



Cochrane
Library

Cochrane Database of Systematic Reviews

Vocabulary interventions for second language (L2) learners up to six years of age (Review)

Hjetland HN, Hofslundsengen H, Klem M, Karlsen J, Hagen ÅM, Engevik LI, Geva E, Norbury C, Monsrud MB, Næss KAB

Hjetland HN, Hofslundsengen H, Klem M, Karlsen J, Hagen ÅM, Engevik LI, Geva E, Norbury C, Monsrud M-B, Næss K-AB.
Vocabulary interventions for second language (L2) learners up to six years of age.
Cochrane Database of Systematic Reviews 2023, Issue 8. Art. No.: CD014890.
DOI: [10.1002/14651858.CD014890.pub2](https://doi.org/10.1002/14651858.CD014890.pub2).

www.cochranelibrary.com

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	4
BACKGROUND	7
OBJECTIVES	11
METHODS	11
Figure 1.	14
RESULTS	17
DISCUSSION	27
AUTHORS' CONCLUSIONS	30
ACKNOWLEDGEMENTS	31
REFERENCES	32
CHARACTERISTICS OF STUDIES	41
RISK OF BIAS	57
DATA AND ANALYSES	59
Analysis 1.1. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 1: Receptive L2 proximal vocabulary	61
Analysis 1.2. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 2: Receptive L2 proximal vocabulary at follow-up	61
Analysis 1.3. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 3: Receptive L2 distal vocabulary	61
Analysis 1.4. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 4: Receptive L2 distal vocabulary at follow-up	62
Analysis 1.5. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 5: Expressive L2 proximal vocabulary	62
Analysis 1.6. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 6: Expressive L2 proximal vocabulary at follow-up	62
Analysis 1.7. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 7: Expressive L2 distal vocabulary	63
Analysis 1.8. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 8: Expressive L2 distal vocabulary at follow-up	63
Analysis 1.9. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 9: L2 listening comprehension	63
Analysis 1.10. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 10: L2 narrative skills	64
Analysis 1.11. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 11: L2 narrative skills at follow-up	64
Analysis 1.12. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 12: L1 receptive proximal vocabulary	64
Analysis 1.13. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 13: L1 receptive distal vocabulary	64
Analysis 1.14. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 14: L1 expressive proximal vocabulary	65
Analysis 1.15. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 15: L1 expressive proximal vocabulary at follow-up	65
Analysis 1.16. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 16: L1 expressive distal vocabulary	65
Analysis 1.17. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 17: L2 grammatical knowledge	65
Analysis 1.18. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 18: L2 grammatical knowledge at follow-up	66
APPENDICES	66
HISTORY	77
CONTRIBUTIONS OF AUTHORS	77

DECLARATIONS OF INTEREST	78
SOURCES OF SUPPORT	78
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	79
INDEX TERMS	79

[Intervention Review]

Vocabulary interventions for second language (L2) learners up to six years of age

Hanne Næss Hjetland^{1,2}, Hilde Hofslundsengen³, Marianne Klem⁴, Jannicke Karlsen⁴, Åste M Hagen¹, Liv Inger Engevik⁵, Esther Geva⁶, Courtenay Norbury^{1,7}, May-Britt Monsrud⁴, Kari-Anne B Næss¹

¹Department of Special Needs Education, University of Oslo, Oslo, Norway. ²Department of Early Childhood Education, Oslo Metropolitan University, Oslo, Norway. ³Department of Language, Literature, Mathematics and Interpreting, Western Norway University of Applied Sciences, Sogndal, Norway. ⁴Department of Speech and Language Disorders, Statped, Oslo, Norway. ⁵Department of Pedagogy, Religion and Social Studies, Western Norway University of Applied Sciences, Bergen, Norway. ⁶Department of Applied Psychology and Human Development, University of Toronto, Toronto, Canada. ⁷Division of Psychology & Language Sciences, University College London, London, UK

Contact: Hanne Næss Hjetland, hanne.hjetland@oslomet.no.

Editorial group: Cochrane Developmental, Psychosocial and Learning Problems Group.

Publication status and date: New, published in Issue 8, 2023.

Citation: Hjetland HN, Hofslundsengen H, Klem M, Karlsen J, Hagen ÅM, Engevik LI, Geva E, Norbury C, Monsrud M-B, Næss K-AB. Vocabulary interventions for second language (L2) learners up to six years of age. *Cochrane Database of Systematic Reviews* 2023, Issue 8. Art. No.: CD014890. DOI: [10.1002/14651858.CD014890.pub2](https://doi.org/10.1002/14651858.CD014890.pub2).

Copyright © 2023 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Second language (L2) learners are a heterogeneous group. Their L2 skills are highly varied due to internal factors (e.g. cognitive development) and external factors (e.g. cultural and linguistic contexts). As a group, their L2 vocabulary skills appear to be lower than their monolingual peers. This pattern tends to persist over time and may have negative consequences for social interaction and inclusion, learning, and academic achievement.

Objectives

To examine the immediate and long-term effects of second language (L2) vocabulary interventions targeting L2 learners up to six years of age on vocabulary and social-emotional well-being.

To examine the associations between L2 vocabulary interventions and the general characteristics of L2 learners (e.g. age, L2 exposure, and L1 skills).

Search methods

We used standard, extensive Cochrane search methods. The latest search date was December 2022.

Selection criteria

We included randomised controlled trials (RCTs) comparing the effects of vocabulary interventions for L2 learners up to six years of age with standard care.

Data collection and analysis

We used standard methodological procedures expected by Cochrane. Our primary outcomes were 1. receptive and 2. expressive L2 vocabulary (both proximal and distal), and 3. mean length of utterance (MLU; which is a measure of potential adverse effects). Our secondary outcomes were 4. L2 narrative skills, 5. L1 receptive vocabulary (both proximal and distal), 6. L1 expressive vocabulary (both

proximal and distal), 7. L1 listening comprehension, 8. L2 grammatical knowledge, 9. L2 reading comprehension (long-term), and 10. socio-emotional well-being (measured with Strengths and Difficulties Questionnaire).

Main results

We found 12 studies involving 1943 participants. Two studies were conducted in Norway, seven in the USA, and single studies conducted in Canada, China, and the Netherlands.

Ten studies were conducted in preschool settings, with a preschool teacher being the most common delivery agent for the intervention. The interventions were mainly organised as small-group sessions, with three or four children per group. The mean dosage per week was 80 minutes and ranged from 24 to 120 minutes.

The studies commonly applied shared book reading (reading aloud with the children), with target words embedded in the books.

Standard care differed based on the setting and local conditions in each country or (pre)school. In some studies, the comparison groups received vocabulary instruction in preschool groups.

Compared to standard care, the effect of L2 vocabulary interventions varied across outcome measures. For vocabulary measures including words that were taught in the intervention (proximal outcome measures), the intervention effects were large for both receptive L2 vocabulary (i.e. understanding of words; standardised mean difference (SMD) 0.97, 95% confidence interval (CI) 0.64 to 1.30; 4 studies, 1973 participants; very low-certainty evidence) and expressive L2 vocabulary (i.e. expressing or producing words; SMD 0.86, 95% CI 0.56 to 1.17; 6 studies, 1121 participants; very low-certainty evidence). However, due to some concerns in the overall risk of bias assessment, substantial heterogeneity, and wide CIs, we have limited confidence in these results.

For language measures that did not include taught vocabulary (distal outcome measures), the intervention effects were small for receptive vocabulary (SMD 0.29, 95% CI 0.02 to 0.55; 6 studies, 1074 participants; low-certainty evidence) and probably made little to no difference to expressive vocabulary (SMD 0.10, 95% CI -0.02 to 0.23; 7 studies, 960 participants; moderate-certainty evidence). There was little to no intervention effect on L2 listening comprehension (SMD 0.19, 95% CI -0.31 to 0.68; 2 studies, 294 participants; very low-certainty evidence), but the evidence was uncertain, and the interventions probably increased L2 narrative skills slightly (SMD 0.37, 95% CI 0.14 to 0.59; 2 studies, 487 participants; moderate-certainty evidence). Only one study reported data on MLU, and we were unable to examine the effect of intervention on this outcome. The level of certainty of the evidence was downgraded mainly due to inconsistency and imprecision.

We were unable to draw conclusions about socio-emotional well-being, or conduct the planned subgroup analyses to examine the second objective, due to lack of data.

Authors' conclusions

Findings from this review suggest that, compared to standard care, vocabulary interventions may benefit children's L2 vocabulary learning but have little to no effect on their listening comprehension, though the evidence is uncertain. Vocabulary interventions probably improve the children's storytelling skills slightly.

Due to the limited number of studies that met our inclusion criteria and the very low- to moderate-certainty evidence as a result of inconsistency and imprecision, implications for practice should be considered with caution.

This review highlights the need for more high-quality trials (e.g. RCTs) of vocabulary interventions for L2 learners, particularly studies of learners outside the USA.

PLAIN LANGUAGE SUMMARY

Vocabulary interventions for second language learners up to six years of age

Key messages

– Studies included in this review suggest that vocabulary interventions for L2 learners may benefit children's L2 vocabulary learning but have little to no effect on their listening comprehension skills, though the evidence is very uncertain. Vocabulary interventions probably improve their storytelling skills.

– There is a need for more high-quality research that follows L2 learners over a longer time to investigate longer-term effects. This should include learners outside the USA.

Why is this review important?

Limited second language (L2) ability may have a negative impact on academic achievement, because vocabulary knowledge in the language of instruction is central to reading comprehension, classroom learning, and inclusion. In adulthood, proficiency in the community language (as well as continued competence in the first language (L1)) is predictive of employment, good relationships, and societal participation.

What was the aim of this review?

The primary objective was to examine the immediate and long-term effects of vocabulary interventions targeting L2 learners up to six years of age on vocabulary and social-emotional well-being (ability to understand and manage emotions, make responsible decisions, build and maintain relationships, and understand and empathise with others). The secondary objective was to examine the associations between L2 vocabulary interventions and the general characteristics of L2 learners (e.g. age, L2 exposure and L1 skills).

What did the review study?

We included randomised controlled trials (RCTs). An RCT is a study in which participants are assigned randomly (for example, by a coin toss) to two or more treatment groups. This is the best way to ensure that groups of participants are similar, and that investigators and participants do not know who is in which group. The comparison group consisted of children receiving standard care. There were no restrictions on either the type of L1 or L2 or the geographical location of the participants.

What were the main results of the review?

This review was based on 12 studies and 1943 participants. Seven studies were conducted in the USA, and the remaining studies were conducted in Canada, China, the Netherlands, and Norway. There were variations in the languages spoken by the L2 learners.

The most common intervention used to teach children L2 vocabulary was shared book reading with a predefined list of target words. Additional language activities included using these words in different contexts, in sentences, and mapping the sound structure of the new words. Occasionally interventions included the same vocabulary items in the child's L1.

Studies compared the intervention with standard care. What standard care entailed differed, depending on the setting and local conditions in each country or (pre)school.

The review suggests that, relative to standard care, it is not clear whether giving L2 learners a vocabulary intervention has any effect on their use of L2 words taught in the intervention, due to our limited confidence in the result. Vocabulary intervention may slightly increase children's understanding of L2 words that were not taught in the intervention, but probably makes little to no difference to their ability to express L2 words that were not taught in the intervention.

It appears to have little to no effect on their L2 listening comprehension, but may increase their storytelling skills slightly.

Our confidence in the evidence is very low to moderate, mainly because there are not enough studies to be certain about the results of our outcomes.

We were unable to examine potential adverse (unwanted) effects because the studies did not report these. We were also unable to draw conclusions about social-emotional well-being, or our second objective, due to a lack of data.

How up-to-date is this review?

The evidence is current until December 2022.

SUMMARY OF FINDINGS

Summary of findings 1. Summary of findings table - Vocabulary interventions compared with standard care for second language (L2) learners up to six years of age

Vocabulary interventions compared with standard care for second language (L2) learners up to six years of age

Patient or population: second language (L2) learners up to six years of age

Setting: preschool. The interventions most commonly taught children's L2 vocabulary through shared book reading activities and included a predefined list of target words. Additional language activities included using these words different contexts, in sentences, and mapping the sound structure of the new words.

Intervention: vocabulary interventions

Comparison: standard care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	Nº of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with standard care	Risk with vocabulary interventions				
Receptive L2 proximal vocabulary (understanding of words taught in the intervention) assessed with: the children were asked to point to a picture in a set of 4 that corresponds to the word uttered by the assessor. The score is number correct words. Timing of outcome measure: first post-test after intervention.	-	SMD 0.97 SD higher (0.64 higher to 1.3 higher)	-	1073 (4 RCTs)	⊕⊕⊕⊕ Very low ^a	A standard deviation (SD) of 0.97 represents a large difference between groups (Cohen 1988).
Receptive L2 distal vocabulary (understanding of words not included in the intervention) assessed with: the children were asked to point to a picture in a set of 4 that corresponds to the word uttered by the assessor. The score is number of correct words. Timing of outcome measure: first post-test after intervention.	-	SMD 0.29 SD higher (0.02 higher to 0.55 higher)	-	1074 (6 RCTs)	⊕⊕⊕⊕ Low ^b	An SD of 0.29 represents a small difference between groups (Cohen 1988).
Expressive L2 proximal vocabulary (expressing or producing words taught in the intervention) assessed with: the children are either asked to provide a definition or a label. The score is the number correct definitions or labels. Timing of outcome measure: first post-test after intervention.	-	SMD 0.86 SD higher (0.56 higher to 1.17 higher)	-	1121 (6 RCTs)	⊕⊕⊕⊕ Very low ^c	An SD of 0.86 represents a large difference between groups (Cohen 1988).
Expressive L2 distal vocabulary (expressing or producing words not taught in the intervention)	-	SMD 0.1 SD higher	-	960 (7 RCTs)	⊕⊕⊕⊕	An SD of 0.1 represents a small dif-

assessed with: the children are either asked to provide a definition or a label. The score is the number correct definitions or labels. Timing of outcome measure: first post-test after intervention.		(0.02 lower to 0.23 higher)			Moderated ^d	ference between groups (Cohen 1988).	
Mean length of utterance (MLU)			(1 RCT)	⊕⊕⊕⊕	Moderate ^e	Not reported.	
L2 listening comprehension assessed with: measures not including taught words. The assessor read short stories, and the child answered a series of questions tapping both literal and inferential understanding of the story. Timing of outcome measure: first post-test after intervention.	-	SMD 0.19 SD higher (0.31 lower to 0.68 higher)	-	294 (2 RCTs)	⊕⊕⊕⊕	Very low ^f	An SD of 0.19 represents a small difference between groups (Cohen 1988).
L2 narrative skills assessed with: retelling of a story not including taught words. Points were awarded if central story elements were included. Timing of outcome measure: first post-test after intervention.	-	SMD 0.37 SD higher (0.14 higher to 0.59 higher)	-	487 (2 RCTs)	⊕⊕⊕⊕	Moderate ^g	An SD of 0.37 represents a small-to-moderate difference between groups.

***The risk in the intervention group** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).

CI: confidence interval; **MD:** mean difference; **SMD:** standardised mean difference

GRADE Working Group grades of evidence

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate: the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited: the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate: the true effect is likely to be substantially different from the estimate of effect.

See interactive version of this table: https://gdt.gradeopro.org/presentations/#/isof/isof_question_revman_web_431591784086563843.

^a Downgraded one level due to 'some concerns' overall in risk of bias assessment, one level due to inconsistency (substantial heterogeneity; I² = 83%), and one level due to imprecision (wide confidence intervals).

^b Downgraded one level due to 'some concerns' overall in risk of bias assessment and one level due to inconsistency (substantial heterogeneity; I² = 74%).

^c Downgraded one level due to 'some concerns' overall in risk of bias assessment, one level due to inconsistency (substantial heterogeneity; I² = 78%), and one level due to imprecision (very wide confidence intervals).

^d Downgraded one level due to 'some concerns' overall in risk of bias assessment.

e Downgraded one level due to 'some concerns' overall in risk of bias assessment.

f Downgraded one level due to 'some concerns' overall in risk of bias assessment, one level due to inconsistency (substantial heterogeneity; $I^2 = 73\%$), and one level due to imprecision (wide confidence intervals).

g Downgraded one level due to 'some concerns' overall in risk of bias assessment.

BACKGROUND

Description of the condition

The vocabulary knowledge of second language (L2) learners includes words in both the first language (L1) and the L2 (Hwang 2020; Monsrud 2022). In this review, we used the term L2 learners to include children with an immigrant background, multilingual children, as well as refugees. Children who speak a language at home that is different from the societal language used in day care, preschool, and school usually have to rely on only part of their vocabulary knowledge – the L2 vocabulary – in these contexts. They are at risk of failing to achieve the same level of vocabulary skills in the societal language as their monolingual peers (August 2008; Farnia 2011; Hoff 2014; Melby-Lervåg 2014; Simos 2014).

Vocabulary skills in this context refer both to the breadth (how many words are known) and depth (how well the words are known – for example, connotations, semantic associates, and morphological options). On group level, L2 learners' breadth and depth of vocabulary knowledge in the societal language are both significantly lower than the vocabulary knowledge of their monolingual peers (Farnia 2011; Jean 2009; Lin 2012; Proctor 2012). In this review important distinctions are also drawn between receptive and expressive vocabulary and between proximal and distal vocabulary. Receptive vocabulary refers to understanding words. Expressive vocabulary refers to expressing and producing words. Proximal vocabulary refers to the words that are directly taught in an intervention. Distal vocabulary includes words that are not directly trained in an intervention. Moreover, it appears that academic vocabulary (i.e. words that are low frequency, more abstract, and potentially more ambiguous), is particularly challenging to acquire for L2 learners (Biemiller 2005; Jean 2009; Lin 2012), though essential for reading comprehension and predictive of education outcomes.

Limited vocabulary may have negative, long-term effects on later language development, reading comprehension, and academic achievement for both monolingual and L2 learners (August 2005; Snow 1995; Stanovich 1986), and may increase risk of dropping out of school (Lervåg 2018). Poor vocabulary has also been associated with negative impacts on mental health (Snow 2016; Toppelberg 2002), and reduced occupational opportunities in adulthood (Johnson 2010). Studies of monolingual children also indicate that children with poor vocabulary are more likely to be rejected by their peers and are less likely to initiate interactions, participate in social interactions, and play (Rice 1991; Stangeland 2017).

At group level, L2 vocabulary appears to be smaller for L2 learners relative to monolingual peers. Melby-Lervåg 2014 found large differences (Cohen's $d = 1.12$ for pooled effect sizes) in oral language (including vocabulary) when comparing language skills between L1 and L2 learners. The poor L2 vocabulary knowledge of L2 learners relative to their monolingual peers is worrisome because the rank-order of children's vocabulary skills is quite stable throughout their early educational years (Lervåg 2010; Storch 2002). In other words, the differences in vocabulary knowledge appear to persist over time (Farnia 2011; Karlsen 2017; Lervåg 2010).

The risk factors associated with poor L2 vocabulary knowledge can be both external and internal to the individual (Paradis 2011). External factors may include the amount and quality of language

exposure; internal factors may refer to the individual differences in child skills such as learning aptitude and phonological short-term memory, where the lower end of the continuum may include children with developmental language disorder (DLD). Exposure to more than one language does not cause or exacerbate DLD, and poor vocabulary in L2 only would not indicate language disorder. Monolingual and multilingual children are probably at the same risk of having DLD, with an estimated prevalence ranging from 3% to 7%, depending on age and definition (Norbury 2016). L2 learners with language disorders can be expected to display persistent language difficulties in any of their languages (Bishop 2017; Farnia 2019; Geva 2015).

It is important to consider that L2 learners are not a homogeneous group (OECD 2019), and identification of external factors that play a role in L2 vocabulary learning need to acknowledge variabilities at the national level that are associated with geography, country of origin, politics, demographics, and immigration factors, as well as variability in the range of policies and opportunities designed to enhance the learning of the societal language by the children of immigrants and refugees. Difficulties academically could also be due to other factors such as trauma; for example, refugees fleeing areas of conflict and cultural dissonance in preschool and school (DeCapua 2005).

Further, the risk factors for poor L2 vocabulary may be related to aspects of low socioeconomic status (SES), parental education, and the amount and quality of exposure to the societal language. In general, compared to their monolingual peers, L2 learners who come from lower SES backgrounds, with more restricted access to high-quality education, are more at risk of having poor vocabulary skills (OECD 2020a). For young children, social interactions with their parents are often one of the main sources of language exposure (Hart 1995; Rowe 2012; Weizman 2001). Research on both monolingual children (Hart 1995), and L2 learners (Hoff 2013) has demonstrated that parents from more privileged SES backgrounds often talk more to their children, use a broader, more abstract, and more precise range of vocabulary, and engage their children in context-independent conversation more often than parents from lower SES backgrounds. This reflects both genetic and environmental factors and their interactions (Hart 2021).

For children who are exposed to different languages at home and at daycare or school, vocabulary learning is often distributed between the two languages (Hwang 2020; Monsrud 2022; Oller 2007). This may mean that a larger percentage of their exposure to new words in L2 occurs mainly at daycare or school, unlike monolingual peers (Bialystok 2010; Paradis 2009; Webb 2020), while at the same time, they are exposed to new vocabulary in L1 at home (Monsrud 2022). Consequently, the time spent in an L2 educational context, such as preschool, may be of particular importance for L2 vocabulary learning. Previous studies have found that both the amount of time spent in preschool (Karlsen 2017), and the teaching quality provided by teachers (Bowers 2011; Rydland 2021), are related to children's L2 vocabulary development. L2 learners may choose to focus on one language exclusively in particular contexts; however, they may also elect to change between both languages in the same setting depending on who they are speaking with or the task they are completing.

Despite challenges associated with L2 acquisition, learning more than one language can have positive effects. Maintaining proficiency in L1 may be important for the child's identity and

cultural affiliation, as well as the family interactions. This in turn has been associated with lower levels of internalising and externalising behaviours by fifth grade (i.e. 10 to 11 years of age) (Han 2010). Furthermore, L2 learners bring languages and cultural resources to preschool and school that may enrich the linguistic environment for all children (Cummins 2000).

Description of the intervention

Systematic vocabulary interventions in preschool settings are effective for enhancing vocabulary development in L2 learners (Leacox 2014; Lugo-Neris 2010; Rogde 2016). One meta-analysis of 43 studies that included both L1 and L2 learners reported a small overall effect size ($g = 0.16$) of broader language interventions on standardised outcomes of linguistic comprehension (Rogde 2019). Although the overall effect size was small, the findings nevertheless suggest that systematic language interventions may increase children's vocabulary skills. As vocabulary has a stable trajectory from a young age (Bornstein 2014; Klem 2015), even a slight early enhancement in vocabulary skills might be beneficial for a particular child. However, specific features of effective vocabulary interventions at different age groups, and the extent to which such interventions generalise to unfamiliar vocabulary or broader language and literacy tasks, has yet to be determined.

Vocabulary interventions for L2 learners between birth and six years of age typically aim to increase experiences with, and exposure to, words and word-meanings in the L2, in order to improve the understanding and use of these targeted words in social communication. In turn, the expectation is that increasing vocabulary should benefit general expressive and receptive language competencies, access to the curriculum, and social engagement (Hagen 2017; Rogde 2016). New vocabulary should also support independent learning of related unfamiliar words, thus sustaining vocabulary growth over time (Rogde 2016).

Interventions can vary in terms of target words, activities, strategies, delivery mode, delivery settings, intervention providers, organisation of the intervention delivery, dosage, and the theories that underpin how the intervention might work.

Approaches

One major difference between the approaches to intervention is the extent to which word meanings are acquired explicitly or implicitly. Explicit instruction requires conscious awareness on the part of the learner about the goals of the intervention, and active participation in learning activities (Baron 2022).

Explicit instruction includes explaining, showing, or testing a hypothesis or phenomenon to achieve conscious awareness about that phenomenon. Such interventions are usually systematically related to at least three elements: 1. a predefined session plan, set of tasks and procedures that gradually increase in complexity and difficulty based on the knowledge of developmental stages; 2. structured activities that target specific words; and 3. the frequency of sessions (Yoder 2014). Implicit (or incidental) learning occurs without conscious awareness, and often employs more naturalistic activities, such as play. Carers may be instructed or coached to engage in interactive behaviours that support vocabulary development; for example, observing the child's communicative attempts, naming and modelling appropriate words, and then extending the child's utterances (Dowdall 2020). Here, the words

to be learned may not be prescribed; instead, the focus is on developing interactions to support the learning of any word.

Explicit and implicit approaches can also be complementary, and successful interventions tend to use a combination of both approaches (Webb 2020).

Target words and activities

Target words

When designing a vocabulary intervention, the starting point is usually selecting target words to be learned throughout the programme. These words may be selected based on:

- the characteristics of the target population, such as age, level of functioning, words that are not known, and words that are meaningful for the children to know;
- aspects of the words, including age of acquisition (Crevecoeur 2014; Vadasy 2015), frequency (Collins 2010; Wood 2018), and phonological complexity (McDaniel 2019; Pearson 2007);
- characteristics of the context, which may entail basic vocabulary for everyday use, such as core or living word vocabulary (tier 1); academic words related to a variety of domains (tier 2); low-frequency subject-specific words (tier 3) (Beck 2013); and words related to cultural values, traditions, or events (Hammer 2016);
- words embedded in existing books or educational materials used in the intervention (Grøver 2020; Restrepo 2013; Rogde 2016);
- methodological aspects, including words that increase the possibility of generalisation and transfer effects to new words not taught in the intervention (words that share morphemes, to increase likelihood of transfer to other words with the same prefixes or suffixes (Torkildsen 2022));
- vocabulary that would not be encountered without direct instruction, thereby yielding long-term intervention effects (Greenwood 2016); and
- the degree to which the words are concrete and thus visually better represented and, therefore, more readily depicted and tested than those that are abstract (Collins 2010; Cycowicz 1997; Leacox 2014; Pollard-Durodola 2016; Restrepo 2013).

The number of target words to be learnt differs between intervention studies; for example, the study by Lugo-Neris 2010 used 20 key words, while Pollard-Durodola 2016 included 94 key words in their intervention.

Content activities/strategies

Internationally, shared picture book reading is considered the most widely used activity for L2 vocabulary intervention in young (preschool) children and can benefit a diverse group of L2 learners (Fitton 2018). Shared book reading interventions often include scaffolding (Rogoff 1990), active listening, and inferencing tasks (Hammer 2016; van Kleeck 1994).

Individual studies have produced contradictory results attributed to design differences; for example, there were large positive effects in studies with non-randomised designs, while studies in which participants were randomly allocated to intervention versus standard care conditions tended to yield null results (Fitton 2018). Design features may be confounded by differences in intervention content, such as the choice of target words, quality

and quantity of book sharing, quality of implementation, or the use of activities or materials originally developed for other purposes and, therefore, only weakly related to the target words (Lawrence 2014). Interventions developed for L1 or monolingual children that were not adapted to a culturally and linguistically diverse population could also contribute to variable outcomes (Larson 2020).

In addition to shared book reading, other common activities in vocabulary interventions are:

- co-construction or retelling of a narrative (Boyce 2010; Hammer 2016; Hargrave 2000);
- perspective taking (Grøver 2020);
- drill and categorisation tasks, including the repetition, sorting, classifying, and defining of target words; sentences requiring the choice of correct words to complete; supporting the use of target words in a broader language context, including activities that target morphology, syntax, and phonology (Hagen 2017; Stahl 1986);
- gaming tasks, which are designed to increase motivation and engagement in order to facilitate learning (Thompson 2020).

Often a range of different activities or strategies is included in an intervention, which is in line with the NICHD 2000 recommendations, which emphasised that a single vocabulary activity or strategy does not result in optimal learning (see also Marulis 2010).

Delivery mode

Vocabulary interventions have traditionally been delivered face-to-face (Rogde 2016), but there is a rapidly increasing interest in, and use of, technology in education (Hassler 2016), with indications that research-based digital educational interventions may be as effective for learning and retention as conventional delivery modes, only (Chauhan 2017; Clark 2016). Relevant vocabulary application (app) strategies include dictionary use or automatic translation (Wood 2018), phonological analysis (De Jong 2000), morphological analysis (Torkildsen 2022), contextual analysis (Nagy 2000), picture book dialogues (Grøver 2020), and narratives (Hur 2012). Apps also provide innovative and personalised opportunities for vocabulary stimulation. Visual and audio exposure, interactive elements, direct feedback, and individual adaptation may lead to both better retention of new words, and improved engagement and learning motivation (Clements 2003; Deng 2015; Hassler 2016; Haugland 1999; Kinash 2012). However, very few of the existing digital interventions target vocabulary, and even fewer have been robustly trialled (Griffith 2020; Hirsh-Pasek 2015). Some apps have been specifically designed and tested to help L2 learners or preschool children to acquire basic academic and cognitive skills (Griffith 2020; Northrop 2019; Shuler 2012), but little is known about vocabulary. There are no systematic reviews or meta-analyses investigating the effects of digital vocabulary interventions on L2 learners, and the question of potential harm from digital tools remains unanswered. This is important considering emerging evidence of harm from excessive consumption of other types of technology, such as television use impacting sleep (McDonald 2014), obesity (Cox 2012), and cognitive development (Zimmerman 2007).

Delivery settings and delivery agents

L2 vocabulary interventions for young children may be conducted in early education and care (nursery/preschool) settings, at home, or in part of healthcare settings. Intervention providers are commonly preschool teachers, teacher assistants, or specialists, such as speech and language pathologists, but they can also be parents or research assistants. The home setting, with parents as providers, may support cross-linguistic connections between L2 and L1 target vocabulary and have a carry-over effect into daily life, thereby improving the maintenance effect. For all providers, pre-intervention training is often necessary before carrying out the intervention programme. Using the children's ordinary preschool teachers or parents implies a more naturalistic intervention than using trained research assistants, although some interventions may include more than one setting. It is unclear if one intervention setting is generally better than two, though studies have found a larger effect of shared book reading interventions if both teachers and parents are involved (Grøver 2020).

Organisation of the intervention delivery

Interventions can be applied one-to-one, in small groups, or in larger groups, which can include the whole classroom. To our knowledge, no review has been conducted on group size for L2 vocabulary interventions specifically, but one review of previous research on linguistic comprehension interventions found that small groups produced larger effects than larger groups or whole classroom (Rogde 2019).

Dosage

The vocabulary intervention dosage varies by the number of sessions, duration, frequency, and length. Optimal dosage for each of these four aspects may be affected by child-related variables, such as age, level of functioning, motivation, concentration, and attention; system-level constraints, such as the available economic and human resources; and intervention-related aspects, such as desired outcomes (Zeng 2012). However, results from different clinical samples suggest that dosage intensity is an important predictor of the intervention effect, indicating that a high frequency is better than a low frequency (Yoder 2014).

Control conditions

Control conditions in vocabulary interventions can include no intervention, a waiting-list control, or standard care. If the control condition includes an active control group receiving an instructional method that targets other aspects of language (e.g. phonological awareness) and that may have a beneficial effect on vocabulary development, its use as a comparison condition is problematic because it is difficult to separate these constructs in early intervention. Notably, comparing a vocabulary intervention with an alternative intervention answers a different research question than one comparing a vocabulary intervention with a group receiving standard care.

According to OECD 2020b, standard care in preschool settings varies greatly between countries in terms of enrolment rates, structures, investment, and governance. Differences in child-staff ratios can affect the types of activities and practices, as well as the quality of interactions between children and staff. Vermeer 2016 investigated the quality and structural features in 23 countries using Environment Rating Scales and found the mean staff-child

ratio to be 8.6, with a range from three to 25 children per member of staff. The Organisation for Economic Co-operation and Development (OECD) survey TALIS Starting Strong reported that the group size of children in preprimary education varies between around 16 children on average in Germany, Iceland, Korea, Norway, and Turkey to more than 20 children in Chile, Israel, and Japan (OECD 2019). The mean number of staff per child working with the same group on the same day reflects large variations from two to three members of staff per 10 children in Israel and Japan to almost nine members of staff per 10 children in Iceland. In addition, OECD 2020b noted children's experience can also vary *within* a country in terms of the preschool setting and the staff working in that setting. In other words, standard care will also vary between contexts. As book reading and talking about word meanings are activities that usually take place in preschools (OECD 2020b), information about standard care is needed to identify the components in both the intervention and control groups that make the intervention different from standard care.

How the intervention might work

Vocabulary interventions for L2 learners are usually broad-based multicomponent programmes consisting of, for example, different oral language components or a combination of oral language and code-related components (Yousefi 2018). Since individual content components have seldom been separated out and used as the basis for randomisation in previous interventions, there is limited knowledge about exactly which component(s) is (are) initiating the change in terms of the breadth, depth, or both. However, the general underlying strategies used in previous interventions can help explain how such vocabulary interventions might work. Explicit or intentional intervention strategies may relate to conscious cognitive processes for understanding and storing new words by committing lexical information to memory (Dixon 2020; Ellis 1994). Implicit intervention strategies may involve an unconscious and gradual accumulation of understanding and remembrance of new words following repeated exposure to the words in different contexts; the learning thus happens implicit (Ellis 1994). When learned implicitly, an increased vocabulary is a 'by-product' of other activities or of different contextual information (e.g. learning words through reading or listening activities). Whether explicit or implicit strategies are most effective for increasing a child's vocabulary remains under discussion (Marulis 2010), and it has also been hypothesised that the strategies are not mutually exclusive – implicit learning can be guided and governed by explicit strategies and explicit learning can be consolidated and reinforced by implicit strategies (Dakun 2000). Therefore, it can be hypothesised that vocabulary develops continuously as a result of both implicit and explicit learning experiences; vocabulary may develop gradually on a continuum from never having heard it before to robust knowledge that has the meaning of the word 'pinned down' and allows it to be used in different contexts and sentences (Bruton 2009; Dale 1965; Stahl 2006).

Assessing the impact of the intervention

Treatment effects for an intervention are usually measured by assessing the participant's vocabulary before the intervention as a baseline measure (pretest) in order to compare them to the results after the intervention (post-test). The effects may be assessed immediately after the intervention or at a specified time after the intervention has ended. In order to determine long-term effects of a

vocabulary intervention, a follow-up assessment may be included; for example, with a reading comprehension measure.

Assessment is typically performed by researchers or trained research assistants (Grøver 2020), but may also be conducted by teachers (Zucker 2019).

Children who are participating in a vocabulary intervention also learn words and develop other language skills naturally outside the intervention programme, and disentangling the direct effect of the intervention and the effects of other contextual factors is challenging. Randomised sampling should ensure the contexts are similar between the two conditions, and an intervention's effect on more distal measures (e.g. not including directly taught words) may also reflect the quality of stimulation outside the intervention. At post-test immediately after an intervention, any treatment effect would be expected to be attributable to the intervention, but following completion of the intervention, the participants would continue to receive instruction independent of the intervention programme. Furthermore, the aim of an intervention is to have a lasting effect, and to achieve this goal, interventionists design interventions such that children learn strategies that they will continue to use after completion. It may be that an intervention has started a learning process that can take time to be expressed in the results, and interventions may also have an impact on the agent of delivery (e.g. parent, preschool teacher) by building their competence in their role in vocabulary learning, which is then positive for the children's development after the intervention. For all these reasons, longer-term assessment of an intervention is important.

