingual education program: Findings, issues, and next steps. *Bilingual Research Journal*, 28(3), 355–377.

- Jesson, J., Matheson, L., & Lacey, F. M. (2011). *Doing your literature review: Traditional and systematic techniques*. SAGE.
- Lindholm-Leary, K. (2014). Bilingual and biliteracy skills in young Spanish-speaking low-SES Children: Impact of instructional language and primary language proficiency. *International Journal of Bilingual Education and Bilingualism*, 17(2), 144–159.
- Loewen, S. (2003). Second language concerns for refugee children. In *Educational intervention for refugee children: Theoretical perspectives and implementing best practice* (pp. 35–52). London: RoutledgeFalmer.
- López, M. G., & Tashakkori, A. (2004). Narrowing the gap: Effects of a two-way bilingual education program on the literacy development of at-risk primary students. *Journal of Education for Students Placed at Risk (JESPAR)*, 9(4), 325–336. doi: 10.1207/s15327671espr0904_1.
- López, M. G., & Tashakkori, A. (2006). Differential outcomes of two bilingual education programs on English language learners. *Bilingual Research Journal*, 30(1), 123–145.
- Marian, V., Shook, A., & Schroeder, S. R. (2013). Bilingual two-way immersion programs benefit academic achievement. *Bilingual Research Journal*, 36(2), 167–186. doi: 10.1080/15235882.2013.818075.
- Murphy, A. F. (2014). The effect of dual-language and transitional-bilingual education instructional models on Spanish proficiency for English language learners. *Bilingual Research Journal*, 37(2), 182–194.
- 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(2), 537–567. doi: 10.1007/s11145-010-9285-4
- Odlin, T. (1989). *Language Transfer: Cross Linguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Rossell, C. H., & Baker, K. (1996). The educational effectiveness of bilingual education. *Research in the Teaching of English*, 30(1), 7–74.
- Ryan, A. M. (2007). Two tests of the effectiveness of bilingual education in preschool. *Journal of Research in Childhood Education*, 21(4), 352.

- Sermier Dessemontet, R., & Bless, G. (2013). The impact of including children with intellectual disability in general education classrooms on the academic achievement of their low-, average-, and high-achieving peers. *Journal of Intellectual & Developmental Disability, 38*(1), 23–30. doi: 10.3109/13668250.2012.757589.
- Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. *Review of Educational Research, 75*(2), 247–284. doi: 10.3102/00346543075002247.
- Slavin, R. E., Madden, N., Calderon, M., Chamberlain, A., & Hennessey, M. (2011). Reading and language outcomes of a multiyear randomized evaluation of transitional bilingual education. *Educational Evaluation and Policy Analysis, 33*(1), 47–58. doi: 10.3102/0162373711398127.
- Smidt, S. (2016). *Multilingualism in the early years: extending the limits of our world*. New York: Routledge.
- Soltero-González, L., Sparrow, W., Butvilofsky, S., Escamilla, K., & Hopewell, S. (2016). Effects of a paired literacy program on emerging bilingual children's biliteracy outcomes in third grade. *Journal of Literacy Research, 48*(1), 80–104.
- Strekalova, E., & Hoot, J. L. (2008). What is special about special needs of refugee children? Guidelines for Teachers. *Multicultural Education, 16*(1), 21–24.
- Szente, J., Hoot, J., & Taylor, D. (2006). Responding to the special needs of refugee children: Practical ideas for teachers. *Early Childhood Education Journal, 34*(1), 15–20. doi: 1007/s10643-006-0082-2.
- Tong, F., Irby, B. J., Lara-Alecio, R., & Mathes, P. G. (2008). English and Spanish acquisition by hispanic second graders in developmental bilingual programs: A 3-year longitudinal randomized study. *Hispanic Journal of Behavioral Sciences, 30*(4), 500–529. doi: 10.1177/0739986308324980.
- United Nations General Assembly. (1989). *Convention on the rights of the child*. New York: United Nations. Retrieved from <http://www.ohchr.org/Documents/ProfessionalInterest/crc.pdf>

Wernig, Lösen & Urban (2008). Cultural and social diversity: An analysis of minority groups in german schools. *The Journal of Special Education*, 42(1), 48-54, doi: 10.1177/0022466907313609.

