

The effectiveness of therapies for dual language children with developmental language disorder: a systematic review of interventional studies.

Hannah Harvey, Hazel Allaway, Sarahjane Jones

*Centre for Social Care, Health and Related Research, Birmingham City University,
Birmingham, UK*

Hannah Harvey (corresponding author) - Hannah.Harvey@bcu.ac.uk

BSc (Hons); Research Assistant.

Centre for Social Care, Health and Related Research, Faculty of Health Education and Life Sciences, Birmingham City University, City South Campus, Westbourne Road, Edgbaston, B15 3TN; 0121 331 6067

Hazel Allaway

MSc, BSc, MRCSLT, MHCPC; Speech and Language Therapist, Birmingham Community Healthcare NHS Foundation Trust.

Sarahjane Jones

PhD, BMedSci (Hons); Senior Research Fellow, Centre for Social Care, Health and Related Research, Faculty of Health Education and Life Sciences, Birmingham City University. Taylor & Francis Word Template for journal articles

Purpose: This study aims to understand the effect of therapies on dual language children with developmental language disorder (DLD) on a range of bilingual language outcomes, compare with second-language-only therapy and determine whether there is any cross-linguistic transfer.

Methods: A systematic review of English articles in 10 electronic databases was conducted. Screening, reviewing and appraising were performed independently by two reviewers. Quality was appraised and findings synthesised in accordance with the research questions.

Results: Nine reports were identified. Five studies were found to be low in bias and therefore high in quality. Two were medium bias and two were high. Key findings were that instruction in the first language is required to support its continued acquisition and that bilingual instruction does not limit second language growth.

Conclusions: There is no identified evidence to suggest that second-language-only is better than bilingual therapy for dual language children with DLD for the development of the second language. There is evidence to suggest that bilingual therapy is equally effective for second language development, and also supports development of the first language. Further work is required to understand the efficacious doses of both languages in order to develop cost effective therapies and achieve optimal outcomes.

Keywords: developmental language disorder; bilingualism; speech and language therapy

Introduction

Developmental language disorder (DLD) is characterised by poor language abilities not attributable to neurological, sensory, cognitive, or motor impairments, or environmental factors (Leonard, 2014; Schwartz, 2017). Dual language children with DLD demonstrate impairment in both languages, compared to aged-matched peers with similar language learning experiences (Caldas 2006). For children with DLD, timely and effective intervention is essential for improving language and by extension academic and social outcomes (Bishop and Leonard, 2014). Through comparison of outcomes in both the first language (L1) and second language (L2), this review intends to assess the effectiveness of approaches used to address DLD in dual language children.

Developmental language disorder

In domains other than language, there is fairly consistent diagnostic terminology to refer to neurodevelopmental disorders. There is no internationally agreed label for unexplained language problems. Different terminology is used within the papers reviewed in this study but they are all referring to unexplained language problems. A recent study employing a Delphi method to achieve consensus on terminology from a professional consortium found ‘developmental language disorder’ (DLD) to be the preferred term (Bishop et al., 2016), which will be used in this article.

Acquisition of Multiple Languages

The number of children with English as a second language is increasing in the UK and U.S., and the range of first languages is changing. Dual language children

might acquire languages together ‘simultaneously’, or one might be introduced later, termed ‘sequential’ (De Houwer, 1995). The U.S. Census Bureau does not track rates of bilingualism, but a 2016 estimate indicated that 21.1% of people (>5 years) speak a language other than English; a slight increase from 19.6% in 2009 (United States Census Bureau, 2018). The 2011 Census taken in England found approximately 9.6% of children (3-15 years) did not have English as a first language (Office for National Statistics 2011). Whilst speaking multiple languages is a positive attribute, add to this a diagnosis of DLD, and a unique challenge is posed to professionals as to how best to support language development. Kohnert (2010) has urged for further research to be conducted into the effectiveness of language intervention for these children.

It is known that across languages, acquisition does not occur identically either in rate or style (Bedore & Pena, 2008). Maital, Dromi, Sagi, and Bornstein (2000) found that during early acquisition, Hebrew speaking preschoolers used more nouns compared with English speaking peers, despite comparable lexical size. Additionally, in an investigation of how children who spoke Welsh, Spanish or English treated novel words, Gathercole, Mon Thomas, and Evans (2000) found that at age 2 only the Welsh and Spanish speakers approached words as collections. By age 4 significantly more Welsh speakers continued to do so which reflects the nature of how the Welsh language treats nouns, compared to English and Spanish.

Dual Language Children with DLD

Not only do children with DLD develop language atypically, but input of first languages is restricted in formal education settings. Early L1 acquisition often begins at home, but exposure may become poor during school years. Early L2 input usually comes from television and contact with peers, while at school, children are abruptly immersed into a curriculum presented in L2.

Research has shown that monolingual and bilingual children with DLD have similar linguistic skills in the shared language. Spoelman and Bol (2012) found no significant difference for subject-verb agreement between 16 monolingual and bilingual children aged 5;11 with DLD. Hakansson, Salameh, and Nettekadt (2003) and Salameh, Hakansson, and Nettekadt (2004) found the level of development of L1 and L2 to be low in 10 children with DLD aged 4-6 years, compared to 10 age-matched bilingual peers without DLD. Crutchley, Botting and Conti-Ramsden, (1997) and Crutchley (1999) used the data of children attending language units in the UK, participating in the Manchester Language Study. Of the total 242 cohort, both studies found that the 26 dual language learners with DLD had more complex and persistent problems with the language of the curriculum (L2) than monolingual peers with DLD for whom it is their L1. Comparing 31 monolingual and eight dual language children, all with DLD and an age span of 6;11-7;7, Paradis et al. (2003) found language development to be very similar, especially for tense bearing morphemes, though it should be noted that participants were simultaneous bilinguals, while participants in other studies were sequential bilinguals.

Cross-Linguistic Transfer

There is evidence to suggest that proficiency of L1 use influences the rate of growth in L2 (Cummins, 1991; Lasagabaster, 2001). The acquisition of one language having influence on the acquisition of another is referred to as “cross-linguistic transfer”. This phenomenon is thought to occur as a result of common underlying cognitive processes that are associated with language development. Working memory and non-verbal intelligence are considered essential cognitive processes for cross-linguistic transfer to occur (Adesope et al., 2010; Barac et al., 2014). Verhoeven, Steenge and van Balkom (2012) query whether children with DLD can experience

cross-linguistic transfer, as the majority demonstrate difficulties with these processes (Kohnert, Kan and Conboy, 2010). Cross-linguistic effects on the vocabulary skills of bilingual children aged 2-5 years with DLD were measured by Lesemen (2000) who found that intervention in L2 resulted in a growth in L2 with L1 remaining stagnant, whilst Schaerlaekens et al. (1995) observed a decline in L1. Both studies indicate a lack of language transfer in these children. Understanding the nature of cross-linguistic transfer is crucial to optimise the educational support for dual language children and can improve the planning of intervention for those with DLD.

Thordardottir (2010) conducted a review of evidence relating to interventions for dual language children with DLD. There were very few studies eligible for review, many of which were considered to be low quality. The main findings were that no monolingual intervention outperformed bilingual interventions, and there are advantages to bilingual approaches, including preservation of L1. Despite the rise in bilingualism, there is a lack of bilingual therapists which impedes research.

Bilingual children with DLD do not acquire language in the same way as their bilingual peers without DLD, or indeed their monolingual peers with DLD. It is important that their language development is supported both for use in the classroom and within their home communities. There is no robust evidence to indicate whether or not intervention in one language can also benefit the other. In response to the growth in the population of multilingual children, this article systematically collates and reviews the evidence generated by studies conducted internationally regarding the nature of bilingual language interventions.

Objectives

The aim of this study was to examine the impact of speech and language therapy for dual language children with DLD. This was achieved by addressing the following

questions:

- How and to what extent do interventions for dual language children with DLD affect the first language (L1)?
- How and to what extent do interventions for dual language children with DLD affect the second language (L2)?
- How do bilingual interventions for dual language children with DLD compare with L2-only interventions?
- Is there a transfer of learning between languages in L2-only and other interventions and if so, to what extent does this occur?

Methods

A systematic review methodology was adopted and a protocol was constructed prior to the initiation of the review.

Search strategy

A literature search strategy was developed using medical subject headings (MeSH) and text words relating to DLD in dual language children: multilingu* OR bilingu* OR “second language”AND “language disorder” OR “language impairment”AND children OR paediatrics. EBSCOhost was used to search CINAHL complete, Medline, PsychINFO, PsychArticles, Ebook collection, British Education Index, Audiobook collection, ERIC, Education Administration Abstracts, Child Development and Adolescent Studies in April 2017. The search included databases across health, psychology and education because of the cross-disciplinary nature of speech and language therapy. The authors felt that the language used to denote

interventions would be too disparate to include in the formal search strategy. Therefore, the nature of the publication, i.e. whether it was an original research report of an intervention or a commentary, was determined at the screening stage.