Additionally, impacts can be assessed with questionnaires of communication skills or surveys of emotional, social, and behavioural skills and functioning based on parent reports, teacher reports, or both. Impacts can also be assessed using scores on language composite tests comprising several language dimensions (e.g. morphology, syntax, narrative skills, listening comprehension). Finally, effects can be measured with tests of L1 vocabulary or a parent's report of a child's L1.

The impact of an intervention is commonly assessed using pre- to post-test gains in outcome measures. However, when the assignment to the control and intervention groups is randomised, the impact can be assessed with post-tests only. This is often the case when assessing longer-term impacts of an intervention, such as when measuring the effects of a preschool intervention on reading in school.

Adverse effects

To our knowledge, few studies have examined the potential adverse effects of vocabulary interventions on children. At most, studies may report no change in children's language skills after the intervention or a control group making more progress in vocabulary than the treatment group, indicating simple failure. However, interventions that take children away from their usual activities may negatively impact learning in other domains because they are not present for planned activities in whole group or to play with other children, but we are unaware of any studies that have measured or reported this. Some children may not wish to participate in an intervention during the school day (e.g. they may find being singled out for intervention stigmatising), or they may find the activities challenging, evoking a negative reaction, such as

irritation or frustration. However, we are unaware of any vocabulary studies that have reported such outcomes. Finally, a higher rate of attrition in the treatment group than in the control group may indicate failure to complete the intervention as planned due to, for example, time issues for the delivery agent or negative child reactions to the demands of the intervention.

Why it is important to do this review

As discussed, limited vocabulary can impede learning in school, leading to academic failure and dropping out. Therefore, it is important to have an updated overview of effective interventions that can help, from an early age, to prevent such difficulties. Although there have been previous reviews on L2 vocabulary interventions for young children (for an overview, see [Appendix 1](#)), they do not have the same objectives and inclusion criteria as applied in this systematic review and meta-analysis.

The primary objective of this review was to examine the effect of L2 vocabulary interventions on L2 learners when a rigorous randomised controlled trial (RCT) design was employed. Although previous reviews have also sought to examine the effect of vocabulary interventions, they have included multiple designs (e.g. single case studies, quasi-experimental designs; [Hur 2020](#); [Larson 2020](#)). RCTs are not always possible in real life, but this design remains the most robust for assessing the relative effects of interventions ([Higgins 2022a](#)). One review on the effects of linguistic comprehension interventions found that quasi-experimental designs yielded larger effect sizes than RCTs, and including different designs would, therefore, make it difficult to determine how effective a vocabulary intervention may be, and for whom, over time ([Rogde 2019](#)).

Previous reviews of L2 vocabulary interventions have included studies conducted exclusively in the English language ([Fitton 2018](#); [Hur 2020](#); [Larson 2020](#)). Therefore, there is a need to summarise studies conducted in different countries and in a variety of languages to get a better idea of the most effective interventions and whether this varies in different contexts. By including all samples of L2 learners, the aim was to examine how different child characteristics were associated with an intervention's effect.

Previous reviews have variously considered a range of different language skills ([Larson 2020](#)), literacy ([Hur 2020](#)), or only shared book reading ([Fitton 2018](#)). The objective of this review was to include studies that were designed with the aim of improving L2 vocabulary skills and to examine how different approaches to vocabulary instruction (e.g. explicit word learning, implicit learning in context) and different variables, such as dosage, setting (home versus school), and provider (teaching assistants, parents, teachers, speech and language pathologists), were associated with effect size. This is important information for practitioners charged with providing young children with the best opportunity for learning, well-being, and future success, and such knowledge will help them to tailor interventions to prevent later academic problems for L2 learners.

As previous primary studies have reported fade-out effects ([Rogde 2016](#)), this review considered the long-term effects of interventions on different primary (i.e. vocabulary) and secondary (e.g. reading comprehension, communication skills, social skills) outcomes. As these programmes are time-consuming and costly, we needed to determine what happened to the children after the intervention

had ended. By compiling information on different approaches, delivery agents, dosages, and child characteristics associated with intervention success, we have the best chance of providing L2 learners, their preschool teachers, speech and language pathologists, and parents with the best methods. This review is thus important for practitioners who are planning interventions and providing counselling and professional development in the area of L2 learning. Moreover, this review provides crucial knowledge for policymakers who are planning for future resources and support needs.

OBJECTIVES

To examine the immediate and long-term effects of second language (L2) vocabulary interventions targeting L2 learners up to six years of age on vocabulary and social-emotional well-being.

To examine the associations between L2 vocabulary interventions and the general characteristics of L2 learners (e.g. age, L2 exposure and L1 skills).

METHODS

Criteria for considering studies for this review

Types of studies

We included RCTs where participants were randomly assigned to either an intervention or a control group. We included studies that used either cluster randomisation or those that used randomisation at the individual level. We excluded all other study designs (e.g. quasi-experimental, within-subjects).

Types of participants

We included studies with children aged five years 11 months or younger at pretest. We did not include studies involving children of six years or older at pretest; this was determined based on the author-reported age range at pretest.

Study authors could have reported eligible participants as being L2 learners with DLD, who thus had language deficits in both L1 and L2. We excluded studies where the participants were diagnosed with learning or developmental disorders (e.g. autism, sensory impairments, disorder of intellectual development).

We did not apply any restrictions to the type of L1 or L2 or the geographical location of the participants.

Types of interventions

Experimental intervention

- We included any vocabulary intervention that aimed to enhance L2 vocabulary skills as one of the main aims of the intervention.
 - Both educational settings (i.e. nursery, kindergarten, preschool, or school) and the home were acceptable, with the delivery agent being a teacher (preschool, kindergarten, or school), teaching assistant, researcher, speech and language pathologist, or carer (e.g. parent).
 - Interventions could be provided face-to-face, digitally, or by other modes (e.g. shared book reading, digital interactive book reading, activities (e.g. making a book), explicit instruction on target vocabulary skills ([Larson 2020](#))), or implicit exposure through different contexts.

- There were no exclusion criteria based on dose, duration, intensity, or different aspects of implementation quality.

Control intervention

- We included inactive control conditions (e.g. waiting list) or standard care.
- We excluded active control interventions (e.g. a different variant of the same intervention, a different type of intervention), as an active control group that focused on other aspects of language (e.g. phonological awareness) may have had an impact on vocabulary skills, given that these constructs are highly correlated at preschool ages (Hjetland 2020).

Types of outcome measures

We included studies that met the above inclusion criteria regardless of whether they reported on the primary and secondary outcomes listed below. For the studies that included the outcomes, we extracted and analysed outcome assessments at the first post-test (assessed immediately after the intervention) and over a longer term (assessed at least one month after the intervention ended). If the studies reported on more than one long-term follow-up, we selected the last reported time point.

Primary outcomes

- Receptive L2 vocabulary (both proximal and distal)
- Expressive L2 vocabulary (both proximal and distal)
- Mean length of utterance (MLU) (potential adverse effects)^a

^aMLU was included as potential adverse effects. Being part of a vocabulary intervention targeting one's L2 can be demanding. This may cause the child to say and speak less because of the emphasis on words that the child does not have command of yet. Notably, all other listed outcome measures could also capture adverse effects of intervention.

Measurement of outcomes

Proximal measures included taught L2 vocabulary measured in terms of either:

- depth of vocabulary (e.g. by asking the child to define the words included in the intervention); or
- breadth of vocabulary (e.g. by determining whether a child can name a word when shown a corresponding picture).

Distal measures assessed L2 vocabulary that was not included amongst the directly trained words in the intervention. Eligible outcomes included standardised tests such as:

- receptive tests (British Picture Vocabulary Scale Third Edition (BPVS-3; Dunn 2009); and Peabody Picture Vocabulary Test Fifth Edition (PPVT-5; Dunn 2018)); or
- expressive tests (Expressive Vocabulary Test Second Edition (EVT-2; Williams 2007)); or
- researcher-developed tests that tap expressive or receptive L2 vocabulary skills or both, covering breadth, depth, or both.

We presented data as mean number of correct responses for both proximal and distal measures.

Secondary outcomes

- L2 listening comprehension
- L2 narrative skills
- L1 receptive vocabulary (both proximal and distal)
- L1 expressive vocabulary (both proximal and distal)
- L1 listening comprehension
- L2 grammatical knowledge
- L2 reading comprehension (long-term)
- Strengths and Difficulties Questionnaire (SDQ; Goodman 1997). SDQ questionnaire measures social and emotional behaviour. As this is a wide concept, we only include this indicator of this theoretical concept.

For listening comprehension and narrative skills, we included only distal measures. This had not been specified in our protocol (Hjetland 2021). The reason for including these outcome measures was to examine the effect of intervention on broader language measures. As these two measures (proximal and distal outcomes) represent different outcome measures, we could not include both in the same analyses.

Timing of outcome measures

We extracted and analysed outcome assessments at the first post-test (assessed immediately after the intervention) and over a longer term (assessed at least one month after the intervention ended). If studies reported on more than one long-term follow-up, we selected the last reported time point.

Hierarchy of outcome measures

If a study reported on more than one measure for an outcome, we selected the most commonly used measure.

Search methods for identification of studies

In September 2021 and December 2022, the Cochrane Information Specialist for Developmental and Psychosocial Learning Problems searched the following databases, apart from Linguistics and Language Behavior Abstracts, which the first review author (HNN) searched. We did not search Education Database Proquest in 2022 because it was no longer available to the review authors or editorial base. The review team conducted the searches reported in [Searching other resources](#).

Electronic searches

The search strategies for each of the electronic databases are provided in [Appendix 2](#).

- Cochrane Central Register of Controlled Trials (CENTRAL; 2022 Issue 11) in the Cochrane Library, which includes the Cochrane Developmental, Psychosocial and Learning Problems Specialised Register. Searched 6 December 2022
- MEDLINE Ovid (1946 to November week 4 2022)
- MEDLINE In-Process, In-Data-Review & PubMed NOT MEDLINE citations Ovid (1946 to 5 December 2022)
- MEDLINE Epub Ahead of Print Ovid (5 December 2022)
- Embase Ovid (1974 to 5 December 2022)
- ERIC EBSCOhost (1966 to 8 December 2022)
- Education Abstracts (H.W. Wilson) EBSCOhost (1983 to 8 December 2022)

- Education Database ProQuest (1988 to 16 September 2021)
- Linguistics and Language Behavior Abstracts ProQuest (LLBA; 1973 to 10 January 2023)
- PsycINFO Ovid (1806 to November week 4 2022)
- Scopus Elsevier (all available years; searched 6 December 2022)
- Science Citation Index-Expanded Web of Science, Clarivate (1970 to 7 December 2022)
- Social Sciences Citation Index Web of Science, Clarivate (1970 to 7 December 2022)
- Conference Proceedings Citation Index-Science Web of Science, Clarivate (1990 to 7 December 2022)
- Conference Proceedings Citation Index-Social Science and Humanities Web of Science, Clarivate (1990 to 7 December 2022)
- Emerging Sources Citation Index Web of Science, Clarivate (2015 to 7 December 2022)
- Sociological Abstracts ProQuest (1952 to 7 December 2022)
- *Cochrane Database of Systematic Reviews* (CDSR; 2022 Issue 11), in the Cochrane Library (searched 8 December 2022)
- Epistemonikos (www.epistemonikos.org; searched 8 December 2022)
- ClinicalTrials.gov (www.clinicaltrials.gov; searched 8 December 2022)
- World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP; trialsearch.who.int; searched 8 December 2022)

We did not limit the searches by year of publication, language of publication, or publication type.

Searching other resources

We identified other eligible candidate studies by searching the reference lists of the already included studies and relevant reviews, as well as searching citations to the included studies. In addition, the review team (HNN, AMH, MBM, KAN) handsearched the following journals.

- *International Journal of Bilingual Education and Bilingualism* (volume 1 1998 to volume 25 2022)
- *Bilingualism Research Journal* (volume 1 1975–1976 to volume 45 2022)
- *Early Education and Development* (volume 1 1989–1990 to volume 33 2022)
- *Journal Early Childhood Research Quarterly* (volume 1 1986 to volume 61 2022)
- *Journal of Speech, Language, and Hearing Research* (volume 1 1958 to volume 65 2022)

The first review author (HNN) searched the following grey literature databases on 23 September 2021 and 14 December 2022.

- OpenGrey (www.opengrey.eu)
- Google Scholar (scholar.google.com/) (as Google Scholar does not have a limit on the number of hits, we screened the first 500 references. We selected the first 500 records, ranked according to relevancy)

The first review author (HNN) contacted relevant researchers identified through this search and from previous relevant reviews (Fitton 2018; Hur 2020; Larson 2020) via email or ResearchGate to ask for other eligible candidate studies.

Data collection and analysis

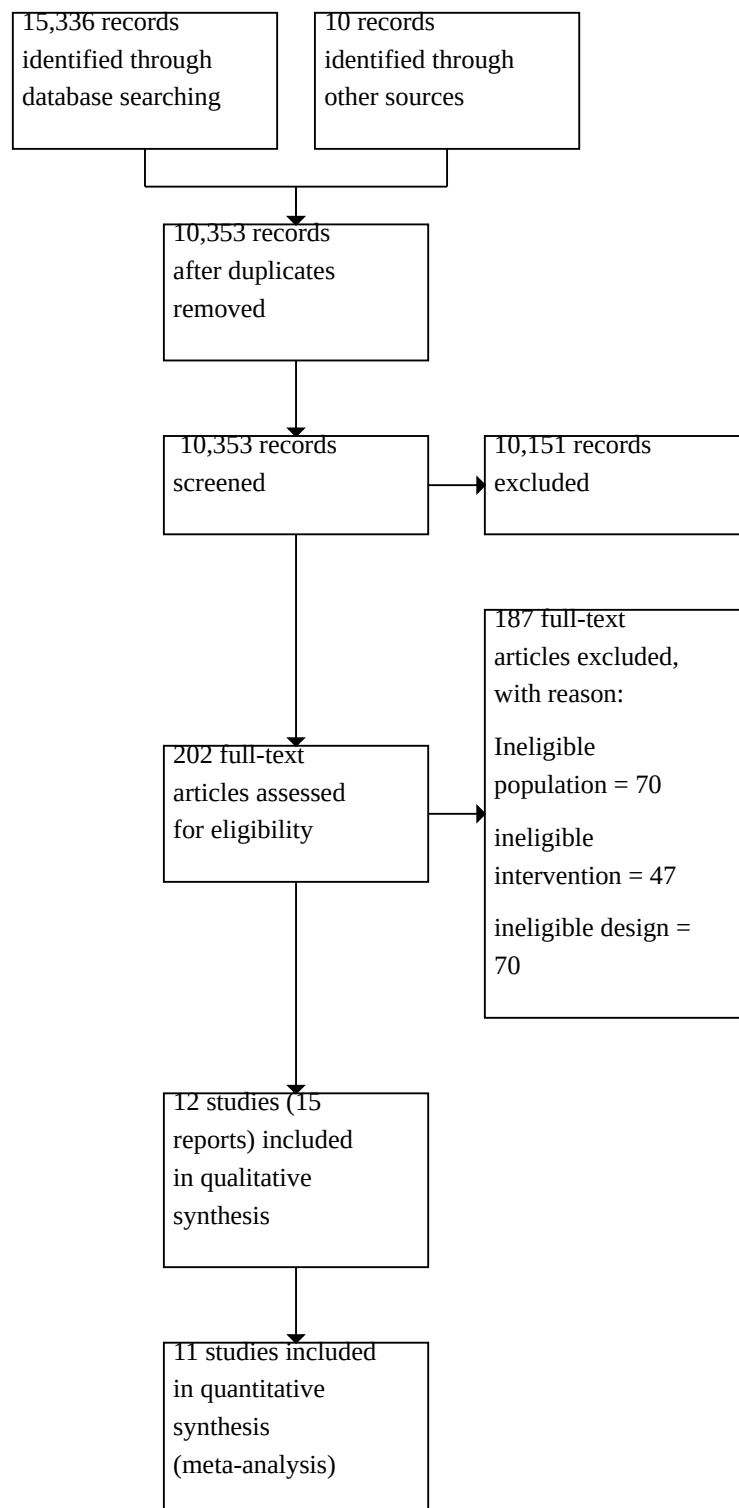
Selection of studies

We imported all records yielded by the searches into EndNote and removed any duplicates. We then exported all records to Covidence (Covidence 2020), where any remaining duplicates were removed before adopting a two-stage approach to screening; we developed a form in Covidence to facilitate the screening process based on the inclusion criteria. In the first stage, two review authors (HNN, HH) independently screened the titles and abstracts of all records against the eligibility criteria (see [Criteria for considering studies for this review](#)). Records deemed potentially eligible, or those that did not provide sufficient information to evaluate eligibility based on the inclusion criteria progressed to the second stage. Inter-rater reliability was strong (Kappa 0.83) (Cohen 1988).

For the next stage, we imported the full texts into Covidence where two review authors (HNN, HH) independently screened the full texts for inclusion based on the selection criteria. We recorded the main reasons for any exclusions in the [Characteristics of excluded studies](#) table. Notably, studies usually were excluded for several reasons; however, we recorded only one reason. Inter-rater reliability was strong (Kappa 0.80) (Cohen 1988). For one German full-text, a German colleague with a PhD and experience with systematic reviews read the publication and provided information about the inclusion criteria.

We resolved disagreements between review authors by discussion and consultation with a third review author (JK). We recorded decisions made throughout the selection process and presented studies in a PRISMA flow diagram (Page 2021; Figure 1), which contained the number of included and excluded studies and the number of studies assessed at each stage.

Figure 1. Study flow diagram.



Data extraction and management

Two review authors (HNN, MK) independently extracted the following categories of information and data from the included studies in Covidence (Covidence 2020). We piloted the data extraction form before starting extracting information from the included studies.

- Information about data extraction from reports (name of data extractors, date of data extraction)
- Study characteristics (title, authors, reference identifier, year of publication, location, source of funding)
- Study method and design (recruitment and sampling procedure, randomisation level, clusters/sites, allocation sequence concealment, blinding, methods used to prevent and address missing data, unit of analysis, statistical methods used, covariates)
- Participant characteristics at baseline (sample size, age, country and region, L1, L2, study eligibility criteria and SES)
- Intervention details (activities, instructional approach, intervention protocols, language of instruction, intervention provider, method of delivery, dosage (frequency and duration), staff qualifications, fidelity, description of standard care control group, etc.)
- Outcomes and outcome measures (any measures related to primary or secondary outcomes (see examples of measures under [Types of outcome measures](#)), timing, standardised or researcher made, expressive or receptive measure, etc.)
- Results (number randomly assigned, number included in the pre-post analysis, number at follow-up, summary data for each group (e.g. 2 × 2 table for dichotomous data, means and standard deviations (SDs) for continuous data), estimate of effect with confidence intervals (CIs), P value, subgroup analyses, etc.)
- Miscellaneous information (key conclusions of primary study authors, correspondence required to retrieve additional data or information, review authors' own comments on study, etc.)

We resolved any disagreements in coding by discussion and consulting a third review author (HH). Once agreement was reached, the first review author (HNN) entered the relevant data into [RevMan Web 2020](#).

Assessment of risk of bias in included studies

Two review authors (HNN, HH) assessed risk of bias using Cochrane's RoB 2 tool for randomised trials (Sterne 2019). We were interested in effects of assignment to intervention, estimated using intention-to-treat (ITT) analyses. Two review authors (HNN, HH) resolved any conflicts in ratings by discussion. If needed, they consulted a third review author (KAN).

RoB 2 includes five domains of bias: 1. bias arising from the randomisation process; 2. bias due to deviation from the intended intervention; 3. bias due to missing outcome data; 4. bias in the measurement of the outcome; and 5. bias in the selection of the reported outcome. We judged the risk of bias in each of the five domains using the RoB 2 signalling questions. We assessed risk of bias in the included primary outcomes at immediate post-test L2 vocabulary (receptive and expressive, both proximal and distal) and MLU (potential adverse effects). In addition to the primary outcomes, we included L2 listening comprehension and narrative skills.

We used the RoB 2 tool for individually randomised parallel-group trials and cluster-randomised trials (Sterne 2019). We used the Excel tool to make decisions for individually randomised parallel-group trials (www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2), and for cluster-randomised parallel-group trials (www.riskofbias.info/welcome/rob-2-0-tool/rob-2-for-cluster-randomized-trials).

For cluster-randomised parallel-group trials, there were some additional considerations when assessing risk of bias in outcomes (Eldridge 2020). While the domains and signalling questions largely follow RoB 2 for parallel-group trials with individual randomisations, we were aware of some differences. For example, in domain 1, randomisation could be based on geography, there could be imbalance in cluster or in participant characteristics. Also, there was a risk of bias if recruitment of individual eligible participants was performed after randomisation of clusters. In domain 2, we assessed if clusters and participants were analysed in their assigned groups.

Based on the domain-level judgement of risk of bias, we reached an overall judgement of risk of bias for the outcomes in each included study: we used 'low risk of bias' to indicate studies with low risk of bias in all domains; 'some concerns of bias' to indicate studies with some concerns of bias in at least one domain; and 'high risk of bias' to indicate studies with at least one high-risk domain or multiple domains with some concerns of bias.

Measures of treatment effect

All outcome data from the included studies were continuous. Because the included studies used different measures to explore the same construct we used the standardised mean difference (SMD) as estimates of the treatment effect for both proximal and distal measures, using means, SDs, and sample sizes to calculate the statistic for each outcome measure and for each group in the study. We used RevMan Web to conduct the random-effects meta-analyses of the treatment effect (RevMan Web 2020), and presented SMD alongside 95% CIs. In instances with only one study, we reported the result from that study using mean difference (MD).

Unit of analysis issues

Cluster-randomised trials

For the cluster-randomised trials, we used corrected data (i.e. intracluster correlation coefficient (ICC)) reported in the studies.

Where studies did not provide ICC data or values, we contacted the authors for further information. If these corrected data were not provided but values that could be used to calculate the intracluster correlation were reported, we estimated corrected data to be used in meta-analyses (Higgins 2022a).

If a cluster-RCT had not been adjusted for clustering in their analysis, we made the adjustment by multiplying the standard errors of the estimates by the square root of the design effect, where the design effect was calculated as $1 + (\text{mean cluster size} - 1) \times \text{ICC}$ (Higgins 2022b).

In cases where ICCs were not reported in the studies or provided by authors on request, we implemented 0.15 as a general rule of thumb. This was considered a plausible value based on the studies that did include ICCs, but also a conservative estimate for the impact of clustering at the classroom level (Schochet 2008).

Studies with multiple treatment groups

For studies that compared multiple treatment groups to a control group, we selected the treatment that had received the highest dose or that had the most vocabulary-based intervention before comparing it with the control group. If there was more than one control group (e.g. one of the control groups included both L1 and L2 learners), we selected the control group with only L2 learners.

Dealing with missing data

To obtain any missing data, the first review author (HNN) contacted the corresponding authors of the publications. We contacted [Baker 2022](#) to request information about the age of the sample, number of participants who completed some of the outcome measures (clearing up information about missing data), and cluster data ([Baker 2022 \[pers comm\]](#)). The corresponding author responded to the e-mails, but, as yet, has not sent the requested information. We contacted [Grøver 2020](#) to ask for detailed information about ICCs. They provided the requested information ([Rydland 2022 \[pers comm\]](#)). We contacted [Chen 2018](#) to ask for information about clusters and ICCs ([Chen 2022 \[pers comm\]](#)). The author responded to the e-mail but has not yet provided the requested information. We contacted [Goodrich 2013](#) to ask for information about group numbers. The corresponding author answered the e-mail and gave the information we requested ([Goodrich 2021 \[pers comm\]](#)).

We reported missing data in the data extraction form and in the risk of bias table.

Assessment of heterogeneity

Potential sources of heterogeneity are related to instruction (e.g. degree of explicit vocabulary instruction), dosage of intervention (e.g. duration and amount), and sample characteristics (e.g. age, L2 exposure, and L1 language). To account for statistical heterogeneity, we tested the heterogeneity of effect sizes using the χ^2 statistic. This establishes the degree to which the variation in effect size is caused by true heterogeneity and not due to chance ([Borenstein 2011](#)). The χ^2 statistic and its P value in a random-effects model reflect whether the variance is significantly different from zero. The null hypothesis is that the studies share a common effect size.

In addition, we reported τ^2 as an indicator of the magnitude of variation in effect sizes between studies. We also assessed the degree of heterogeneity across studies using the I^2 statistic to quantify the amount of true variability in the effect sizes. Specifically, the I^2 statistic indicates the proportion of variance in effects that can be attributed to true heterogeneity versus random error. When interpreting the I^2 statistic, we used the recommended rules of thumb in Section 10.10.2 of the *Cochrane Handbook for Systematic Reviews of Interventions* ([Deeks 2022](#)), which state:

- 0% to 40% might not be important;
- 30% to 60% may represent moderate heterogeneity;
- 50% to 90% may represent substantial heterogeneity; and
- 75% to 100% indicates considerable heterogeneity.

Consequently, we interpreted significant unexplained heterogeneity in the results with caution.

Assessment of reporting biases

Because the meta-analyses included fewer than 10 studies, we were unable to create and examine a funnel plot to assess reporting bias (see [Appendix 3](#) for unused methods).

Data synthesis

We conducted meta-analyses using RevMan Web where outcome data were available from at least two RCTs ([RevMan Web 2020](#)). We pooled the data using random-effects models, as these models take into account that the variation in effect sizes between studies may be due to both random error and systematic differences in the study characteristics ([Borenstein 2011](#)). We conducted meta-analyses of treatment effects for effect sizes obtained at both immediate postintervention and after a longer period of follow-up (e.g. six months after the intervention or one year after the intervention) to examine any long-term maintenance of the treatment effect.

In the event that we were unable to perform a meta-analysis, we provided a narrative summary of the available data.

Subgroup analysis and investigation of heterogeneity

We were unable to conduct subgroup analyses as planned (see [Appendix 3](#)), because there were fewer than 10 studies in the meta-analyses that reported on the moderators ([Deeks 2022](#)). Descriptions of the characteristics of the participants, interventions, and outcomes are provided in [Included studies](#) and the [Characteristics of included studies](#) table.

Sensitivity analysis

We conducted sensitivity analyses with alternative meta-analyses using the fixed-effect model to assess the robustness of the results to decisions made throughout the review process as planned in our protocol ([Hjetland 2021](#)). Because none of the studies included in the quantitative analysis had an overall high risk of bias, we did not conduct the planned sensitivity analysis excluding studies with a high risk of bias ([Appendix 3](#)). In our protocol ([Hjetland 2021](#)), we stated that other sensitivity analyses than the one listed could be considered. Therefore, we decided to run a sensitivity analysis where we excluded the one study that had participants with identified DLD because only this study had this characteristic.

Summary of findings and assessment of the certainty of the evidence

We created a summary of findings table for vocabulary interventions compared with standard care for L2 learners up to six years of age. The table includes the primary outcomes at immediate post-test: L2 vocabulary (both receptive and expressive vocabulary; proximal and distal) and MLU. In addition to the primary outcomes, we included L2 listening comprehension and L2 narrative skills in the summary of findings table. As planned in the protocol ([Hjetland 2021](#)), we chose these secondary outcomes as they are broader language measures that measure mastering of language ability in a more real-life context.

We assessed the certainty of the evidence using GRADEpro GDT ([GRADEpro GDT 2020](#)). Two review authors (HNN, HH) independently evaluated the certainty of the evidence as high, moderate, low, or very low and a third review author (MK) assisted in settling any disagreements. The rating was based on the five domains (inconsistency, indirectness, imprecision, publication

bias, and overall risk of bias) to determine how confident we were that the estimated effect reflected the true effect (GRADEpro GDT 2020). We presented the GRADE ratings in the table and provided reasons for downgrading the certainty in the footnotes. High-certainty evidence meant that the true effect would be close to the estimated effect. Moderate-certainty evidence implied that the true effect was likely close to the estimate but with the possibility that the effect could have been substantially different. Low-certainty evidence reflected that the true effect could have been different from the estimated effect. Very low-certainty evidence implied that the true effect was likely to be different from the estimated effect.

RESULTS

Description of studies

Results of the search

We identified 15,336 records through the electronic databases, 10 through other sources (i.e. through handsearches and screening for included studies in previous reviews). We imported them into EndNote, which identified and removed 4265 duplicates. The remaining 11,081 studies were imported to Covidence (Covidence 2020). Covidence identified and removed an additional 728 duplicates, leaving 10,353 studies ready for title and abstract screening. After screening abstracts for eligibility, we excluded 10,151 irrelevant records, leaving 202 records for full-text screening. After evaluating the full texts for eligibility based on the inclusion criteria, we excluded 187 articles with reasons, leaving 12 included studies from 15 reports (see the flow of studies, Figure 1). We provided four examples of excluded studies in the Characteristics of excluded studies table.

Included studies

For an overview of the 12 included studies, see Characteristics of included studies table.

Study design

All 12 studies reported random assignment to intervention and control conditions. Notably, only five studies were defined as RCTs by the authors (Chen 2018; Grøver 2020; Rogde 2016; Spencer 2020; Thordardottir 2015). Five studies were individually randomised studies (Goodrich 2013; Pérez 2019; Rogde 2016; Thordardottir 2015; Verhallen 2006). Seven studies used cluster-randomised designs; six were clustered at the classroom level (Baker 2022; Chen 2018; Grøver 2020; Hermanns 2011; Pollard-Durodola 2016; Spencer 2020), and one at the school level (Tong 2008).

Baker 2022 described their study as an experimental design; Goodrich 2013 called theirs an experimental intervention study; Hermanns 2011 reported an efficacy trial; Tong 2008 stated that their study was a quasi-experimental design at the student level and an experimental design at the school level; and Pollard-Durodola 2016, Pérez 2019, and Verhallen 2006 stated that their studies were designed to experimentally examine the effect of an intervention with random assignment.

Location

Seven studies were conducted in the USA (Baker 2022; Goodrich 2013; Hermanns 2011; Pérez 2019; Pollard-Durodola 2016; Spencer 2020; Tong 2008), two in Norway (Grøver 2020; Rogde 2016), one

in Canada (Thordardottir 2015), one in the Netherlands (Verhallen 2006), and one in China (Chen 2018).

Setting

Ten studies were conducted in preschool settings; Thordardottir 2015 was conducted in a clinical setting with speech-language pathologists and parents, and Verhallen 2006 was implemented in an experimental setting.

Sample size

The 12 studies included 1943 children. The mean number of children in a study was 162 and ranged from 20 (Tong 2008) to 464 (Grøver 2020). The mean number in the intervention group was 68 children, ranging from 10 (Tong 2008) to 219 (Grøver 2020). The mean number in the control group was 67 children, ranging from nine (Thordardottir 2015) to 190 (Grøver 2020) children.

Participants

Age

The 12 studies included children aged 5 years and 11 months or younger at pretest. The mean age was 4.9 years (58.3 months) in the intervention group and 4.8 years (57.6 months) in the control group. Notably, two studies reported age for the whole sample and not for each group (Pérez 2019; Tong 2008). Baker 2022 confirmed via e-mail that the whole sample was under 5 years and 11 months but has not, at this time, sent the specifics on age that we requested (Baker 2022 [pers comm]).

First and second language

Based on the 12 studies, there were variations in both the L1 and L2 of the participants. Three studies varied L1 language background of the sample (Grøver 2020; Rogde 2016; Thordardottir 2015). In Grøver 2020, the children spoke a variety of L1s (e.g. Albanian, Arabic, Bosnian, Sorani Kurdish, Polish, Russian, Somalian, Tamil, Turkish, Urdu, and Vietnamese). Rogde 2016 reported that the children had different mother tongues from around the world but did not include the different languages. Thordardottir 2015 reported that L1s included Arabic, Bangla, Bengali, Chinese, Dutch, English, Japanese, Kabyl, Punjabi/Urdu, Russian, Sinhalese, Spanish, and Tamil. In the remaining nine studies, the samples involved English language learners (ELL) with Spanish as their L1.

The second language (L2; i.e. the language of the vocabulary interventions) was English in seven studies (Baker 2022; Goodrich 2013; Hermanns 2011; Pérez 2019; Pollard-Durodola 2016; Spencer 2020; Tong 2008), Norwegian in two studies (Grøver 2020; Rogde 2016), French in one study (Thordardottir 2015), Dutch in one study (Verhallen 2006), and Chinese in one study (Chen 2018).

Language skills as eligibility criteria

Seven of the 12 studies used results from a language screening or the previous identification of low language proficiency to select the children who were eligible to participate (Baker 2022; Pérez 2019; Pollard-Durodola 2016; Spencer 2020; Thordardottir 2015; Tong 2008; Verhallen 2006). The studies reported different eligibility criteria.

- Baker 2022 required the children to score below the 30th percentile on an English receptive vocabulary measure to participate. Thus, the sample was considered at risk for

language and learning difficulties. However, children who scored below the second percentile were excluded because the authors did not consider they would benefit from a tier 2 English intervention. The participating children were also required to have some level of receptive Spanish vocabulary.

- [Pérez 2019](#) required children to pass the Spanish speech, language, and hearing screening administered by a Spanish–English bilingual speech–language pathologist, score within the normal limits (standard score 85 or greater) on a Spanish expressive vocabulary measure, and score within the normal limits on a non-verbal intelligence assessment.
- [Pollard-Durodola 2016](#) screened and selected children for the study based on their scores at the prefunctional and beginning levels of English proficiency.
- [Spencer 2020](#) included children with scores within the expected range for their age on an English language screening (listening retell and expressive vocabulary). Thus, this sample included children who displayed low English (L2) language proficiency and low, moderate, or high Spanish (L1) language skills.
- [Thordardottir 2015](#) recruited children with a primary language impairment previously identified by a certified speech–language pathologist. Here, the term 'primary language impairment' referred to the population of individuals who, until recently, were referred to as children with 'specific language impairment'. The present term for this is DLD and the term we used in this review ([Bishop 2017](#)). The children in [Thordardottir 2015](#) study also needed to have significant bilingual exposure – a minimum of six months of regular L2 exposure (e.g. in a daycare setting). Diagnosis of primary language impairment was defined as scores at or below -1.5 SD of the normative monolingual mean on at least one L2 language measure.
- [Tong 2008](#) identified participants with limited English language proficiency, and home language surveys indicated that Spanish was the primary language spoken at home.
- [Verhallen 2006](#) specified that the children did not have language impairments or special educational needs. The participants scored in the bottom 50% on a language test standardised for Dutch kindergarten children (i.e. L2) and had non-verbal intelligence scores within the normal range.

In addition to these seven studies, two studies reported that participants were recruited from Head Start centre classrooms ([Goodrich 2013](#); [Hermanns 2011](#)), which provide comprehensive early childhood education, health, nutrition, and parent involvement services to low-income children and families.