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

Wright, W. E. (2012) Bilingual education, in *The handbook of bilingualism and multilingualism*, second edition (eds T. K. Bhatia and W. C. Ritchie), John Wiley & Sons, Ltd, Chichester, UK. doi: 10.1002/9781118332382.ch24.

Appendix

A. Table displaying search words used in each database during the search procedure.

| Database | Search Terms | | |
|--------------------|---|---|--|
| | <i>Preschool and Primary School Education</i> | <i>Bilingual Program/Intervention</i> | <i>Educational outcomes</i> |
| <i>ERIC</i> | <ul style="list-style-type: none"> • Kindergarten • Elementary Schools • Preschool Learning • Early Childhood Education • Elementary Education • Early Intervention • Young Children | <ul style="list-style-type: none"> • Migrant Education • Bilingual Education Programs • Immersion Programs • Bilingual Schools • Multilingualism • Bilingualism • Intercultural Programs • Bilingual Education • Multicultural Education | <ul style="list-style-type: none"> • Outcomes of Education • Program Evaluation • Program Effectiveness • Student Development • School Effectiveness • Educational Benefits • Instructional Effectiveness |
| <i>PI</i> | <ul style="list-style-type: none"> • Elementary School Students • Preschool Education • Primary School Students | <ul style="list-style-type: none"> • Multilingualism • Multicultural Education • Bilingualism • Bilingual Education | <ul style="list-style-type: none"> • Educational Program Evaluation • Educational Quality |
| <i>SC & WS</i> | <ul style="list-style-type: none"> • Early Childhood • Preschool Education • Primary Education | <ul style="list-style-type: none"> • Bilingual Programs • Bilingual Education • Dual Language Education • Multilingualism • Multicultural Education | <ul style="list-style-type: none"> • Outcomes • Educational Outcomes • Program Evaluation • Educational Benefits • Instructional Effectiveness |

Note: ERIC = Education Resources Information Center, PI = PsychInfo, SC = Scopus, WS = Web of Science.

B. Quality assessment tool.

| Population and Sample |
|---|
| <p>1. Population. Does the population that was eligible to be selected for the study include the entire population of interest?</p> <p>[1] Eligible population includes entire population of interest or a substantial portion of it. [0] Population represents a limited, atypical, or selective subgroup of the population of interest. [-1] No description of the population.</p> <p>2. Randomized selection of participants. Were study participants randomly selected for the study?</p> <p>[1] Random selection. [0] Nonrandom selection. [-1] No description of sample selection procedure.</p> <p>3. Sample size.</p> <p>[1] Sample size larger than similar studies. [0] Sample size the same as similar studies. [-1] Sample size smaller than similar studies.</p> <p>4. Response and Attrition Rate.</p> <p>[1] High response or participation rate. [0] Moderate to low response rate. [-1] No information on response rate.</p> |
| Measurement |
| <p>5. Main variables or concepts. Are each of the main variables or concepts of interest described fully?</p> <p>[1] Accurately described and can be matched [0] Vague definition or cannot be matched [-1] No definition of main variables or concepts.</p> <p>6. Operationalization of concepts. Did the authors choose variables that make sense as good measures of the main concepts in the study?</p> <p>[1] Key concepts are measured with variables that make sense. [0] Key concepts are measured with variables that do not make sense, and variables have not been used in previous research studies. [-1] Variable operationalization is not discussed.</p> |
| Analysis |
| <p>7. Numeric tables. Are the means and standard deviation/standard errors for all numeric variables presented?</p> <p>[1] Means and standard deviations/standard errors are presented. [0] Means, but no standard deviation/standard errors presented. [-1] Neither mean nor standard deviations/standard errors presented.</p> |

8. Missing data. Are the number of cases with missing data specified? Is the statistical procedure(s) for handling missing data described?