Eligibility criteria:

Studies were selected according to the following inclusion and exclusion criteria:

Inclusion criteria:

Types of studies - all types of interventional studies with no exclusion placed on study design or data type.

- Types of articles - original, peer-review research articles.
- Types of participants - dual language children with DLD.
- Types of measures - all outcome measures were considered.
- Location of research - no restrictions placed on location.
- Date of publication - no restrictions on date of publication.
- Language of publication – research articles published in English.

Exclusion criteria:

Articles not reporting an original, peer-reviewed interventional study of interventions for dual language children with DLD.

- Articles not published in English.

Study selection

Two authors (HA and HH) independently screened the titles and abstracts yielded by the search against the eligibility criteria. Articles were categorised as ‘relevant’ or ‘not relevant’, first on title, then by abstract and finally by full article (see Appendix A). Full reports were obtained for all titles that appeared to meet the eligibility criteria. Where uncertainty existed, studies were included for screening at the next level of review.

Data collection process

Data from the included articles were extracted using a pre-constructed form. Reviewers (HA and HH) extracted data from articles independently, resolving disagreements through discussion. A third person (SJ) was utilised to oversee this.

Outcomes

The unknown outcomes of studies examining the effect of interventions for dual language children with DLD rendered a priori determination impossible. This review reports all outcomes included in the studies.

Risk of bias

The search strategy imposed no study design restriction; therefore, a non-design-specific quality assessment tool has been selected to assess bias, as presented by Baxter et al. (2015), adapted from the Cochrane Collaboration (2011), to appraise observational and experimental studies (Table 1). As acknowledged by Higgins, Altman and Sterne (2018), there is bias associated with studies which do not employ randomisation or systematic methods. In emerging areas of research, particularly those which are clinical in nature, designs typically considered low quality are important to consider. “Quality”

can be defined as measure of methodological strength, or the extent to which the design and conduct of a study prevents systematic errors, or bias. "Bias" can be defined as the potential hindrances to evidence quality, determined through assessing features of study design including: measures for blinding, length of follow up and statistical analyses.

Bias was assessed independently by two authors (SJ and HH); no discrepancies occurred. Each design characteristic was given a classification of low, medium or high. Overall categorisation was achieved by aggregation, and where numbers were equal the higher bias classification was recorded.

Table 1: Quality appraisal for individual studies

Citation	1. Selection Bias	2. Performance Bias	3. Detection Bias	4. Reporting Bias	Overall risk of bias	Details of Concerns
	Method used to generate the allocation sequence, methods used to conceal allocation sequence. Presence of control, characteristics of participants, +/- 10 sample	Measures used to blind participants and personnel and outcome assessors, presence of other potential threats to validity. Collection and assessment of speech sample.	Accuracy of measurement of outcomes, length of follow up. Reliable tool used, adequate speech sample, outside laboratory recording, immediate versus longer-term follow-up.	Selective reporting, accuracy of reporting. Use of inferential versus descriptive statistics, pooled or individual reporting.	Lower/higher	
Gutierrez-Clellen <i>et al.</i> (2012)	Low	Low	Low	Medium	Lower	Dubious reporting of statistical significance when p was equal to or greater than 0.05.
Simon-Cerejido <i>et al.</i> (2013)	Low	Low	Low	Medium	Lower	Dubious reporting of statistical significance when p was equal to or greater than 0.05.
Ebert <i>et al.</i> (2014)	Medium	Low	Low	Low	Lower	Small sample size & no concealment to condition.
Pham <i>et al.</i> (2014)	Medium	Low	Low	Low	Lower	Small sample size & no concealment to condition.
Thordardottir <i>et al.</i> (2015)	Low	Low	Low	Low	Lower	Small sample size and no follow up for control group.
Gutierrez-Clellen and Simon-Cerejido (2014)	Medium	High	Low	Medium	Medium	No concealment to condition. Broad confidence intervals.
Restrepo <i>et al.</i> (2013)	Low	Medium	Medium	Low	Medium	Unclear as to whether blinding of outcome assessors occurred. Self-developed outcome measures.
Ebert <i>et al.</i> (2012)	High	High	Low	Medium	Higher	Small sample of two. No randomisation. Overstating of NCP findings.
Thordartottir <i>et al.</i> (1997)	High	High	High	High	Higher	Small sample of one. Small number of outcomes, devised by research team.

Data synthesis

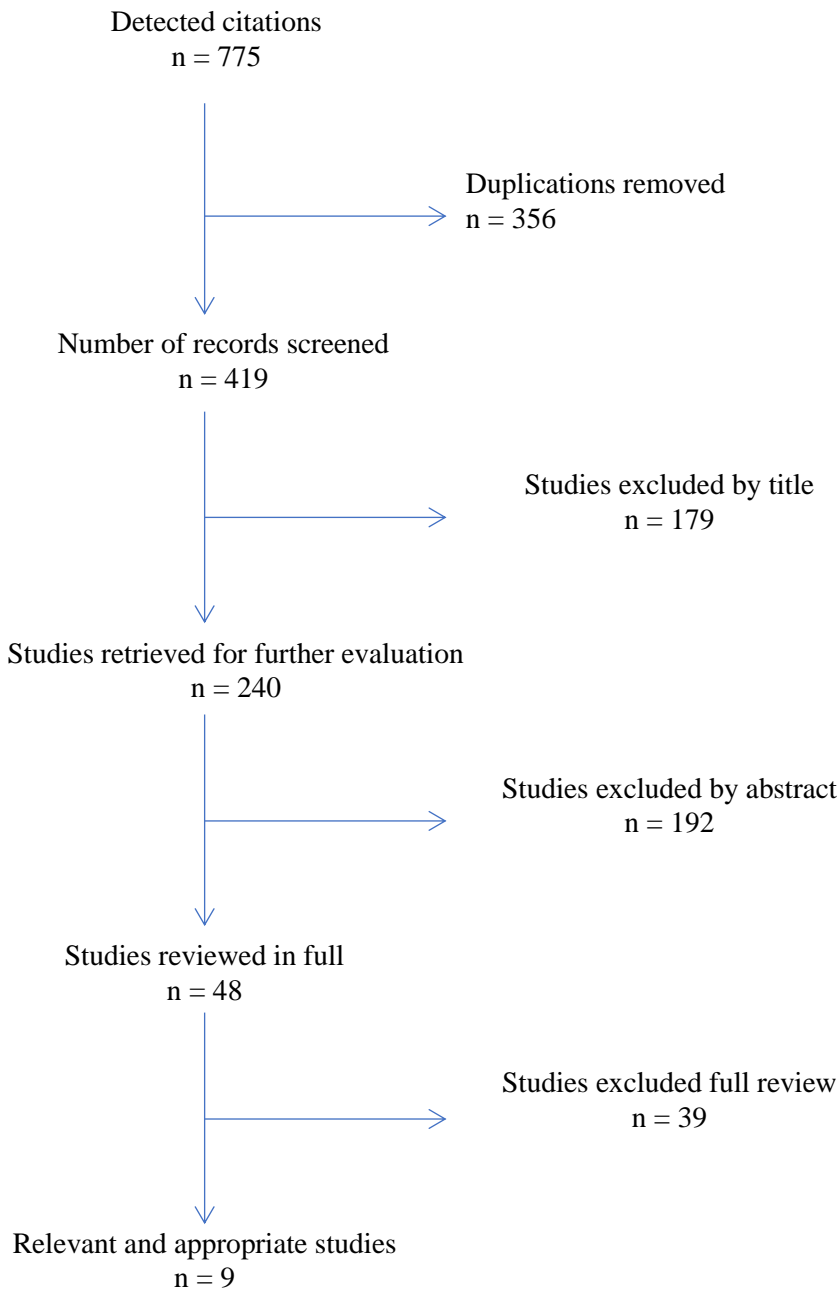
The likelihood of data synthesis for a meta-analysis was not pre-determinable. Kohnert (2013) acknowledges that relatively few high-quality interventional studies have been conducted in this area. Instead, a data reporting approach with conclusion synthesis was taken i.e. the findings are reported in relation to the research questions, and the conclusions across studies are synthesised. Interventions are compared for effectiveness irrespective of individually reported outcome measures.

Results

Following identification of 419 citations (excluding duplications), nine met the inclusion criteria and were reviewed in full (process depicted in Figure 1). The paucity of studies meant that all studies reporting intervention outcomes were included. Eight studies were conducted in the U.S. and one in Canada (Thordardottir et al., 2015). Seven investigated Spanish-English dual language children, one investigated an Icelandic-English child (Thordardottir, Weismer and Smith, 1997) and one had participants with French as their L2 and a range of first languages (Thordardottir et al., 2015).

Two studies report on the same intervention but present outcomes specific to L1 and L2 independently (Simon-Cerejido et al., 2013; Gutierrez-Clellen et al., 2012). Pham, Dunahy and Kohnert (2014) offers longitudinal follow up data to the study conducted by Ebert et al. (2014).