In the remaining studies, the only selection criterion was that the children were L2 learners ([Chen 2018](#); [Grøver 2020](#); [Rogde 2016](#)).

Gender

All 12 studies included both girls and boys in their samples. For details on the number of boys and girls in the studies, see [Characteristics of included studies](#) table.

Funding sources

All 12 studies reported on funding sources for their respective studies. All included studies, except [Pérez 2019](#), reported receiving funds from grants.

Interventions

Experimental intervention

The included studies employed several strategies for teaching vocabulary.

- Shared book reading
- Explicit preteaching of vocabulary before the story was read
- Explicit recap of taught vocabulary after the story was finished
- Having children retell the story or produce the target words, or both
- Included additional games and activities to reinforce book reading vocabulary
- Engaging parents/families in these activities

Elements of explicit vocabulary instruction

The included studies generally had detailed (manualised) intervention programmes with a session plan and frequent activities. [Pérez 2019](#), [Rogde 2016](#), [Spencer 2020](#), and [Tong 2008](#) included detailed scripts for the book reading activities and other lessons. [Grøver 2020](#) noted that they did not include scripts in their intervention. The remaining studies did not report whether they included scripts. Six studies described the intervention as a curriculum ([Baker 2022](#); [Chen 2018](#); [Goodrich 2013](#); [Pollard-Durodola 2016](#); [Spencer 2020](#); [Tong 2008](#)).

The interventions included shared book reading as a core element of the vocabulary intervention. Only [Thordardottir 2015](#) did not include this. [Verhallen 2006](#) used multimedia as a medium for book reading, while the others used more traditional paper books. The multimedia stimuli included action videos, music, and sounds that were related to the main aspects of the story, such as characters, location, time, problem, goal, story events, resolution, and theme. [Spencer 2020](#) used wordless picture books.

Elements of implicit vocabulary instruction

Implicit vocabulary instruction may be defined as a 'naturalistic approach' in which word and language learning takes place during naturalistic, everyday activities. Learning is unintentional in that the learner is not explicitly aware of a learning context or specific targets. Most studies in this review had explicit targets, although implicit learning was often possible. For example, [Grøver 2020](#) encouraged teachers to use play and other activities that expanded on the themes covered in the books, thereby introducing possibilities to learn additional words or a greater depth of meaning of target words. [Spencer 2020](#) included suggested activities that allowed for incorporation of the target vocabulary into circle time and playground games, and a list of additional storybooks that contained the target vocabulary. [Chen 2018](#) included activities such as a field trip and an art activity.

Target words

Ten studies used a selection of target words to support the children's vocabulary ([Baker 2022](#); [Chen 2018](#); [Grøver 2020](#); [Hermanns 2011](#); [Pérez 2019](#); [Pollard-Durodola 2016](#); [Rogde 2016](#); [Spencer 2020](#); [Thordardottir 2015](#); [Tong 2008](#)). Two studies did not include target words ([Goodrich 2013](#); [Verhallen 2006](#)).

The criteria used to select the words were root word meanings ([Hermanns 2011](#)) that preschoolers with limited vocabulary were unlikely to know ([Hermanns 2011](#); [Rogde 2016](#)). Similarly, [Pérez](#)

2019 and Rogde 2016 selected tier 2 words, as these are more abstract and not easily learned without instruction in a general preschool setting. However, these words are highly relevant for building a more abstract language and for learning in school (Beck 2013). Thordardottir 2015 mixed familiar and unfamiliar words, such that five of the target words were words that the child was reported to understand but not produce, while the remaining targets were unfamiliar to the child. Importantly, the children in this study had primary language impairments.

Some interventions included concrete words because they could be easily demonstrated (Hermanns 2011) or represented with a picture (Pérez 2019). In contrast, Chen 2018 included abstract words (such as 'boundary') in addition to the more concrete nouns (e.g. animal names). Chen 2018 selected words related to social life (e.g. wall, bell, passport), and Hermanns 2011 included words that were thought to be interesting for children and useful in everyday situations (e.g. colour, size, sound).

Other considerations were that the words occurred in the books used for instruction (image or text) (Pérez 2019), or in school textbooks (Rogde 2016). Finally, the words that could be presented in both languages (e.g. Spanish and English) were selected. Grøver 2020, Hermanns 2011, Pérez 2019, and Spencer 2020 also provided the target words in the children's L1. Pollard-Durodola 2016 noted that 15% of the words were English/Spanish cognates or words with similar spelling, meaning, and origin; however, they did not specify whether the words were also presented in the L1.

The number of target words explicitly taught each week ranged from two to 10 words. Spencer 2020 included two words for each story. Tong 2008 included three in kindergarten and four to six in grade 1. Rogde 2016 selected three to four words from each book. Pérez 2019 included four new words per week. Grøver 2020 included four or five target words in each book, Baker 2022 included five words each week, Hermanns 2011 and Pollard-Durodola 2016 included six words per week, and Thordardottir 2015 targeted 10 words in each session. Chen 2018 did not specify the number of target words per week.

Content activities and strategies

Understanding new words

Shared book reading was commonly used across studies, with target words embedded in the books. Baker 2022 noted that the lessons in the intervention began with a review of the target word definitions using an instructional routine developed by the research team. Similarly, Pollard-Durodola 2016 included the strategy of reviewing the taught concepts by identifying/naming the picture/concept cards. This took place after book reading. Tong 2008 reviewed the content words as an introduction on the second day.

To support the learning of the content words, the interventions included child-friendly definitions and picture cards that illustrated the words, and anchor sentences that provided meaningful context-aided comprehension were read to them (Baker 2022). Pérez 2019 reported that their programme included the strategy of defining, naming, and repeating the words. Thordardottir 2015 included games and activities that allowed the modelling of target words in meaningful contexts in the part of the sessions devoted to vocabulary, and repetition was also emphasised in the activities.

In the multimedia intervention by Verhallen 2006, the story was repeated four times.

Using new words in expressive language

Another teaching technique used was asking open-ended questions (Chen 2018; Goodrich 2013; Hermanns 2011; Rogde 2016; Tong 2008). Pollard-Durodola 2016 used comprehension questions where the children used their new vocabulary and acquired knowledge and talked about the characters and the story. Hermanns 2011 listed prompting the children with questions as a dialogical reading technique. Rogde 2016 noted that teachers used specific questioning techniques in shared book reading (dialogical reading) to encourage the children to develop more context-independent language and expand their vocabulary while jointly paying attention to a book.

At least six studies reported having additional activities that included broader language components (e.g. phonological awareness, morphology, literacy) and employing tasks as a tool to support the use of target words (Baker 2022; Chen 2018; Goodrich 2013; Rogde 2016; Thordardottir 2015; Tong 2008). Goodrich 2013 included activities related to phonological awareness skills (e.g. word games using pictures to help children better understand that words are made up of individual units of sound) and print knowledge (recognition of the letters in the children's names and later introducing the alphabet as well as the sounds that corresponded to the letters). Baker 2022 noted that reviewing the target words provided the children with explicit practice pronouncing the vocabulary words and repeating the definition. Chen 2018 reported using a rhyming song as one of the activities. Rogde 2016 had a broad linguistic comprehension approach, with activities also targeting grammar and narratives; however, the main focus was on expressive vocabulary. The sessions did not include phonological skills, letter knowledge, and print awareness. Syntactic knowledge was also an important component for Thordardottir 2015 and included the production of subject-verb-object sentence structures. Part of the intervention curriculum used by Tong 2008 included lessons on phonemic awareness, phonics, reading fluency, and reading comprehension in addition to vocabulary development.

Another strategy employed was encouraging the children to speak more (Chen 2018). Baker 2022 reported that the delivery agents of the intervention supported (scaffolded) the children as they used words in novel sentences, with particular attention given to helping children elaborate and extend their sentences. Similarly, Hermanns 2011 reported evaluating and expanding the children's responses as a strategy.

Pollard-Durodola 2016 implemented a discussion activity to encourage the children to relate new words and concepts to their own lived experiences. The discussions were framed within the context of a science or social studies topic and theme. The children were encouraged to use their new vocabulary at home, and the lessons were organised to prime background knowledge. Discussions were also part of the interventional programme in Chen 2018, which used picture book-related discussions. Tong 2008 included role-play conversation as an activity.

Retelling was also a strategy that was used to support word learning (Grøver 2020; Hermanns 2011; Spencer 2020). In Grøver 2020, the children retold stories from a different perspective. Spencer

2020 included packages for each story with additional materials and extension exercises. These included five illustrations, wordless picture books, and story schema icons to scaffold retells, story games, and objects related to target words and icons.

Chen 2018 and Hermanns 2011 used giving positive feedback as a strategy. Play activities were included to ensure that children had fun doing these activities. Spencer 2020 reported that supportive prompts were given along with corrections, modelling, and leading. Thordardottir 2015 also included positive reinforcement and responding to children's production attempts.

Bilingual focus and home settings

Our main focus is vocabulary learning in the L2. However, these children are exposed to and learn two or more languages at the same time. Therefore, some of the studies included activities to support and acknowledge vocabulary development in the children's L1 languages (Chen 2018; Grøver 2020; Spencer 2020; Thordardottir 2015). Chen 2018 included discussions about the picture books in the children's L1 language to scaffold and help them understand new vocabulary and concepts. Grøver 2020 and Spencer 2020 provided books and target words in the L1 that the parents could use with the children at home. Parents were present at all sessions in Thordardottir 2015. In these multilingual sessions, parents were asked to model target words and sentence structures in the L1. The study authors reported that maintaining parent engagement in therapy activities was challenging. Tong 2008 included parent training sessions in which parents received take-home literacy activity books and activities. Notably, Goodrich 2013, Hermanns 2011, Pérez 2019, Thordardottir 2015, and Tong 2008 all included an additional intervention group in which the focus was on both L1 and L2. However, in these studies, the intervention group selected for inclusion in the review was the one that had the largest L2 focus, as planned in our protocol (Hjetland 2021).

Organisation of the intervention delivery

The interventions were mainly organised as small-group sessions (usually with three or four children). Three studies included a combination of group sessions and individual sessions (Pérez 2019; Rogde 2016; Tong 2008). Two studies delivered the intervention only individually (Thordardottir 2015; Verhallen 2006). In Thordardottir 2015, the parent, child, and speech-language pathologist were in the room together. In Verhallen 2006, the sessions took place in a room containing a computer and a digital camera. Some studies combined small-group sessions with a whole-class activity (Chen 2018), or with a larger group, including other children from the classroom (Grøver 2020).

Delivery agent

(Pre)school teachers delivered the intervention in eight studies (Baker 2022; Chen 2018; Grøver 2020; Hermanns 2011; Pollard-Durodola 2016; Rogde 2016; Spencer 2020; Tong 2008). The remaining studies delivered the intervention by bilingual graduate research assistants (Goodrich 2013), monolingual English-speaking first-year graduate student clinicians (Pérez 2019), speech-language pathologists (Thordardottir 2015), or experimenters (Verhallen 2006).

Dosage

The mean dosage per week was 80 minutes and ranged from 24 to 120 minutes. This was calculated based on the nine studies that reported the information (Baker 2022; Chen 2018; Goodrich 2013; Pérez 2019; Pollard-Durodola 2016; Rogde 2016; Thordardottir 2015; Tong 2008; Verhallen 2006). The remaining three studies did not report the number of minutes (Grøver 2020 held three shared reading sessions per week, Hermanns 2011 and Spencer 2020 held daily sessions four days a week).

Duration

The interventions lasted between one week and two years. The duration in six studies ranged from 16 to 22 weeks (Baker 2022; Goodrich 2013; Grøver 2020; Pollard-Durodola 2016; Rogde 2016; Thordardottir 2015). The studies with a longer duration included Spencer 2020, whose intervention was organised in three blocks, with each block lasting eight to 10 weeks (24 to 30 weeks' total duration). The longest durations were in Chen 2018 (one school year) and Tong 2008 (two school years). The studies with the shortest durations were Verhallen 2006 (1 week), Pérez 2019 (six weeks), and Hermanns 2011 (12 weeks).

Control intervention

All studies compared the intervention with standard care. However, typical practice differed based on the setting and local conditions in each country or (pre)school. Thordardottir 2015 employed a waiting list control design in which L2 learners in the control group (all of whom met the criteria for primary language impairment) were offered treatment at the end of the study. However, the outcomes for this delayed treatment group were not reported.

Distinguishing standard care from an active control group may be difficult. For example, in Baker 2022, the comparison group received tier 1 vocabulary instruction in whole groups, while in Tong 2008, the comparison group had the typical practice of structured English immersion (SEI). The intervention groups in these two studies received supplemental tier 2 instruction or enhanced SEI instruction.

In some studies, the control group received materials to work with (Grøver 2020; Hermanns 2011). Grøver 2020 supplied the comparison group with one or two books (seven in total) within each thematic unit. These books were topically linked to the unit but not identical to the books that the intervention group received. Similarly, Hermanns 2011 introduced control classrooms to Spanish and English versions of the same books during the same weeks; however, the preschool teachers used them as they normally would and did not include the additional dialogical reading techniques that were employed in the intervention condition. In addition, the teachers in the control classroom did not explicitly discuss the target vocabulary words.

Outcomes

All 12 studies included data on outcome measures at post-test. As planned in the protocol (Hjetland 2021), we included the outcomes at the first available time point postintervention from each study. Four studies also reported on follow-up assessments (Baker 2022; Pérez 2019; Rogde 2016; Thordardottir 2015). Thordardottir 2015 included information about a follow-up assessment two months after the completion of the training. However, this was applied only to the intervention group because the wait list control group

received the intervention. [Baker 2022](#) followed the participants one year after spring in grade 1. [Rogde 2016](#) included a seven-month follow-up. [Pérez 2019](#) had follow-up testing six weeks after instruction.

Primary outcomes

Receptive L2 proximal vocabulary

Five studies reported an outcome measure of L2 learners' receptive knowledge of the taught vocabulary items ([Baker 2022](#); [Grøver 2020](#); [Hermanns 2011](#); [Pollard-Durodola 2016](#); [Spencer 2020](#)). This was assessed by showing bespoke illustrated sets of four words, which contained one of the target words and three distractors. The assessor uttered the target word, and the child was asked to point to the illustration that corresponded to the word. One study assessed and reported the vocabulary taught after each of the three units of the intervention ([Spencer 2020](#)). For this study, we included the results for the first unit. The scores reflected the number of correct answers.

Receptive L2 distal vocabulary

Seven studies reported outcomes measuring L2 learners' receptive vocabulary on standardised tests comprising words that were not directly taught in the intervention ([Chen 2018](#); [Goodrich 2013](#); [Grøver 2020](#); [Hermanns 2011](#); [Pollard-Durodola 2016](#); [Rogde 2016](#); [Thordardottir 2015](#)). [Grøver 2020](#) and [Rogde 2016](#) reported a Norwegian translation of the BPVS (second edition) ([Dunn 1997](#) adapted to Norwegian by [Lyster 2010](#)), [Pollard-Durodola 2016](#) included the Peabody Picture Vocabulary Test Fourth Edition (PPVT-4; [Dunn 2007](#)), [Thordardottir 2015](#) used the French version of the PPVT-4 (EVIP; Échelle de Vocabulaire en Images Peabody), [Chen 2018](#) reported using the Peabody Picture Vocabulary Test Revised Edition (PPVT-R) translated to Chinese, [Hermanns 2011](#) used the Receptive One-Word Picture Vocabulary Test (ROWPVT) ([Brownell 2000](#)), and [Goodrich 2013](#) used a subtest from the Preschool Comprehensive Test of Phonological and Print Processing (P-CTOPPP; [Lonigan 2002a](#)). In all cases, the children were asked to point to a picture in a set of four that corresponded to the word uttered by the assessor. The scores reflected the number of correct answers.

Expressive L2 proximal vocabulary

Six studies reported on L2 learners expressive word knowledge of the taught vocabulary words using bespoke, researcher-designed assessments ([Baker 2022](#); [Grøver 2020](#); [Pérez 2019](#); [Pollard-Durodola 2016](#); [Rogde 2016](#); [Verhallen 2006](#)). Three studies assessed expressive vocabulary by asking the child to provide a definition of the target word ([Baker 2022](#); [Grøver 2020](#); [Rogde 2016](#)). [Pollard-Durodola 2016](#) performed the assessment by asking the child to name or label what was depicted in a picture of an object. [Pérez 2019](#) included a definition task and a naming task; only the naming task was included in the meta-analysis. Finally, [Verhallen 2006](#) asked L2 learners to fill in the last word to complete a sentence spoken by the assessor while the child saw a matching picture (e.g. The cat lies on the ... (floor)). The scores reflected the number of correct answers.

Expressive L2 distal vocabulary

Seven studies reported children's expressive word knowledge using standardised tests comprising words not directly taught in the interventions ([Chen 2018](#); [Goodrich 2013](#); [Pollard-Durodola](#)

[2016](#); [Rogde 2016](#); [Spencer 2020](#); [Thordardottir 2015](#); [Tong 2008](#)). [Goodrich 2013](#) and [Rogde 2016](#) included word definitions, while the remaining five studies included picture-naming tasks in which pictures included a variety of objects, actions, or concepts presented in pictures ([Chen 2018](#); [Pollard-Durodola 2016](#); [Spencer 2020](#); [Thordardottir 2015](#); [Tong 2008](#)). The scores reflected the number of correct answers.

Mean length of utterance

One study included information about the MLU for both L1 and L2 ([Thordardottir 2015](#)). The study recorded language samples in the home languages, which trained interpreters transcribed and analysed for MLU. The scores represented the number of correct answers.

Secondary outcomes

L2 listening comprehension

Three studies reported secondary outcomes on measures of L2 listening comprehension ([Baker 2022](#); [Spencer 2020](#); [Tong 2008](#)). In [Spencer 2020](#), the assessor read short stories, and the child answered a series of questions that tapped both literal and inferential understanding of the story. In the standardised measure included in [Tong 2008](#) (Listening Comprehension subtest of the Woodcock Language Proficiency Battery – Revised), the child heard a passage read aloud and was required to say a single word that was missing at the end of the passage. [Baker 2022](#) measured the effect on listening comprehension with a bespoke measure containing the content words taught in the intervention. The assessor read short passages aloud and asked the child to answer a series of questions using the target words. The scores reflected the number of correct answers.

L2 narrative skills

Four studies included an outcome measure assessing L2 narrative skills ([Grøver 2020](#); [Rogde 2016](#); [Spencer 2020](#); [Verhallen 2006](#)). The L2 learners were asked to retell a story, and were awarded points if they included central story elements. The measures included in [Grøver 2020](#) and [Rogde 2016](#) did not include stories or words that were explicitly taught in the intervention. However, the measures included in [Spencer 2020](#) and [Verhallen 2006](#) were similar to the stories included in the intervention and thus were not included in the meta-analysis on this outcome measure. The score reflected the number of correct answers.

L1 receptive proximal vocabulary

Three studies included a receptive measure of the taught words in the L1 ([Grøver 2020](#); [Hermanns 2011](#); [Spencer 2020](#)). The children were asked to identify and point to the correct picture/illustration that corresponded to the word (out of four) given by the assessor. The scores reflect the number of correct answers.

L1 receptive distal vocabulary

Three studies included a receptive outcome measure on the distal (i.e. not taught) vocabulary in L1 ([Chen 2018](#); [Goodrich 2013](#); [Hermanns 2011](#)). All three standardised measures asked the children to point to the correct picture out of four that corresponded to the word uttered by the assessor. The scores reflected the number of correct answers.

L1 expressive proximal vocabulary

One study included a measure of the expressive vocabulary taught in the L1 (Pérez 2019). The study assessed both naming and definition. In naming, the L2 learners were asked to label a picture of the target vocabulary. If the L2 learners provided the correct label, they were asked to provide a definition of the word. Definitions were not requested if the L2 learners could not label the word. Statistics on the naming task were included in the coding for the meta-analysis. The scores reflected the number of correct answers.

L1 expressive distal vocabulary

Three studies included a measure of expressive vocabulary in the L1, not including taught words (Chen 2018; Goodrich 2013; Spencer 2020). All three studies asked the L2 learners to label pictures of objects or actions. In the measure used by Goodrich 2013, the L2 learners were also asked to describe its important features. The scores reflected the number of correct answers.

L1 listening comprehension

None of the studies reported L1 listening comprehension.

L2 grammatical knowledge

Four studies reported L2 grammatical knowledge gains following the intervention (Grøver 2020; Rogde 2016; Spencer 2020; Verhallen 2006). Grøver 2020 and Rogde 2016 included the same receptive measure (TROG-2; Bishop 2003). Spencer 2020 included a similar receptive measure (sentence structure, CELF-P) that required L2 learners to point to pictures corresponding to a spoken sentence as well as an expressive word structure measure. To ensure comparable measures in the meta-analysis, we included only the CELF-P outcomes in the meta-analysis. Verhallen 2006 measured grammatical knowledge by having the L2 learners repeat a selection of sentences based on the text used in the intervention; therefore, this was considered a proximal measure and not included in the meta-analysis. The scores reflected the number of correct answers.

L2 reading comprehension

None of the studies reported secondary outcomes on reading comprehension.

Strengths and Difficulties Questionnaire

None of the studies reported secondary outcomes on the children's social, emotional, or behavioural development, as measured using the SDQ.

Excluded studies

We excluded 187 full-text articles because they did not meet the [Criteria for considering studies for this review](#).

The main reasons for exclusion were ineligible population (70), ineligible intervention (47), and ineligible design (70).

Four of the excluded studies are described in [Characteristics of excluded studies](#) table. These were selected because that they are likely to be considered eligible and relevant by readers of the review.

Risk of bias in included studies

Risk of bias judgements are summarised for the outcomes in [Analysis 1.1](#); [Analysis 1.3](#); [Analysis 1.5](#); [Analysis 1.7](#); [Analysis 1.9](#); [Analysis 1.10](#). The Excel tool used to implement RoB 2, with detailed risk of bias assessment data, including consensus responses to the signalling questions, can be accessed separately for [individual randomised studies](#) and for [cluster randomised studies](#).

Bias arising from the randomisation process

Only one study included information about the methods used for randomisation. Thordardottir 2015 reported that random assignment was accomplished using the 'Research Randomizer' website generating six sets of six numbers per set (www.randomizer.org). The remaining studies included information about the level of randomisation and numbers in each group but not the method used or how it was conducted.

Bias due to deviations from intended interventions

The studies did not include information on whether the participants, delivery agents, or parents were aware that they were taking part in an intervention. However, it is reasonable to assume that all parties were aware that they were taking part in a vocabulary intervention. Appropriate analyses (e.g. analysis of variance) were conducted to estimate the effect of assignment to intervention.

Bias due to missing outcome data

The studies generally reported on missing data and the reasons for this. However, only five studies used robust analysis to handle missing data (e.g. maximum likelihood estimation) (Chen 2018; Grøver 2020; Pollard-Durodola 2016; Rogde 2016; Spencer 2020).

Bias in measurement of the outcome

All studies used appropriate methods to measure outcomes, and the same measures were applied to both the intervention and standard care comparison groups. Most studies did not specify whether the assessors were blind to the group assignment. However, Pérez 2019 noted that the assessors were blind to the participants' assignments, and Spencer 2020 reported that the assessors were *not* blind to the assignments.

Bias in selection of the reported result

Although the studies did not report on a prespecified analysis plan (e.g. a protocol), there was no reason to suspect selective reporting.

Overall bias

There were some risk-of-bias concerns in the included studies. However, all studies used random assignment. The concerns were mainly related to information that was not explicitly reported, for example, handling of missing data, methods used for randomisation, and absence of protocols.

Effects of interventions

See: [Summary of findings 1 Summary of findings table - Vocabulary interventions compared with standard care for second language \(L2\) learners up to six years of age](#)

See [Summary of findings 1](#).

For each outcome, we first present the comparison between the group that received the vocabulary intervention versus standard care assessed immediately after the intervention. Next, we present the results from the four studies that included a follow-up assessment (Baker 2022; Pérez 2019; Rogde 2016; Thordardottir 2015). Follow-up data in Thordardottir 2015 are not included in the meta-analyses as they reported neither the means nor the SDs for the follow-up assessment two months postintervention. In addition, this study employed a cross-over, wait-list design in which children in the control group were later offered intervention.

Primary outcomes

Receptive L2 proximal vocabulary

Post-test

Based on four studies with 1073 participants (all cluster-randomised studies), the overall SMD between the intervention and control groups on the taught receptive L2 vocabulary was 0.97 (95% CI 0.64 to 1.30; $P < 0.001$; very low-certainty evidence; Analysis 1.1). These results reflect a mean increase of 0.97 SDs in the taught receptive word knowledge across different measures for the children in the intervention groups measured immediately after the intervention. The result suggests that children learned the words that were explicitly taught in the different programmes. The effect was large, but the certainty of this evidence was very low.

Heterogeneity

There was a considerable level of statistical heterogeneity ($Tau^2 = 0.09$; $Chi^2 = 17.81$, degrees of freedom (df) = 3; $P < 0.001$; $I^2 = 83%$). Due to an insufficient number of studies, we were unable to explore this heterogeneity in subgroup analyses. Although the studies shared similar characteristics (e.g. participants were L2 learners, outcome measures were similar in design and task demands, similar age range), there were variations in content, languages, and length of intervention, which could contribute to explaining the heterogeneity.

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in the four studies, we reran the analysis using either the ICC reported in the studies, or 0.15 as a general rule, producing an adjusted SMD of 0.94 (95% CI 0.62 to 1.26). The analysis indicated that adjusting for the effect of clustering had a minimal effect on the result. There was evidence of substantial statistical heterogeneity ($Tau^2 = 0.07$; $Chi^2 = 10.41$, df = 3; $P = 0.02$; $I^2 = 71%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results, we reran the analysis using a fixed-effect model. The overall SMD between the intervention and control groups on the taught receptive L2 vocabulary was 0.98 (95% CI 0.85 to 1.11; $P < 0.001$; 4 studies, 1073 participants). There was evidence of substantial statistical heterogeneity ($Chi^2 = 17.81$, df = 3; $P < 0.001$; $I^2 = 83%$).

Follow-up

One study with 351 participants assessed the effects of the intervention on receptive L2 proximal vocabulary at follow-up (Baker 2022). There was a difference between the groups in favour of the intervention group (mean difference (MD) 2.51, 95% CI 1.88 to 3.14; $P < 0.001$; Analysis 1.2).

Receptive L2 distal vocabulary

Post-test

Based on the random-effects model from six studies with 1074 participants (three were cluster-randomised studies), the overall SMD between the intervention and control groups on receptive L2 vocabulary that was not taught was 0.29 (95% CI 0.02 to 0.55; $P = 0.03$; low-certainty evidence; Analysis 1.3). These results reflect a mean increase of 0.29 SDs in receptive vocabulary that was not directly taught across different measures for the children participating in the intervention group measured immediately after the intervention. The result suggests that intervention gains may transfer to words not targeted in the intervention; however, the effect was small and the certainty of the evidence low.

Heterogeneity

There was evidence of considerable statistical heterogeneity ($Tau^2 = 0.07$; $Chi^2 = 19.02$, df = 5; $P < 0.001$; $I^2 = 74%$). Due to an insufficient number of studies, we were unable to explore this heterogeneity in subgroup analyses. All studies shared similar characteristics (e.g. the participants were L2 learners, outcomes were similarly constructed, and age range of participants was similar). However, there were variations in the length of intervention, content, and languages, which could contribute to explaining the heterogeneity.

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in three studies (Chen 2018; Grøver 2020; Pollard-Durodola 2016), we reran the analysis using either the ICC reported in the studies, or 0.15 as a general rule, yielding an SMD of 0.28 (95% CI 0.04 to 0.52). The analysis indicated that adjusting for the effect of clustering had a minimal effect on the results. There was evidence of moderate statistical heterogeneity ($Tau^2 = 0.05$; $Chi^2 = 11.01$, df = 5; $P = 0.05$; $I^2 = 55%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results, we reran the analysis using a fixed-effect model. Based on this model, the overall SMD between the intervention and control group on distal receptive L2 vocabulary was 0.30 (95% CI 0.18 to 0.42; $P < 0.001$; 6 studies, 1074 participants). There was evidence of considerable statistical heterogeneity ($Chi^2 = 19.02$, df = 5; $P = 0.002$; $I^2 = 74%$). We reran the analysis excluding Thordardottir 2015, since the participants in this study were children with DLD. This yielded an overall SMD of 0.32 (95% CI 0.04 to 0.59; $P = 0.03$; 5 studies). There was evidence of considerable statistical heterogeneity ($Tau^2 = 0.08$; $Chi^2 = 18.20$, df = 4; $P = 0.001$; $I^2 = 78%$).

Follow-up

One study with 113 participants assessed the effects of intervention on receptive L2 distal vocabulary at follow-up (Rogde 2016). There was no evidence of a difference between groups (MD 4.05, 95% CI -1.54 to 9.64; $P = 0.16$; Analysis 1.4).

Expressive L2 proximal vocabulary

Post-test

Based on the random-effects model from six studies with 1121 participants (three were cluster-randomised studies), the overall SMD between the intervention and control groups on the taught expressive L2 vocabulary was 0.86 (95% CI 0.56 to 1.17; $P <$

0.001; very low-certainty evidence; [Analysis 1.5](#)). These results reflect a mean increase of 0.86 SD in expressive vocabulary that was directly taught across different measures for the children participating in the intervention group assessed immediately after the intervention. The result suggests that children in the intervention group learned how to define and label words they were exposed to. The effect was large, but the certainty very low.

Heterogeneity

There was evidence of considerable statistical heterogeneity ($Tau^2 = 0.10$; $Chi^2 = 18.49$, $df = 5$; $P = 0.002$; $I^2 = 78\%$). Due to an insufficient number of studies, we were unable to explore this heterogeneity in subgroup analyses. Although the studies shared similar characteristics, there were some variations in how the outcome measures were designed. This, together with variations in duration and other factors, could contribute to explaining the heterogeneity.

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster-randomisation in three studies ([Baker 2022](#); [Grøver 2020](#); [Pollard-Durodola 2016](#)), we reran the analysis using either the ICC reported in the studies, or 0.15 as a general rule, which produced an SMD of 0.87 (95% CI 0.54 to 1.19). The analysis indicated that adjusting for the effect of clustering had a minimal effect on the results. There was evidence of substantial statistical heterogeneity ($Tau^2 = 0.10$; $Chi^2 = 18.49$, $df = 5$; $P = 0.002$; $I^2 = 73\%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results we reran the analysis using a fixed-effect model. Based on this model, the overall SMD between the intervention and control group on proximal expressive L2 vocabulary was 0.77 (95% CI 0.65 to 0.89; $P < 0.001$; 6 studies, 1121 participants). There was evidence of considerable statistical heterogeneity ($Chi^2 = 22.24$, $df = 5$; $P < 0.001$; $I^2 = 78\%$).

Follow-up

Based on the random-effects model from three studies with 472 participants (one was a cluster-randomised study), the overall SMD between the intervention and control groups on taught expressive L2 vocabulary at follow-up was 0.74 (95% CI 0.38 to 1.11; $P < 0.001$; [Analysis 1.6](#)). These results reflect a mean increase of 0.74 SD in the children's expressive L2 vocabulary across different measures for the children participating in the intervention group measured at follow-up. The result suggests that the children in the intervention group maintained vocabulary knowledge for the words targeted in the intervention for 1.5 to 12 months after the intervention ended. This effect was large.

Heterogeneity

There was evidence of moderate-to-substantial statistical heterogeneity ($Tau^2 = 0.06$; $Chi^2 = 4.88$, $df = 2$; $P = 0.09$; $I^2 = 59\%$). Due to an insufficient number of studies, we were unable to explore this heterogeneity in subgroup analyses. Although the studies shared similar characteristics, there were variations in, for example, the period after the intervention ended and the design of the outcome measures, which could contribute to explaining the heterogeneity.

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in one study ([Baker 2022](#)), we reran the analysis using the ICC reported in the studies. This resulted in an SMD of 0.74 (95% CI 0.55 to 0.92). The analysis indicated that adjusting for the effect of clustering has a minimal effect on the results. There was evidence of moderate statistical heterogeneity ($Tau^2 = 0.06$; $Chi^2 = 4.77$, $df = 2$; $P = 0.09$; $I^2 = 58\%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results, we reran the analysis using a fixed-effect model. Based on the fixed-effect model, the overall SMD between the intervention and control groups on the taught expressive L2 vocabulary at follow-up was 0.74 (95% CI 0.55 to 0.92; $P < 0.001$; 3 studies, 472 participants). There was evidence of moderate statistical heterogeneity ($Chi^2 = 4.88$, $df = 2$; $P = 0.09$; $I^2 = 59\%$).

Expressive L2 distal vocabulary

Post-test

Based on the random-effects model from seven studies with 960 participants (four were cluster-randomised studies), the overall SMD between the intervention and comparison groups on distal expressive L2 vocabulary was 0.10 (95% CI -0.02 to 0.23; $P = 0.11$; moderate-certainty evidence; [Analysis 1.7](#)). These results reflect a mean increase of 0.10 SD in the expressive vocabulary of words that were not directly taught across different measures for the children participating in the intervention group measured immediately after the intervention. The result suggests that transfer from what the children learned on how to define and label novel words was minimal, given that the effect was small.

Heterogeneity

There was little evidence of statistical heterogeneity ($Tau^2 = 0.00$; $Chi^2 = 5.92$, $df = 6$; $P = 0.43$; $I^2 = 0\%$). Thus, this may reflect the true effect.

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in four studies ([Chen 2018](#); [Pollard-Durodola 2016](#); [Spencer 2020](#); [Tong 2008](#)), we reran the analysis using either the ICC reported in the studies, or 0.15 as a general rule, which produced an SMD of 0.16 (95% CI -0.02 to 0.33). The analysis indicated that adjusting for the effect of clustering has a minimal effect on the results. There was no evidence of statistical heterogeneity ($Tau^2 = 0.00$; $Chi^2 = 4.61$, $df = 6$; $P = 0.59$; $I^2 = 0\%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results we reran the analysis using a fixed-effect model. Based on this model, the overall SMD between the intervention and control groups on distal expressive L2 vocabulary was 0.10 (95% CI -0.02 to 0.23; $P = 0.11$; 7 studies, 960 participants). There was no evidence of statistical heterogeneity ($Chi^2 = 5.92$, $df = 6$; $P = 0.43$; $I^2 = 0\%$). We also tested the robustness of the results by excluding [Thordardottir 2015](#), since the participants in this study were children with DLD. This yielded an overall SMD of 0.11 (95% CI -0.02 to 0.25; $P = 0.09$; 6 studies). There was little evidence of statistical heterogeneity ($Tau^2 = 0.00$; $Chi^2 = 5.36$, $df = 5$; $P = 0.37$; $I^2 = 7\%$).