[1] Number of cases with missing data are specified and the strategy for handling missing data is described.

[0] Number of cases with missing data specified, but these cases are removed from the analysis.

[-1] Missing data issues not discussed.

9. Appropriateness of Statistical Techniques. Does the study describe the statistical technique used? Does the study explain why the statistical technique was chosen? Does the study include caveats about the conclusions that are based on the statistical technique?

[1] Statistical techniques, reasons for choosing technique, and caveats are fully explained.

[0] Statistical technique is explained, but the reasons for choosing technique or the caveats are not included.

[-1] Statistical technique, reasons for choosing technique, and caveats are not explained.

10. Omitted Variable Bias. Could the results of the study be due to alternative explanations that are not addressed in the study?

[1] All important explanations are included in the analysis.

[0] Important explanations are omitted from the analysis.

[-1] Variables and concepts included in the analysis are not described in sufficient detail to determine whether key alternative explanations have been omitted.

11. Analysis of Main Effect Variables. Are coefficients for the main effect variables in the statistical models presented? Are the standard errors of these coefficients presented? Are significance levels or the results of statistical tests presented?

[1] Model coefficients and standard errors or hypothesis tests for the main effects variables are presented.

[0] Either model coefficients or hypothesis tests for the main effects variables are presented.

[-1] Neither estimated.

Note. Adapted from CCERC (2013).

C. Quality assessment scores of the reviewed studies.

| | Study Number | | | | | | | | | | | | | | | | |
|---|--------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1. Population selection | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2. Randomized Selection of participants | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | -1 | 0 | 0 | 1 | 0 | 1 |
| 3. Sample size | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| 4. Response and attrition rate | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 5. Main variable | 1 | -1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 6. Operationalization | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
| 7. Numeric tables | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 8. Missing data | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 9. Statistical techniques | 1 | 0 | -1 | 0 | -1 | 1 | 0 | -1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| 10. Omitted Variables | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 11. Analysis of main effect | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| TOTAL | 8 | 4 | 4 | 4 | 5 | 8 | 3 | 3 | 4 | 4 | 5 | 3 | 6 | 3 | 6 | 6 | 8 |
| ASSESSMENT | HQ | MQ | MQ | MQ | MQ | HQ | MQ | MQ | MQ | MQ | MQ | MQ | HQ | MQ | HQ | HQ | HQ |

Note. MQ = Medium Quality, HQ = High Quality.

D. Data extraction and selection protocol.

| Data Extraction Protocol | |
|--------------------------|--|
| General information | Title of the study: |
| | Authors: |
| | Journal and year of publication: |
| | Country: |
| | Aim: |
| | Research question(s): |
| Population | Number: |
| | School (preschool vs. primary school): |
| | L1: |
| | L2: |
| Bilingual Program(s) | Program denomination: |
| | Program description: |
| | Program duration: |
| | L1:L2 ratio: |
| Methodology | Characterization of comparison group(s): |
| | Instruments used: |
| | Pre-/post-test: |
| | Statistical analysis: |
| | Control of language of instruction: |
| | Language dimensions assessed (L1): |
| | Language dimensions assessed (L2): |
| Results | L1: |
| | L2: |
| | Other: |
| Conclusion | TRANSCRIBE: |