Figure 1: Review Flow Diagram



Eight of the nine presented findings from 532 participants, of which 54 were typically developing dual language controls in one study. Of the 532, 196 (34%) participants were girls and 336 (66%) were boys. The remaining study (Simon-Cereijido and Gutierrez-Clellen, 2014) reported 74 participants in the abstract, 107 in the methodology and varying participant numbers in the results, which the authors attributed to participants' inability to undertake assessments rather than attrition.

Seven of the nine reported at least a bilingual intervention in comparison to an L2-only intervention; three reported additional comparator conditions including non-linguistic cognitive processing, deferred treatment, a bilingual mathematics intervention and an L2-only mathematics intervention. Ebert, Rentmeester-Disher and Kohnert (2012) report a non-linguistic cognitive processing treatment only. Five studies report L1 and L2 outcome measures, two reported L2 outcome measures only, and one reported L1 measures only.

Appraisal of study designs

Quality appraisal data is available in Table 1, and a summary of the study designs can be found in Table 2.

Six of the nine studies reported a randomised controlled trial (RCT) design with between two and five treatment conditions; three of which did not state randomisation methods (Ebert et al., 2014; Pham et al., 2014; Restrepo et al., 2013). Simon-Cereijido and Gutierrez-Clellen (2014) report a non-randomised controlled group study. Ebert, Rentmeester-Disher and Kohnert (2012) used a single-subject experimental design with two participants, and Thordardottir, Weismer and Smith (1997) used a single-case alternating design.

Table 2: Study designs

Study	Gutierrez-Clellen <i>et al.</i> (2012) and Simon-Cerejido <i>et al.</i> (2013)	Ebert <i>et al.</i> (2014)	Pham <i>et al.</i> (2014)	Thordardottir <i>et al.</i> (2015)	Restrepo <i>et al.</i> (2013)	Simon-Cerejido and Gutierrez-Clellen (2014)	Ebert <i>et al.</i> (2012)	Thordardottir <i>et al.</i> (1997)
Design	RCT	RCT	RCT	RCT	RCT	Randomised group comparison	Single-case alternating conditions	Single-subject experimental
Method of randomisation (if applicable)	Groups of 2-3 according to school Schools with >3 participants: random allocation to groups; random allocation to treatment condition	did not elaborate	did not elaborate	Participants randomised by computer programme, condition concealed until start of treatment	did not elaborate	n/a	n/a	n/a
No. of treatment conditions	2	3	3	3	4	2	1	2
Types of treatments	1) Bilingual 2) English-only	1) Bilingual 2) English-only 3) Non-linguistic cognitive processing	1) Bilingual 2) English-only 3) Non-linguistic cognitive processing	1) Bilingual 2) English-only 3) No (delayed) treatment	1) Bilingual vocabulary 2) English-only vocabulary 3) Bilingual maths 4) English-only maths	1) Bilingual 2) English-only	1) Non-linguistic cognitive processing	1) Bilingual 2) English-only
No. of participants	188	59	48	29	256	107 (data on 98 only)	2	1
Age of participants (years; months)	4;5 (mean value) SD =4 months	5;6 – 11;2	5;6 – 11;3	5;0 (mean value)	3;7 – 5;8	4;5 (mean value) SD =4 months	8;4 – 7;5	4;11
Attrition (%)	1.5	7.8	33.3	9.3	44.1 (at final measure)	8.5	0	0
Recruitment Criteria	Scored below cut-off on language measures No hearing loss, mental retardation, emotional disturbances or neurological impairment Normal nonverbal IQ.	Scored below average on language tests. No other diagnosis.	Recruited from school-based special education services. No primary health concerns.	Recruited from multiple sites providing SLT treatment. Language score >1.5 SD below mean. Noverbal IQ >70. Hearing within normal limits.	No hearing impairment, cognitive delays or neurological deficit Non-verbal IQ score >70	No hearing impairment, mental retardation, motor difficulties, neurological deficits or emotional disturbances Non-verbal cognitive scores in normal range.	Delayed language, academic difficulty. No frank sensory, motor, cognitive or social-emotional deficits.	n/a
Timeframe of Treatment Conditions	45 minutes 4 days per week 12 weeks	75 minutes 4 days per week 6 weeks	75 minutes 4 days per week 6 weeks	50 minutes 1 day per week 16 weeks	45 minutes 4 days per week 12 weeks	45 minutes 4 days per week 9 weeks	75 minutes 4 days per week 5 weeks	50 minutes 2 days per week 7 weeks

Recruitment

All studies excluded children with other diagnoses associated with DLD. In seven studies, hearing screenings were conducted. Three studies recruited children in receipt of school-based special education services, and five recruited children who met referral criteria for special education services. Simon-Cerejido and Gutierrez-Clellen (2014) recruited in collaboration with other studies (Gutiérrez-Clellen and Restrepo 2005–2009). Participants continued to receive standard speech and language therapy during participation in four of the studies, four did not state, and one was conducted during school holidays, during which participants did not receive simultaneous treatment. The age of participants ranges across studies, which is important to consider as language development accelerates at different rates; at early ages language skills of monolingual children with LI are comparable to bilinguals (Kohnert, Windsor, & Ebert, 2009).

Interventions

Interventions lasted 5-16 weeks, and were delivered in 14-48 sessions. Treatment sessions individually lasted 45-90 minutes. Given this variance, the details of each intervention are presented.

Gutierrez-Clellen, Simon-Cerejido and Sweet (2012) and Simon-Cerejido, Gutierrez-Clellen and Sweet (2013) report a two condition RCT comparing a bilingual intervention with an L2-only intervention for 188 children (average age: 4;5). Simon-Cerejido, Gutierrez-Clellen and Sweet (2013) investigated factors associated with rates of L1 development in children with DLD, whilst Gutierrez-Clellen, Simon-Cerejido and Sweet (2012) evaluated the influence of various factors on the same children's acquisition of L2. Both interventions were delivered four days a week, 45 minutes per day for 12 weeks in groups of no more than four. In the L2-only intervention, all lessons

were delivered in English. For the bilingual intervention the first lesson of the week was delivered in Spanish; the consecutive lesson was repeated in English. The language of the remaining lessons was alternated weekly.

Ebert et al. (2014) present an RCT with three interventions and a deferred control to 59 children aged 5;6 – 11;2. Similar to Gutierrez-Clellen, Simon-Cerejido and Sweet (2012) and Simon-Cerejido, Gutierrez-Clellen and Sweet (2013), participants continued to receive language therapy. The study sought to evaluate the impact of bilingual, L2-only, and non-linguistic cognitive processing interventions for children with DLD. All conditions contained 75 minutes of activity four times per week for six weeks; half the study period of Gutierrez-Clellen, Simon-Cerejido and Sweet (2012) and Simon-Cerejido, Gutierrez-Clellen and Sweet (2013). The number of treatment sessions varied from 17 to 24 depending on the time of year, with participants completing 13-24 sessions. The same intervention was reported in the follow up (Pham et al. 2014).

Simon-Cerejido and Gutierrez-Clellen (2014) present a comparison of standard preschool (L2-only) teaching to 107 children with (average age: 4;5) using the Vocabulary, Oral Language and Academic Readiness (VOLAR) curriculum, for 45 minutes, four days a week, for nine weeks. VOLAR was designed to facilitate language development in dual language learners with DLD, and the authors examine the effect of its implementation within preschool curriculum. The first day of each week was taught in L1, with the same content in L2 on the consecutive day. For the remaining two days, the languages alternated. The control group attended an L2 speaking preschool.

Restrepo, Morgan and Thompson (2013) describe an intervention delivery similar to that of Gutierrez-Clellen, Simon-Cerejido and Sweet (2012) and Simon-Cerejido, Gutierrez-Clellen and Sweet (2013); a 12 week programme, 4 days per week

for 45 minutes per day providing an L2-only vocabulary intervention, or a bilingual vocabulary intervention that delivered L1 on day one, L2 on day two, and alternated on days three and four. Groups consisted of between two and five children; a total of 256 children aged 3;7 – 5;8 participated. This study included two additional treatment conditions – an L2-only maths intervention and a bilingual maths intervention – both focused on language through teaching maths activities. The study's purpose was to compare the efficacy of each intervention for dual language learners with DLD.

In the RCT presented by Thordardottir et al. (2015), 29 children (average age: 5;0) received 16 intervention sessions provided weekly, each lasting 50 minutes. A speech pathologist addressed vocabulary and syntax for 20 minutes each. Parents attended sessions, and those in the bilingual condition participated to support L1. It was reported to be challenging to sustain parent involvement, with some appearing uncomfortable participating. The control group were not seen during this period. The study focused on the clinical effectiveness of monolingual versus bilingual interventions.

Thordardottir, Weismer and Smith (1997) present a single-case alternating treatment design to investigate differences between a bilingual and L2-only intervention for a child aged 4;11. Fourteen 50 minute sessions were delivered twice weekly in a randomly determined order. This study was solely concerned with determining the effect this had on acquisition of L2 words. In the L2-only intervention the participant's L1 utterances were not responded to, and reminders were given to use L2. In the bilingual intervention, the participant's utterances were responded to in either language. Vocabulary was presented through semi-structured play activities.