Follow-up

One study with 113 participants assessed the effects of intervention on expressive L2 distal vocabulary at follow-up (Rogde 2016), and observed no evidence of a difference between the groups (MD 1.57, 95% CI -1.33 to 4.47; $P = 0.29$; Analysis 1.8). In the original article, study authors used latent variables, including a factor reflecting the common variance of expressive grammar, expressive vocabulary, and narrative retelling. Rogde 2016 examined the effect of the intervention at both post-tests, controlling for initial skills at the pretest. At follow-up seven months after the immediate post-test, the effect found at the immediate post-test ($d = 0.55$; $P < 0.001$) was still present, but was reduced in size ($d = 0.26$; $P = 0.037$).

Mean length of utterance

Post-test

One study included information about MLU (Thordardottir 2015). There were no differences in the change scores for either MLU in L2 or L1. In L1, MLU increased only for the intervention group and decreased slightly for the control group. Thordardottir 2015 reported that because of the small number of children for whom data were available in the home language, group differences were compared using non-parametric testing. We rated the certainty of this evidence as moderate.

Follow-up

The study did not report MLU at follow-up.

Secondary outcomes

L2 listening comprehension

Based on the random-effects model from two studies with 294 participants (both were cluster randomised studies), the overall SMD between the intervention and control groups on L2 listening comprehension was 0.19 (95% CI -0.31 to 0.68; $P = 0.46$; very low certainty evidence; Analysis 1.9). These results reflect a mean increase of 0.19 SD in the children's listening comprehension across different measures for the children participating in the intervention group assessed immediately after the intervention. However, the effect was small, and we are very uncertain about the evidence.

Heterogeneity

There was evidence of considerable statistical heterogeneity ($\text{Tau}^2 = 0.09$; $\text{Chi}^2 = 3.65$, $\text{df} = 1$; $P = 0.06$; $I^2 = 73\%$).

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in the two studies, we reran the analysis using either the ICC reported in the studies, or 0.15 as a general rule, which produced an SMD of 0.33 (95% CI -0.12 to 0.77). The analysis indicated that adjusting for the effect of clustering had an effect on the results. There was little evidence of statistical heterogeneity ($\text{Tau}^2 = 0.01$; $\text{Chi}^2 = 1.09$, $\text{df} = 1$; $P = 0.30$; $I^2 = 8\%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results, we reran the analysis using a fixed-effect model. Based on this model, the overall SMD between the intervention and control groups on L2 listening comprehension was 0.10 (95% CI -0.13 to 0.34; $P = 0.38$; 2 studies, 294 participants). There was evidence of considerable statistical heterogeneity ($\text{Chi}^2 = 3.65$, $\text{df} = 1$; $P = 0.06$; $I^2 = 73\%$).

Follow-up

None of the included studies reported on this outcome at follow-up.

L2 narrative skills

Post-test

Based on the random-effects model from two studies with 487 participants (one study used cluster randomisation), the overall SMD between the intervention and control groups on L2 narrative skills was 0.37 (95% CI 0.14 to 0.59; $P = 0.002$; moderate-certainty evidence; Analysis 1.10). These results reflect a mean increase of 0.37 SD in narrative skills across different measures for the children participating in the intervention group measured immediately after the intervention. The result suggests that the children in the intervention group learned how to retell a narrative (story) by including important details such as the different characters, story line, and plot. We note that this effect may reflect the mode of intervention delivery, which most often occurred in the context of shared narrative. The effect was moderate.

Heterogeneity

There was little evidence of statistical heterogeneity ($\text{Tau}^2 = 0.01$; $\text{Chi}^2 = 1.35$, $\text{df} = 1$; $P = 0.002$; $I^2 = 26\%$).

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in one of the two studies (Grøver 2020), we reran the analysis using the ICC reported in the study, which produced an SMD of 0.37 (95% CI 0.14 to 0.60). The analysis indicated that adjusting for the effect of clustering had a minimal effect on the results. There was little evidence of statistical heterogeneity ($\text{Tau}^2 = 0.01$; $\text{Chi}^2 = 1.30$, $\text{df} = 1$; $P = 0.25$; $I^2 = 23\%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results, we reran the analysis using a fixed-effect model. Based on the fixed-effect model, the overall SMD between the intervention and control groups on L2 narrative skills was 0.35 (95% CI 0.17 to 0.53; $P = 0.0002$; 2 studies, 487 participants). There was little evidence of statistical heterogeneity ($\text{Chi}^2 = 1.35$, $\text{df} = 1$; $P = 0.24$; $I^2 = 26\%$).

Follow-up

One study with 113 participants assessed the effects of intervention on L2 narrative skills at follow-up (Rogde 2016), and observed no evidence of a difference between the groups (MD 2.14, 95% CI -0.69 to 4.97; $P = 0.14$; Analysis 1.11). In the original article, study authors used latent variables, including a factor reflecting the common variance of expressive grammar, expressive vocabulary, and narrative retelling. Rogde 2016 examined the effect of the intervention at both post-tests, controlling for initial skills at the pretest. At follow-up seven months after the immediate post-test, the effect found at the immediate post-test ($d = 0.55$; $P < 0.001$) was still present, but was reduced in size ($d = 0.26$; $P = 0.037$).

L1 receptive proximal vocabulary

Post-test

Based on the random-effects model from two studies with 373 participants (both studies used cluster randomisation), the overall SMD between the intervention and control groups on the L1 receptive vocabulary taught was 0.53 (95% CI -0.15 to 1.21; $P =$

0.13; [Analysis 1.12](#)). These results reflect a mean increase of 0.53 SD in L1 receptive vocabulary taught across different measures for the children participating in the intervention group measured immediately after the intervention. The result suggests that the children in the intervention group may learn the words that were targeted in L1. However, given the small number of studies, the effect should be interpreted with caution.

Heterogeneity

There was evidence of considerable statistical heterogeneity ($Tau^2 = 0.21$; $Chi^2 = 7.08$, $df = 1$; $P = 0.008$; $I^2 = 86\%$). Due to an insufficient number of studies, we were unable to explore this heterogeneity in subgroup analyses.

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in the two studies ([Grøver 2020](#); [Spencer 2020](#)), we reran the analysis using the ICC reported in the two studies. This produced an SMD of 0.52 (95% CI -0.16 to 1.20). The analysis indicated that adjusting for the effect of clustering had a minimal effect on the results. There was evidence of considerable statistical heterogeneity ($Tau^2 = 0.20$; $Chi^2 = 5.97$, $df = 1$; $P = 0.01$; $I^2 = 83\%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results, we reran the analysis using a fixed-effect model. Based on this model, the overall SMD between the intervention and control groups on the L1 receptive vocabulary taught was 0.35 (95% CI 0.14 to 0.55; $P = 0.0010$; 2 studies, 373 participants). There was evidence of considerable of statistical heterogeneity ($Chi^2 = 7.08$, $df = 1$; $P = 0.008$; $I^2 = 86\%$).

Follow-up

None of the included studies reported on this outcome at follow-up.

L1 receptive distal vocabulary

Post-test

Based on the random-effects model from two studies with 301 participants (one study was cluster randomised), the overall SMD between the intervention and control groups on distal L1 receptive vocabulary was 0.25 (95% CI 0.02 to 0.48; $P = 0.03$; [Analysis 1.13](#)). This result reflects a mean increase of 0.25 SD in distal L1 receptive vocabulary across different measures for the children participating in the intervention group measured immediately after the intervention. The result suggests that the children in the intervention group may have increased their vocabulary knowledge of words in their L1 that were not directly taught in the intervention; however, this result was based on only two studies and should be interpreted with caution. The effect was small.

Heterogeneity

There was no evidence of statistical heterogeneity ($Tau^2 = 0.00$; $Chi^2 = 0.07$, $df = 1$; $P = 0.80$; $I^2 = 0\%$). Thus, this may reflect the true effect.

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in one of the two studies ([Chen 2018](#)), we reran the analysis using an ICC of 0.15 as a general rule when information was not provided. This resulted in an SMD of 0.24 (95% CI -0.05 to 0.53). The analysis indicated that adjusting for the effect of clustering had

minimal effects on the results. There was no evidence of statistical heterogeneity ($Tau^2 = 0.00$; $Chi^2 = 0.05$, $df = 1$; $P = 0.82$; $I^2 = 0\%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results, we reran the analysis using a fixed-effect model. Based on this model, the overall SMD between the intervention and control groups on the L1 receptive vocabulary that was not taught was 0.25 (95% CI 0.02 to 0.48; $P = 0.03$; 2 studies, 301 participants). There was no evidence of statistical heterogeneity ($Chi^2 = 0.07$, $df = 1$; $P = 0.80$; $I^2 = 0\%$).

Follow-up

None of the included studies reported on this outcome at follow-up.

L1 expressive proximal vocabulary

Post-test

One study with 29 participants reported effects of intervention on expressive L1 proximal vocabulary ([Pérez 2019](#)). [Pérez 2019](#) reported no differences between groups on expressive L1 proximal vocabulary (MD 0.47, 95% CI -1.19 to 2.13; $P = 0.58$; [Analysis 1.14](#)).

Follow-up

[Pérez 2019](#) (29 participants) also reported the effects of the intervention on expressive L1 proximal vocabulary at follow-up, observing no evidence of a difference between the groups (MD 0.68, 95% CI -0.98 to 2.34; $P = 0.42$; [Analysis 1.15](#)).

L1 expressive distal vocabulary

Based on the random-effects model from three studies with 382 participants (two were cluster randomised studies), the overall SMD between the intervention and control groups on distal L1 expressive vocabulary was 0.13 (95% CI -0.07 to 0.34; $P = 0.19$; [Analysis 1.16](#)). These results reflect a mean increase of 0.13 SD in distal L1 expressive vocabulary across different measures for the children participating in the intervention group measured immediately after the intervention. The result suggests that transfer from the taught vocabulary to standardised L1 measures was minimal, as the effect was small.

Heterogeneity

There was no evidence of statistical heterogeneity ($Tau^2 = 0.00$; $Chi^2 = 0.70$, $df = 2$; $P = 0.70$; $I^2 = 0\%$).

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in two of the studies ([Chen 2018](#); [Spencer 2020](#)), we reran the analysis using either the ICC reported in the studies, or 0.15 as a general rule. This resulted in an SMD of 0.10 (95% CI -0.14 to 0.35). The analysis indicated that adjusting for the effect of clustering has a minimal effect on the results. There was no evidence of statistical heterogeneity ($Tau^2 = 0.00$; $Chi^2 = 0.51$, $df = 2$; $P = 0.78$; $I^2 = 0\%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results we reran the analysis using a fixed-effect model. Based on this model, the overall SMD between the intervention and control groups on the L1 expressive vocabulary that was not taught was 0.13 (95% CI -0.07 to 0.38; $P = 0.19$; 3 studies, 382 participants). There was no evidence of statistical heterogeneity ($Chi^2 = 0.70$, $df = 2$; $P = 0.70$; $I^2 = 0\%$).

L1 listening comprehension

None of the included studies reported on L1 listening comprehension.

L2 grammatical knowledge

Post-test

Based on the random-effects model from three studies with 601 participants (two were cluster randomised studies), the overall SMD between the intervention and control groups on grammatical knowledge was 0.31 (95% CI 0.13 to 0.49; $P = 0.0007$; [Analysis 1.17](#)). These results reflect a mean of 0.31 SD increase in children's grammatical knowledge in their L2 across different measures for the children participating in the intervention group when assessed immediately after the intervention. The result suggests that the children in the intervention group increased their morphological and syntactical knowledge after the intervention. This effect was small to moderate.

Heterogeneity

There was little evidence of statistical heterogeneity ($\text{Tau}^2 = 0.00$; $\text{Chi}^2 = 2.23$, $\text{df} = 2$; $P = 0.33$; $I^2 = 10\%$).

Adjusting for cluster randomisation

To assess the effects of adjusting for cluster randomisation in two of the studies ([Grøver 2020](#); [Spencer 2020](#)), we reran the analysis using the ICC reported in the studies. This resulted in an SMD of 0.30 (95% CI 0.11 to 0.49). The analysis indicated that adjusting for the effect of clustering had a minimal effect on the results. There was little evidence of statistical heterogeneity ($\text{Tau}^2 = 0.00$; $\text{Chi}^2 = 2.17$, $\text{df} = 2$; $P = 0.34$; $I^2 = 8\%$).

Sensitivity analysis: fixed-effect model

To assess the robustness of the results we reran the analysis using a fixed-effect model. Based on this model, the overall SMD between the intervention and control groups on L2 grammatical knowledge was 0.31 (95% CI 0.15 to 0.47; $P = 0.0001$; 3 studies, 601 participants). There was little evidence of statistical heterogeneity ($\text{Chi}^2 = 2.23$, $\text{df} = 2$; $P = 0.33$; $I^2 = 10\%$).

Follow-up

One study with 113 participants assessed the effects of intervention on L2 grammatical knowledge at follow-up ([Rogde 2016](#)), and observed no evidence of a difference between the groups (MD 1.46, 95% CI -4.04 to 6.96; $P = 0.60$; [Analysis 1.18](#)). In the original article, study authors used latent variables, including a factor reflecting the common variance of expressive grammar, expressive vocabulary, and narrative retelling. [Rogde 2016](#) examined the effect of the intervention at both post-tests, controlling for initial skills at the pretest. At the follow-up seven months after the immediate post-test, the effect found at the immediate post-test ($d = 0.55$; $P < 0.001$) was still present, but was reduced in size ($d = 0.26$; $P = 0.037$).

L2 reading comprehension

None of the included studies reported on L2 reading comprehension.

Strengths and Difficulties Questionnaire

None of the included studies reported used the SDQ.

DISCUSSION

Summary of main results

We identified 12 studies for inclusion. We conducted four meta-analyses with L2 vocabulary measured immediately after L2 intervention (receptive proximal and distal and expressive proximal and distal) comparing the intervention to standard care. We excluded active comparison interventions as these would answer a different research question.

It is unclear if vocabulary interventions have an effect on L2 learners' understanding of the words taught in the intervention, and their ability to use them. The evidence suggests that there may be a slight increase in receptive distal vocabulary compared with standard care. L2 vocabulary interventions probably result in a slight increase in expressive distal vocabulary. We conducted meta-analyses to investigate the effect of the intervention on other L2 language skills and found that the evidence is very uncertain about the effects of L2 vocabulary interventions on L2 listening comprehension, but L2 vocabulary interventions likely result in a slight increase in L2 narrative skills.

Four studies reported follow-up, but there was only sufficient evidence to conduct a meta-analysis on expressive L2 proximal vocabulary. The follow-up evidence suggests that there may be a large increase in expressive L2 proximal vocabulary and that the children remembered the words learned in the intervention. None of the included studies provided data about the participants' social-emotional well-being.

The studies differed in terms of the vocabulary-learning intervention programmes. The interventions differed in dosage and duration. With one exception, all the intervention programmes included shared book reading as a core element of the intervention. At least seven studies used a selection of content words to support the L2 learners vocabulary knowledge, and used different criteria to select the words (e.g. abstract, concrete, possible to illustrate, tier 2 words that the children did not know, root words). The strategies used in these interventions were reviewing target words, providing definitions, asking open-ended questions, and discussing. At least six studies included activities on broader language components as an approach to support the use of the target words (e.g. phonological awareness, morphology).

The control groups were described as standard care; however, some of the children in the classroom were given books and had tier 1 interventions (i.e. research-based vocabulary instruction given to all children).

Overall completeness and applicability of evidence

The overall completeness of our review depended on the number of studies that met our inclusion criteria. The low number of studies reporting on the same outcome variables limited opportunities to examine the heterogeneity found and prevented subgroup analysis from being conducted. Therefore, we do not report evidence of the impact of different potential moderators, such as characteristic variables of L2 vocabulary interventions or sample characteristics. Finally, the completeness of the evidence on long-term effects was limited. Only four studies provided information about follow-up, which represents insufficient evidence to draw conclusions about the long-term effects of vocabulary interventions.

Regarding the context and applicability of the results, the included studies were conducted in different locations. However, they predominantly had samples with the same L1 language, and studies from the USA were most common. The fact that only a few studies included children from other linguistic backgrounds may affect the generalisation and applicability of the results to classrooms with a range of multilingual communities. Further, only one study included children with DLD (Thordardottir 2015). Notably, the study authors use the term 'primary language impairment' in the article. In this review we use DLD, also to describe the sample in this study, due to the change in term and DLD being broader term. L2 learners may, on average, have limited vocabulary in both their L1 and L2 compared to their monolingual peers (Karlsen 2017; Melby-Lervåg 2014). This may be because their L1 differs from the societal language to which they may have insufficient exposure. Unless the child lacks age-appropriate skills in both languages, this should not be regarded as a language disorder (Bishop 2017). On the one hand, the small number of studies including children with DLD could appear problematic due to these children's need for interventions. On the other hand, the low age of the children included in this review may be seen in light of terminology changes highlighting the uncertainties related to the prediction of prognosis and the persistence of the disorder in young children (Bishop 2017). The results from the young children without any identified language disorder included in our review may represent normal variations in young children and simply indicate the diversity in language proficiency amongst L2 learners of preschool age. Several of the studies used a language screening to select participants for the intervention. The results were almost the same when we excluded the study with children who had DLD (Thordardottir 2015), which could indicate that some of the participants in the other studies had unidentified DLD.

Most of the included studies were multicomponent vocabulary interventions that also involved other language components. It could be questioned whether these multicomponent programmes are vocabulary interventions per se or whether they should be termed 'language interventions', since, in addition to vocabulary components, they may have included an intentional focus on language variables such as syntax (Thordardottir 2015) and morphology (Rogde 2016). However, the dimensionality of language skills in young children at four years of age loaded on one language factor confirms the considerable inter-relationships between vocabulary and other language components (Klem 2015). Since vocabulary can change meaning across contexts, including the use and comprehension of words in specific contexts is necessary for children to both interpret and use the word in a correct way. Learning words in context is also important to achieve the aims of a vocabulary intervention, which is usually for learners to increase their vocabulary and make use of it in daily life (Snow 2017).

A fundamental question raised by this review is how to measure young L2 children's vocabulary. For the proximal expressive measures, two studies included naming tasks, which provided information about how many words the children knew (vocabulary breadth) (Pérez 2019; Pollard-Durodola 2016). Three studies administered definition tasks, which provided information about how well they knew the words (vocabulary depth) (Baker 2022; Grøver 2020; Rogde 2016). For distal expressive measures, only two studies included word definitions (Goodrich 2013; Rogde 2016), while five studies included picture-naming tasks (Chen 2018;

Pollard-Durodola 2016; Spencer 2020; Thordardottir 2015; Tong 2008). This skewed outcome focus towards naming tasks may have increased the magnitude of the effect size in comparison to the possible effects of studies that mainly focused on definition or explanation tasks. As such, our data may represent a simple and narrow definition of what it means to know a word. There is a debate about whether it is possible to distinguish between the breadth and depth of vocabulary or whether depth simply reflects a larger breadth (Vermeer 2001).

Another measurement issue is spillover from the specific trained vocabulary to the untrained vocabulary captured by standardised vocabulary measures. It is possible that learning new words has a cumulative effect on the learning of other words not targeted in the training phase. However, while theoretically plausible, evidence of such transfer to other language skills is limited (see review by Rogde 2016; Rogde 2019 for grammatical knowledge). Some of the studies showed transfer to the extent that enhanced vocabulary was also associated with enhanced narrative comprehension (Grøver 2020; Rogde 2016) and listening comprehension (Spencer 2020). The possibility of transfer taking place is plausible. At the same time, it is important to acknowledge that teaching preschool children new vocabulary does not occur in a vacuum and rarely takes place in isolation. As part of the intervention, the children did not simply learn the meanings of new, isolated words. Instead, as part of the interventions, they were exposed to these new words in various contextualised communicative contexts (e.g. conversation, following directions, language games, storytelling, and story comprehension). Authentic activities, including narratives and listening comprehension, are an integral vehicle for broader, multicomponent language interventions, resulting in enhanced narrative comprehension following vocabulary interventions, which reflects transfer. It may be that these language components are intertwined during the intervention as well as when measuring the outcomes and are difficult to separate; therefore, 'high road' (i.e. morally based) claims for transfer should be avoided.

While all the participants in the studies were L2 learners, there were some significant differences between the studies. Seven studies used screening to select participants. The researchers used different screening measures, which most often targeted L2 vocabulary, such as the lowest 50% in an L2 measure (Verhallen 2006), or below the 30th percentile in L2 receptive vocabulary (Baker 2022). Some of the studies also screened children's L1 skills. Most studies included children with the same L1 background; for example, Spanish as the home language combined with exposure to English at preschool, indicating a rather homogenous sample. Only three studies included participants with varied linguistic backgrounds.

Quality of the evidence

Based on GRADE, we assessed the overall certainty of the evidence as very low for three of the outcomes: receptive L2 proximal vocabulary, expressive L2 proximal vocabulary, and L2 listening comprehension (see [Summary of findings 1](#)). We downgraded one level owing to inconsistencies resulting from substantial heterogeneity, one level due to imprecision, and one level due to 'some concern' overall in the risk of bias assessment. We assessed the certainty of evidence as low for receptive L2 distal vocabulary, which was downgraded one level due to inconsistency resulting from substantial heterogeneity and one level due to 'some concern'

overall in the risk of bias assessment. We assessed the certainty of evidence as moderate for three outcomes: expressive L2 distal vocabulary, mean length of utterance and L2 narrative skills. We downgraded one level due to 'some concern' overall in the risk of bias assessment.

We were unable to examine the potential causes of heterogeneity (by subgroup analyses) due to the insufficient number of studies in the meta-analyses.

The risk of publication bias is unclear, as we were unable to draw a funnel plot due to the insufficient number of studies (fewer than 10) included in the meta-analyses (Deeks 2022).

Potential biases in the review process

The strength of this review is its use of a broad search strategy with no linguistic restrictions. In addition to databases, we searched the grey literature to find potentially relevant studies. All screening for eligibility and coding was performed independently by two review authors to minimise bias.

This review has some limitations. First, by only including RCT designs, we limited the number of vocabulary interventions for L2 learners. Second, by using a standard care control group, we excluded studies with active control groups or control interventions. In the current review, we wanted to assess the intervention programmes against a standard preschool setting to understand whether they increased learners' L2 vocabulary skills. Including an active control group could make this difficult to answer and demand a different research question. The definition of a typical practice control group in an educational setting was less clear than we first anticipated. Third, we restricted the age group to six years or less and intervention effects may differ for older children. Fourth, we focused on L2 acquisition in this review and did not include the full language competence of the participants. While some studies included elements of L1 in their interventions, a review investigating both L1 and L2 is warranted.

Our main outcome and focus in this review is the effect of vocabulary intervention on children's L2 vocabulary acquisition. When exploring vocabulary interventions, vocabulary is the main component, but in some instances, it is difficult to know where to differentiate between vocabulary skills and related language skills. Vocabulary is not targeted in isolation, and vocabulary skills develop within a context (Snow 2017). While vocabulary is the target skill in this review, it is also a component of various embedded skills. This was shown in the included studies, which had a range of dependent variables. The vocabulary interventions in the included studies may have targeted multiple and related skills.

Agreements and disagreements with other studies or reviews

Three previous reviews are specifically relevant to the results of this review (Fitton 2018; Hur 2020; Larson 2020; see Appendix 1 for an overview). As described in [Why it is important to do this review](#), these differ from our review in several ways. First, the previous reviews only included studies conducted in English-speaking countries, whereas ours had no restrictions on language or setting. Second, the three reviews included research designs other than RCTs, whereas ours included only studies with random assignment to the intervention and control groups. Finally, the previous reviews focused on other types of interventions (e.g.

only on shared book reading), whereas our review included all vocabulary interventions. Taking these differences into account, the results from these reviews are still relevant to discuss in relation to this one.

Research design

Previous systematic reviews (Hur 2012; Larson 2020) and meta-analyses (Fitton 2018) have included research designs other than RCTs, but Fitton 2018 conducted an analysis that included only nine studies with random assignment. The analyses resulted in an overall combined effect size of $g = 0.47$. Fitton 2018 combined effect sizes measuring oral language, literacy, or combinations of both. This makes it difficult to directly compare the results to our review, since we analysed each of these outcomes separately. Hur 2012 found effect size point estimates for receptive vocabulary, which ranged from $d = -0.61$ to 1.39 and for expressive vocabulary, which ranged from $d = -1.43$ to 0.86. These effect sizes are *similar* in size to the proximal vocabulary measures in our review. In contrast, Hur 2020 included a wider range of research designs beyond RCTs, making it difficult to compare their results to ours.

Characteristics of the participants

Due to the insufficient number of studies in the meta-analyses, we were unable to conduct the subgroup analyses planned in our protocol (Hjetland 2021). These analyses could have provided information about factors that might explain some of the heterogeneity in the results. Interestingly, Fitton 2018 reported that neither age (measured in years) nor SES (low versus mean or mixed) moderated the effect size. Fitton 2018 also conducted an analysis of children's developmental status. Children with speech disorders or DLD exhibited lower levels of growth from shared book-reading interventions than typically developing children. The effect size estimates were 0.48 for typically developing children and 0.17 for children with speech disorders or DLD. Only one of the studies in our review included children with DLD (Thordardottir 2015).

The results of the sensitivity analysis after we excluded Thordardottir 2015 showed a slight increase in effect sizes. Importantly, although only Thordardottir 2015 included participants with primary language impairments (now termed DLD), several of the included studies in our review screened for language to find participants for their study. In light of unknown selection criteria, it is reasonable to assume that some of the participants in these studies were at risk for DLD, since they were at the lower end of the distribution of language skills. Hur 2020 presented evidence supporting bilingual instruction over English instruction for children at risk of language disabilities. The most common disability or risk category reported in the studies included in Hur 2020 was language or speech delays or disabilities.

Content

With regard to the content of the vocabulary interventions, the review by Larson 2020 identified four categories. These categories were similar to the intervention content identified in our review.

The first category was the explicit instruction of target skills. These interventions took place in groups of one to five children and were usually based at school. Larson 2020 exemplified a study by Collins 2010 that provided children with rich descriptions of researcher-selected English target words while reading. Most of the studies in

our review also included target words usually included in the books used.

The second category involved classroom curriculum interventions. Here, the studies examined the effects of exposing children to a specific curriculum. This was seen less in the studies included in our review, in which five studies reported using a particular curriculum. Since these five studies were from the US and China, this difference may be a result of different curricula in preschool settings and country-specific educational settings for children under the age of six years (Baker 2022; Chen 2018; Goodrich 2013; Pollard-Durodola 2016; Spencer 2020).

The third category from Larson 2020 consisted of interactive book reading or bookmaking interventions (or both). In line with our results, these interventions included shared book reading in combination with strategies designed to engage children in the text and story. One technique reported in both reviews was asking open-ended questions. Hur 2020 also reported that most of the studies in their review (84%) used interactive book reading as a context for implementing the interventions.

The fourth category can be characterised as naturalistic, routine-based interventions. In these interventions, the focus was on caregiver-child oral interactions in everyday routines and activities. This was not a large focus of the studies included in our review. However, some of the studies included unstructured play activities (e.g. Grøver 2020).

Dose and duration

The previous reviews provided limited information about the dose and duration of interventions. Unfortunately, neither Hur 2020 nor Fitton 2018 included this information in their analyses. Larson 2020 described great variations in dose and duration, but, because of the different inclusion criteria, it is difficult to compare their results to ours.

AUTHORS' CONCLUSIONS

Implications for practice

Due to the limited number of studies that met our inclusion criteria and the very low-to-moderate uncertainty of the evidence as a result of inconsistency and imprecision, implications for practice should be considered with caution.

Findings from this review suggest that compared to standard care, second language (L2) vocabulary interventions increase young L2 learners' vocabulary skills, but the evidence is very uncertain. The magnitude of the effects of L2 vocabulary interventions varied from minimal to large, partly reflecting whether or not the outcome measures included taught vocabulary. Outcome measures including taught vocabulary showed a large increase at immediately post-test.

There was also evidence of a small generalisation effect to some broader L2 language measures not reflecting taught vocabulary, such as narrative skills and listening comprehension with little to no evidence of a difference. We are unable to determine whether these results reflect the influence of improved vocabulary per se or whether it represents more regular and structured exposure to narrative and listening tasks as vehicles for strengthening new vocabulary within the interventions. Nevertheless, the large effects

on taught vocabulary and the small generalisation effects suggest a need for careful considerations when selecting target words ensuring that they are relevant and useful for young L2 children.

The available data were insufficient to investigate the heterogeneity via subgroup analyses, such as dosage and L2 participant characteristics. This may imply a lack of research to guide evidence-based practice regarding these aspects.

Only a few studies reported follow-up effects, highlighting that there is limited evidence on whether the immediate positive effects of vocabulary interventions compared to standard care will last long term, and how to best provide L2 support after the intervention.

Notably, only one study included children with documented language deficits in both L1 (first language) and L2 (Thordardottir 2015). Thordardottir 2015 was also unique in that a speech-language pathologist administered the intervention and involved parents directly in therapy activities. Thordardottir 2015 reported challenges in keeping parents engaged. Thus, practitioners should be mindful of how best to involve families in interventions. The success of the interventions reviewed here largely reflects cohorts with no known language or learning difficulties, and therefore, may not replicate in populations where there is evidence of language disorder in L1.

Implications for research

This review highlights the need for more high-quality trials of L2 vocabulary interventions for L2 learners and children who are learning and exposed to multiple languages. Specifically, efforts should also be directed towards studying the effects of vocabulary interventions with children who vary in their L1 skills, particularly with learners outside the US and learners with more diverse language abilities in their L1.

So far, there is little research on how variety in the children's context affects L2 learning. More research is needed to separate the roles of classroom demographics, societal bias, bias towards less prestigious languages, teacher training policies, educational policies, and socio-economic correlates, such as parental education of language communities, that may contribute to the outcomes of interventions designed to enhance L2 learning.

Only one study in our review included children who met the diagnostic criteria for developmental language disorders (DLD) (Thordardottir 2015). Therefore, it is not possible to determine whether these interventions would have similar effects on children with DLD or whether the delivery agent (teaching staff, speech-language pathologist, parent) makes a meaningful difference to treatment outcomes.

We note the wide range of intervention content, strategies, dosage, sample sizes, and outcome measures represented by the small number of studies included in this review. Only two studies included a follow-up assessment more than six months after the intervention. We have little knowledge regarding the longer-term maintenance of these initial treatment gains. This is concerning given that 'fade-out', in which treatment effects attenuate over time, is widely reported for early language interventions (Abenavoli 2019). For treatment gains to be maintained over time, the learned vocabulary needs to facilitate the learning of new words and generalise to other measures of language. We currently lack well-

specified theories of how generalisation occurs. In this review, improvements in broader language tests, such as narrative tasks, could reflect either the generalisation of new vocabulary to narrative contexts or the learning of narrative structures through repeated exposure to narratives in the intervention activities. Designing tasks that tap into the transfer and generalisation of specific vocabulary targets could elucidate these mechanisms (Melby-Lervåg 2019).

Therefore, we recommend that future researchers consider large-scale replications of existing trials and include long-term follow-up of treatment gains and theoretically motivated measures of transfer and generalisation.

ACKNOWLEDGEMENTS

The review authors would like to thank Dr Joanne Duffield (Managing Editor), Dr Sarah Davies (Deputy Managing Editor),

Margaret Anderson (Information Specialist), and Cochrane Developmental, Psychosocial and Learning Problems (CDPLP) for their expert help and support throughout the preparation of this review. The review authors also thank all the reviewers for their thoughtful comments and suggestions, as well as study authors for providing information when contacted.

The Cochrane Review Group is grateful to the following reviewers for their time and comments: Jin Hee Hur, California State University, Long Beach, USA; Nuala Livingstone, Cochrane Evidence Production and Methods Directorate, UK; Brook Sawyer, USA; and Genna White, UK; as well as three other reviewers who chose to remain anonymous.

REFERENCES

References to studies included in this review

Baker 2022 {published and unpublished data}

* Baker DL, McCoach BD, Ware S, Coyne MD, Rattan SM. Effects of Spanish vocabulary knowledge on the English word knowledge and listening comprehension of bilingual students. *International Journal of Bilingual Education and Bilingualism* 2022;**25**:2269-83. [DOI: [10.1080/13670050.2021.1908219](https://doi.org/10.1080/13670050.2021.1908219)]

Coyne MD, McCoach DB, Ware S, Austin CR, Loftus-Rattan SM, Baker DL. Racing against the vocabulary gap: Matthew effects in early vocabulary instruction and intervention. *Exceptional Children* 2019;**85**(2):163-170. [DOI: [10.1177/0014402918789162](https://doi.org/10.1177/0014402918789162)]

Chen 2018 {published data only}

Chen S, Lawrence JF, Zhou J, Min L, Snow CE. The efficacy of a school-based book-reading intervention on vocabulary development of young Uyghur children: a randomized controlled trial. *Early Childhood Research Quarterly* 2018;**44**(3):206-19. [DOI: [10.1016/j.ecresq.2017.12.008](https://doi.org/10.1016/j.ecresq.2017.12.008)]

Goodrich 2013 {published data only}

Farver JA, Lonigan CJ, Eppe S. Effective early literacy skill development for young Spanish-speaking English language learners: an experimental study of two methods. *Child Development* 2009;**80**(3):703-19. [DOI: [10.1111/j.1467-8624.2009.01292.x](https://doi.org/10.1111/j.1467-8624.2009.01292.x)] [PMID: 19489898]

* Goodrich JM, Lonigan CJ, Farver JM. Do early literacy skills in children's first language promote development of skills in their second language? An experimental evaluation of transfer. *Journal of Educational Psychology* 2013;**105**(2):414-26. [DOI: [10.1037/a0031780](https://doi.org/10.1037/a0031780)] [PMCID: PMC3764594] [PMID: 24019555]

Grøver 2020 {published and unpublished data}

Grøver V, Rydland V, Gustafsson J-E, Snow, CE. Shared book reading in preschool supports bilingual children's second-language learning: a cluster-randomized trial. *Child Development* 2020;**91**(6):2192-210. [DOI: [10.1111/cdev.13348](https://doi.org/10.1111/cdev.13348)] [PMID: 31943173]

Hermanns 2011 {published and unpublished data}

Hermanns CB. Leveling the Playing Field: Investigating Vocabulary Development in Latino Preschool-Age English Language Learners [EdD thesis]. Boston (MA): Harvard University, 2011.

Pérez 2019 {published data only}

* Pérez JF, Creaghead NA, Washington K, Guo Y, Raisor-Becker L, Combs S. Using audio prompting to assist monolingual speech-language pathologists to teach English-Spanish vocabulary to English learners. *Communication Disorders Quarterly* 2019;**41**(1):3-11. [DOI: [10.1177/1525740118819659](https://doi.org/10.1177/1525740118819659)]

Pérez JF. The Use of Text-to-Speech to Teach Vocabulary to English Language Learners [Doctoral thesis]. Cincinnati (OH): University of Cincinnati, 2016.