E. Table displaying general information about the reviewed studies.

| SN | Study | Purpose/Hypothesis | School | Program(s) | Language Related Outcomes |
|----|----------------------------|--|-----------|------------|---|
| 1 | Bae (2007) | This study assesses the progress of the program toward meeting the goal of English development, with a special focus given to the early phase of immersion (grades K to 2) (p. 306) | Primary | KETWIP | No difference in L2 writing proficiency, compared to peers in EO programs. |
| 2 | Barnett et al. (2007) | Programs were compared on measures of children's growth on language, emergent literacy, and mathematics. (p. 277) | Preschool | TWI EI | Results show no statistical difference between the groups. |
| 3 | Berens et al. (2013) | We study reading and language performance in second- and third-grade children (matched for age, and socioeconomic status [SES]) who were in 50:50 Simultaneous or 90:10 Sequential (Spanish/English) dual-language learning contexts (p. 38) | Primary | TBE TWI | Children in bilingual program have an advantage in reading compared to monolingual children learning in a single-language context. Comparison between TBE and TWI showed differences in reading performance according to L1:L2 ratio. |
| 4 | Carlisle and Beeman (2000) | (a) Do the EI and SI classes differ in English and Spanish listening and reading comprehension in the spring of Grade 1? (b) Are there differences in the development of writing skills in English and Spanish for the EI and SI classes in the spring of Grade 1? (c) Does the language of instruction (English or Spanish) contribute significantly to the prediction of reading comprehension in English and Spanish in the fall of second grade, after differences in language proficiency are accounted for? (p. 377) | Primary | EI SI | SI class as strong as the EI class on reading and writing, but stronger on Spanish reading and writing. The advantage in listening for EI class. |

| | | | | | |
|---|---------------------------------|--|-----------|-----------------------|--|
| 5 | Carlo et al. (2014) | The goal of this longitudinal study was to investigate the relationship of initial levels of English oral language proficiency and Spanish reading comprehension with growth in English reading comprehension and the role that language of instruction plays in moderating these relationships (p. 288) | Primary | TBE (late/early exit) | The fastest growth in reading comprehension in English among student in early-exit TBE. At the end of the study, English reading comprehension skills of the late-exit TBE group were significantly lower than the other groups. |
| 6 | Duran et al. (2013, 2010, 2015) | The purpose of the present study was to provide a longitudinal, experimental comparison of EO and TBE instruction on the language and literacy development of low-income Spanish-speaking preschoolers (p. 208) | Preschool | TBE EO | Year 1: Results supported the prediction that TBE would be associated with enhanced Spanish language and literacy development without significant cost to English development. No significant loss for TBE. Year 2: Strong gains in TBE on Spanish receptive vocabulary, No similar gains in English for PE. Stronger gains in both languages in the TBE program. Year 3: significant evidence of the validity of native-language support. No inhibition of English acquisition. |
| 7 | Hofstetter (2004) | This study examined the effectiveness of the transitional bilingual ALA (Academic Language Acquisition) program in enhancing K-5 students' English-language proficiency, as well as their English performance in academic subject areas, in comparison with the SEI process (p. 359) | Primary | TBE SEI | Only nominal differences between the groups were found. |

| | | | | | |
|----|-----------------------------|--|---------|------------------|--|
| 8 | Lindholm-Leary (2014) | The purpose of this study was to better understand the bilingual and biliteracy skills of Spanish-speaking low-SES children entering preschool through second grade (p. 146) | Primary | TWI TBE EO | Students in bilingual group had more growth in language skills than other peers in immersion. |
| 9 | Lopéz and Tashakkori (2004) | The purpose of this study is to investigate the impact of such an early bilingual education program. The effects of an extended two-way bilingual education program (Extended Foreign Language, or EFL) on the language development of at-risk kindergarten students with limited English proficiency were studied by comparing them with a group of students who were more proficient English speakers (p. 327) | Primary | TWI | At the end of first grade, there were no statistical differences between the two groups in English reading. |
| 10 | Lopéz and Tashakkori (2006) | The Purpose of this study was to compare these two different approaches -- TWBE (maintenance and instruction in the L1) and TBE (transitional instruction in L1) -- among children with different L2 proficiency levels (p. 125) | Primary | TWBE TBE | Type of program was statistically significant on level of Spanish reading skills. No significant differences between the two groups. TWBE doesn't increase the amount of time to be proficient in English. TWBE students acquired oral English at a faster rate. |
| 11 | Marian et al. (2013) | Our aim in the current study was to examine whether two-way immersion programs benefit academic achievement in both minority- and majority-language students. (p. 168) | Primary | TWI-S | Significant progress in reading for the TWI-S group. The analysis suggests that TWI can lead to significant improvements for minority-language students. |
| 12 | Murphy (2014) | A study on this point has potential to determine which approach best leverages students' home-language skills to facilitate second-language and content learning—a crucial endeavor in | Primary | DL TBE | DL group had some advantage in oral expression in both L1 and L2. |