Ebert, Rentmeester-Disher and Kohnert (2012) present a single-subject multiple baseline design in the delivery of a non-linguistic cognitive processing intervention to

two bilingual participants with DLD aged 7;5 and 8;4. Pre and post-treatment measures were recorded, and repeated measures were taken throughout. Sessions lasting 90 minutes were delivered four days a week. Across five weeks, 14 intervention sessions and five testing sessions were conducted. One participant received 13 of 14 sessions and the other received 7 of 14. Sessions consisted of five activities each lasting 15 minutes, a 15 minute break, with the remaining 15 minutes dedicated to conducting repeated measures. One day a week, no measures were taken and the time was used for additional intervention. Like Ebert et al. (2014), six treatment activities were included; three computer-based and three interactive activities

Outcome Measures

Ebert et al. (2014) present pre and post-treatment outcome measures; Pham, Dunahy and Kohnert (2014) present 3 month follow up data. Gutiérrez-Clellen et al, 2012 report pre and post-treatment measures plus 3 and 5 month follow up data for L2 outcomes; Simon-Cerejido, Gutierrez-Clellen and Sweet (2013) report the respective L1 outcomes. Thordardottir et al. (2015) undertook a pre and post-test evaluation of L1 and L2 measures with an additional 2 month follow up period. Restrepo, Morgan and Thompson (2013) report pre and post-treatment measures within 2-3 weeks with 4 and 8 month follow up data.

Ebert, Rentmeester-Disher and Kohnert (2012) used a repeated measures design, collecting data three times per week during the intervention period. One study did not clearly identify the timing of their data collection, but it was assumed to be pre and post-treatment given the analyses conducted (Ebert et al., 2014). Simon-Cerejido and Gutierrez-Clellen (2014) report two episodes of follow up data four and seven months after baseline testing.

In the nine studies reviewed, 43 pre and post-test outcome measures were reported (see Table 3). Only in four did any overlap exist for reported measures; one of which was a follow up and was therefore not suitable for aggregation. Given the range of measures, the findings have not been meta-analysed and instead are presented in the context of each measure.

Table 3: Outcome Measures

Vocabulary Measure	Description	Studies
ROW-E	Receptive One-Word Picture Vocabulary Test - English: measures of receptive vocabulary for English	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)
ROW-S	Receptive One-Word Picture Vocabulary Test - Bilingual Edition: measures of receptive vocabulary for Spanish	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)
EOW-E	Expressive One-Word Picture Vocabulary Test - English: measures of expressive vocabulary for English	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)
EOW-S	Expressive One-Word Picture Vocabulary Test - Bilingual Edition. Measures of expressive vocabulary for Spanish	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)
ROWPVT-E	Receptive One Word Picture Vocabulary Test – English: measures of receptive vocabulary for English	Ebert <i>et al.</i> (2012)
ROWPVT-S	Receptive One Word Picture Vocabulary Test - Spanish: measures of receptive vocabulary for Spanish	Ebert <i>et al.</i> (2012)
EOWPVT-E	Expressive One Word Picture Vocabulary Test – English: measures of expressive vocabulary for English	Ebert <i>et al.</i> (2012)
EOWPVT-S	Bilingual version of the Expressive One Word Picture Vocabulary Test: Measures of expressive vocabulary for Spanish	Ebert <i>et al.</i> (2012)
EOWPVT (adapted for French)	Expressive One Word Picture Vocabulary Test adapted for French speakers	Thordardottir <i>et al.</i> (2015)
CELF-(4)E	Clinical Evaluation of Language Fundamentals, 4 th Edition: measures of global language skills, English	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)
CELF-(4)S	Clinical Evaluation of Language Fundamentals, 4 th Edition: measures global language skills, Spanish	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)
CELF-CD-E	Concepts and Following Directions subtest of CELF-(4)E	Ebert <i>et al.</i> (2012)
CELF-CD-S	Concepts and Following Directions subtest of CELF-(4)S	Ebert <i>et al.</i> (2012)
CELF-WS-E	Word Structure subtest of CELF-(4)E	Ebert <i>et al.</i> (2012)
CELF-WS-S	Word Structure subtest of CELF-(4)S	Ebert <i>et al.</i> (2012)
CELF-FS-E	Formulated Sentences subtest of CELF-(4)E	Ebert <i>et al.</i> (2012)
CELF-FS-S	Formulated Sentences subtest of CELF-(4)S	Ebert <i>et al.</i> (2012)
CELF-RS-S	Recalling Sentences subtest of CELF-(4)S	Ebert <i>et al.</i> (2012)
Échelle de vocabulaire en images Peabody (EVIP)	French version of the Peabody Picture Vocabulary Test	Thordardottir <i>et al.</i> (2015)
Spanish receptive vocabulary	Assesses receptive vocabulary knowledge of target words taught	Restrepo <i>et al.</i> (2013)
English expressive vocabulary	Assess expressive vocabulary knowledge of target words taught	Restrepo <i>et al.</i> (2013)
Spanish expressive vocabulary	Assesses expressive vocabulary knowledge of target words taught	Restrepo <i>et al.</i> (2013)
Conceptual receptive vocabulary	Measure of the total concepts known regardless of language	Restrepo <i>et al.</i> (2013)
Conceptual expressive vocabulary	Measure of the total concepts known regardless of language	Restrepo <i>et al.</i> (2013)

Repetition Measure	Description	Studies
English NWR	English nonword repetition	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)
Spanish NWR	Spanish nonword repetition	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)
Spanish Sentence Repetition Task (SRT)	Measure of ability to repeat a sentence spoken by an evaluator	Simon-Cerejido <i>et al.</i> (2013)

Length of Production Measure	Description	Studies
English MLU _w	Mean length of utterance, measured in words, within a spontaneous narrative sample, English	Gutierrez-Clellen <i>et al.</i> (2012) Simon-Cerejido and Gutierrez-Clellen (2014)
Spanish MLU _w	Mean length of utterance, measured in words, within a spontaneous narrative sample, Spanish	Simon-Cerejido <i>et al.</i> (2013) Simon-Cerejido and Gutierrez-Clellen (2014)
French MLU _w	Mean length of utterance, measured in words, within a spontaneous narrative sample, French	Thordardottir <i>et al.</i> (2015)
English MLU _m	Mean length of utterance, measured in morphemes, within a spontaneous narrative sample, English	Gutierrez-Clellen <i>et al.</i> (2012)
English TNV	Total number of verbs used in a narrative sample, English	Gutierrez-Clellen <i>et al.</i> (2012) Simon-Cerejido and Gutierrez-Clellen (2014)
Spanish TNV	Total number of verbs used in a narrative sample, Spanish	Simon-Cerejido <i>et al.</i> (2013) Simon-Cerejido and Gutierrez-Clellen (2014)
English NDW	Number of different words used in a narrative sample, English	Gutierrez-Clellen <i>et al.</i> (2012) Simon-Cerejido and Gutierrez-Clellen (2014)
Spanish NDW	Number of different words used in a narrative sample, Spanish	Simon-Cerejido <i>et al.</i> (2013) Simon-Cerejido and Gutierrez-Clellen (2014)
Production of target vocabulary in school language	Production of target vocabulary in school language (L2)	Thordardottir <i>et al.</i> (1997)
Production of target vocabulary in home words	Production of target vocabulary in home words (L1)	Thordardottir <i>et al.</i> (1997)

Cognitive Measure	Description	Studies
Reynell Developmental Language Scales (RDLS)	Assessment for identifying speech and language delays and impairments in very young children	Thordardottir <i>et al.</i> (2015)
Auditory Serial Memory (ASM)	Measure of working memory for nonverbal auditory information	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)
Sustained Selective Attention (SSA)	Measure of selective attention in an auditory task	Ebert <i>et al.</i> (2014); Pham <i>et al.</i> (2014)

Morphosyntactic Measures	Description	Studies
English Picture Description Task	Verb and argument structure assessment, English	Gutierrez-Clellen <i>et al.</i> (2012)
Spanish Picture Description Task	Used as an assessment for verb and argument structure, Spanish	Simon-Cerejido <i>et al.</i> (2013)

Study Findings

Effect on first language outcomes

Simon-Cerejido, Gutierrez-Clellen and Sweet (2013) describe the language of intervention to be a ‘marginal’ predictor of change on the Spanish sentence repetition task (SRT); however, since this did not meet the significance level set by the authors, it is not reported here as significant. Use of English in the classroom was a significant predictor of performance on Spanish SRT ($F(1,80)=5.51, p=0.02$); children using more English made smaller gains during the follow up period. Baseline conceptual vocabulary was found to be a predictor of change for the vocabulary measures Spanish Number of Different Words and Total Number of Verbs ($f(1,94)=4.69, p=0.03$ and $f(1, 94) = 6.05, p=0.02$ respectively), with lower scores associated with greater gains immediately following intervention. This was not sustained at follow up.