Pollard-Durodola 2016 {published data only}

Pollard-Durodola SD, Gonzalez JE, Saenz L, Soares D, Resendez N, Kwok O, et al. The effects of content-related shared book reading on the language development of preschool dual language learners. *Early Childhood Research Quarterly* 2016;**36**(3):106-21. [DOI: [10.1016/j.ecresq.2015.12.004](https://doi.org/10.1016/j.ecresq.2015.12.004)]

Rogde 2016 {published data only}

Rogde K, Melby-Lervåg M, Lervåg A. Improving the general language skills of second-language learners in kindergarten: a randomized controlled trial. *Journal of Research on Educational Effectiveness* 2016;**9**(Suppl 1):150-70. [DOI: [10.1080/19345747.2016.1171935](https://doi.org/10.1080/19345747.2016.1171935)]

Spencer 2020 {published data only}

Spencer TD, Moran M, Thompson MS, Petersen DB, Restrepo MA. Early efficacy of multitiered dual-language instruction: promoting preschoolers' Spanish and English oral language. *AERA Open* 2020;**6**(1):1-16. [DOI: [10.1177/2332858419897886](https://doi.org/10.1177/2332858419897886)]

Thordardottir 2015 {published data only}

Thordardottir E, Cloutier G, Ménard S, Pelland-Blais E, Rvachew S. Monolingual or bilingual intervention for primary language impairment? A randomized control trial. *Journal of Speech, Language, and Hearing Research* 2015;**58**(2):287-300. [DOI: [10.1044/2014_JSLHR-L-13-0277](https://doi.org/10.1044/2014_JSLHR-L-13-0277)] [PMID: 25381447]

Tong 2008 {published data only}

Tong F, Lara-Alecio R, Irby B, Mathes P, Kwok O-M. Accelerating early academic oral English development in transitional bilingual and structured English immersion programs. *American Educational Research Journal* 2008;**45**(4):1011-44. [PMID: 10.3102/0002831208320790]

Verhallen 2006 {published data only}

Verhallen MJ, Bus AG, De Jong MT. The promise of multimedia stories for kindergarten children at risk. *Journal of Educational Psychology* 2006;**98**(2):410-19. [DOI: [10.1037/0022-0663.98.2.410](https://doi.org/10.1037/0022-0663.98.2.410)]

References to studies excluded from this review

Gagarina 2018 {published data only}

Gagarina N, Topaj N, Posse D, Czapka S. Acquisition of German by bilingual children with L1 Russian or Turkish: is there an influence of language support programs after all? [Der Erwerb des Deutschen bei türkisch-deutsch und russisch-deutsch bilingualen Kindern: gibt es doch einen Einfluss von Sprachfördermaßnahmen?]. *Journal of Childhood and Adolescence Research* 2018;**2**:191-210. [DOI: [10.3224/diskurs.v13i2.05](https://doi.org/10.3224/diskurs.v13i2.05)]

Park 2019 {published data only}

Park HW, Grover I, Spaulding S, Gomez L, Breazeal C. A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education. *Proceedings of the AAAI Conference*

on *Artificial Intelligence* 2019;**33**(1):687-94. [DOI: [10.1609/aaai.v33i01.3301687](https://doi.org/10.1609/aaai.v33i01.3301687)]

Simon-Cerejido 2014 {published data only}

Simon-Cerejido G, Gutiérrez-Clellen VF. Bilingual education for all: Latino dual language learners with language disabilities. *International Journal of Bilingual Education and Bilingualism* 2014;**17**(2):235-54. [DOI: [10.1080/13670050.2013.866630](https://doi.org/10.1080/13670050.2013.866630)]

Wood 2018 {published data only}

Wood C, Fitton L, Petscher Y, Roderiguez E, Sunderman G, Lim T. The effect of e-book vocabulary instruction on Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research* 2018;**61**(8):1945-69. [DOI: [10.1044/2018_JSLHR-L-17-0368](https://doi.org/10.1044/2018_JSLHR-L-17-0368)] [PMID: 30073307]

Additional references

Abenavoli 2019

Abenavoli RM. The mechanisms and moderators of "fade-out": towards understanding why the skills of early childhood program participants converge over time with the skills of other children. *Psychological Bulletin* 2019;**145**(12):1103-27. [DOI: [10.1037/bul0000212](https://doi.org/10.1037/bul0000212)] [PMID: 31714103]

August 2005

August D, Carlo M, Dressler C, Snow C. The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice* 2005;**20**(1):50-7. [DOI: [10.1111/j.1540-5826.2005.00120.x](https://doi.org/10.1111/j.1540-5826.2005.00120.x)]

August 2008

August D, Shanahan T, editor(s). *Developing Reading and Writing in Second-Language Learners: Lessons from the Report of the National Literacy Panel on Language-Minority Children and Youth*. New York (NY): Routledge, 2008.

Baker 2022 [pers comm]

Baker DL. Question about sample size for a meta-analysis [personal communication]. Email to: HN Hjetland 2 February 2022.

Baron 2022

Baron LS, Arbel Y. An implicit-explicit framework for intervention methods in developmental language disorder. *American Journal of Speech-Language Pathology* 2022;**31**(4):1557-73. [DOI: [10.1044/2022_AJSLP-21-00172](https://doi.org/10.1044/2022_AJSLP-21-00172)]

Beck 2013

Beck IL, McKeown MG, Kucan L. *Bringing Words to Life: Robust Vocabulary Development*. 2nd edition. New York (NY): Guilford Press, 2013.

Bialystok 2010

Bialystok E, Luk G, Peets KF, Yang S. Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism* 2010;**13**(4):525-31. [DOI: [10.1017/S1366728909990423](https://doi.org/10.1017/S1366728909990423)] [PMCID: PMC4349351] [PMID: 25750580]

Biemiller 2005

Biemiller AJ, Boote C. Selecting useful word meanings for instruction in the primary grades. In: *Demography and Democracy in the Era of Accountability [La démographie et la démocratie à l'ère de l'imputabilité]*. American Educational Research Association (AERA) 2005 Annual Meeting; 2005 Apr 11-15; Montréal (QC). 2005:183. [WEB PAGE: Available at www.aera.net/Portals/38/docs/Annual_Meeting/AM_2005_000_Full%20Program.pdf]

Bishop 2003

Bishop DV. *Test for Reception of Grammar, Version 2: TROG-2*. London (UK): Pearson, 2003.

Bishop 2017

Bishop DV, Snowling MJ, Thompson PA, Greenhalgh T, CATALISE-2 Consortium. Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: terminology. *Journal of Child Psychology and Psychiatry* 2017;**58**(10):1068-80. [DOI: [10.1111/jcpp.12721](https://doi.org/10.1111/jcpp.12721)] [PMCID: PMC5638113] [PMID: 28369935]

Borenstein 2011

Borenstein M, Hedges LV, Higgins JP, Rothstein HR. *Introduction to Meta-Analysis*. New York (NY): John Wiley & Sons, 2011.

Bornstein 2014

Bornstein MH, Hahn C-S, Putnick DL, Suwalsky JT. Stability of core language skill from early childhood to adolescence: a latent variable approach. *Child Development* 2014;**85**(4):1346-56. [DOI: [10.1111/cdev.12192](https://doi.org/10.1111/cdev.12192)] [PMCID: PMC4286341] [PMID: 25165797]

Bowers 2011

Bowers EP, Vasilyeva M. The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics* 2011;**32**(1):221-41. [DOI: [10.1017/S0142716410000354](https://doi.org/10.1017/S0142716410000354)]

Boyce 2010

Boyce LK, Innocenti MS, Roggman LA, Jump Norman VK, Ortiz E. Telling stories and making books: evidence for an intervention to help parents in migrant Head Start families support their children's language and literacy. *Early Education and Development* 2010;**21**(3):343-71. [DOI: [10.1080/10409281003631142](https://doi.org/10.1080/10409281003631142)]

Brownell 2000

Brownell R. *Receptive One-Word Picture Vocabulary Test: Manual*. 2nd edition. Novato (CA): Academic Therapy Publications, 2000.

Brownell 2001

Brownell R. *Receptive One-Word Picture Vocabulary Test Manual: Spanish Bilingual Edition*. Novato (CA): Academic Therapy Publications, 2001.

Bruton 2009

Bruton A. The Vocabulary Knowledge Scale: a critical analysis. *Language Assessment Quarterly* 2009;**6**(4):288-97. [DOI: [10.1080/15434300902801909](https://doi.org/10.1080/15434300902801909)]

Chauhan 2017

Chauhan S. A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education* 2017;**105**:14-30. [DOI: [10.1016/j.compedu.2016.11.005](https://doi.org/10.1016/j.compedu.2016.11.005)]

Chen 2022 [pers comm]

Chen S. Intra-cluster correlation coefficient (ICC) for a meta-analysis [personal communication]. Email to: HN Hjetland 15 February 2022.

Clark 2016

Clark DB, Tanner-Smith EE, Killingsworth SS. Digital games, design, and learning: a systematic review and meta-analysis. *Review of Educational Research* 2016;**86**(1):79-122. [DOI: [10.3102/0034654315582065](https://doi.org/10.3102/0034654315582065)] [PMCID: PMC4748544] [PMID: 26937054]

Clements 2003

Clements DH, Sarama J. Strip mining for gold: research and policy in educational technology – a response to "Fool's gold". *Educational Technology Review* 2003;**11**(1):7-69. [ERIC #: EJ673505]

Cohen 1988

Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2 edition. Hillsdale (NJ): Lawrence Erlbaum Associates, 1988.

Collins 2010

Collins MF. ELL preschoolers' English vocabulary acquisition from story-book reading. *Early Childhood Research Quarterly* 2010;**25**(1):84-97. [DOI: [10.1016/j.ecresq.2009.07.009](https://doi.org/10.1016/j.ecresq.2009.07.009)]

Covidence 2020 [Computer program]

Covidence. Version accessed 15 January 2020. Melbourne, Australia: Veritas Health Innovation, 2020. Available at covidence.org.

Cox 2012

Cox R, Skouteris H, Rutherford L, Fuller-Tyszkiewicz M, Dell'Aquila D, Hardy LL. Television viewing, television content, food intake, physical activity and body mass index: a cross-sectional study of preschool children aged 2–6 years. *Health Promotion Journal of Australia* 2012;**23**(1):58-62. [DOI: [10.1071/he12058](https://doi.org/10.1071/he12058)] [PMID: 22730942]

Crevecoeur 2014

Crevecoeur YC, Coyne MD, McCoach DB. English language learners and English-only learners' response to direct vocabulary instruction. *Reading & Writing Quarterly* 2014;**30**(1):51-78. [DOI: [10.1080/10573569.2013.758943](https://doi.org/10.1080/10573569.2013.758943)]

Cummins 2000

Cummins J. *Language, Power, and Pedagogy. Bilingual Children in the Crossfire*. Clevedon (UK): Multilingual Matters, 2000.

Cycowicz 1997

Cycowicz YM, Friedman D, Rothstein M, Snodgrass JG. Picture naming by young children: norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology* 1997;**65**(2):171-237. [DOI: [10.1006/jecp.1996.2356](https://doi.org/10.1006/jecp.1996.2356)] [PMID: 9169209]

Psychology 1997;**65**(2):171-237. [DOI: [10.1006/jecp.1996.2356](https://doi.org/10.1006/jecp.1996.2356)] [PMID: 9169209]

Dakun 2000

Dakun W. Vocabulary acquisition: implicit learning and explicit teaching. *REACT* 2000;**2**:15-22.

Dale 1965

Dale E. Vocabulary measurement: techniques and major findings. *Elementary English* 1965;**42**(8):895-901, 948. [WEB PAGE: Available at www.jstor.org/stable/41385916]

De Jong 2000

De Jong PF, Seveke MJ, van Veen M. Phonological sensitivity and the acquisition of new words in children. *Journal of Experimental Child Psychology* 2000;**76**(4):275-301. [DOI: [10.1006/jecp.1999.2549](https://doi.org/10.1006/jecp.1999.2549)] [PMID: 10882476]

DeAvila 2000

DeAvila EA, Duncan SE. *PreLAS2000: English and Spanish Technical Notes*. Monterey (CA): CTB/McGraw-Hill, 2000.

DeCapua 2005

DeCapua A, Marshall HW. Reframing the conversation about students with limited or interrupted formal education: from achievement gap to cultural dissonance. *NASSP Bulletin* 2005;**99**(4):1-15. [DOI: [10.1177/0192636515620662](https://doi.org/10.1177/0192636515620662)]

Deeks 2022

Deeks JJ, Higgins JP, Altman DG, editor(s). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3 (updated February 2022)*. Cochrane, 2022. Available from www.training.cochrane.org/handbook.

Deng 2015

Deng Q, Trainin G. Learning vocabulary with apps: from theory to practice. *Nebraska Educator: a Student-Led Journal* 2015;**29**:49-69. [digitalcommons.unl.edu/nebeducator/29]

Dixon 2020

Dixon C, Thomson J, Fricke S. Evaluation of an explicit vocabulary teaching intervention for children learning English as an additional language in primary school. *Child Language Teaching and Therapy* 2020;**36**(2):91-108. [DOI: [10.1177/0265659020925875](https://doi.org/10.1177/0265659020925875)]

Dowdall 2020

Dowdall N, Melendez-Torres GJ, Murray L, Gardner F, Hartford L, Cooper PJ. Shared picture book reading interventions for child language development: a systematic review and meta-analysis. *Child Development* 2020;**91**(2):383-99. [DOI: [10.1111/cdev.13225](https://doi.org/10.1111/cdev.13225)] [PMID: 30737957]

Dunn 1986

Dunn LM, Lugo DE, Padilla ER, Dunn LM. *Test de Vocabulario en Imágenes Peabody: Adaptación Hispanoamericana [Peabody Image Vocabulary Test: Hispanic American Adaptation]*. Circle Pines (MS): AGS, 1986.

Dunn 1993

Dunn L, Thériault-Whalen C, Dunn L. Échelle de Vocabulaire en Images Peabody: Adaptation Française du Pea-body Picture Vocabulary Test [Peabody Picture Vocabulary Test: French adaptation]. Toronto (ON): PsyCan, 1993.

Dunn 1997

Dunn LM, Dunn DM, Whetton C, Burley J. British Picture Vocabulary Scale. 2nd edition. Windsor (UK): NEFER-Nelson, 1997.

Dunn 2007

Dunn LM, Dunn DM. Peabody Picture Vocabulary Test. 4th edition. Minneapolis (MN): Pearson Assessment, 2007.

Dunn 2009

Dunn LM, Dunn DM, Styles B, Sewell J. British Picture Vocabulary Scale. 3rd edition. Windsor (UK): GL Assessment, 2009.

Dunn 2018

Dunn DM. Peabody Picture Vocabulary Test. 5th edition. Bloomington (MN): Pearson Assessment, 2018.

Ehrler 2008

Ehrler DJ, McGhee RL. PTONI: Primary Test of Nonverbal Intelligence. Austin (TX): Pro-Ed, 2008.

Eldridge 2020

Eldridge S, Campbell MK, Campbell MJ, Drahota AK, Giraudeau B, Reeves B, et al. Revised Cochrane risk of bias tool for randomized trials (RoB 2); November 2020. Available from www.sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool (accessed prior to 20 April 2023).

Ellis 1994

Ellis NC. Vocabulary acquisition: the implicit ins and outs of explicit cognitive mediation. In: Ellis NC, editors(s). *Implicit and Explicit Learning of Languages*. London (UK): Academic Press, Harcourt Brace & Company Publishers, 1994:211-82.

Farnia 2011

Farnia F, Geva E. Cognitive correlates of vocabulary growth in English language learners. *Applied Psycholinguistics* 2011;**32**(4):711-38. [DOI: [10.1017/S0142716411000038](https://doi.org/10.1017/S0142716411000038)]

Farnia 2019

Farnia F, Geva E. Late-emerging developmental language disorders in English-speaking monolinguals and English-language learners: a longitudinal perspective. *Journal of Learning Disabilities* 2019;**52**(6):468-79. [DOI: [10.1177/0022219419866645](https://doi.org/10.1177/0022219419866645)] [PMID: 31387462]

Fitton 2018

Fitton L, McIlraith AL, Wood CL. Shared book reading interventions with English learners: a meta-analysis. *Review of Educational Research* 2018;**88**(5):712-51. [DOI: [10.3102/0034654318790909](https://doi.org/10.3102/0034654318790909)]

Gagarina 2012

Gagarina N, Klop D, Kunnari S, Tantele K, Välimaa T, Balčiūnienė I, et al. MAIN – Multilingual Assessment Instrument for Narratives. *ZAS Papers in Linguistics* 2012;**56**:155. [DOI: [10.21248/zaspil.56.2019.414](https://doi.org/10.21248/zaspil.56.2019.414)]

Gardner 1983

Gardner MF. Expressive One-Word Picture Vocabulary Test. Novato (CA): Academic Therapy Publications, 1983.

Geva 2015

Geva E, Wiener J. Psychological Assessment of Culturally and Linguistically Diverse Children and Adolescents: a Practitioner's Guide. New York (NY): Springer, 2015.

Goodman 1997

Goodman R. The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry* 1997;**38**(5):581-6. [DOI: [10.1111/j.1469-7610.1997.tb01545.x](https://doi.org/10.1111/j.1469-7610.1997.tb01545.x)] [PMID: 9255702]

Goodrich 2021 [pers comm]

Goodrich M. Information about group size for a systematic review [personal communication]. Email to: HN Hjetland 17 November 2021.

GRADEpro GDT 2020 [Computer program]

GRADEpro GDT. Version accessed 15 January 2020. Hamilton (ON): McMaster University (developed by Evidence Prime), 2020. Available from gradepr.org.

Greenwood 2016

Greenwood CR, Carta JJ, Kelley ES, Guerrero G, Kong NY, Atwater J, et al. Systematic replication of the effects of a supplementary technology-assisted storybook intervention for preschool children with weak vocabulary and comprehension skills. *Elementary School Journal* 2016;**116**(4):574-99. [DOI: [10.1086/686223](https://doi.org/10.1086/686223)]

Griffith 2020

Griffith SF, Hagan MB, Heymann P, Heflin BH, Bagner DM. Apps as learning tools: a systematic review. *Pediatrics* 2020;**145**(1):e20191579. [DOI: [10.1542/peds.2019-1579](https://doi.org/10.1542/peds.2019-1579)] [PMID: 31871246]

Grøver 2020

Grøver V, Rydland V, Gustafsson J-E, Snow CE. Shared book reading in preschool supports bilingual children's second-language learning: a cluster-randomized trial. *Child Development* 2020;**91**(6):2192-210. [DOI: [10.1111/cdev.13348](https://doi.org/10.1111/cdev.13348)] [PMID: 31943173]

Hagen 2017

Hagen ÅM, Melby-Lervåg M, Lervåg A. Improving language comprehension in preschool children with language difficulties: a cluster randomized trial. *Journal of Child Psychology and Psychiatry* 2017;**58**(10):1132-40. [DOI: [10.1111/jcpp.12762](https://doi.org/10.1111/jcpp.12762)] [PMID: 28671266]

Hammer 2016

Hammer CS, Sawyer B. Effects of a culturally responsive interactive book-reading intervention on the language abilities of preschool dual language learners: a pilot study. *NHSA Dialog* 2016;**18**(4):59-79. [WEB PAGE: Available at journals.uncc.edu/dialog/article/view/421]

Han 2010

Han W-J, Huang C-C. The forgotten treasure: bilingualism and Asian children's emotional and behavioral health. *American Journal of Public Health* 2010;**100**(5):831-8. [DOI: [10.2105/AJPH.2009.174219](https://doi.org/10.2105/AJPH.2009.174219)]

Hargrave 2000

Hargrave AC, Sénéchal M. A book reading intervention with preschool children who have limited vocabularies: the benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly* 2000;**15**(1):75-90. [DOI: [10.1016/S0885-2006\(99\)00038-1](https://doi.org/10.1016/S0885-2006(99)00038-1)] [ERIC #: EJ633360]

Hart 1995

Hart B, Risley TR. *Meaningful Differences in the Everyday Experience of Young American Children*. Baltimore (MD): Paul H Brookes Publishing Co, 1995.

Hart 2021

Hart SA, Little C, van Bergen E. Nurture might be nature: cautionary tales and proposed solutions. *NPJ Science of Learning* 2021;**6**(1):1-12. [DOI: [10.1038/s41539-020-00079-z](https://doi.org/10.1038/s41539-020-00079-z)]

Hassler 2016

Hassler B, Major L, Hennessy S. Tablet use in schools: a critical review of the evidence for learning outcomes. *Journal of Computed Assisted Learning* 2016;**32**(2):139-15. [DOI: [10.1111/jcal.12123](https://doi.org/10.1111/jcal.12123)]

Haugland 1999

Haugland SW. What role should technology play in young children's learning? Part 1. *Young children* 1999;**54**(6):26-31. [ERIC #: EJ597728]

Higgins 2022a

Higgins JP, Lasserson T, Chandler J, Tovey D, Thomas J, Flemyng E, et al. *Methodological Expectations of Cochrane Intervention Reviews*. London (UK): Cochrane, 2022.

Higgins 2022b

Higgins JP, Eldridge S, Li T, editor(s). Chapter 23: Including variants on randomized trials. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor(s). *Cochrane Handbook for Systematic Reviews of Interventions Version 6.3* (updated February 2022). Cochrane, 2022. Available from training.cochrane.org/handbook.

Hirsh-Pasek 2015

Hirsh-Pasek K, Zosh JM, Golinkoff RM, Gray JH, Robb MB, Kaufman J. Putting education in "educational" apps: lessons from the science of learning. *Psychological Science in the Public Interest* 2015;**16**(1):3-34. [DOI: [10.1177/1529100615569721](https://doi.org/10.1177/1529100615569721)] [PMID: 25985468]

Hjetland 2020

Hjetland HN, Brinchmann EI, Scherer R, Hulme C, Melby-Lervåg M. Preschool pathways to reading comprehension: a systematic meta-analytic review. *Educational Research Review* 2020;**30**:100323. [DOI: [10.1016/j.edurev.2020.100323](https://doi.org/10.1016/j.edurev.2020.100323)]

Hoff 2013

Hoff E. Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental Psychology* 2013;**49**(1):4-14. [DOI: [10.1037/a0027238](https://doi.org/10.1037/a0027238)] [PMCID: PMC4061698] [PMID: 22329382]

Hoff 2014

Hoff E. *Language Development*. 5th edition. Belmont (CA): Wadsworth Cengage Learning, 2014.

Hur 2012

Hur JW, Suh S. Making learning active with interactive whiteboards, podcasts, and digital storytelling in ELL classrooms. *Computers in the Schools* 2012;**29**(4):320-38. [DOI: [10.1080/07380569.2012.734275](https://doi.org/10.1080/07380569.2012.734275)]

Hur 2020

Hur JH, Snyder P, Reichow B. Systematic review of English early literacy interventions for children who are dual language learners. *Topics in Early Childhood Special Education* 2020;**40**(1):6-23. [DOI: [10.1177/02711121419894623](https://doi.org/10.1177/02711121419894623)]

Hwang 2020

Hwang JK, Mancilla-Martinez J, McClain JB, Oh MH, Flores I. Spanish-speaking English learners' English language and literacy skills: the predictive role of conceptually scored vocabulary. *Applied Psycholinguistics* 2020;**41**(1):1-24. [DOI: [10.1017/s0142716419000365](https://doi.org/10.1017/s0142716419000365)] [PMCID: PMC7117096] [PMID: 32255882]

Jean 2009

Jean M, Geva E. The development of vocabulary in English as a second language children and its role in predicting word recognition ability. *Applied Psycholinguistics* 2009;**30**(1):153-95. [DOI: [10.1017/S0142716408090073](https://doi.org/10.1017/S0142716408090073)]

Johnson 2010

Johnson CJ, Beitchman JH, Brownlie EB. Twenty-year follow-up of children with and without speech-language impairments: family, educational, occupational, and quality of life outcomes. *American Journal of Speech-Language Pathology* 2010;**19**(1):51-65. [DOI: [10.1044/1058-0360\(2009\)08-0083](https://doi.org/10.1044/1058-0360(2009)08-0083)] [PMID: 19644128]

Karlsen 2017

Karlsen J, Lyster S-AH, Lervåg A. Vocabulary development in Norwegian L1 and L2 learners in the kindergarten-school transition. *Journal of Child Language* 2017;**44**(2):402-26. [DOI: [10.1017/S0305000916000106](https://doi.org/10.1017/S0305000916000106)] [PMID: 26951479]

Kinash 2012

Kinash S, Brand J, Mathew T. Challenging mobile learning discourse through research: student perceptions of Blackboard

Mobile Learn and iPads. *Australasian Journal of Educational Technology* 2012;**28**(4):639-55. [DOI: [10.14742/ajet.832](https://doi.org/10.14742/ajet.832)]

Klem 2015

Klem M, Melby-Lervåg M, Hagtvet B, Lyster SAH, Gustafsson J-E, Hulme C. Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science* 2015;**18**(1):146-54. [DOI: [10.1111/desc.12202](https://doi.org/10.1111/desc.12202)] [PMCID: PMC4309482] [PMID: 24986395]

Korky 1996

Korky P, Thomas V. Heksenpul met Hennie de heks en de kat Helmer [Winnie the witch] [CD-ROM]. Nieuwegein (the Netherlands): Bombilla, 1996.

Larson 2020

Larson AL, Cyclic LM, Carta JJ, Hammer CS, Baralt M, Uchikoshi Y, et al. A systematic review of language-focused interventions for young children from culturally and linguistically diverse backgrounds. *Early Childhood Research Quarterly* 2020;**50**(Part 1):157-78. [DOI: [10.1016/j.ecresq.2019.06.001](https://doi.org/10.1016/j.ecresq.2019.06.001)]

Lawrence 2014

Lawrence JF, Rolland RG, Branum-Martin L, Snow CE. Generating vocabulary knowledge for at-risk middle school readers: contrasting program effects and growth trajectories. *Journal of Education for Students Placed at Risk* 2014;**19**(2):76-97. [DOI: [10.1080/10824669.2014.958836](https://doi.org/10.1080/10824669.2014.958836)] [ERIC #: EJ1045000]

Leacox 2014

Leacox L, Wood Jackson C. Spanish vocabulary-bridging technology-enhanced instruction for young English language learners' word learning. *Journal of Early Childhood Literacy* 2014;**14**(2):175-97. [DOI: [10.1177/1468798412458518](https://doi.org/10.1177/1468798412458518)]

Lervåg 2010

Lervåg A, Aukrust VG. Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *Journal of Child Psychology and Psychiatry* 2010;**51**(5):612-20. [DOI: [10.1111/j.1469-7610.2009.02185.x](https://doi.org/10.1111/j.1469-7610.2009.02185.x)] [PMID: 19878367]

Lervåg 2018

Lervåg A, Hulme C, Melby-Lervåg M. Unpicking the developmental relationship between oral language skills and reading comprehension: it's simple, but complex. *Child Development* 2018;**89**(5):1821-38. [DOI: [10.1111/cdev.12861](https://doi.org/10.1111/cdev.12861)] [PMID: 28605008]

Lin 2012

Lin DM, Ramírez G, Wilson JS, Geva E. Bridging lexical knowledge and literacy. In: Cumming AH, editors(s). *Adolescent Literacies in a Multicultural Context*. London (UK): Routledge, 2012:120-35. [ERIC #: ED530905]

Lonigan 2002a

Lonigan CJ, Wagner RK, Torgesen JK, Rashotte CA. Preschool Comprehensive Test of Phonological & Print Processing (P-

CTOPPP). Tallahassee (FL): Department of Psychology, Florida State University, 2002.

Lonigan 2002b

Lonigan CJ, Farver JM, Eppe S. Preschool Comprehensive Test of Phonological & Print Processing: Spanish version (P-CTOPPP-S). Tallahassee (FL): Department of Psychology, Florida State University, 2002.

Lu 1998

Lu L, Liu HS. The Peabody Picture Vocabulary Test-Revised in Chinese. Taipei (Taiwan): Psychology Publishing, 1998.

Lugo-Neris 2010

Lugo-Neris MJ, Jackson CW, Goldstein H. Facilitating vocabulary acquisition of young English language learners. *Language, Speech, and Hearing Services in Schools* 2010;**41**(3):314-27. [DOI: [10.1044/0161-1461\(2009/07-0082\)](https://doi.org/10.1044/0161-1461(2009/07-0082))] [PMID: 20421611]

Lyster 2009

Lyster S-AH, Horn E. Test for Reception of Grammar-2 (TROG-2): Norwegian Version [Norsk håndbok for Test for Reception of Grammar-2]. Stockholm (Sweden): Pearson, 2009.

Lyster 2010

Lyster SA, Horn E, Rygvold AL. Vocabulary and vocabulary development in Norwegian children and youth: results from testing with British Picture Vocabulary Scale II [Ordforråd og ordforrådsutvikling hos norske barn og unge: resultater fra en utprøving av British Picture Vocabulary Scale II, second edition]. *Spesialpedagogikk* 2010;**9**:37-45. [URL: www.utdanningsnytt.no/files/2019/08/21/Spesialpedagogikk%209%202010.pdf]

Martin 2013

Martin NA. Expressive One-Word Picture Vocabulary Test-4: Spanish Bilingual Edition (EOWPVT-4: Spanish-Bilingual Edition). Novato (CA): Academic Therapy, 2013.

Marulis 2010

Marulis LM, Neuman SB. The effects of vocabulary intervention on young children's word learning: a meta-analysis. *Review of Educational Research* 2010;**80**(3):300-35. [DOI: [10.3102/0034654310377087](https://doi.org/10.3102/0034654310377087)]

McDaniel 2019

McDaniel J, Benítez-Barrera CR, Soares AC, Vargas A, Camarata S. Bilingual versus monolingual vocabulary instruction for bilingual children with hearing loss. *Journal of Deaf Studies and Deaf Education* 2019;**24**(2):142-60. [DOI: [10.1093/deafed/eny042](https://doi.org/10.1093/deafed/eny042)] [PMCID: PMC6422237] [PMID: 30597033]

McDonald 2014

McDonald L, Wardle J, Llewellyn CH, van Jaarsveld CH, Fisher A. Predictors of shorter sleep in early childhood. *Sleep Medicine* 2014;**15**(5):536-40. [DOI: [10.1016/j.sleep.2014.01.005](https://doi.org/10.1016/j.sleep.2014.01.005)] [PMCID: PMC4038745] [PMID: 24726571]

Melby-Lervåg 2014

Melby-Lervåg M, Lervåg A. Reading comprehension and its underlying components in second-language learners: a meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin* 2014;**140**(2):409-33. [DOI: [10.1037/a0033890](https://doi.org/10.1037/a0033890)] [PMID: 23937316]

Melby-Lervåg 2019

Melby-Lervåg M, Hagen Å, Lervåg A. Disentangling the far transfer of language comprehension gains using latent mediation models. *Developmental Science* 2019;**23**(4):1-14. [DOI: [10.1111/desc.12929](https://doi.org/10.1111/desc.12929)]

Monsrud 2022

Monsrud M-B, Rydland V, Geva E, Thurmann-Moe AC, Lyster SA. The advantages of jointly considering first and second language vocabulary skills among emergent bilingual children. *International Journal of Bilingual Education and Bilingualism* 2022;**25**:42-58. [DOI: [10.1080/13670050.2019.1624685](https://doi.org/10.1080/13670050.2019.1624685)]

Murphy 1997

Murphy J. Slaap Lekker, Meneer Beer [Sleep Well, Mr Bear]. Haarlem (the Netherlands): Holland BV, Uitgeversmaatschappij, 1997.

Nagy 2000

Nagy WE, Scott JA. Vocabulary processes. In: Kamil ML, Mosenthal PB, Pearson PD, Barr R, editors(s). *Handbook of Reading Research*. Vol. 3. Mahwah (NJ): Lawrence Erlbaum Associates, 2000:269-84.

NICHD 2000

National Institute of Child Health and Human Development (NICHD). National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction; April 2000. Available at www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf (accessed prior to 20 April 2023).

Norbury 2016

Norbury CF, Gooch D, Wray C, Baird G, Charman T, Simonoff E, et al. The impact of nonverbal ability on prevalence and clinical presentation of language disorder: evidence from a population study. *Journal of Child Psychology and Psychiatry* 2016;**57**:1247-57.

Northrop 2019

Northrop L, Andrei E. More than just word of the day: vocabulary apps for English learners. *Reading Teacher* 2019;**72**(5):623-30. [DOI: [10.1002/trtr.1773](https://doi.org/10.1002/trtr.1773)] [ERIC #: EJ1207060]

OECD 2019

OECD. PISA 2018 Results (Volume II): Where All Students Can Succeed. Paris (France): OECD Publishing, 2019. [DOI: [10.1787/b5fd1b8f-en](https://doi.org/10.1787/b5fd1b8f-en)]

OECD 2020a

OECD. International Migration Outlook 2020. Paris (France): OECD Publishing, 2020. [DOI: [10.1787/ec98f531-en](https://doi.org/10.1787/ec98f531-en)]

OECD 2020b

OECD. Quality Early Childhood Education and Care for Children Under Age 3: Results from the Starting Strong Survey 2018, TALIS. Paris (France): OECD Publishing, 2020. [DOI: [10.1787/99f8bc95-en](https://doi.org/10.1787/99f8bc95-en)]

Oller 2007

Oller DK, Pearson BZ, Cobo-Lewis AB. Profile effects in early bilingual language and literacy. *Applied Psycholinguistics* 2007;**28**(2):191-230. [DOI: [10.1017/S0142716407070117](https://doi.org/10.1017/S0142716407070117)] [PMCID: PMC3358777] [PMID: 22639477]

Page 2021

Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;**372**:1-9. [DOI: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)]

Paradis 2009

Paradis J. Early bilingual and multilingual acquisition. In: Auer P, Wei L, editors(s). *Handbook of Multilingualism and Multilingual Communication*. New York (NY): Mouton de Gruyter, 2009:15-44.

Paradis 2011

Paradis J. Individual differences in child English second language acquisition: comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism* 2011;**1**(3):213-37. [DOI: [10.1075/lab.1.3.01par](https://doi.org/10.1075/lab.1.3.01par)]

Pearson 2007

Pearson PD, Hiebert EH, Kamil ML. Vocabulary assessment: what we know and what we need to learn. *Reading Research Quarterly* 2007;**42**(2):282-96. [DOI: [10.1598/RRQ.42.2.4](https://doi.org/10.1598/RRQ.42.2.4)]

Petersen 2016

Petersen DB, Spencer TD. CUBED. Laramie (WY): Language Dynamics Group, 2016.