light of research demonstrating that home-language proficiency is associated with second-language and content learning (p. 186)

| | | | | | |
|----|-------------------------|---|-----------|--------------------------------------|--|
| 13 | Nakamoto et al. (2012) | In the present study, we report longitudinal analyses of the prediction of reading comprehension, in both English and Spanish, from kindergarten to grade 3 for three instructional groups, a transitional Spanish–English bilingual group (n = 210), a dual-Spanish–English immersion group (n = 207), and an English immersion group (n = 85) (p. 541) | Primary | TBE | TBE and DL performed better on tests in comparison to EI students. Phonological/decoding skills in Spanish was significant predictor to English reading comprehension. |
| 14 | Ryan (2007) | This article will use the changes in the Even Start program in the 2003-04 and 2004-05 program years to conduct two separate tests of the effectiveness of bilingual education in preschool, using follow-up kindergarten data on the original rest wave cohort as well as additional data from the new, second wave cohort of preschoolers (p. 355) | Preschool | NS | Effectiveness of the bilingual program, but only at marginally significant levels of statistical inference. |
| 15 | Slavin et al. (2011) | The present article reports on final findings of a study in which three successive cohorts of kindergarteners were randomly assigned to bilingual or English-only conditions and then followed for a period up to 5 years. It compares early-exit TBE and structured English immersion (SEI) – the most common instructional approaches today for ELLs (p. 49-50) | Primary | TBE SEI | No statistically significant results in favor of bilingual programs. |
| 16 | Soltéro-González (2016) | Research Question 1: How do the biliteracy outcomes of students in a paired literacy model compare to those of students | Primary | Paired vs. Sequential Literacy | Students in the paired language program outperformed the sequential literacy students in writing, reading in both L1 and L2. |

in a sequential literacy model in Grade 3 as measured by informal writing and reading assessments and what effect does it have?

Research Question 2: How do the reading outcomes of students in a paired literacy model compare to those of students in a sequential literacy model in Grade 3 as measured by a state standardized assessment? (p. 82)

| | | | | | |
|----|--------------------|---|---------|-------------------------|--|
| 17 | Tong et al. (2008) | The purpose of our study was to evaluate the effectiveness of student outcomes in language and literacy acquisition in both Spanish and English longitudinally from kindergarten through second grade, under controlled conditions, of an enhanced 70/30 Spanish/English late-exit TBE model (specifically, a DBE instructional model, which will be referred to as such hereafter) implementing an intervention of systematic English curriculum compared with an 80/20 Spanish/English late-exit TBE model implementing typical practice in the same district. (p. 504) | Primary | 70:30 TBE vs. 80:20 TBE | Treatment group (70:30 TBE) showed advantages in reading and oral proficiency in both L1 (Spanish) and L2 (English). |
|----|--------------------|---|---------|-------------------------|--|

Note. For an explanation of the abbreviation, see Appendix G.

F. List of abbreviations.

| | |
|--------|--|
| BICS | Basic Interpersonal Communicative Skills |
| CALP | Cognitive/Academic Language Proficiency |
| DL | Dual Language |
| DLL | Dual Language Learner |
| EB | Emergent Bilingual |
| ELL | English Language Learner |
| EO | English Only |
| KETWIP | Korean/English Two-Way Immersion Program |
| LEP | Limited English Proficient |
| TBE | Transitional Bilingual Education |
| TPI | Transitional Program of Instruction |
| TWI | Two-Way Immersion |
| SEI | Structured English Immersion |
| SI | Spanish Immersion |