Ebert et al. (2014) found that L1 non-word repetition (NWR) measures did improve in the non-linguistic cognitive processing (NCP) group (12.0PPC points improvement, $t(15)=2.61, p=0.02, d=0.54$), but not in bilingual or L2-only groups, although not statistically significant with respect to absolute effectiveness. The authors state that “absolute effectiveness” was used to show the within group change from pre to post testing. The bilingual group reached statistical significance in the Spanish-Bilingual Edition of the Expressive One-Word Picture Vocabulary Test: (EOWPVT-S), with an average improvement of 5.1 standard points, ($t(14)=2.37, p=0.032, d=0.43$) and 7.1 raw points ($t(14)=4.87, p<0.001, d=0.61$), although not statistically significant with respect to relative effectiveness. The authors used “relative effectiveness” to show changes between groups, using analysis of covariance, with the pre-test score acting as the covariate. In the bilingual group, only the raw score on the fourth edition of the Spanish Clinical Evaluation of Language Fundamentals (CELF-4S) improved

significantly (on average by 4.9 raw points, $t(14)=2.44$, $p=0.028$, $d=0.19$) but this relationship was not present in relative effectiveness.

Pham, Dunahy and Kohnert (2014) found Spanish NWR to be the only L1 outcome measure that continued to improve at follow up, demonstrating a positive change in phonological processing at a rate of 5.9%/testing time in the L2-only group. This appears to have been modest as no difference was detected between groups. All other L1 outcome measures were maintained.

Thordardottir et al. (2015) report that mean length of utterance for words (MLU_w) increased for the L2-only group, but decreased slightly for the bilingual and control groups, however this was not significant ($p=0.79$).

Simon-Cerejido and Gutierrez-Clellen (2014) found that participants in the bilingual group performed better overall than the control group. For L1 outcomes, 53-60% of the control group performed below the average for the children in the bilingual intervention group (MLU_w -Spanish, $U_3 = 0.60$, NDW Spanish, $U_3 = 0.53$ and TNV-Spanish, $U_3 = 0.53$).

Restrepo, Morgan and Thompson (2013) found that children in the bilingual group did significantly better in L1 receptive (5-7 words higher) and expressive (18-21 points higher) measures of vocabulary immediately following intervention, which persisted to follow up, but rate of language growth was significantly lower for this group.

Ebert, Rentmeester-Disher and Kohnert (2012) found that Participant 1 made significant improvements on Spanish vocabulary measures, Receptive and Expressive One-Word Picture Vocabulary Tests (ROWPVT-S and EOWPVT-S); but made no other gains on the pre and post-test measures. Participant 2 demonstrated significant improvement in the ROWPVT-S measure only. With respect to during-intervention

repeated measures, both participants made significant gains on the L1 NWR task (P1 $d=2.01$ and P2 $d=1.12$).

Effect on second language outcomes

In examining the effect of bilingual against L2-only intervention on L2 outcomes, Gutierrez-Clellen, Simon-Cerejido and Sweet (2012) assessed narrative samples for mean length of utterance for words (MLU_w) and morphemes (MLU_m), and found greater rates of improvement in L2 for the bilingual group ($MLU_w, d = 1.79$), showing language of intervention to be a predictor of MLU_w ($\beta = 1.28, SE = 0.58$), $F(1, 21) = 4.94, p = 0.04$). Greater baseline scores in Spanish MLU_w were associated with greater gains in English MLU_w ($\beta = 0.52, SE = 0.23$), $F(1, 21) = 5.24, p = 0.03$ and MLU_m ($\beta = 0.54, SE = 0.24$), $F(1, 21) = 5.02, p = 0.04$). L2 vocabulary scores at baseline were a predictor of growth on the English picture description task, used to assess spontaneous production of language, with children improving at greater rates if they demonstrated better baseline L2 vocabulary scores ($\beta = 0.74, SE = 0.37$), $F(1, 145) = 4.01, p = 0.047$), or more proficient L2 use at baseline ($F(1, 145) = 6.18, p = 0.01$).

With regards to absolute effectiveness, Ebert et al. (2014) found that all three treatment conditions significantly improved scores in EOWPVT-E (L2-only group: $\uparrow 7.2$ standard points, $t(16)=5.76, p<0.001, d=0.82$; and 10.1 raw points, $t(16)=6.78, p<0.001, d=0.79$; bilingual group: $\uparrow 4.3$ standard points, $t(14)=2.73, p<0.017, d=0.55$; and 5.7 raw points, $t(14)=2.95, p=0.011, d=0.36$; NCP group: $\uparrow 4.6$ raw points, $t(15)=3.02, p=0.009, d=0.29$). No difference was found between the groups when assessing relative effectiveness. All groups significantly improved scores on the English Clinical Evaluation of Language Fundamentals (CELF-4E) (L2-only group: $\uparrow 6.8$ standard points, $t(16)=3.96, p=0.001, d=0.60$; and 17.1 raw points, $t(16)=6.93, p<0.001, d=0.45$;

bilingual group: $\uparrow 6.5$ standard points, $t(14)=2.81$, $p=0.014$, $d=0.72$; and 15.5 raw points, $t(14)=2.81$, $p<0.001$, $d=0.50$; NCP group: $\uparrow 3.9$ standard points, $t(15)=2.60$, $p=0.020$, $d=0.33$; and 8.4 raw points, $t(15)=3.29$, $p=0.005$, $d=0.25$). The L2-only and bilingual groups improved significantly more than the NCP group with regard to relative effectiveness of raw scores (L2-only vs. NCP, $p=0.011$, $d=1.37$; bilingual vs. NCP $p=0.027$, $d=1.45$).

Pham, Dunahy and Kohnert (2014) found statistically significant improvements in the English language outcome measures ROWPVT, EOWPVT and CELF in the L2-only and bilingual treatment groups (L2-only group, 4.1% 3.5% and 3.9% for each measure respectively; bilingual group, 3.3%, 3.5% and 4.8% respectively). Also, the bilingual condition demonstrated significance for positive change for the English NWR (5.3%). Statistical significance was recorded as $p<0.05$ or a z-score $> +/-1.96$.

Thordardottir et al. (2015) conducted an ANOVA to measure the differences pre and post-test for L2 receptive vocabulary ($F(2, 26) = 10.362$, $p = 0.000$, $n^2 = 0.463$) and expressive vocabulary ($F(2, 27) = 14.186$, $p = 0.000$, $n^2 = 0.532$), both reached significance. Post hoc testing showed that the control group had significantly lower scores than the treatment groups for receptive ($p = 0.000$, and $p = 0.03$) and expressive vocabulary ($p = 0.000$ and $p = 0.001$). No significant differences were found between bilingual and L2-only interventions ($p = 0.0362$ and $p = 0.203$) or between any groups for the story retell syntactic probe ($p = 0.230$). The MLU_w , group means increased for all groups, but the difference between groups was not significant ($p=0.517$). Upon follow up, time had a significant effect on receptive vocabulary ($F(2, 32) = 97.734$, $p = 0.000$, $n^2 = 0.859$) and expressive vocabulary ($F(2, 34) = 98.694$, $p = 0.000$, $n^2 = 0.853$) with significant effects found between Time 1 and 2 and Time 1 and 3 but not between

Time 2 and 3. For story retell, time was again significant ($F(2, 28) = 18.745, p = 0.000, \eta^2 = 0.572$) however, only between Time 1 and 3 ($p = 0.009$).

Simon-Cerejido and Gutierrez-Clellen (2014) found that between 62.5 and 87.5% of the control group scored lower than the average of the bilingual group on measures of expressive vocabulary (MLU_w-English, $U_3 = 0.625$, NDW English, $U_3 = 0.875$ and TNV-English, $U_3 = 0.875$).

Following intervention, Restrepo, Morgan and Thompson (2013) found that the L2-only vocabulary group did not exceed the bilingual vocabulary group in the receptive measure, but scored statistically significantly higher than either mathematics group or the control group. Although both mathematics groups demonstrated significant improvements at follow up, they did not exceed final measures of either vocabulary group. The growth rate for the L2-only vocabulary group was not sustained at follow up. For the L2 expressive measure, the L2-only vocabulary group outperformed all other groups, but no statistical significance was found between the vocabulary groups. The vocabulary groups did not sustain growth at follow up.

Ebert, Rentmeester-Disher and Kohnert (2012) found that participant one (P1) made statistically significant improvements on two elements of the CELF-4E: the receptive measure 'concepts and following directions' and syntax measure 'word structure'. Participant 2 (P2) made no gains in L2 vocabulary pre and post-test measures. With respect to the during-intervention repeated measures, P1 made significant gains on L2 measures (sentence repetition, $d=2.03$, rapid automatic naming, $d=2.3$ and nonword repetition English, $d=2.06$). P2 made significant gains on two measures (sentence repetition, $d=1.8$ and nonword repetition English, $d=1.27$)

Thordardottir, Weismer and Smith (1997) found no difference in the total number of L2 words learned between the bilingual or L2-only intervention.

An overview of the findings in the context of their effect on L1 and L2 can be located in Table 4. Table 5 synthesises the key factors relating to study design and outcomes.