Pollard-Durodola 2016

Pollard-Durodola SD, Gonzalez JE, Saenz L, Soares D, Resendez N, Kwok O, et al. The effects of content-related shared book reading on the language development of preschool dual language learners. *Early Childhood Research Quarterly* 2016;**36**:106-21. [DOI: [10.1016/j.ecresq.2015.12.004](https://doi.org/10.1016/j.ecresq.2015.12.004)]

Proctor 2012

Proctor CP, Silverman RD, Harring JR, Montecillo C. The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing* 2012;**25**(7):1635-64. Erratum in: *Reading and Writing* 2012;**25**:1795. [DOI: [10.1007/s11145-011-9336-5](https://doi.org/10.1007/s11145-011-9336-5)]

Renfrew 1997

Renfrew CE. Bus Story Test: a Test of Narrative Speech. 4th edition. Bicester (UK): Speechmark, 1997.

Restrepo 2013

Restrepo MA, Morgan GP, Thompson MS. The efficacy of a vocabulary intervention for dual-language learners with language impairment. *Journal of Speech,*

Language, and Hearing Research 2013;**56**(2):748-65. [DOI: [10.1044/1092-4388\(2012/11-0173\)x](https://doi.org/10.1044/1092-4388(2012/11-0173)x)] [PMID: 23690568]

RevMan Web 2020 [Computer program]

Review Manager Web (RevMan Web). Version 1.22.0. The Cochrane Collaboration, 2020. Available at revman.cochrane.org.

Rice 1991

Rice ML, Sell MA, Hadley PA. Social interactions of speech- and language-impaired children. *Journal of Speech and Hearing Research* 1991;**34**(6):1299-307. [DOI: [10.1044/jshr.3406.1299](https://doi.org/10.1044/jshr.3406.1299)] [PMID: 1787712]

Rogde 2016

Rogde K, Melby-Lervåg M, Lervåg A. Improving the general language skills of second-language learners in kindergarten: a randomized controlled trial. *Journal of Research on Educational Effectiveness* 2016;**9**(Suppl 1):150-70. [DOI: [10.1080/19345747.2016.1171935](https://doi.org/10.1080/19345747.2016.1171935)]

Rogde 2019

Rogde K, Hagen ÅM, Melby-Lervåg M, Lervåg A. The effect of linguistic comprehension instruction on generalized language and reading comprehension skills: a systematic review. *Campbell Systematic Reviews* 2019;**15**(4):e1059. [DOI: [10.1002/cl2.1059](https://doi.org/10.1002/cl2.1059)]

Rogoff 1990

Rogoff B. *Apprenticeship in Thinking: Cognitive Development in Social Context*. New York (NY): Oxford University Press, 1990.

Rowe 2012

Rowe ML. A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development* 2012;**83**(5):1762-74. [DOI: [10.1111/j.1467-8624.2012.01805.x](https://doi.org/10.1111/j.1467-8624.2012.01805.x)] [PMCID: PMC3440540] [PMID: 22716950]

Rydland 2021

Rydland V, Grøver V. Language use, home literacy environment, and demography: predicting vocabulary skills among diverse young dual language learners in Norway. *Journal of Child Language* 2021;**48**(7):717-36. [DOI: [10.1017/S0305000920000495](https://doi.org/10.1017/S0305000920000495)] [PMID: 33023680]

Rydland 2022 [pers comm]

Rydland V (forwarded from first author Grøver V). Intra-cluster correlation coefficient (ICC) til meta-analyse [personal communication]. Email to: HN Hjetland 21 February 2022.

Schochet 2008

Schochet PZ. Statistical power for random assignment evaluations of education programs. *Journal of Educational and Behavioral Statistics* 2008;**33**(1):62-87. [DOI: [10.3102/1076998607302714](https://doi.org/10.3102/1076998607302714)]

Semel 2004

Semel EM, Wiig EH, Secord WA. CLEF 4: Clinical Evaluation of Language Fundamentals® 4 Screening Test. 4th edition. Toronto (ON): Psychological Corporation, 2004.

Shuler 2012

Shuler C. *iLearn II; an Analysis of the Education Category of the iTunes App Store*. New York (NY): Joan Ganz Cooney Center at Sesame Workshop, 2012.

Simos 2014

Simos PG, Sideridis GD, Mouzaki A, Chatzidaki A, Tzevelekiou M. Vocabulary growth in second language among immigrant school-aged children in Greece. *Applied Psycholinguistics* 2014;**35**(3):621-47. [DOI: [10.1017/S0142716412000525](https://doi.org/10.1017/S0142716412000525)]

Snow 1995

Snow CE, Tabors PO, Nicholson PA, Kurland BF. SHELL: oral language and early literacy skills in kindergarten and first-grade children. *Journal of Research in Childhood Education* 1995;**10**(1):37-48. [DOI: [10.1080/02568549509594686](https://doi.org/10.1080/02568549509594686)]

Snow 2016

Snow PC, Woodward M, Mathis M, Powell MB. Language functioning, mental health and alexithymia in incarcerated young offenders. *International Journal of Speech-Language Pathology* 2016;**18**(1):20-31. [DOI: [10.3109/17549507.2015.1081291](https://doi.org/10.3109/17549507.2015.1081291)] [PMID: 28425363]

Snow 2017

Snow CE. The role of vocabulary versus knowledge in children's language learning: a fifty-year perspective [El papel del vocabulario frente al conocimiento en el aprendizaje lingüístico de los niños: una perspectiva de cincuenta años]. *Journal for the Study of Education and Development* 2017;**40**(1):1-18. [DOI: [10.1080/02103702.2016.1263449](https://doi.org/10.1080/02103702.2016.1263449)]

Spencer 2019

Spencer TD, Goldstein H. *Assessment of Story Comprehension*. Baltimore (MD): Paul H Brookes Publishing Co, 2019.

Stahl 1986

Stahl SA, Fairbanks MM. The effects of vocabulary instruction: a model-based meta-analysis. *Review of Educational Research* 1986;**56**(1):72-110. [DOI: [10.3102/00346543056001072](https://doi.org/10.3102/00346543056001072)]

Stahl 2006

Stahl SA, Nagy WE. *Teaching Word Meanings*. Mahwah (NJ): Lawrence Erlbaum, 2006.

Stangeland 2017

Stangeland EB. The impact of language skills and social competence on play behaviour in toddlers. *European Early Childhood Education Research Journal* 2017;**25**(1):106-21. [DOI: [10.1080/1350293X.2016.1266224](https://doi.org/10.1080/1350293X.2016.1266224)] [ERIC #: EJ1130487]

Stanovich 1986

Stanovich KE. Matthew effects in reading: some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly* 1986;**21**(4):306-407. [WEB PAGE: Available at www.jstor.org/stable/747612]

Sterne 2019

Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias

in randomised trials. *BMJ* 2019;**366**:l4898. [DOI: [10.1136/bmj.l4898](https://doi.org/10.1136/bmj.l4898)] [PMID: 31462531]

Storch 2002

Storch SA, Whitehurst GJ. Oral language and code-related precursors to reading: evidence from a longitudinal structural model. *Developmental Psychology* 2002;**38**(6):934-47. [DOI: [10.1037/0012-1649.38.6.934](https://doi.org/10.1037/0012-1649.38.6.934)] [PMID: 2428705]

Thompson 2020

Thompson CG, von Gillern S. Video-game based instruction for vocabulary acquisition with English language learners: a Bayesian meta-analysis. *Educational Research Review* 2020;**30**:100332. [DOI: [10.1016/j.edurev.2020.100332](https://doi.org/10.1016/j.edurev.2020.100332)]

Toppelberg 2002

Toppelberg CO, Medrano L, Morgens LP, Nieto-Castañon A. Bilingual children referred for psychiatric services: associations of language disorders, language skills, and psychopathology. *Journal of the American Academy of Child & Adolescent Psychiatry* 2002;**41**(6):712-22. [DOI: [10.1097/00004583-200206000-00011](https://doi.org/10.1097/00004583-200206000-00011)] [PMID: 12049446]

Torkildsen 2022

Torkildsen JV, Bratlie SS, Kristensen JK, Gustafsson J-E, Lyster S-A, Snow CE, et al. App-based morphological training produces lasting effects on word knowledge in primary school children: a randomized controlled trial. *Journal of Educational Psychology* 2022;**114**(4):833-54. [DOI: [10.1037/edu0000688](https://doi.org/10.1037/edu0000688)]

Vadasy 2015

Vadasy PF, Sanders EA. Incremental learning of difficult words in story contexts: the role of spelling and pronouncing new vocabulary. *Reading and Writing* 2015;**28**:371-94. [DOI: [10.1007/s11145-014-9529-9](https://doi.org/10.1007/s11145-014-9529-9)]

van Kleeck 1994

van Kleeck A. Potential cultural bias in training parents as conversational partners with their children who have delays in language development. *American Journal of Speech-Language Pathology* 1994;**3**(1):67-78. [DOI: [10.1044/1058-0360.0301.67](https://doi.org/10.1044/1058-0360.0301.67)]

Vermeer 2001

Vermeer A. Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics* 2001;**22**(2):217-34. [DOI: [10.1017/S0142716401002041](https://doi.org/10.1017/S0142716401002041)]

Vermeer 2016

Vermeer HJ, van IJzendoorn MH, Cárcamo RA, Harrison LJ. Quality of child care using the Environment Rating Scales: a meta-analysis of international studies. *International Journal of Early Childhood* 2016;**48**:33-60. [DOI: [10.1007/s13158-015-0154-9](https://doi.org/10.1007/s13158-015-0154-9)]

Webb 2020

Webb S. *The Routledge Handbook of Vocabulary Studies*. London (UK): Routledge, 2020.

Wechsler 2003

Wechsler D. Wechsler Intelligence Scale for Children III. Norwegian Administration Manual. 3rd edition. Assessment (Norway): Psychological Corporation, 2003.

Wechsler 2008

Wechsler D. Wechsler Preschool and Primary Scale of Intelligence III: Norwegian Manual. 3rd edition. Stockholm (Sweden): Pearson Assessment, 2008.

Weizman 2001

Weizman ZO, Snow CE. Lexical output as related to children's vocabulary acquisition: effects of sophisticated exposure and support for meaning. *Developmental Psychology* 2001;**37**(2):265-79. [DOI: [10.1037/0012-1649.37.2.265](https://doi.org/10.1037/0012-1649.37.2.265)] [PMID: 11269394]

Wiig 2009

Wiig EH, Secord W, Semel EM. *Clinical Evaluation of Language Fundamentals: Preschool-2 Spanish*. 2nd edition. San Antonio (TX): Psychological Corporation, 2009.

Williams 2007

Williams KT. *Expressive Vocabulary Test*. 2nd edition. Minneapolis (MN): Pearson, 2007.

Wood 2018

Wood C, Fitton L, Petscher Y, Rodriguez E, Sunderman G, Lim T. The effect of e-book vocabulary instruction on Spanish-English speaking children. *Journal of Speech, Language, and Hearing Research* 2018;**61**(8):1945-69. [DOI: [10.1044/2018_JSLHR-L-17-0368](https://doi.org/10.1044/2018_JSLHR-L-17-0368)] [PMID: 30073307]

Woodcock 1991

Woodcock RW. *Woodcock Language Proficiency Battery-Revised*. Chicago (US): DLM/Riverside Publishers, 1991.

Yoder 2014

Yoder N. *Teaching the Whole Child: Instructional Practices that Support Social-Emotional Learning in Three Teacher Evaluation Frameworks*. Research-to-Practice Brief. Revised Edition. Washington (DC): Center on Great Teachers and Leaders, American Institutes for Research, 2014. [ED581718]

Yousefi 2018

Yousefi MH, Biria R. The effectiveness of L2 vocabulary instruction: a meta-analysis. *Asian-Pacific Journal of Second and Foreign Language Education* 2018;**3**(21):1-19. [DOI: [10.1186/s40862-018-0062-2](https://doi.org/10.1186/s40862-018-0062-2)]

Zeng 2012

Zeng B, Law J, Lindsay G. Characterizing optimal intervention intensity: the relationship between dosage and effect size in interventions for children with developmental speech and language difficulties. *International Journal of Speech-Language Pathology* 2012;**14**(5):471-7. [DOI: [10.3109/17549507.2012.720281](https://doi.org/10.3109/17549507.2012.720281)] [PMID: 22974106]

Zimmerman 2007

Zimmerman FJ, Christakis DA, Meltzoff AN. Associations between media viewing and language development in children

under age 2 years. *Journal of Pediatrics* 2007;**151**(4):364-8. [DOI: [10.1016/j.jpeds.2007.04.071](https://doi.org/10.1016/j.jpeds.2007.04.071)] [PMID: 17889070]

Zucker 2019

Zucker TA, Carlo MS, Landry SH, Masood-Saleem SS, Williams JM, Bhavsar V. Iterative design and pilot testing of the Developing Talkers Tiered Academic Language Curriculum for pre-kindergarten and kindergarten. *Journal of Research on Educational Effectiveness* 2019;**12**(2):274-306. [DOI: [10.1080/19345747.2018.1519623](https://doi.org/10.1080/19345747.2018.1519623)]

References to other published versions of this review

Hjetland 2021

Hjetland HN, Hofslundsengen H, Klem M, Karlsen J, Hagen ÅM, Engevik LI, et al. Vocabulary interventions for second language (L2) learners up to six years. *Cochrane Database of Systematic Reviews* 2021, Issue 9. Art. No: CD014890. [DOI: [10.1002/14651858.CD014890](https://doi.org/10.1002/14651858.CD014890)]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Baker 2022

Study characteristics

Methods	<p>Design: cluster-RCT</p> <p>Unit of randomisation: created clusters of 3–4 at-risk students that were matched by initial PPVT scores within classrooms. These clusters of students were randomly assigned to the treatment or control groups.</p>
Participants	<p>Location: USA</p> <p>Languages: L1 = Spanish; L2 = English</p> <p>Sample characteristics: bilingual children who spoke Spanish only or Spanish and English at home, and they all had some level of receptive Spanish and English vocabulary as measured by the TVIP (Dunn 1986) and the PPVT-4 (Dunn 2007).</p> <p>Students who scored below the 30th percentile on the PPVT (i.e. score 70–92) were considered at-risk for language and learning difficulties and therefore eligible to participate. However, students who scored below the 2nd percentile were excluded because the study authors did not consider they would benefit from a Tier 2 English intervention.</p> <p>This group was drawn from a larger study of multitiered vocabulary instruction conducted with 1601 students in 284 kindergarten classrooms across 3 states.</p> <p>Mean age at pretest: not reported. We contacted authors who confirmed all children were aged < 5 years and 11 months, but did not provide mean age.</p> <p>Gender: not reported</p> <p>Sample size: 351 (intervention = 179; control = 172)</p> <p>Attrition: 21 did not complete the ETW measure at post-test. This represents an attrition rate of 6.03%. There were also students who did not take the listening comprehension assessment because this measure was not administered across all years of the study in all sites. Analysis of listening comprehension included only 224 bilingual students with complete pretest and post-test kindergarten data.</p> <p>Clusters: 3 or 4 children matched by initial PPVT scores within classrooms. Information about number of classroom not included (only number of classroom in the larger study where the sample was drawn from). ICC not reported, requested from authors, but not received</p>
Interventions	<p>Intervention (179 children): supplemental Tier 2 Early Vocabulary Intervention instruction, in addition to Tier 1 vocabulary instruction. Early Vocabulary Intervention was designed to supplement</p>

Baker 2022 (Continued)

Tier 1 classroom vocabulary lessons and to incorporate elements of instruction that have been shown to support vocabulary learning.

The interventionists supported (scaffolded) the children as they used words in novel sentences, with particular attention given to helping students elaborate and extend their sentences. Children reviewed the target vocabulary and concepts taught in Tier 1 classroom instruction.

Format of delivery: groups of 3 or 4 students

Delivery agent: the interventionists implementing the tier 2 intervention were selected by the schools from their staff. Kindergarten teachers provided Tier 1 classroom vocabulary instruction.

Dosage: 30 minutes per day, 4 days per week

Duration: 22 weeks

Control (172 children): Tier 1 instruction. Whole-group L2 vocabulary instruction consisting of 24 vocabulary lessons delivered weekly for 15–20 minutes a day in a 5-day sequence. The children were introduced to 5 target words (e.g. memorable, perilous, inquire, solitude, transform) that were included in a story the teacher read aloud. The children were taught student-friendly definitions; picture cards that illustrated the word were used and anchor sentences that provided meaningful context and supported understanding was read to them. Children were given multiple opportunities to use the words and definitions in a series of oral, listening, and workbook-based activities. Each lesson began with a review of the target word definitions using an instructional routine taught to them by the research team. This provided all children with explicit practice pronouncing the vocabulary words and repeating the definition.

Outcomes

Primary outcomes

- Receptive L2 vocabulary proximal: RTW. Timing: immediate post-test and follow-up. Scale: not reported, higher score was favourable
- Expressive L2 vocabulary proximal: ETW. Timing: immediate post-test. Scale: 0–16 (number of items), higher score was favourable

Secondary outcomes

- L2 listening comprehension: listening comprehension containing target words developed by the researchers. Timing: immediate post-test and follow-up. Scale: not reported, higher score was favourable

Other measures

- Receptive L2 vocabulary distal: PPVT-4 (Dunn 2007) (middle of kindergarten, mean and SD not reported for the whole sample, not included in analyses)
- L1 receptive vocabulary distal: TVIP (Dunn 1986) (middle of kindergarten, mean and SD not reported for the whole sample, not included in analyses)

Notes

Funding: supported by grant number R324110135, titled Early Vocabulary Instruction and Intervention funded by the IES, US Department of Education to the University of Connecticut.

Conflict of interest: no potential conflicts of interest

Chen 2018
Study characteristics

Methods

Design: cluster-RCT

Unit of randomisation: classroom

Chen 2018 (Continued)

Participants	<p>Location: China</p> <p>Languages: L1 = Uyghur; L2 = Chinese</p> <p>Sample characteristics: ethnic minority children. Uyghur-speaking children in China. Recruited classes from 2 age cohorts: entry age 4 years and entry age 5 years</p> <p>Mean age at pretest: intervention = 54.0 (SD 7.8) months; control = 54.0 (SD 7.3) months</p> <p>Gender: reported as mean number of girls in each class: intervention = 22.3 (SD 6.1); control = 22.1 (SD 5.7)</p> <p>Sample size: 256 (intervention = 134; control = 122)</p> <p>Attrition: 17 children in the entry age 5 years cohort were lost in the second and third wave of data collection because of families relocating.</p> <p>Clusters: intervention = 16; control = 15</p>
Interventions	<p>Intervention (134 children): programme included talk and discussion about the picture books to scaffold and help children understand new vocabulary and concepts.</p> <p>Bilingual picture books, bilingual posters, and word cards related to the books were included.</p> <ul style="list-style-type: none"> • Picture-book-related discussions and activities in classroom (e.g. a rhyming song, a field trip, an art activity) • Teaching techniques: focusing on pre-identified target words, asking open-ended questions (guiding questions provided for each book), encouraging students to speak more, and give positive feedback <p>Pre-identified target words were included. Quote: "The key vocabulary items included abstract terms such as boundary, concrete nouns, such as animal names (e.g. beaver, wolf), and words related to social life (e.g. wall, bell, passport)" (p 210).</p> <p><i>Format of delivery:</i> small-group discussions and whole-class activity</p> <p><i>Delivery agent:</i> teachers</p> <p><i>Dosage:</i> 1.67 hours per week (100 minutes)</p> <p><i>Duration:</i> 1 year (4 × 20-minute small-group discussions and 1 whole-class activity each week, for 1 year)</p> <p>Control (122 children): standard care. Typically, there were few picture books, reading materials, or shared book reading activities.</p>
Outcomes	<p>Primary outcomes</p> <ul style="list-style-type: none"> • Receptive L2 vocabulary distal: Chinese receptive vocabulary (Chinese version of the PPVT-R) (Lu 1998). Timing: immediate post-test. Scale: 0–115, higher score was favourable • Expressive L2 vocabulary distal: Chinese expressive vocabulary (Chinese version of the EVT). Timing: immediate post-test. Scale: 0–87, higher score was favourable <p>Secondary outcomes</p> <ul style="list-style-type: none"> • L1 receptive vocabulary distal: Uyghur receptive vocabulary (Uyghur version of the PPVT). Timing: immediate post-test. Scale: 0–97, higher score was favourable • L1 expressive vocabulary distal: Uyghur expressive vocabulary (Uyghur version of the EVT). Timing: immediate post-test. Scale: 0–74, higher score was favourable
Notes	<p>Funding: grants from Ministry of Education of China (11JJD740024) and the Young Scholar of Distinction Fellowship of Ministry of Education, China (to first study author)</p>

Chen 2018 (Continued)

Conflict of interest: not reported

Goodrich 2013
Study characteristics

Methods	Design: RCT Unit of randomisation: children were randomly assigned to 1 of 3 intervention conditions.
Participants	Location: USA Languages: L1 = Spanish; L2 = English Sample characteristics: children from 10 classes in a Head Start centre in Los Angeles, California Gender: only reported for the whole sample (3 groups, 2 groups are included in the review) Mean age at pretest: intervention = 54.0 (SD 4.19) months; control = 54.4 (SD 5.6) months Sample size: 62 (intervention = 31; control = 31) Attrition: about midway through the school year, 2 children moved away and left the programme, which made the final sample 94 children (all 3 groups).
Interventions	Intervention (31 children): Literacy Express Preschool Curriculum <ul style="list-style-type: none"> • Oral language skills: dialogical reading, including open-ended questions and retelling. Simple questions at first that focused on the pictures in the book. Later, more complex questions (e.g. how pictures and other elements of the book related to each other and to other literary elements such as plot) • Phonological awareness skills: word games using pictures to help children better understand that words are made up of individual units of sound • Print knowledge: recognition of the letters in the children's names and gradually moved to introduce the names of all letters as well as the sounds that correspond to letters <p><i>Format of delivery:</i> small-group instruction</p> <p><i>Delivery agent:</i> bilingual graduate research assistants</p> <p><i>Dosage:</i> approximately 1 hour 20 minutes (1.4 times × approximately 20 minutes per week)</p> <p><i>Duration:</i> 21 weeks</p> <p>Control (31 children): standard care. Classroom curriculum (High/Scope Curriculum), which takes an approach called "active participatory learning". Children build knowledge through a learning experience that involves direct interactions with people and objects (highscope.org).</p>
Outcomes	Primary outcomes <ul style="list-style-type: none"> • Receptive L2 vocabulary distal: receptive vocabulary, a subtest from the P-CTOPPP (Lonigan 2002a). Timing: immediate post-test. Scale: 0–40, higher score was favourable • Expressive L2 vocabulary distal: definitional vocabulary, a subtest from the P-CTOPPP. Timing: immediate post-test. Scale: 0–80, higher score was favourable Secondary outcomes <ul style="list-style-type: none"> • L1 receptive vocabulary distal: receptive vocabulary from the P-CTOPPP-Spanish (Lonigan 2002b). Timing: immediate post-test. Scale: 0–40, higher score was favourable

Goodrich 2013 (Continued)

- L1 expressive vocabulary distal: definitional vocabulary from the P-CTOPPP-Spanish. Timing: immediate post-test. Scale: 0–80, higher score was favourable

Notes

Funding: grant from the National Science Foundation (REC-0128970). Preparation of work supported by grants from the National Institute of Child Health and Human Development (HD060292) and IES (R305B090021)

Conflict of interest: not reported

Grøver 2020
Study characteristics

Methods

Design: cluster-RCT

Unit of randomisation: classroom

Participants

Location: Norway

Languages: L1 = children spoke a variety of first languages (Albanian 5.2%, Arabic 9.3%, Bosnian 3.0%, Sorani Kurdish 4.7%, Polish 9.7%, Russian 1.5%, Somalian 14.0%, Tamil 7.5%, Turkish 6.0%, Urdu 20.3%, and Vietnamese 4.5%, while 14.2% of the sample reported other first languages); L2 = Norwegian

Sample characteristics: for a classroom to be included in the study, a minimum of 2 families with children identified as bilingual by their parents and with parents speaking a non-Scandinavian language had to agree to participate. Sample described as DLLs, aged 3–5 years, acquiring Norwegian as L2 in preschool and speaking a variety of L1s at home

Mean age at pretest: intervention = 53.8 (SD 9.5) months; control = 51.2 (SD 9.6) months

Gender: 230 girls and 234 boys

Sample size: 471 at randomisation reported in the supplemental material. Number reported in the article was 464 (due to attrition right after randomisation). Number of children in each group not reported (only number of classrooms, clusters). However, numbers of participants who completed the different measures were reported.

Attrition: immediately following randomisation (in online supplement, p 2): (quote) "The recruitment procedure resulted originally in 471 children in 124 classrooms in 60 preschools. Immediately following randomization, one intervention classroom with two children in a three-classroom school withdrew due to staff illness. Simultaneously, parents of five children in five different classrooms that had agreed to participate in the study informed us that their plans had changed and that they were returning to the family's country of origin (two from intervention classrooms, three from control classrooms), leaving us with the sample of 464 children in 123 classrooms."

Clusters: 98 classroom (intervention = 49; control = 49). ICC calculated at pretest sent by request.

Interventions

Intervention (not reported): sessions included dialogical reading of 15 selected books, 4–5 target words in each book were explicitly taught, with open-ended questions and retelling. Retelling from a different perspective.

Structured activities and frequent sessions. Session plans included materials with suggestions and examples.

Teachers were asked to invite play and other activities that expanded on the books' themes. 4 books were sent home and parents asked to read with their children in their home language.

Format of delivery: small-group instruction and larger groups

Grøver 2020 (Continued)

Delivery agent: preschool teacher

Dosage: 3 shared reading sessions per week

Duration: 16 weeks

Control (not reported): standard care. The control group received 1 or 2 books (7 in total) within each thematic unit that were topically linked to the unit, but not identical with the books that the intervention group received.

Outcomes
Primary outcomes

- Receptive L2 vocabulary proximal: targeted receptive vocabulary. Timing: immediate post-test. Scale: 0–46, higher score was favourable
- Receptive L2 vocabulary distal: BPVS–2 (Dunn 1997; adapted to Norwegian by Lyster 2010). Timing: immediate post-test. Scale: not reported, higher score was favourable
- Expressive L2 vocabulary proximal: targeted expressive vocabulary. Timing: immediate post-test. Scale: 0–18, higher score was favourable

Secondary outcomes

- L2 narrative skills: the MAIN (MAIN_NARRATIVE; Gagarina 2012); The Baby Goat version. Timing: immediate post-test. Scale: 0–16, higher score was favourable
- L1 receptive vocabulary proximal: first-language receptive vocabulary 43 items. Timing: immediate post-test. Scale: 0–43, higher score was favourable
- L2 grammatical knowledge: TROG–2 (Bishop 2003; adapted to Norwegian by Lyster 2009). Timing: immediate post-test. Scale: 0–12, higher score was favourable

Notes

Funding: Norwegian Research Council (grant number 218280)

Conflict of interest: not reported

Hermanns 2011
Study characteristics
Methods

Design: cluster-RCT (efficacy trial)

Unit of randomisation: classroom

Participants

Location: USA

Languages: L1 = Spanish; L2 =English

Sample characteristics: Head Start Latino ELLs

Mean age at pretest: intervention = 55.4 (SD 5.0) months; control = 54.6 (SD 5.2) months

Gender: 33 girls and 47 boys (intervention = 22 girls and 30 boys; control = 11 girls and 17 boys)

Sample size: 80 (intervention = 52; control = 28)

Attrition: not reported. All children participated in the intervention and assessments.

Clusters: intervention = 4 classrooms; control = 2 classrooms

Interventions

Intervention (52 children): children read and discussed 2 books with their teachers and were introduced to 6 target vocabulary books.

Dialogical reading techniques included:

Hermanns 2011 *(Continued)*

- prompt the child with questions;
- evaluate and expand the child's responses;
- praise the child's efforts to retell the story and name objects and actions in the book; and
- have fun!

The researcher chose 100 target words from the books and then narrowed the list to 72 words in collaboration with the teachers, based on these criteria:

- "root-word meanings that preschoolers with limited vocabularies (e.g. ELLs) were not likely to know
- words that were not abstract, i.e. they refer "to objects that can be seen, actions that can be carried out, and modifiers that can be apprehended directly (e.g. color, size, sound, etc.)"
- words that could be easily demonstrated
- words that were interesting and for which the children would be able to find uses in every-day situations" (p 17).

Format of delivery: group sessions (not specified)

Delivery agent: classroom teachers and teacher aides

Dosage: daily sessions, 4 each week, exact dose not reported

Duration: 12 weeks

Control (28 children): control classrooms read the Spanish and English versions of the same books during the same weeks, but how they normally read with the children. They did not include the additional dialogical reading techniques.

Outcomes
Primary outcomes

- Receptive L2 vocabulary proximal: ETV. Timing: immediate post-test. Scale: 0–30, higher score was favourable
- Receptive L2 vocabulary distal: ROWPVT (Brownell 2000). Timing: immediate post-test. Scale: not reported, higher score was favourable

Secondary outcomes

- L1 receptive vocabulary proximal: SPTV assessment. Timing: immediate post-test. Scale: not reported, higher score was favourable
- L1 receptive vocabulary distal: ROWPVT-SP (Brownell 2001). Timing: immediate post-test. Scale: not reported, higher score was favourable

Notes

Mean and SD not reported in dissertation. Author did not respond to the data request. Study not included in the meta-analyses.

Funding: (quote) "My dissertation study was funded in part by a Dissertation Research Fellowship from The Achievement Gap Initiative at Harvard University, and a Head Start Graduate Student Research grant from the Department of Health & Human Services, Administration for Children and Families (award #90YR0001)" (p ii).

Conflict of interest: not reported

Pollard-Durodola 2016
Study characteristics
Methods

Design: RCT

Pollard-Durodola 2016 (Continued)

Unit of randomisation: classroom and individual level

Participants	<p>Location: USA</p> <p>Languages: L1 = Spanish; L2 = English</p> <p>Sample characteristics: Spanish-speaking preschool children who were learning English as a L2/ Spanish-speaking DLLs. Children were screened using the English Language Proficiency Assessment for Early Learners (preLAS; DeAvila 2000), and were selected for the study based on their scores at the prefunctional and beginning level of English proficiency.</p> <p>Mean age at pretest: intervention = 56.9 (SD 3.7) months; control = 56.1 (SD 5.6) months</p> <p>Gender: 126 girls and 126 boys (intervention = 69 girls and 69 boys; control = 57 girls and 57 boys)</p> <p>Sample size: 252 (intervention = 126; control = 126)</p> <p>Attrition: numbers of participants were reported for each outcome and time point. 3 children did not complete PPVT-4 (Dunn 2007) assessments and 14 children did not complete EVT-2 (Williams 2007) assessments.</p> <p>Clusters: 48 teachers/classrooms at randomisation, reduced to 42 before intervention (intervention = 23; control = 19). ICC data reported in the article.</p>
Interventions	<p>Intervention (126 children): Words of Oral Reading and Language Development (WORLD) interactive shared book reading approach. After book reading:</p> <ul style="list-style-type: none"> • review taught concepts by identifying the picture/concept cards; • discussion: encouraged children to relate the concepts to their own lived experiences; and • comprehension questions, where the children used their new vocabulary and acquired knowledge and talked about the characters and the story. <p>Discussions were framed within the context of the science or social studies topic and theme. The children were encouraged to use their new vocabulary at home. Lessons were organised to prime background knowledge.</p> <p><i>Format of delivery:</i> small-group instruction</p> <p><i>Delivery agent:</i> preschool teachers</p> <p><i>Dosage:</i> 1.67 hours a week (100 minutes)</p> <p><i>Duration:</i> 18 weeks</p> <p>Control (126 children): comparison teachers engaged in standard care shared book reading activities for approximately 20 minutes with a small group of children.</p>
Outcomes	<p>Primary outcomes</p> <ul style="list-style-type: none"> • Receptive L2 vocabulary proximal: RDRPVT. Timing: immediate post-test. Scale: 0–18, higher score was favourable • Receptive L2 vocabulary distal: PPVT-4. Timing: immediate post-test. Scale: 0–228, higher score was favourable • Expressive L2 vocabulary proximal: RDEPVT. Timing: immediate post-test. Scale: 0–36, higher score was favourable • Expressive L2 vocabulary distal: EVT-2. Timing: immediate post-test. Scale: not reported, higher score was favourable
Notes	<p>Funding: supported in part by Project WORLD, Grant R305A110638, Early Learning and Policies Research, U.S. Department of Education, Institute of Education Services</p> <p>Conflict of interest: not reported</p>

Pérez 2019

Study characteristics

Methods	<p>Design: randomised block design</p> <p>Unit of randomisation: individual</p>
Participants	<p>Location: USA</p> <p>Languages: L1 = Spanish; L2 = English</p> <p>Sample characteristics: English learners/Spanish-speaking preschool children. Children met the following inclusion criteria: passed the Spanish speech, language, and hearing screening administered by a Spanish-English bilingual speech-language pathologist; scored within normal limits (standard score ≥ 85) on the EOWPVT-4: Spanish-Bilingual Edition (Martin 2013); and scored within normal limits (standard score ≥ 90) on the PTONI (Ehrler 2008).</p> <p>Mean age at pretest: reported for the whole sample (3 groups): 55.2 (SD 6.6) months</p> <p>Gender: reported for the whole sample (3 groups): 17 girls and 26 boys</p> <p>Sample size: 29 (intervention = 14; control = 15)</p> <p>Attrition: not reported</p>
Interventions	<p>Intervention (English-only instruction group; 14 children): a book-reading context. 24 English words were targeted (4 new words per week).</p> <p>The sessions included the following 4 steps.</p> <ul style="list-style-type: none"> • Introduction • Vocabulary explanation • Book reading: show paper photograph and say word, say definition, say and repeat each word 3 times in unison, show paper photograph and activate word, activate definition, activate word and repeat 3 times in unison • Individual activities (labelling, definition, naming, repetition) <p>The target words were presented by the clinicians in English. Quote: "The books used in the instruction were of narrative genre, appropriate for the age range, with colourful illustrations on each page, and the target words appeared in the text and/or images in the book. One book was selected for each week of instruction" (p 6).</p> <p>"Target words were selected by the research team based on the following criteria.</p> <ul style="list-style-type: none"> • Words occurred in the book used for instruction (image or text). • Words are Tier 2 vocabulary for preschoolers 3 to 5 years of age. • Words have translation equivalents in Spanish and English. • Words can be represented with a picture." <p>"Four target words (two nouns and two verbs) were selected per week/book for the English instruction" (p 5).</p> <p><i>Format of delivery:</i> small-group and individual instruction (3 or 4 children).</p> <p><i>Delivery agent:</i> 2 monolingual, English-speaking, first-year graduate student clinicians</p> <p><i>Dosage:</i> 1.17 hours per week (e.g. 2 \times 35 minutes = 70 minutes per week. Children received vocabulary instruction during 35-minute sessions twice per week over the same 6-week period).</p> <p><i>Duration:</i> 6 weeks</p>

Pérez 2019 (Continued)

	Control (15 children): received no instruction
Outcomes	<p>Primary outcomes</p> <ul style="list-style-type: none"> Expressive L2 vocabulary proximal: English naming. Timing: immediate post-test and follow-up. Scale: 0–24, higher score was favourable <p>Secondary outcomes</p> <ul style="list-style-type: none"> L1 expressive vocabulary proximal: Spanish naming. Timing: immediate post-test and follow-up. Scale: 0–24, higher score was favourable <p>Other outcomes (not included)</p> <ul style="list-style-type: none"> Expressive L2 vocabulary proximal: English definition. Timing: immediate post-test and follow-up. Scale: 0–24, higher score was favourable L1 expressive vocabulary proximal: Spanish definition. Timing: immediate post-test and follow-up. Scale: 0–24, higher score was favourable
Notes	<p>Funding: no financial support</p> <p>Conflict of interest: none</p>

Rogde 2016

	Study characteristics
Methods	<p>Design: RCT</p> <p>Unit of randomisation: individual</p>
Participants	<p>Location: Norway</p> <p>Languages: L1 = varied (languages not reported; different mother tongues from all over the world); L2 = Norwegian</p> <p>Sample characteristics: L2 learners. Eligible participants were children for whom Norwegian was an L2, born in 2006 and attended kindergartens in 2 suburban municipalities in Norway. To be included as a participant in the study, both parents had to have an L1 other than Norwegian.</p> <p>Mean age at pretest: intervention = 66.3 (SD 3.3) months; control = 66.1 (SD 3.7) months</p> <p>Gender: 47 girls and 68 boys (intervention = 24 girls and 34 boys; control = 23 girls and 34 boys)</p> <p>Sample size: 115 (intervention = 58; control = 57)</p> <p>Attrition: lost 2 participants (1 from each group) due to relocation at post-test 1</p>
Interventions	<p>Intervention (58 children): programme based on dialogical storybook reading: language teaching tasks included classification of words and concepts, listening activities, exercises related to grammatical knowledge, story structuring, and story sequencing.</p> <p>The intervention included a broad scope of activities, including training in vocabulary, grammar, and narratives, particularly focusing on expressive vocabulary.</p> <p>3 or 4 words were selected from each book and directly taught to the children. The selection of these words was based on the concept of Tier-2 words that were domain-general, sophisticated labels for concepts with which young learners were already familiar. These were more abstract words that children will not easily learn by themselves in a general kindergarten setting. These words were also selected on the basis of school textbooks.</p>

Rogde 2016 (Continued)

Format of delivery: 2 small-group sessions (40–45 minutes and 1 individual session of 10 minutes) each week

Delivery agent: preschool teacher

Dosage: 1.67 hours/1 hour + 40 minutes (e.g. 100 minutes a week)

Duration: 18 weeks

Control (57 children): business-as-usual in which the children received the ordinary kindergarten programme for L2 learners. Regular practice was based on reading and language activities but were less structured and in larger groups than the intervention programme.