Table 4: Key Findings

Study	Effect on first language (L1) outcomes	Effect on second language (L2) outcomes
Gutierrez-Clellen et al. (2012)	n/a	Greater improvement in L2 MLU _w and MLU _m in bilingual intervention group. Greater baseline L1 MLU _w scores predictor of gains in L2 MLU _w . Baseline L2 vocabulary scores predictor of growth in L2 picture description task. Use of L2 predictor of growth in L2 picture description task.
Simon-Cerejido et al. (2013)	Lower baseline conceptual vocabulary predictor of greater gains on L1 scores for subtests <i>number of different words</i> and <i>total number of verbs</i> (not sustained at follow up). Greater use of L2 in the classroom context predictor of smaller change in sentence repetition task.	n/a
Ebert et al. (2014)	L1 nonword repetition improvements in nonlinguistic cognitive processing intervention (not with relative effectiveness). Improvement in L1 expressive measures in bilingual intervention (not with relative effectiveness). Improvement in CELF-4S in bilingual intervention (not with relative effectiveness).	All intervention groups reached statistical significance for measures of L2 expressive language All intervention groups reached statistical significance for CELF-4E measures.
Pham et al. (2014)	All L1 pre-intervention scores sustained or improved at follow up. L1 nonword repetition continued to improve at follow up for L2-only intervention group. No overall difference between groups.	Improvements in L2 receptive and expressive language, CELF scores and L2 nonword repetition in bilingual intervention Improvements in L2 receptive and expressive language and CELF scores in L2-only intervention.
Thordardottir et al. (2015)	Post treatment MLU _w scores increased for the L2 only group; decreased for bilingual and no treatment groups (not significant).	L2 only and bilingual groups performed significantly better on expressive and receptive language probes than the control group.
Restrepo et al. (2013)	Improvement in L1 expressive measures in bilingual intervention (sustained at follow up). Improvement in L1 receptive measures in bilingual intervention (sustained at follow up). Rate of growth between intervention and follow-up significantly higher in other intervention groups.	Significantly higher scores in L2-only vocabulary group than other groups in receptive L2 (growth rate not sustained at follow-up). Significantly higher scores in L2-only vocabulary group than L2-only and bilingual mathematics groups in L2 expressive measures (growth rate not sustained at follow-up). Significant improvements in L2-only and bilingual mathematics groups but overall scores did not exceed either vocabulary group.
Simon-Cerejido and Gutierrez-Clellen (2014)	Bilingual group performed better overall than the control group on post-treatment language measures.	Bilingual programme had greater effects on improving L2 outcomes for post-treatment language measures compared to control group.
Ebert et al. (2012)	Participant 1: Significant gains on L1 receptive and expressive vocabulary measures Participant 2: Significant gains on L1 receptive vocabulary measures	Participant 1: Significant gains on three L2 measures during intervention. Improvement in some aspects of CELF-4E post-intervention. Participant 2: Significant gains on two L2 measures during intervention.
Thordardottir et al. (1997)	n/a	No difference in number of L2 words learnt between interventions.

Table 5: Synthesis of Key Factors

Citation	Intervention	Bilingual and L2 or L2 only Intervention	Risk of Bias	Outcomes
Gutierrez-Clellen <i>et al.</i> (2012)	RCT	Bilingual and L2	Lower	Bilingual intervention group demonstrated greater improvement in L2 for measures of MLU _w and MLU _m than L2-only group
Simon-Cerejido <i>et al.</i> (2013)	RCT	Bilingual and L2	Lower	No significant effect detected interventions for L1 outcomes.
Ebert <i>et al.</i> (2014)	RCT	Bilingual and L2	Lower	Bilingual intervention group showed improvement in L1 expressive measures and CELF-4S. All groups reached statistical significance for measures of L2 expressive language and CELF-4E
Pham <i>et al.</i> (2014)	RCT	Bilingual and L2	Lower	L2-only intervention group demonstrated improvements in L2 receptive and expressive language and CELF scores. They continued to improve L1 nonword repetition but no overall difference detected between intervention groups. Bilingual intervention group demonstrated improvements in L2 receptive and expressive language, CELF scores and L2 nonword repetition.
Thordardottir <i>et al.</i> (2015)	RCT	Bilingual and L2	Lower	L2 only group scored higher in MLU _w , which decreased slightly for bilingual and no treatment groups (not significant). Both L2 only and bilingual groups performed significantly better on expressive and receptive language probes than the control group.
Restrepo <i>et al.</i> (2013)	RCT	Bilingual and L2	Medium	Improvement in L1 expressive measures in bilingual intervention
Gutierrez-Clellen and Simon-Cerejido (2014)	Non-randomised controlled group study	Bilingual and L2	Medium	Bilingual intervention group performed better than children in the control group on both L1 and L2 measures.
Ebert <i>et al.</i> (2012)	Single-subject experimental design	L2 only	Higher	Bilingual intervention group demonstrated significant gains on L1 and L2 receptive and expressive vocabulary measures L2-only vocabulary group scored higher in receptive and expressive L2 than all other groups.
Thordartottir <i>et al.</i> (1997)	Single-case alternating treatment	Bilingual and L2	Higher	No difference in number of L2 words learnt in either intervention.

Discussion

The search generated a small number of articles (n=9) representing seven interventions. The findings of these studies have been presented according to their effect on L1 and L2 outcomes to reflect the research questions. The nature of language intervention, whether it is L2-only or bilingual, is also compared. Study quality varied by research design and sample size.

Effect on first language

The evidence relating to the impact on L1 outcomes is variable. No improvements in L1 outcomes in the L2-only group were achieved. Some studies were able to demonstrate improvement over time on some L1 outcomes across interventions. Ebert et al. (2014) demonstrated improvements in L1 NWR in the control group and EOWPVT-S (standard and raw) and CELF-4S (raw only) in the bilingual group. There is no evidence to indicate that these improvements were a result of the intervention (Ebert et al. 2014) and none were sustained at follow up (Pham et al. 2014). Simon-Cerejido, Gutierrez-Clellen and Sweet (2013) also observed no relationship between intervention type and improvements in L1. The comparative improvements to L1 are insufficient to determine any advantageous treatment with regard to L1 outcomes.

The application of general test measures, rather than measures that specifically test the targeted treatment domains, could have implications for the size of the effect seen. These general measures could be insensitive to the accurate size of gains seen in the directly targeted treatment domains. This is further supported by the substantial gains made on treated exemplars (Ebert et al. 2014, Pham et al. 2014). The presence of gains pre and post-treatment suggests that dual language children with DLD are able to generalise some learning to untreated exemplars examined in the general test measures.

Restrepo, Morgan and Thompson (2013) found that children in the bilingual group did significantly better in L1 receptive and expressive outcomes immediately following intervention. This persisted at follow up, in contrast with the findings of Ebert et al. (2014) and Simon-Cerejido, Gutierrez-Clellen and Sweet (2013). The post-intervention growth rates did not differ significantly between groups for receptive outcomes. It should be noted that a small decline in the expressive outcome measure at follow up resulted in a nonsignificant growth rate for the bilingual group, and significantly greater growth rates for the other treatment groups in comparison to the bilingual group. The continued growth in L1 expressive vocabulary in the other treatment groups, at a similar rate, could reflect natural growth over time. This suggests that the bilingual group experienced some attrition of the target vocabulary, but likely share similar natural growth seen in the other groups, resulting in the persistent significant difference in the expressive vocabulary measure at follow up in comparison to the other groups.

Attrition in the first language appears to be predicted by the extent of use of the second language, and is also observed in typically developing dual language learners in the absence of instruction of L1 (Jia and Aaronson, 2003). Over a 12 week study, Restrepo, Morgan and Thompson (2013) found that attrition of L1 occurred, yet Pham, Dunahy and Kohnert (2014) did not find any attrition in a study lasting 6 weeks. If L1 attrition can be limited following dual language instruction, children could make a bigger contribution to their home environment in which L1 is the dominant language. The importance of understanding the loss of the first language is particularly pertinent if interventions like bilingual vocabulary instruction, which are designed to enhance first language acquisition, are to be successful. There is naturally occurring attrition of L1 in typically developing dual language learners as they age through the education system

and consequently the growth of L2 exceeds the change in L1 (Jia and Anderson 2003). This could increase the meaningfulness of even small gains in L1 as they are harder to obtain, even in the typically developing population.

The absence of a strong, clear relationship between the language of intervention and L1 outcomes may be related to the dosage of intervention. In the RCTs, L1 was delivered for half of the bilingual intervention, which was half of the amount of L2 provided in the L2-only intervention. The types of tasks being measured may require longer and/or more intensive instruction to see significant improvement from the L2-only group.

The presence of other factors that predict L1 growth, namely use of L2, baseline conceptual vocabulary and their impact immediately post intervention and at follow up, highlight the complexity of L1 language growth in dual language children with DLD.

The RCT studies showed some significant improvements in comprehension and use of the first language across various measures following intervention. These were observed for language interventions in both bilingual and L2-only programmes. Gains in L1 were not statistically attributable to participation in bilingual intervention.

Effect on second language

There is inconsistency regarding the impact on L2 outcomes, and it is not possible to assign an overall improvement in L2 to any one of the interventions.

Ebert et al. (2014) found that the bilingual group performed equivalently in L2 outcomes (EOWPVT-E and CELF-4E) as the L2-only group, but the difference was not statistically significant. This was sustained at 3 months follow up for both groups; but again was not significant between groups (Pham, Dunahy and Kohnert, 2014). Similar findings were attained by Restrepo, Morgan and Thompson (2013). Gutierrez-Clellen, Simon-Cerejido and Sweet (2012) found that the bilingual group performed better on

MLU_w than the L2-only group. The absence of a significant difference between the bilingual and L2-only interventions indicates that bilingual intervention does not hinder progress in L2. Gutierrez-Clellen, Simon-Cerejido and Sweet (2012) found participants with higher baseline scores for L1 and L2 experienced greater growth in L2. Although greater baseline scores could reflect less severe language impairment, such outcomes have also been found in typically developing children (Rolstand, Mahoney and Glass, 2005).

L2-only vs bilingual intervention

In both the L2-only and bilingual interventions in the six RCT studies, children received an equal number of total hours instruction, with half the content administered in L2 and repeated in L1 for the bilingual condition. The relative outcomes can be compared between conditions. Some authors reported non-statistically significant results as if they had met significance using terms such as ‘marginal predictor’ (Simon-Cerejido et al. 2013), potentially enhancing the effect of interventions. All authors set the significance level to be $p < 0.05$, where this was not met, findings are not reported here.

There is some evidence to suggest that bilingual intervention improves some L2 outcomes (Gutierrez-Clellen, Simon-Cerejido and Sweet, 2012) but none to suggest that L2-only intervention yields better L2 outcomes than bilingual interventions. For L1 outcomes, Simon-Cerejido, Gutierrez-Clellen and Sweet (2013) reported that the bilingual condition promoted greater growth than L2-only, but no differences reached statistical significance. The authors stated that within-study factors were the likely causation for this, such as insufficient dosage or intensity of L1 exposure. No other study found significant differences between a bilingual and L2-only intervention for measured outcomes of L2. Simon-Cerejido and Gutierrez-Clellen (2014) found that the

bilingual programme had greater effects on improving L2 outcomes, but this was in comparison to a control group who were not receiving intervention. Thordardottir et al. (2015) found no difference between groups.

The implication is that no differences exist between L2-only and bilingual programmes for L2 outcomes, and that L2 outcomes of bilingual interventions are no worse than the L2-only condition, despite half the instruction. If, as Simon-Cereijido et al. (2013) hypothesise, future studies are able to detect an additional benefit of preserving L1, it may be an advantageous approach for dual language learners with DLD. Gutierrez-Clellen, Simon-Cereijido and Sweet (2012) suggest testing an L1-only intervention for a more valid comparison with regard to dosage.

Cross-linguistic transfer

There was limited evidence supporting the notion of cross-linguistic transfer from L2 to L1. Despite only half the instruction, improvements in L2 outcomes in bilingual groups were at least equivalent to those in L2-only groups which might indicate L1 to L2 transfer (Ebert et al. 2014; Pham et al. 2014). Specifically, Gutierrez-Clellen, Simon-Cereijido and Sweet (2012) found MLU_w scores for L1 predicted growth in L2 MLU_w for children receiving bilingual intervention, indicating possible syntactic transfer. Kohnert and Derr (2004) also found that proficiency in L1 correlated with acquisition of L2 words with similar phonological features in typically developing children.

The evidence supports the notion of cross-linguistic transfer from L1 to L2. This aligns with the language-in-contact proposal (Döpke, 2000), which posits that languages with common syntactic features such as Spanish and English facilitate cross-linguistic transfer (Castilla, Restrepo and Perez-Leroux, 2009; Paradis, 2010). Alternatively, equivalent L2 outcomes between interventions could indicate saturation

in the rate of L2 acquisition, i.e. “double the instruction” yields no benefit. This is important because it raises questions about dose of L2 instruction. Dose is an important issue to consider in terms of the intensity requirements for input of L2 intervention compared to L1, as L2 is reinforced in the same classroom environment. Moreover, half the dosage of L2 intervention in addition to L1 intervention had the greatest effect on L2 growth across studies, therefore it is possible that there may be some interaction or additive effect of the bilingual intervention that promotes L2 growth.

All studies showed minimal or no L1 gains in L2-only conditions. This indicates that instruction in both languages, as opposed to L2-only, is necessary for significant growth in L1. In a review of evidence, Thordardottir (2010) also concluded that incorporating L1 to intervention aids acquisition of L2, and that a focus on both languages has potential benefits.

Further work

No studies compared an intervention in which L1 was the only language of instruction. Ebert et al. (2014) suggested that the addition of an L1-only group in receipt of an equal number of instruction hours to the L2-only group would provide a more valid comparison. Previous evidence suggests that an L1-only intervention promotes growth in L1 for typically developing dual language learners, but the effects on L2 were not measured (Restrepo et al., 2010). Since the evidence suggests that L2 outcomes are similar with half the instruction, dosage of instruction should be assessed further.

Understanding the capacity for acquisition over time is important in ensuring maximum effectiveness. With bilingualism on the rise, there is a clear need for more high-quality studies of intervention with dual language children.

Limitations of the study

The application of bilingual interventions to children with DLD is a relatively under-researched area. The search procedure generated a limited number of articles for review, despite placing no restrictions on quality or study design. Although the data did facilitate conclusions, the validity is constrained by the overall lack of studies. It is possible that some valid results may be unduly omitted, consequential to the current state of evidence available.

There was little overlap in outcome measures across studies, and the variability precluded meaningful statistical comparison. This eliminated the possibility of conducting a purely numerical meta-analysis.

Conclusions

This systematic review has highlighted two key findings with respect to the effectiveness of bilingual and L2-only interventions, and where future research priorities should be focused. The search identified some high-quality studies with low bias designs, good sample sizes and lengthy follow up periods.

To improve L1 outcomes, direct instruction in the first language is required. Given that rates of improvement of L1 outcomes do not persist following intervention, sustained L1 language instruction may be necessary. Targeting difficulties with bilingual interventions provides an opportunity to practise, use, and develop the first language, which is important in delivering holistic outcomes for the child.

There is evidence to suggest that dual language learners with DLD receiving bilingual intervention have similar L2 outcomes than those in the L2-only group (Ebert et al., 2014), persisting up to 5 months. Gutierrez-Clellen, Simon-Cerejido and Sweet (2012) indicate that children receiving bilingual intervention scored higher in some

outcome measures. Therefore, it can be assumed that children receiving bilingual intervention do just as well, if not better in L2, than children receiving L2-only interventions.

The studies reviewed demonstrate the complexity of language acquisition in children with DLD. The predictive factors to improvements in L1 and L2 suggest that children will respond variably to intervention based on the presence of predisposed factors. In particular, the child's prior language skills should be considered when choosing a language of intervention.

References

- ADESOPE, O. O., LAVIN, T., THOMPSON, T. & UNGERLEIDER, C. (2010). A Systematic Review and Meta-Analysis of the Cognitive Correlates of Bilingualism. *Review of Educational Research*, 80(2), 207-245.
- BARAC, R., BIALYSTOK, E., CASTRO, D. C., & SANCHEZ, M. (2014). The Cognitive Development of Young Dual Language Learners: A Critical Review. *Early Childhood Research Quarterly*, 29(4), 699-714.
- BAXTER, S., JOHNSON, M., BLANK, L., CANTRELL, A., BRUMFITT, S., ENDERBY, P. & GOYDER, E. (2015). The state of the Art in Non-Pharmacological Interventions for Developmental Stuttering. Part 1: A Systematic Review of Effectiveness. *International Journal of Language and Communication Disorders*, 50(5), pp.676-718.
- BEDORE, L. M. & PENA, E. D. 2008. Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism*, 11, 1-29.
- BISHOP, D. V. & LEONARD, L. (Eds.). (2014). Speech and Language Impairments in Children: Causes, Characteristics, Intervention and Outcome. *Psychology Press*.