Outcomes

Primary outcomes

- Receptive L2 vocabulary distal: BPVS-2 (Dunn 1997). The Norwegian version of this test has been translated and adapted for Norwegian conditions (Lyster 2010). Timing: immediate post-test and follow-up. Scale: not reported, higher score was favourable
- Expressive L2 vocabulary proximal: taught vocabulary (word definition). Timing: immediate post-test and follow-up. Scale: 0–84, higher score was favourable
- Expressive L2 vocabulary distal: vocabulary subtests of the WPPSI-III (Wechsler 2008) and the WISC-III (Wechsler 2003). Timing: immediate post-test and follow-up. Scale: not reported, higher score was favourable

Secondary outcomes

- L2 narrative skills: the Renfrew Bus Story test (Renfrew 1997). Timing: immediate post-test and follow-up. Scale not reported, higher score was favourable
- L2 grammatical knowledge: the Norwegian version of the TROG-2 (Bishop 2003; translated and adapted to Norwegian conditions by Lyster 2009). Timing: immediate post-test and follow-up. Scale: not reported, higher score was favourable

Notes

Funding: Research Council of Norway, Utdanning 2020, Grant No.: 203335

Conflict of interest: not reported

Spencer 2020

Study characteristics

Methods

Design: cluster-randomised group study

Unit of randomisation: classroom

Participants

Location: USA

Languages: L1 = Spanish; L2 = English

Sample characteristics: participants with scores within age expectations for English on screening measure of listening retell and English EV subtest (CELF-P) were not included. Thus, children who displayed low English skills and low, moderate, or high Spanish language skills were included as participants.

Mean age at pretest: intervention = 50 (range 39–59) months; control = 49 (range 37–59) months

Gender: 46 girls and 29 boys (intervention = 25 girls and 16 boys; control = 21 girls and 13 boys). Authors reported of incomplete demographic survey data

Sample size: 81 (intervention = 43; control = 38 children)

Spencer 2020 (Continued)

Attrition: shortly after pretest, 2 children from the control group moved away. < 1% missing scores overall

Clusters: 25 classrooms (12 in intervention and 13 in control group)

Interventions

Intervention (43 children): Tier 2 intervention conducted in HeadStart preschool classrooms. It focused on the Puente de Cuentos (Bridge Made of Stories) curriculum. The curriculum comprised 72 stories (36 English; 36 Spanish) delivered over 3 units of assessment, each lasting 8–10 weeks and including 12 stories.

There were 2 target vocabulary words (adjectives or verbs) for each story. In addition, subordinating and co-ordinating conjunctions were introduced, and each story contained additional low frequency nouns and academic concept words.

The intervention included large-group, scripted lessons in English in which the teacher read the story to the class, and then engaged the children in activities that required group or individual responses and retell attempts. In classrooms in which the teacher or teaching assistant spoke Spanish, small-group activities with selected children were conducted in Spanish, and then again in English. Teachers followed a set of principles for explicit teaching that included modelling, leading, supportive prompts, and corrections.

Each story included a package of additional materials and extension exercises. These included 5 illustrations, wordless picture books, and story schema icons to scaffold retells, story games, and objects related to target words and icons. Suggested activities included incorporating target vocabulary into circle time and playground games, and a list of additional storybooks that contained target vocabulary.

Parents were provided with an overview of the project and suggested activities to engage children in the stories and target words in Spanish.

Format of delivery: small-group lessons

Delivery agent: Head Start (preschool) teachers together with teaching assistant

Dosage: lessons 4 days a week

Duration: approximately 24–30 weeks (3 units of instruction (Units A–C), with each unit lasting 8–10 weeks)

Control (38 children): 'business as usual' condition. Centre directors reported that teachers used small-group instruction to differentiate individual students but most consistently delivered instruction in large groups. Instruction was completed in English with occasional directions or explanations in Spanish.

Outcomes
Primary outcomes

- Receptive L2 vocabulary proximal: Puente de Cuentos Picture Vocabulary Assessment (receptive picture vocabulary assessment) (English unit A). Timing: immediate post-test. Scale: not reported, higher score was favourable
- Expressive L2 vocabulary distal: expressive vocabulary (CELF-P, English, [Semel 2004](#)). Timing: immediate post-test. Scale: not reported, higher score was favourable

Secondary outcomes

- L2 listening comprehension: ASC ([Spencer 2019](#)). Timing: immediate post-test. Scale: 0–17, higher score was favourable
- L2 narrative skills: 8 NLM-Listening (a subtest of the Decoding, Language, Reading (CUBED) Assessment; [Petersen 2016](#)). Similar to the stories directly taught (proximal). Timing: immediate post-test. Scale: not reported, higher score was favourable
- L1 receptive vocabulary proximal: Puente de Cuentos Picture Vocabulary Assessment (receptive picture vocabulary assessment) (Spanish unit A). Timing: immediate post-test. Scale: not reported, higher score was favourable

Spencer 2020 (Continued)

- L1 expressive vocabulary distal: expressive vocabulary (CELF-Preschool, Spanish, [Wiig 2009](#)). Timing: immediate post-test. Scale: not reported, higher score was favourable
- L2 grammatical knowledge: sentence structure (receptive subtest, CELF-P, [Semel 2004](#)). Timing: immediate post-test. Scale: not reported, higher score favourable

Notes

Funding: IES, U.S. Department of Education, through Grant R305A140093 for US dollars 1,481,960

Conflict of interest: not reported

Thordardottir 2015
Study characteristics

Methods

Design: RCT

Unit of randomisation: individual

Participants

Location: Montréal, Québec, Canada

Languages: L1 = varied (included Arabic (1 participant), Bangla (1 participant), Bengali (3 participants), Chinese (2 participants), Dutch (1 participant), English (1 participant), Japanese (1 participant), Kabyl (1 participant), Punjabi/Urdu (8 participants), Russian (2 participants), Sinhalese (1 participant), Spanish (4 participants), and Tamil (3 participants)). L2 = French

Sample characteristics: in addition to speaking a minority native language, the children needed to have PLI, previously clinically identified by a certified speech–language pathologist. The term PLI referred to the population of individuals that was until recently referred to as specific language impairment.

As part of the L1 criterion, participating children also needed to have had significant bilingual exposure. At a minimum, participants had to have had ≥ 6 months of regular L2 exposure (e.g. in a day-care setting). For diagnostic status (i.e. PLI) to be confirmed, children had to score at or below -1.5 SD of the mean on at least 1 measure, as per normative data for monolingual children.

Mean age at pretest: intervention = 61.6 (SD 4.7) months; control = 58.6 (SD 7.8) months

Gender: 3 girls and 26 boys (reported only for the whole sample, 3 groups. 2 groups included in the review)

Sample size: 20 (intervention = 11; control = 9)

Attrition: 5 children who qualified for the intervention phase withdrew from the study. 3 of these children had started treatment. 1 of these 3 was included in the monolingual treatment (the intervention group included in this review).

Interventions

Intervention (11 children): intervention comprised 16 individual sessions targeting vocabulary and syntactic skills in French only or bilingually, conducted by qualified SLPs and involving parent collaboration during the clinical sessions.

Target vocabulary items were identified according to the child's profile and drawn from the McArthur-Bates CDI (French) and 10 items were targeted in each therapy session: 4 verbs and 6 nouns. 5 of the target words were words that the child was reported to understand but not produce, the remaining targets were unfamiliar to the child. 20 minutes of each therapy session were devoted to vocabulary and included games and activities that allowed modelling of target words in meaningful contexts. Activities emphasised repetition and positive reinforcement and responding to children's production attempts.

Syntactic targets varied but generally included production of subject–verb–object sentence structures. As before, games (e.g. bingo) and manipulation of toys were used to model and elicit the target structure.

Thordardottir 2015 (Continued)

Parents were present at all sessions. In the bilingual sessions, parents were asked to model target words and sentence structures in the L1. Study authors reported that maintaining parent engagement in therapy activities was challenging.

Format of delivery: individual sessions with parents

Delivery agent: 2 certified and experienced speech–language pathologists

Dosage: 0.83 hours/50 minutes (weekly sessions each lasting 50 minutes. Vocabulary for 20 minutes and syntactic for 30 minutes)

Duration: 16 weeks

Control (9 children): received no clinical language intervention but were offered the treatment at the end of the control period.

Outcomes	<p>Primary outcomes</p> <ul style="list-style-type: none"> • Receptive L2 vocabulary distal: EVIP (Dunn 1993). Timing: immediate post-test. Scale: not reported, higher score was favourable • Expressive L2 vocabulary distal: EOWPVT (Gardner 1983). Timing: immediate post-test. Scale: not reported, higher score was favourable • Mean length of utterance in words and morphemes. Timing: immediate post-test and follow-up. Scale: not reported, higher score was favourable
Notes	<p>Funding: research grant from The Canadian Language and Literacy Research Network (CLLRNet 27061801), awarded to principal investigator and 2 collaborators, with the Montréal Children's Hospital and the Jewish Rehabilitation Hospital in Laval as partners.</p> <p>Conflict of interest: not reported</p>

Tong 2008

Study characteristics

Methods	<p>Design: cluster-RCT</p> <p>Unit of randomisation: school (quasi-experiment at student level)</p>
Participants	<p>Location: USA</p> <p>Languages: L1 = Spanish; L2 = English</p> <p>Sample characteristics: identified as limited English proficient and had home-language surveys indicating that Spanish was the primary language spoken at home</p> <p>Mean age at pretest: overall sample (4 groups) = 67.1 (SD 4.9) months</p> <p>Gender: not reported</p> <p>Sample size: 213 (intervention = 88; control = 125)</p> <p>Attrition: not reported</p> <p>Clusters: schools; 4 clusters (intervention = 1; control = 3). In addition, there are 2 groups not included in this review.</p>
Interventions	<p>Intervention (88 children): structured as a tier intervention</p> <p>Tier I was the regular instruction given to all students in English.</p>

Tong 2008 (Continued)

Tier II was English as an L2 language intervention that included 1. daily sessions in the Santillana Intensive English programme, a research-based curriculum in teaching Spanish speakers content areas, including maths, science, and social studies in English; 2. storytelling and retelling with higher-order thinking skills for English-language and English-literacy acquisition, 3. teacher-conducted AOL activity using Question of the Day.

Tier III of the instructional intervention was provided for the very lowest performing students receiving the Santillana Intensive English programme. This instruction was composed of communication games.

Format of delivery: individual, in pairs or small groups

Delivery agent: teachers

Dosage: 75 minutes in kindergarten and 90 minutes in first grade

Duration: 12 + 12 (24 months = 2 school years)

Control (125 children): typical practice of SEI for kindergarten and first grade in the school district. All children were taught in English.

Outcomes	<p>Primary outcomes</p> <ul style="list-style-type: none"> Expressive L2 vocabulary distal: Picture Vocabulary subtest of the Woodcock Language Proficiency Battery–Revised (Woodcock 1991). Timing: immediate post-test. Scale: not reported, higher score was favourable <p>Secondary outcomes</p> <ul style="list-style-type: none"> L2 listening comprehension: Listening Comprehension subtest of the Woodcock Language Proficiency Battery–Revised. Timing: immediate post-test. Scale: not reported, higher score was favourable
Notes	<p>Funding: supported by Grant R305P030032, funded by the IES</p> <p>Conflict of interest: not reported</p>

Verhallen 2006
Study characteristics

Methods	<p>Design: RCT</p> <p>Unit of randomisation: individual</p>
Participants	<p>Location: the Netherlands</p> <p>Languages: L1 = Turkish, Moroccan-Arabic, or Berber; L2 = Dutch</p> <p>Sample characteristics: no special language impairments or special educational needs. Scored at the level of the lowest 50% on a language test standardised for Dutch kindergarten children, had non-verbal intelligence in the normal range, and were unfamiliar with the focal story.</p> <p>Mean age at pretest: intervention = 67.6 (SD 2.0) months; control = 67.5 (SD 2.8) months</p> <p>Gender: 10 girls and 10 boys (intervention = 5 girls and 5 boys; control = 5 girls and 5 boys)</p> <p>Sample size: 20 (intervention = 10; control = 10)</p> <p>Attrition: not reported</p>

Verhallen 2006 (Continued)

Interventions

Intervention (10 children): children heard the story about Winnie the Witch (Korky 1996), but there were differences in format (static images vs multimedia) and frequency of story encounters (once vs 4 times). The intervention included in this review is the one where they heard the story 4 times. Both versions, the multimedia and static, had an identical text, were told in the same voice, and both were presented on a computer screen. The multimedia stimulus included action video, music, and sounds that were related to main aspects of the story such as characters, location, time, problem, goal, story events, resolution, and theme.

Format of delivery: individual

Delivery agent: experimenter

Duration: 4 sessions

Dosage: 24 minutes (the 4 treatment sessions took, on average, about 6 minutes.)

Control (10 children): pre- and post-test only

Outcomes

Primary outcomes

- Expressive L2 vocabulary proximal: expressive vocabulary was pre- and post-tested with 42 content words from the Winnie the Witch story. Timing: immediate post-test. Scale: not reported (40 words), higher score was favourable

Secondary outcomes

- L2 narrative skills: story comprehension. Retelling of Winnie the Witch and Peace at last (Murphy 1997) (combination of proximal and distal). Timing: immediate post-test. Scale: not reported, higher score was favourable
- L2 grammatical knowledge: syntax was pre- and post-tested by having children repeat a selection of sentences from the text of Winnie the Witch (proximal). Timing: immediate post-test. Scale: not reported, higher score was favourable

Notes

Funding: grant support from the Dutch Reading Association (Stiching Lezen)

Conflict of interest: not reported

AOL: Academic Oral Language; ASC: Assessment of Story Comprehension; BPVS-2: British Picture Vocabulary Scale, Second Edition; CELF-P: Clinical Evaluations of Language Fundamentals/-Preschool; DLL: dual language learner; ELL: English language learner; EOWPVT/-4: Spanish-Bilingual: Expressive One-Word Picture Vocabulary Test/-Fourth Edition: Spanish-Bilingual Edition; ETW: Expressive measure of Target Words; EV: expressive vocabulary; EVIP: Échelle de Vocabulaire en Images Peabody; EVT/-2: Expressive Vocabulary Test/-Second Edition; ICC: intracluster correlation coefficient; IES: Institute of Education Sciences; L1: first language; L2: second language; McArthur-Bates CDI: McArthur-Bates Communicative Development Inventories; NLM: Narrative Language Measures; P-CTOPPP/-Spanish: Preschool Comprehensive Test of Phonological and Print Processing/-Spanish version; PLI: primary language impairment; PPVT-4/R: Peabody Picture Vocabulary Test, Fourth Edition/Revised; PTONI: Primary Test of Nonverbal Intelligence; RCT: randomised controlled trial; RDEPVT: Researcher-Developed Expressive Picture Vocabulary Test; RDRPVT: Researcher-Developed Receptive Picture Vocabulary Test; ROWPVT/-SP: Receptive One-Word Picture Vocabulary Test/-Spanish Bilingual Edition; RTW: Receptive measure of Target Words; SD: standard deviation; SEI: structured English immersion; SLP: speech-language pathologists; SPTV: Spanish Target Vocabulary; TROG-2: Norwegian version of the Test for Reception of Grammar, Version 2; TVIP: Test de Vocabulario en Imágenes; WISC-III: Wechsler Intelligence Scale for Children, Third Edition; WPPSI-III: Wechsler Preschool and Primary Scale of Intelligence, Third Edition.

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Gagarina 2018	Wrong study design (not an RCT)
Park 2019	Wrong study design (not an RCT)

























Study	Reason for exclusion
Simon-Cereijido 2014	Wrong intervention (not targeting second language)
Wood 2018	Wrong population (children aged > 5 years and 11 months; mean 72 (SD 9.8) months)

RCT: randomised controlled trial; SD: standard deviation.

























RISK OF BIAS

Legend:  Low risk of bias  High risk of bias  Some concerns

Risk of bias for analysis 1.1 Receptive L2 proximal vocabulary

Study	Bias					Overall
	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported results	
Grøver 2020						
Pollard-Durodola 2016						
Spencer 2020						
Baker 2022						

Risk of bias for analysis 1.3 Receptive L2 distal vocabulary

Study	Bias					Overall
	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported results	
Chen 2018						
Goodrich 2013						
Grøver 2020						
Pollard-Durodola 2016						

Study	Bias					Overall
	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported results	
Rogde 2016						
Thordardottir 2015						

Risk of bias for analysis 1.5 Expressive L2 proximal vocabulary

Study	Bias					Overall
	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported results	
Baker 2022						
Grøver 2020						
Pérez 2019						
Pollard-Durodola 2016						
Rogde 2016						
Verhallen 2006						

Risk of bias for analysis 1.7 Expressive L2 distal vocabulary

Study	Bias					Overall
	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported results	
Chen 2018						
Goodrich 2013						
Pollard-Durodola 2016						

Bias						
Study	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported results	Overall
Rogde 2016	⚠	✓	✓	✓	⚠	⚠
Spencer 2020	⚠	⚠	✓	✓	⚠	⚠
Thordardottir 2015	✓	✓	✓	✓	⚠	⚠
Tong 2008	⚠	✓	✓	⚠	⚠	⚠

Risk of bias for analysis 1.9 L2 listening comprehension

Bias						
Study	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported results	Overall
Spencer 2020	⚠	⚠	✓	✓	⚠	⚠
Tong 2008	⚠	✓	✓	⚠	⚠	⚠

Risk of bias for analysis 1.10 L2 narrative skills

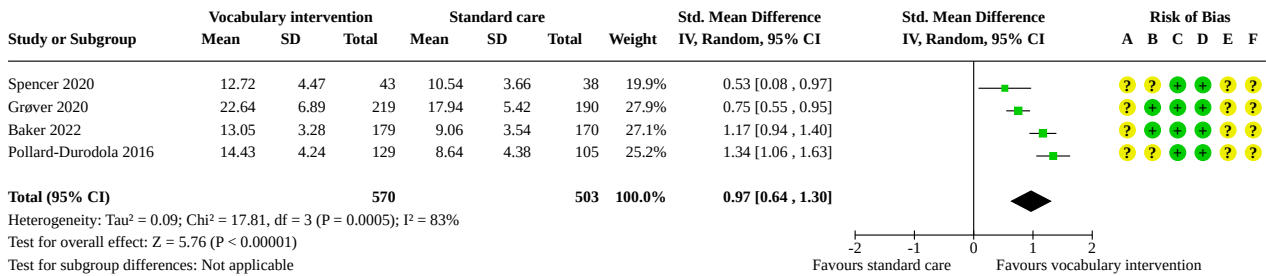
Bias						
Study	Randomisation process	Deviations from intended interventions	Missing outcome data	Measurement of the outcome	Selection of the reported results	Overall
Grøver 2020	⚠	✓	✓	✓	⚠	⚠
Rogde 2016	⚠	✓	✓	✓	⚠	⚠

DATA AND ANALYSES

Comparison 1. Vocabulary interventions compared with standard care for second language (L2) learners

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.1 Receptive L2 proximal vocabulary	4	1073	Std. Mean Difference (IV, Random, 95% CI)	0.97 [0.64, 1.30]
1.2 Receptive L2 proximal vocabulary at follow-up	1	351	Mean Difference (IV, Random, 95% CI)	2.51 [1.88, 3.14]
1.3 Receptive L2 distal vocabulary	6	1074	Std. Mean Difference (IV, Random, 95% CI)	0.29 [0.02, 0.55]
1.4 Receptive L2 distal vocabulary at follow-up	1	113	Mean Difference (IV, Random, 95% CI)	4.05 [-1.54, 9.64]
1.5 Expressive L2 proximal vocabulary	6	1121	Std. Mean Difference (IV, Random, 95% CI)	0.86 [0.56, 1.17]
1.6 Expressive L2 proximal vocabulary at follow-up	3	472	Std. Mean Difference (IV, Random, 95% CI)	0.74 [0.38, 1.11]
1.7 Expressive L2 distal vocabulary	7	960	Std. Mean Difference (IV, Random, 95% CI)	0.10 [-0.02, 0.23]
1.8 Expressive L2 distal vocabulary at follow-up	1	113	Mean Difference (IV, Random, 95% CI)	1.57 [-1.33, 4.47]
1.9 L2 listening comprehension	2	294	Std. Mean Difference (IV, Random, 95% CI)	0.19 [-0.31, 0.68]
1.10 L2 narrative skills	2	487	Std. Mean Difference (IV, Random, 95% CI)	0.37 [0.14, 0.59]
1.11 L2 narrative skills at follow-up	1	113	Mean Difference (IV, Random, 95% CI)	2.14 [-0.69, 4.97]
1.12 L1 receptive proximal vocabulary	2	373	Std. Mean Difference (IV, Random, 95% CI)	0.53 [-0.15, 1.21]
1.13 L1 receptive distal vocabulary	2	301	Std. Mean Difference (IV, Random, 95% CI)	0.25 [0.02, 0.48]
1.14 L1 expressive proximal vocabulary	1	29	Mean Difference (IV, Random, 95% CI)	0.47 [-1.19, 2.13]
1.15 L1 expressive proximal vocabulary at follow-up	1	29	Mean Difference (IV, Random, 95% CI)	0.68 [-0.98, 2.34]
1.16 L1 expressive distal vocabulary	3	382	Std. Mean Difference (IV, Random, 95% CI)	0.13 [-0.07, 0.34]
1.17 L2 grammatical knowledge	3	601	Std. Mean Difference (IV, Random, 95% CI)	0.31 [0.13, 0.49]
1.18 L2 grammatical knowledge at follow-up	1	113	Mean Difference (IV, Random, 95% CI)	1.46 [-4.04, 6.96]

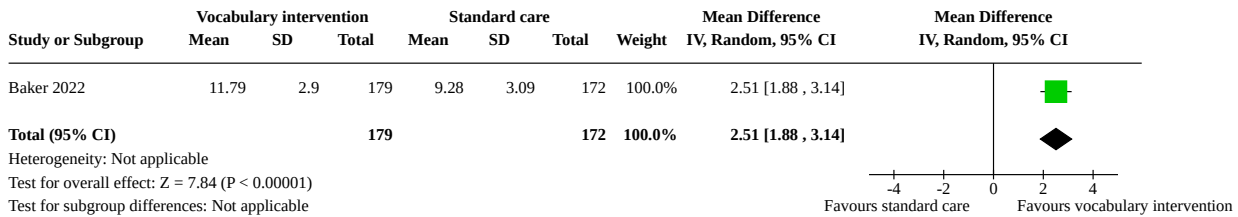
Analysis 1.1. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 1: Receptive L2 proximal vocabulary



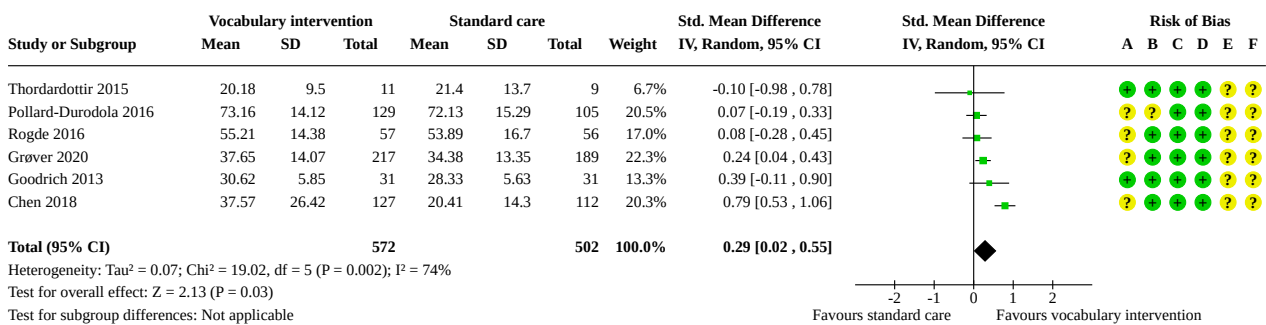
Risk of bias legend

- (A) Bias arising from the randomization process
- (B) Bias due to deviations from intended interventions
- (C) Bias due to missing outcome data
- (D) Bias in measurement of the outcome
- (E) Bias in selection of the reported result
- (F) Overall bias

Analysis 1.2. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 2: Receptive L2 proximal vocabulary at follow-up



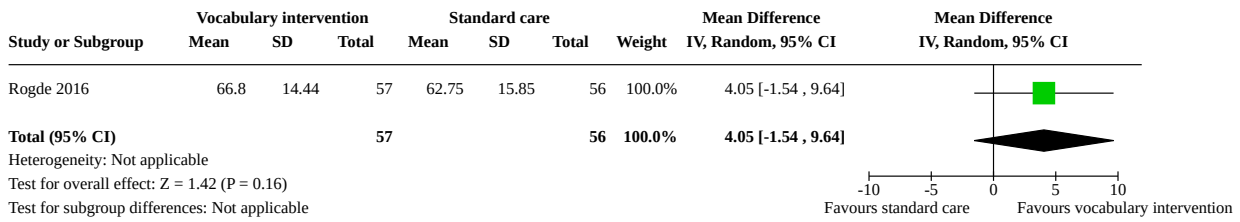
Analysis 1.3. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 3: Receptive L2 distal vocabulary



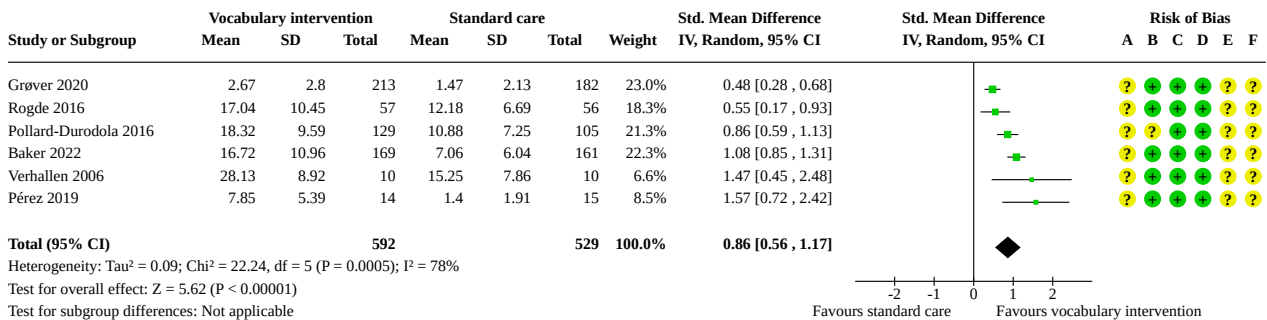
Risk of bias legend

- (A) Bias arising from the randomization process
- (B) Bias due to deviations from intended interventions
- (C) Bias due to missing outcome data
- (D) Bias in measurement of the outcome
- (E) Bias in selection of the reported result
- (F) Overall bias

Analysis 1.4. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 4: Receptive L2 distal vocabulary at follow-up



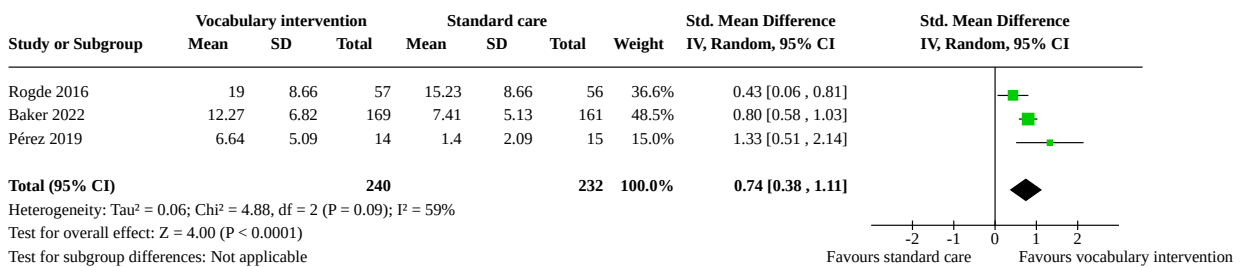
Analysis 1.5. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 5: Expressive L2 proximal vocabulary



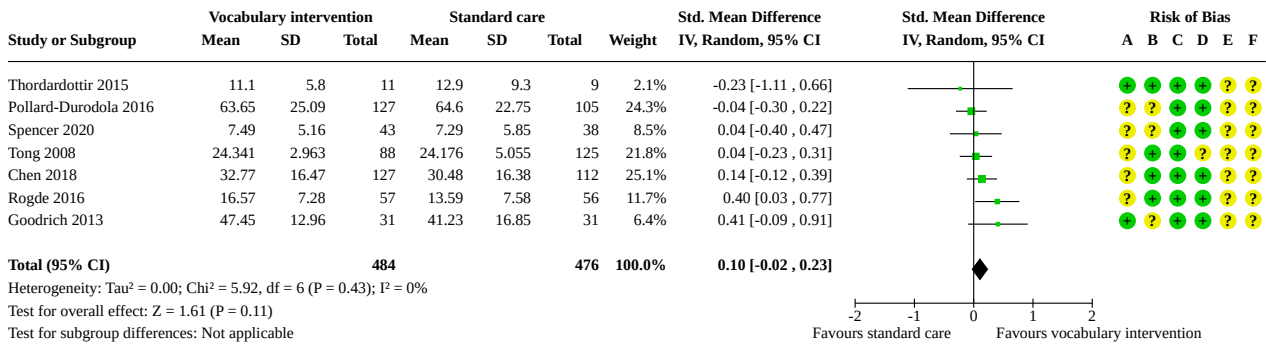
Risk of bias legend

- (A) Bias arising from the randomization process
- (B) Bias due to deviations from intended interventions
- (C) Bias due to missing outcome data
- (D) Bias in measurement of the outcome
- (E) Bias in selection of the reported result
- (F) Overall bias

Analysis 1.6. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 6: Expressive L2 proximal vocabulary at follow-up



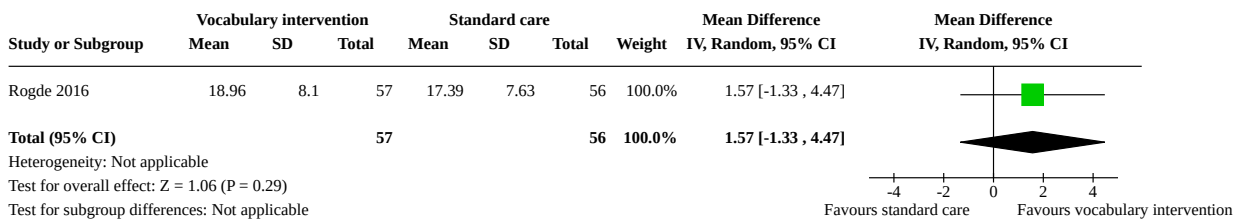
Analysis 1.7. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 7: Expressive L2 distal vocabulary



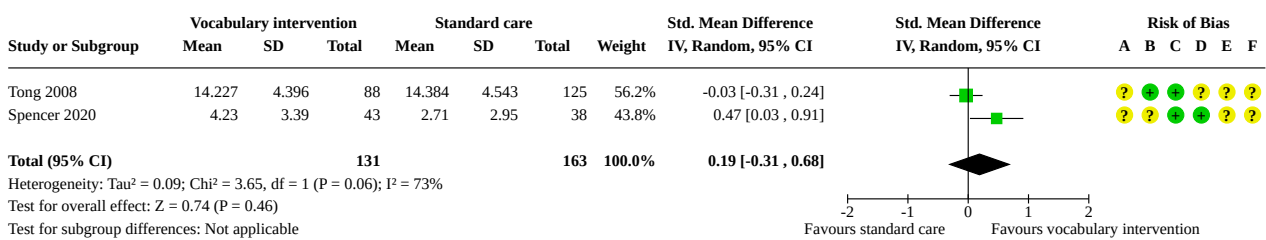
Risk of bias legend

- (A) Bias arising from the randomization process
- (B) Bias due to deviations from intended interventions
- (C) Bias due to missing outcome data
- (D) Bias in measurement of the outcome
- (E) Bias in selection of the reported result
- (F) Overall bias

Analysis 1.8. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 8: Expressive L2 distal vocabulary at follow-up