- BISHOP, D.V., SNOWLING, M.J., THOMPSON, P.A. & GREENHALGH, T. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLoS One*, 11(7), p. e0158753.
- CALDAS, S.J. (2006). Raising Bilingual-Biliterate Children in Monolingual Cultures. *Clevedon, UK: Multilingual Matters*.
- CASTILLA, A.P., RESTREPO, M.A. & PEREZ-LEROUX, A.T. (2009). Individual Differences and Language Interdependence: A Study of Sequential Bilingual Development in Spanish-English Preschool Children. *International Journal of Bilingual Education and Bilingualism*, 12, pp.1-16.
- COCHRANE COLLABORATION (2011). Cochrane Handbook for Systematic Reviews of Interventions, *Cochrane Collaboration, Version 5.1.0 March 2011*. Retrieved 18 February 2016, from <http://handbook.cochrane.org/>
- CRUTCHLEY, A., BOTTING, N. & CONTI-RAMSDEN, G. (1997), Bilingualism and Specific Language Impairment in Children Attending Language Units. *European Journal of Disorders of Communication*, 32, 267-276.
- CRUTCHLEY, A. (1999). Bilingual Children with SLI Attending Language Units: Getting the Bigger Picture. *Child Language Teaching and Therapy*, 15, pp.201 – 217.
- CUMMINS, J. (1991). Interdependence of First- And Second Language Proficiency in Bilingual Children. In E. Bialystok (ed.), *Language Processing in Bilingual Children* (Cambridge: Cambridge University Press), pp. 70–89
- DE HOUWER, A. (1995). Bilingual Language Acquisition. In: P. Fletcher and B. MacWhinney, ed., *The Handbook of Child Language*, 1st ed. Oxford, UK: Blackwell Publishing Ltd, pp.219-240.
- DÖPKE, S. (2000). Generation of and Retraction from Cross-Linguistically Motivated Structures in Bilingual First Language Acquisition. *Bilingualism: Language and Cognition*, 3(3), 209-226.
- EBERT, K.D., RENTMEESTER-DISHER, J. & KOHNERT, K. (2012). Nonlinguistic Cognitive Treatment for Bilingual Children with Primary Language Impairment. *Clinical Linguistics & Phonetics*, 26(6), pp.485-501.

- EBERT, K.D., KOHNERT, K., PHAM, G., DISHER, J.R. & PAYESTEH, B. (2014). Three Treatments for Bilingual Children with Primary Language Impairment: Examining Cross-Linguistic and Cross-Domain Effects. *Journal of Speech, Language, and Hearing Research*, 57(1), pp.172-186.
- GATHERCOLE, V. C., MON THOMAS, E. & EVANS, D. 2000. What's in a noun? Welsh-, English-, and Spanish-speaking children see it differently. *First Language*, 20, 055-90.
- GUTIÉRREZ-CLELLEN, V., & RESTREPO, M. A. (2005–2009). *Vocabulary, Oral Language and Academic Readiness (VOLAR): A Language Intervention for Latino Preschool English Language Learners with Language Disorders*.
- GUTIÉRREZ-CLELLEN, V., SIMON-CEREIJIDO, G. & SWEET, M. (2012). Predictors of Second Language Acquisition in Latino Children with Specific Language Impairment. *American Journal of Speech-Language Pathology*, 21(1), pp.64-77.
- HAKANSSON, G., SALAMEH, E.K., & NETTEKBKADT, U. (2003). Measuring Language Development in Bilingual Children: Swedish-Arabic Children With and Without Language Impairment. *Linguistics*, 41, pp.255-288.
- HIGGINS, J., ALTMAN, D. and STERNE, J. (2018). 8 Assessing Risk of Bias in Included Studies. [online] Handbook-5-1.cochrane.org. Available at: http://handbook-5-1.cochrane.org/chapter_8/8_assessing_risk_of_bias_in_included_studies.htm [Accessed 20 Feb. 2018].
- JIA, G. & AARONSON, D.V.M. (2003). A Longitudinal Study of Chinese Children and Adolescents Learning English in the United States. *Applied Psycholinguistics*, 24, pp.131-161.
- KAN, P.F. & KOHNERT, K. (2005). Preschoolers Learning Among and English Lexical-Semantic Skills in L1 and L2. *Journal of Speech, Language, and Hearing Research*, 48(2), pp.372-383.
- KOHNERT, K. (2013). *Language Disorders in Bilingual Children and Adults* (2nd ed.). San Diego: Plural Publishing.

- KOHNERT, K. (2010). Bilingual Children with Primary Language Impairment: Issues, Evidence and Implications for Clinical Actions. *Journal of Communication Disorders*, 43(6), 456-473.
- KOHNERT, K. & DERR, A. (2004). *Language Intervention with Bilingual Children*. In: Goldstein, B., Editor. *Bilingual Language Development and Disorders in Spanish-English Speakers*. Baltimore: Brookes. pp.315-343.
- KOHNERT, K., KAN, P. F., & CONBOY, B. T. (2010). Lexical and Grammatical Associations in Sequential Bilingual Preschoolers. *Journal of Speech, Language, and Hearing Research*, 53(3), 684-698.
- KOHNERT, K., WINDSOR, J., & EBERT, K. (2009). Primary or “specific” language impairment and children learning a second language. *Brain and Language*, 109(2-3), 101-111.
- LASAGABASTER, D. (2001). The Effect of Knowledge about the L1 on Foreign Language Skills and Grammar. *International Journal of Bilingual Education and Bilingualism*, 4(5), 310-331
- LEONARD, L. B. (2014). Children with Specific Language Impairment and Their Contribution to the Study of Language Development. *Journal of Child Language*, 40(Supp 1), pp. 38-47.
- MAITAL, S. L., DROMI, E., SAGI, A. & BORNSTEIN, M. H. 2000. The Hebrew Communicative Development Inventory: Language specific properties and cross-linguistic generalizations. *Journal of Child Language*, 27, 43-67.
- OFFICE FOR NATIONAL STATISTICS (2011). DC2104EW - Main Language by Sex by Age. Retrieved 27 March 2016, from http://www.nomisweb.co.uk/census/2011/DC2104EW/view/2092957699?rows=c_age&cols=mainlangprf11
- PARADIS, J., CRAGO, M., GENESEE F. & RICE, M. (2003). French-English Bilingual Children with SLI: How do they compare with their Monolingual Peers? *Journal Of Speech, Language, and Hearing Research*, 46, 113-127.
- PARADIS, J. (2010). The Interface between Bilingual Development and Specific Language Impairment. *Applied Psycholinguistics*. 31, pp.227-252.

- PHAM, G., DUNAHY, K. & KOHNERT, K. (2014). Bilingual Children with Primary Language Impairment: 3 Months after Treatment. *International Journal of Language and Communication Disorders*, 50(1), pp.94-105.
- RESTREPO, M.A., MORGAN, G.P. & THOMPSON, M.S. (2013). The Efficacy of a Vocabulary Intervention for Dual-Language Learners with Language Impairment. *Journal of Speech, Language, and Hearing Research*, 56(2), pp.748-765.
- RESTREPO, M.A., CASTILLA, A.P., SCHWANENFLUGEL, P.J., NEUHARTH-PRITCHETT, S., HAMILTON, C.E. & ARBOLEDA, A. (2010). Effects of a Supplemental Spanish Oral Language Program on Sentence Length, Complexity, and Grammaticality in Spanish-Speaking Children Attending English-Only Preschools. *Language, Speech, and Hearing Services in Schools*, 41(1), pp.3-13.
- ROLSTAND, K., MAHONEY, K. & GLASS, G.V. (2005). The Big Picture: A Meta-Analysis of Program Effectiveness Research on English Language Learners. *Educational Policy*, 19, pp.572-594
- SALAMEH, E. K., HAKANSSON, G., & NETTEKBAKADT, U. (2004). Developmental Perspectives on Bilingual Swedish-Arabic Children with and without Language Impairment: A Longitudinal Study. *International Journal of Language and Communication Disorders*, 39, 65-91.
- SCHWARTZ, R.G. (Ed.) (2017). *Handbook of Child Language Disorders* (2nd ed.) New York: Routledge.
- SIMON-CEREJIDO, G. & GUTIEREZ-CLELLEN V.F. (2014) Bilingual Education for All: Latino Dual Language Learners with Language Disabilities. *International Journal of Bilingual Education and Bilingualism*. 17 pp. 235-254.
- SIMON-CEREJIDO, G., GUTIÉRREZ-CLELLEN, V.F. & SWEET, M. (2013). Predictors of Growth or Attrition of the First Language in Latino Children with Specific Language Impairment. *Applied Psycholinguistics*, 34(06), pp.1219-1243.

SPOELMAN, M. & BOL, G. W. 2012. The use of subject–verb agreement and verb argument structure in monolingual and bilingual children with specific language impairment. *Clinical Linguistics & Phonetics*, 26, 357-379.

THORDARDOTTIR, E. T. (2010). Towards Evidence-Based Practice in Language Intervention for Bilingual Children. *Journal of Communication Disorders*, 43(6), 523-537.

THORDARDOTTIR, E. T., CLOUTIER, G., MÉNARD, S., PELLAND-BLAIS, E. & RVACHEWA, S. (2015). Monolingual or Bilingual Intervention for Primary Language Impairment? A Randomized Control Trial. *Journal of Speech, Language, and Hearing Research*, 58, pp. 287–300.

THORDARDOTTIR, E.T., WEISMER, S.E. & SMITH, M.E. (1997). Vocabulary Learning in Bilingual and Monolingual Clinical Intervention. *Child Language Teaching and Therapy*, 13(3), pp.215-227.

UNITED STATES CENSUS BUREAU (2018). *American FactFinder*. [online] Factfinder.census.gov. Available at: <https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=CF> [Accessed 15 Feb. 2018].

VERHOEVEN, L., STEENGE, J., & van BALKHOM, H. (2012). Linguistic Transfer in Bilingual Children with Specific Language Impairment. *International Journal of Language and Communication Disorders*, 47(2), 176-183.