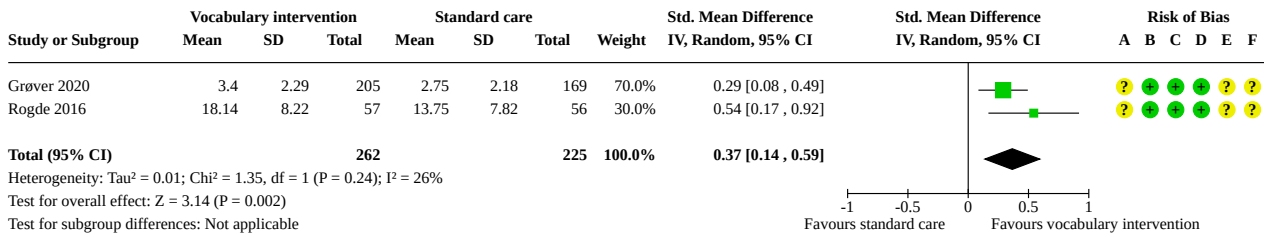
Analysis 1.9. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 9: L2 listening comprehension



Risk of bias legend

- (A) Bias arising from the randomization process
- (B) Bias due to deviations from intended interventions
- (C) Bias due to missing outcome data
- (D) Bias in measurement of the outcome
- (E) Bias in selection of the reported result
- (F) Overall bias

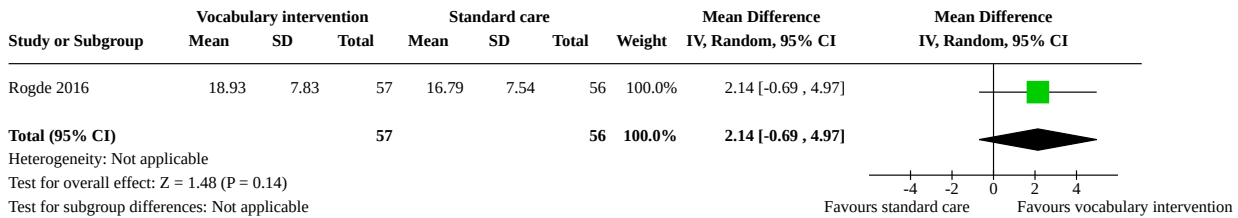
Analysis 1.10. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 10: L2 narrative skills



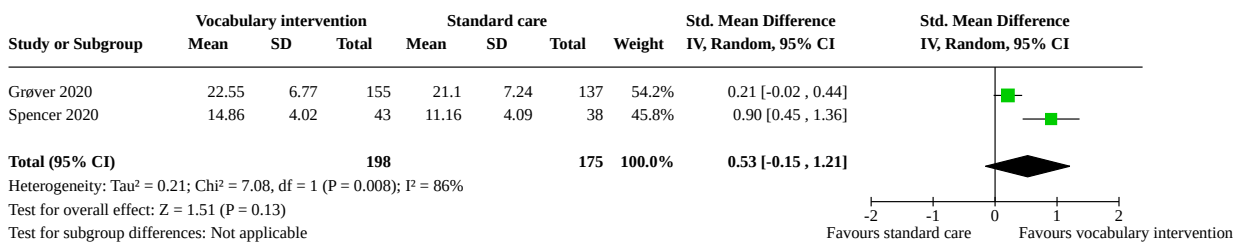
Risk of bias legend

- (A) Bias arising from the randomization process
- (B) Bias due to deviations from intended interventions
- (C) Bias due to missing outcome data
- (D) Bias in measurement of the outcome
- (E) Bias in selection of the reported result
- (F) Overall bias

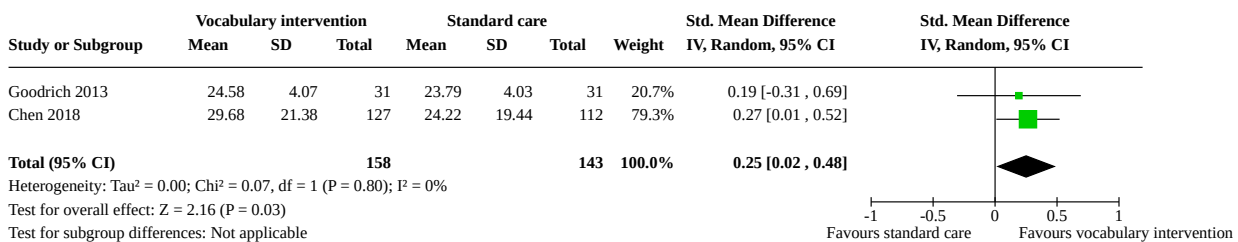
Analysis 1.11. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 11: L2 narrative skills at follow-up



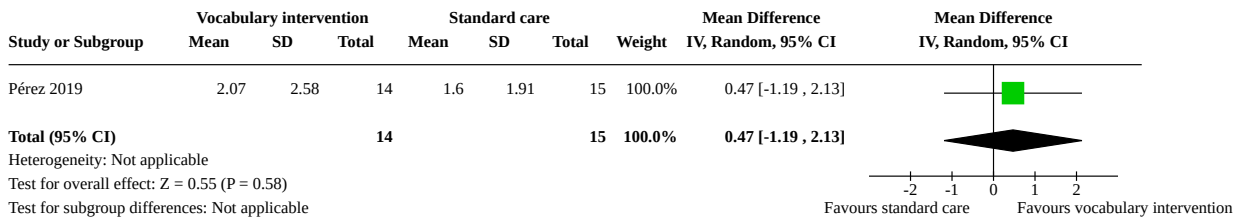
Analysis 1.12. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 12: L1 receptive proximal vocabulary



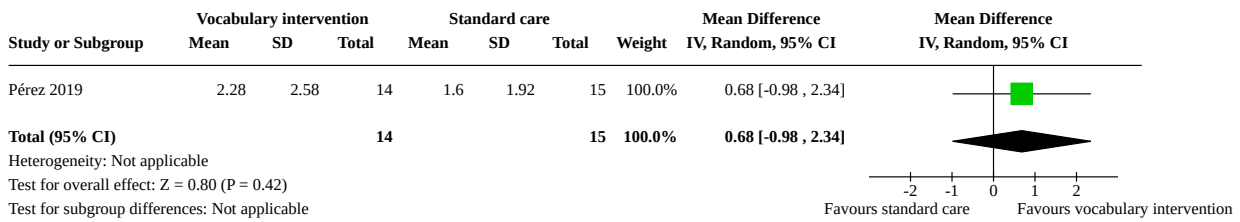
Analysis 1.13. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 13: L1 receptive distal vocabulary



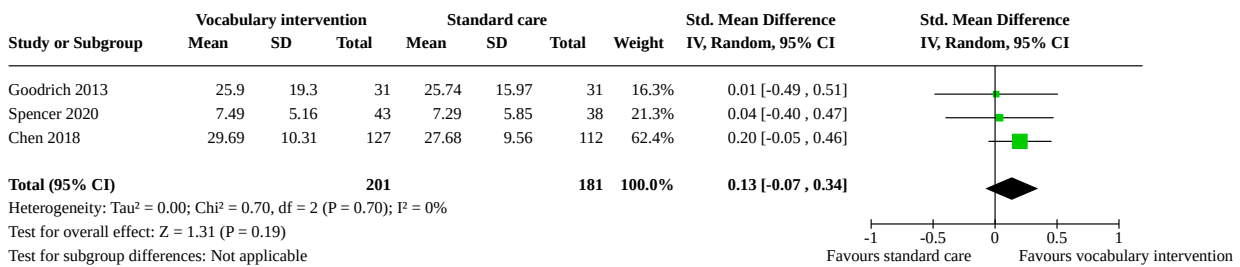
Analysis 1.14. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 14: L1 expressive proximal vocabulary



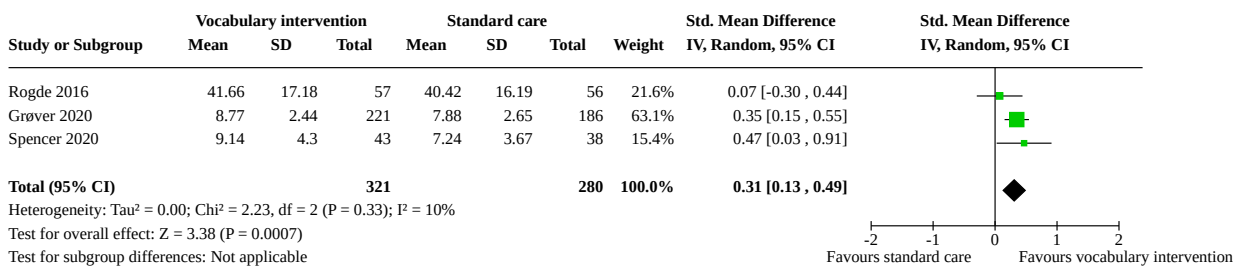
Analysis 1.15. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 15: L1 expressive proximal vocabulary at follow-up



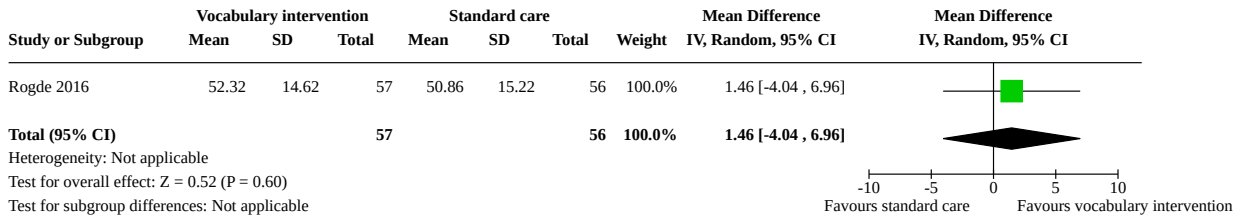
Analysis 1.16. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 16: L1 expressive distal vocabulary



Analysis 1.17. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 17: L2 grammatical knowledge



Analysis 1.18. Comparison 1: Vocabulary interventions compared with standard care for second language (L2) learners, Outcome 18: L2 grammatical knowledge at follow-up



APPENDICES

Appendix 1. Overview of systematic reviews and meta-analyses of second language (L2) vocabulary interventions for L2 learners

Review			Included studies			
Author and year of publication	Type of review	Aim of study	Search	Types of study design	Types of interventions	Types of participants
Fitton 2018	Meta-analysis	"... to examine the impact of shared book reading on language and literacy outcomes among ELs, and to evaluate potential moderators that influence the impact of shared book reading on ELs' outcomes." (quote)	PsycINFO, ERIC, MEDLINE, Academic Search Premier, and ProQuest Social Sciences Studies published in English between 1 January 1981 and 30 April 2017	Empirical studies that used experimental design (quantitative)	Shared book reading	<ul style="list-style-type: none"> Participant sample comprising ≥ 80% ELs Study conducted in the US Age: ≤ 12 years
Hur 2020	Systematic review	"... to describe key features of English early literacy interventions provided to children who were DLLs and their effects on English early literacy skills." (quote)	PsycINFO, ERIC, MEDLINE, Academic Search Premier, and ProQuest Social Sciences Peer-reviewed journals Published in English, through to August 2016	Group experimental research designs in which participants were randomly assigned to conditions Quasi-experimental research designs SCEDs	English early literacy interventions	<ul style="list-style-type: none"> DLLs (English learners) Children with or without disability Age: 0–5 years
Larson 2020	Systematic review	"... to discuss how cultural and linguistic factors were addressed in the interventions, examine the methodological rigor of the stud-	PsycINFO, ERIC, MEDLINE, PubMed, Academic Search Complete, Web	Randomised or non-randomised experimental designs, including	Interventions focused on 4 areas:	<ul style="list-style-type: none"> Young children from CLD

(Continued)

ies, identify the outcomes and measures used, determine the efficacy of the interventions on language skills in English and in children's home language(s), and describe the reported social validity of the interventions." (quote)	of Science, and Google Scholar Peer-reviewed articles Published in English between 1975 and 2015	SCEDs with ≥ 2 participants and adequate experimental control	<ul style="list-style-type: none"> • explicit instruction on targeted skills; • classroom curriculum interventions; • interactive book reading or book making (or both) interventions; and • naturalistic, routines-based interventions 	back-grounds <ul style="list-style-type: none"> • Age: 0–5 years
--	--	---	---	---

Footnotes

CLD: culturally and linguistically diverse; DLL: dual language learner; ELs: English learners; ERIC: Education Resource Information Center; SCED: single case experimental design.

Appendix 2. Search strategies

Cochrane Central Register of Controlled Trials (CENTRAL)

Searched 13 September 2021 (262 records. A large volume of irrelevant trials register records were retrieved from CENTRAL so these were excluded. Trials registers were searched separately using ClinicalTrials.gov and WHO ICTRP)

Searched 6 December 2022. Limited to records added to CENTRAL from 13 September 2021 to 8 December 2022 (41 records)

ID Search

- #1 [mh multilingualism]
- #2 ((mh refugees] and (language* or read* or vocabular*)) OR ((immigrant* or migrant* or refugee*) near/5 (language* or read* or vocabular*)) in Trials
- #3 (multilingual* or multi next lingual*)
- #4 first next language*
- #5 second* next language*
- #6 ((L1 or L2 or 1L or 2L) near/10 (language* or read* or vocabular*))
- #7 dual next language*
- #8 native next language*
- #9 (minority near/2 language*)
- #10 (bilingual* or bi next lingual*)
- #11 (first near/2 second) near/5 language*
- #12 (both next languages or two next languages)
- #13 (host near/1 language*)
- #14 {or #1-#13}
- #15 [mh infant] or [mh child]
- #16 (baby or babies or infant* or toddler* or child* or boy* or girl* or day care or early next years or foundation next stage* or key next stage or kindergarten* or nurser* or play next group or play next school or pre next kindergarten* or prekindergarten* or pre next K or pre next primary or preschool* or pre next school* or elementary next grade* or elementary next school*)
- #17 (1st next year* or 2nd next year* or 1st next Grade* or 2nd next Grade* or First next Grade* or Second next Grade* or Primary next One or Primary next Two or Primary next "1" or Primary next "2")

Vocabulary interventions for second language (L2) learners up to six years of age (Review)

Copyright © 2023 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

#18 {or #15-#17}
 #19 #14 and #18 in Trials
 #20 (IRCT* OR RBR* OR JPRN* OR TCTR* OR SLCTR* OR CTRI* OR EUCT* OR NCT* OR ISRCTN* OR ACTRN* OR DRKS* OR PACT*):AU
 #21 #19 not #20 in Trials

MEDLINE Ovid

Searched 13 September 2021 (960 records)

Searched 6 December 2022 (131 records)

1 multilingualism/
 2 (refugees/ and (language\$ or read\$ or vocabular\$).tw,kf.) or ((immigrant\$ or migrant\$ or refugee\$) adj5 (language\$ or read\$ or vocabular\$)).tw,kf.
 3 (multilingual\$ or multi-lingual\$).tw,kf.
 4 first language\$.tw,kf.
 5 second\$ language\$.tw,kf.
 6 ((L1 or L2 or 1L or 2L) and (language\$ or read\$ or vocabular\$)).tw,kf.
 7 dual language\$.tw,kf.
 8 native language\$.tw,kf.
 9 (minority adj2 language\$).tw,kf.
 10 (bilingual\$ or bi-lingual\$).tw,kf.
 11 (first adj2 second adj5 language\$).tw,kf.
 12 (both languages or two languages).tw,kf.
 13 (host adj1 language\$).tw,kf.
 14 or/1-13
 15 exp child/ or exp infant/
 16 (baby or babies or infant\$ or toddler\$ or child\$ or boy\$ or girl\$ or day care or early years or foundation stage\$ or key stage or kindergarten \$ or nurser\$ or play group or play school or pre-kindergarten\$ or prekindergarten\$ or pre-K or pre-primary or preschool\$ or pre-school\$ or elementary grade\$ or elementary school\$).tw,kf.
 17 ("1st year" or "2nd year" or "1st Grade\$" or "2nd Grade\$" or "First Grade\$" or "Second Grade\$" or Primary One or Primary Two or "Primary 1" or "Primary 2").tw,kf.
 18 or/15-17
 19 14 and 18
 20 randomized controlled trial.pt.
 21 controlled clinical trial.pt.
 22 randomi#ed.ab.
 23 placebo\$.ab.
 24 drug therapy.fs.
 25 randomly.ab.
 26 trial.ab.
 27 groups.ab.
 28 or/20-27
 29 exp animals/ not humans.sh.
 30 28 not 29
 31 19 and 30

MEDLINE(R) In-Process, In-Data-Review & PubMed NOT MEDLINE Citations Ovid

Searched 10 September 2021 (424 records)

Searched 6 December 2022 (86 records)

1 (multilingual\$ or multi-lingual\$).tw,kf.
 2 ((immigrant\$ or migrant\$ or refugee\$) adj5 (language\$ or read\$ or vocabular\$)).tw,kf.
 3 (first language\$ or second\$ language\$).tw,kf.
 4 ((L1 or L2 or 1L or 2L) and (language\$ or read\$ or vocabular\$)).tw,kf.
 5 dual language\$.tw,kf.
 6 native language\$.tw,kf.
 7 (minority adj2 language\$).tw,kf.
 8 (bilingual\$ or bi-lingual\$).tw,kf.
 9 (first adj2 second adj5 language\$).tw,kf.
 10 (both languages or two languages).tw,kf.
 11 (host adj1 language\$).tw,kf.

Vocabulary interventions for second language (L2) learners up to six years of age (Review)

Copyright © 2023 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

12 or/1-11
13 (baby or babies or infant\$ or toddler\$ or child\$ or boy\$ or girl\$ or day care or early years or foundation stage\$ or key stage or kindergarten\$ or nurser\$ or play group or play school or pre-kindergarten\$ or prekindergarten\$ or pre-K or pre-primary or preschool\$ or pre-school\$ or elementary grade\$ or elementary school\$).tw,kf.
14 ("1st year" or "2nd year" or "1st Grade\$" or "2nd Grade\$" or First Grade\$ or Second Grade\$ or Primary One or Primary Two or "Primary 1" or "Primary 2").tw,kf.
15 or/13-14
16 12 and 15
17 (random\$ or control\$ or group\$ or cluster\$ or placebo\$ or trial\$ or assign\$ or allocat\$ or prospectiv\$ or meta-analysis or systematic review or longitudinal\$).tw,kf.
18 16 and 17
19 limit 18 to ("in data review" or in process or "pubmed not medline")

MEDLINE Epub Ahead of Print Ovid

Searched 10 September 2021 (71 records)

Searched 6 December 2022 (53 records)

1 (multilingual\$ or multi-lingual\$).tw,kf.
2 ((immigrant\$ or migrant\$ or refugee\$) adj5 (language\$ or read\$ or vocabular\$)).tw,kf.
3 (first language\$ or second\$ language\$).tw,kf.
4 ((L1 or L2 or 1L or 2L) and (language\$ or read\$ or vocabular\$)).tw,kf.
5 dual language\$.tw,kf.
6 native language\$.tw,kf.
7 (minority adj2 language\$).tw,kf.
8 (bilingual\$ or bi-lingual\$).tw,kf.
9 (first adj2 second adj5 language\$).tw,kf.
10 (both languages or two languages).tw,kf.
11 (host adj1 language\$).tw,kf.
12 or/1-11
13 (baby or babies or infant\$ or toddler\$ or child\$ or boy\$ or girl\$ or day care or early years or foundation stage\$ or key stage or kindergarten\$ or nurser\$ or play group or play school or pre-kindergarten\$ or prekindergarten\$ or pre-K or pre-primary or preschool\$ or pre-school\$ or elementary grade\$ or elementary school\$).tw,kf.
14 ("1st year" or "2nd year" or "1st Grade\$" or "2nd Grade\$" or First Grade\$ or Second Grade\$ or Primary One or Primary Two or "Primary 1" or "Primary 2").tw,kf.
15 or/13-14
16 12 and 15
17 (random\$ or control\$ or group\$ or cluster\$ or placebo\$ or trial\$ or assign\$ or allocat\$ or prospectiv\$ or meta-analysis or systematic review or longitudinal\$).tw,kf.
18 16 and 17

Embase Ovid

Searched 14 September 2021 (1189 records)

Searched 6 December 2022 (245 records)

1 multilingualism/ or bilingualism/
2 (exp refugee/ and (language\$ or read\$ or vocabular\$).tw,kw.) or ((immigrant\$ or migrant\$ or refugee\$) adj5 (language\$ or read\$ or vocabular\$)).tw,kw. (1741)
3 (multilingual\$ or multi-lingual\$).tw,kw.
4 first language\$.tw,kw.
5 second\$ language\$.tw,kw.
6 ((L1 or L2 or 1L or 2L) and (language\$ or read\$ or vocabular\$)).tw,kw.
7 dual language\$.tw,kw.
8 native language\$.tw,kw.
9 (minority adj2 language\$).tw,kw.
10 (bilingual\$ or bi-lingual\$).tw,kw.
11 (first adj2 second adj5 language\$).tw,kw.
12 (both languages or two languages).tw,kw.
13 (host adj1 language\$).tw,kw.
14 or/1-13
15 exp child/

16 (baby or babies or infant\$ or toddler\$ or child\$ or boy\$ or girl\$ or day care or early years or foundation stage\$ or key stage or kindergarten\$ or nurser\$ or play group or play school or pre-kindergarten\$ or prekindergarten\$ or pre-K or pre-primary or preschool\$ or pre-school\$ or elementary grade\$ or elementary school\$).tw,kw.
 17 ("1st year" or "2nd year" or "1st Grade\$" or "2nd Grade\$" or "First Grade\$" or Second Grade\$ or Primary One or Primary Two or "Primary 1" or "Primary 2").tw,kw.
 18 or/15-17
 19 14 and 18
 20 Randomized controlled trial/
 21 Controlled clinical study/
 22 random\$.ti,ab.
 23 randomization/
 24 intermethod comparison/
 25 placebo.ti,ab.
 26 (compare or compared or comparison).ti.
 27 ((evaluated or evaluate or evaluating or assessed or assess) and (compare or compared or comparing or comparison)).ab.
 28 (open adj label).ti,ab.
 29 ((double or single or doubly or singly) adj (blind or blinded or blindly)).ti,ab.
 30 double blind procedure/
 31 parallel group\$1.ti,ab.
 32 (crossover or cross over).ti,ab.
 33 ((assign\$ or match or matched or allocation) adj5 (alternate or group\$1 or intervention\$1 or patient\$1 or subject\$1 or participant \$1)).ti,ab.
 34 (assigned or allocated).ti,ab.
 35 (controlled adj7 (study or design or trial)).ti,ab.
 36 (volunteer or volunteers).ti,ab.
 37 human experiment/
 38 trial.ti.
 39 or/20-38
 40 (random\$ adj sampl\$ adj7 ("cross section\$" or questionnaire\$1 or survey\$ or database\$1)).ti,ab. not (comparative study/ or controlled study/ or randomi?ed controlled.ti,ab. or randomly assigned.ti,ab.)
 41 Cross-sectional study/ not (randomized controlled trial/ or controlled clinical study/ or controlled study/ or randomi?ed controlled.ti,ab. or control group\$1.ti,ab.)
 42 (((case adj control\$) and random\$) not randomi?ed controlled).ti,ab.
 43 (Systematic review not (trial or study)).ti.
 44 (nonrandom\$ not random\$).ti,ab.
 45 "Random field\$".ti,ab.
 46 (random cluster adj3 sampl\$).ti,ab.
 47 (review.ab. and review.pt.) not trial.ti.
 48 "we searched".ab. and (review.ti. or review.pt.)
 49 "update review".ab.
 50 (databases adj4 searched).ab.
 51 (rat or rats or mouse or mice or swine or porcine or murine or sheep or lambs or pigs or piglets or rabbit or rabbits or cat or cats or dog or dogs or cattle or bovine or monkey or monkeys or trout or marmoset\$1).ti. and animal experiment/
 52 Animal experiment/ not (human experiment/ or human/)
 53 or/40-52
 54 39 not 53
 55 19 and 54

ERIC EBSCOhost

Searched 15 September 2021 (1490 records)

Searched 8 December 2022 (370 records)

S1 (DE "Second Language Instruction" OR DE "Second Language Learning" OR DE "Second Language Programs" OR DE "Second Languages") AND (DE "Vocabulary" OR DE "Vocabulary Development" OR DE "Vocabulary Skills" OR DE "Reading" OR DE "Reading Skills" OR DE "Literacy" OR DE "Literacy Education" or DE "Verbal Development")

S2 (DE "Bilingual Education" OR DE "Bilingual Education Programs" OR DE "Bilingual Instructional Materials") AND (DE "Vocabulary" OR DE "Vocabulary Development" OR DE "Vocabulary Skills" OR DE "Reading" OR DE "Reading Skills" OR DE "Literacy" OR DE "Literacy Education" or DE "Verbal Development")

S3 (DE "Refugees" or DE "immigrants") AND (DE "Vocabulary" OR DE "Vocabulary Development" OR DE "Vocabulary Skills" OR DE "Reading" OR DE "Reading Skills" OR DE "Literacy" OR DE "Literacy Education" OR TI(language* or read* or vocabular*))

S4 TI((immigrant* or migrant* or refugee*) N5 (language* or read* or vocabular*)) or AB ((immigrant* or migrant* or refugee*) N5 (language* or read* or vocabular*))
 S5 TI("first language*" or "second language*")
 S6 TI((L1 or L2 or 1L or 2L) AND (language* or read* or vocabular*))
 S7 TI("dual language*" or "native language*" or "minority language*" or "both languages" or "two languages" or "host language*")
 S8 TI((bilingual* or "bi lingual*" OR multilingual* or "multi lingual*") AND (language* or read* or vocabular*))
 S9 TI((first N2 second) N5 language*) OR AB((first N2 second) N5 language*)
 S10 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8 OR S9
 S11 (DE "Infants" OR DE "Young Children" OR DE "Toddlers" OR DE "Preschool Children" OR (DE "Preschool Education" OR DE "Kindergarten" OR DE "Primary Education")
 S12 TI ("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2") OR AB ("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2")
 S13 TI(baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or pre-kindergarten* or prekindergarten* or pre-K or pre-primary or preschool* or pre-school* or "elementary grade*" or "elementary school*") OR AB(baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or pre-kindergarten* or prekindergarten* or pre-K or pre-primary or preschool* or pre-school* or "elementary grade*" or "elementary school*")
 S14 S11 OR S12 OR S13
 S15 S10 AND S14
 S16 DE "Meta Analysis" OR DE "Evaluation Research" OR DE "Control Groups" OR DE "Experimental Groups" OR DE "Longitudinal Studies" OR DE "Followup Studies" OR DE "Program Effectiveness" OR DE "Program Evaluation"
 S17 TI (randomi#ed or randomly or trial* or experiment* or PROSPECTIVE* OR longitudinal or BLIND* or CONTROL*) OR AB (randomi#ed or randomly or trial* or experiment* or PROSPECTIVE* OR longitudinal or BLIND* or CONTROL*)
 S18 S16 OR S17
 S19 S15 AND S18

Education Abstracts (H.W. Wilson) EBSCOhost

Searched 15 September 2021 (630 records)

Searched 8 December 2022 (81 records)

Education Source EBSCOhost: Searched 8 December 2022 (335 records)

S1 (DE "Second language acquisition" OR DE "Bilingual education" OR DE "Bilingual education -- Aids & devices" OR DE "Multilingual education" OR DE "Multilingual teaching aids & devices" or DE "English as a foreign language" OR DE "Foreign language education" OR DE "Native language") AND (DE "Vocabulary education in elementary schools" OR DE "Vocabulary education" OR DE "Vocabulary" OR DE "Reading" OR DE "Reading (Early childhood)" OR DE "Reading (Elementary)" OR DE "Reading (Preschool)" OR DE "Literacy")
 S2 (DE"Education of refugees" OR DE"Education of immigrants" OR DE"Refugee children" OR DE "Children of immigrants") AND (DE"Language aquisition" OR TI(language* or read* or vocabular*) or AB(language* or read* or vocabular*))
 S3 TI((immigrant* or migrant* or refugee*) N5 (language* or read* or vocabular*)) or AB ((immigrant* or migrant* or refugee*) N5 (language* or read* or vocabular*))
 S4 TI("first language*" or "second language*")
 S5 TI((L1 or L2 or 1L or 2L) AND (language* or read* or vocabular*)) OR AB ((L1 or L2 or 1L or 2L) N5 (language* or read* or vocabular*))
 S6 TI("dual language*" or "native language*" or "minority language*" or "both languages" or "two languages" or "host language*") OR AB("dual language*" or "native language*" or "minority language*" or "both languages" or "two languages" or "host language*")
 S7 TI((bilingual* or "bi lingual*" OR multilingual* or "multi lingual*") AND (language* or read* or vocabular*)) OR AB((bilingual* or "bi lingual*" OR multilingual* or "multi lingual*") N5 (language* or read* or vocabular*))
 S8 TI((first N2 second) N5 language*) OR AB((first N2 second) N5 language*)
 S9 S1 OR S2 OR S3 OR S4 OR S5 OR S6 OR S7 OR S8
 S10 (DE "Infant school education (Great Britain)" OR DE "Infant schools (Great Britain)" OR DE "Preschool children" OR DE "Preschool education" OR DE "Kindergarten" OR DE "Kindergarten children" OR DE "Children" OR DE "School children")

S11 TI ("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2") OR AB ("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2")

S12 TI (baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or pre-kindergarten* or prekindergarten* or pre-K or pre-primary or preschool* or pre-school* or "elementary grade*" or "elementary school*") OR AB (baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or pre-kindergarten* or prekindergarten* or pre-K or pre-primary or preschool* or pre-school* or "elementary grade*" or "elementary school*")

S13 S10 OR S11 OR S12

S14 S9 AND S13

S15 (DE "Control groups" OR DE "Evaluation research" OR DE "Experimental groups" OR DE "Longitudinal method" OR DE "Evaluation research" OR DE "Educational evaluation" OR DE "Program effectiveness (Education)")

S16 TI (randomi#ed or randomly or trial* or experiment* or PROSPECTIVE* OR longitudinal or BLIND* or CONTROL*) OR AB (randomi#ed or randomly or trial* or experiment* or PROSPECTIVE* OR longitudinal or BLIND* or CONTROL*)

S17 S15 OR S16

S18 S14 AND S17

Proquest Education Database

Searched 16 September 2021 (1004 records)

Searched 8 December 2022 (ProQuest Dissertations & Theses Global; 148 records)

((MAINSUBJECT.EXACT("Multilingualism") OR MAINSUBJECT.EXACT("Bilingualism") OR MAINSUBJECT.EXACT("Bilingual materials") OR MAINSUBJECT.EXACT("Second language learning") OR MAINSUBJECT.EXACT("French as a second language") OR MAINSUBJECT.EXACT("English as a second language")) OR ((MAINSUBJECT.EXACT("Noncitizens") OR MAINSUBJECT.EXACT("Refugees")) AND (MAINSUBJECT.EXACT("Vocabulary development") OR MAINSUBJECT.EXACT("Language acquisition") OR TI(language* or read* or vocabular*) OR AB(language* or read* or vocabular*))) OR (((immigrant* or migrant* or refugee*) NEAR/3 (language* or read* or vocabular*)) OR AB ((immigrant* or migrant* or refugee*) NEAR/3 (language* or read* or vocabular*))) OR TI("first language*" or "second language*") OR TI((L1 or L2 or 1L or 2L) AND (language* or read* or vocabular*)) OR AB ((L1 or L2 or 1L or 2L) Near/5 (language* or read* or vocabular*))) OR TI(("dual language*" or "native language*" or "minority language*" or "both languages" or "two languages" or "host language*") AND (language* or read* or vocabular*)) OR AB(("dual language*" or "native language*" or "minority language*" or "both languages" or "two languages" or "host language*") NEAR/3 (language* or read* or vocabular*)) OR TI((bilingual* or "bi lingual*" OR multilingual* or "multi lingual*") AND (language* or read* or vocabular*)) OR AB((bilingual* or "bi lingual*" OR multilingual* or "multi lingual*") Near/3 (language* or read* or vocabular*)) OR TI((first Near/2 second) Near/5 language*) OR AB((first Near/2 second) Near/5 language*)) AND (((TI(baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or "pre-kindergarten*" or prekindergarten* or "pre-K" or "pre-primary" or preschool* or "pre-school*" or "elementary grade*" or "elementary school*") OR AB(baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or "pre-kindergarten*" or prekindergarten* or "pre-K" or "pre-primary" or preschool* or "pre-school*" or "elementary grade*" or "elementary school*")) OR TI ("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2") OR AB ("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2")) OR (MAINSUBJECT.EXACT("Babies") OR MAINSUBJECT.EXACT("Preschool children") OR MAINSUBJECT.EXACT("Preschool education") OR MAINSUBJECT.EXACT("Toddlers") OR MAINSUBJECT.EXACT("Children & youth")))) AND (MAINSUBJECT.EXACT("Clinical trials") OR MAINSUBJECT.EXACT("Effectiveness studies") OR MAINSUBJECT.EXACT("Educational evaluation") OR MAINSUBJECT.EXACT("Experiments") OR TI (randomi*ed or randomly or "random allocation" or trial* or experiment* or prospective* OR longitudinal or blind* or control* or "follow up") OR AB (randomi*ed or randomly or "random allocation" or trial* or experiment* or prospective* OR longitudinal or blind* or control* or "follow up"))

Linguistics and Language Behavior Abstracts ProQuest (LLBA)

Searched 23 September 2021 (432 records)

Searched 10 January 2023 (40 records)

((MAINSUBJECT.EXACT("Multilingualism") OR MAINSUBJECT.EXACT("Bilingualism") OR MAINSUBJECT.EXACT("Bilingual materials") OR MAINSUBJECT.EXACT("Second language learning") OR MAINSUBJECT.EXACT("French as a second language") OR MAINSUBJECT.EXACT("English as a second language")) OR ((MAINSUBJECT.EXACT("Noncitizens") OR MAINSUBJECT.EXACT("Refugees"))

AND (MAINSUBJECT.EXACT("Vocabulary development") OR MAINSUBJECT.EXACT("Language acquisition") OR TI(language* or read* or vocabular*) OR AB(language* or read* or vocabular*)) OR (((immigrant* or migrant* or refugee*) NEAR/3 (language* or read* or vocabular*)) OR TI("first language*" or "second language*") OR TI((L1 or L2 or 1L or 2L) AND (language* or read* or vocabular*)) OR AB ((L1 or L2 or 1L or 2L) Near/5 (language* or read* or vocabular*)) OR TI(("dual language*" or "native language*" or "minority language*" or "both languages" or "two languages" or "host language*") AND (language* or read* or vocabular*)) OR AB(("dual language*" or "native language*" or "minority language*" or "both languages" or "two languages" or "host language*") NEAR/3 (language* or read* or vocabular*)) OR TI((bilingual* or "bi lingual*" OR multilingual* or "multi lingual*") AND (language* or read* or vocabular*)) OR AB((bilingual* or "bi lingual*" OR multilingual* or "multi lingual*") Near/3 (language* or read* or vocabular*)) OR TI((first Near/2 second) Near/5 language*) OR AB((first Near/2 second) Near/5 language*)) AND (((TI(baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or "pre-kindergarten*" or prekindergarten* or "pre-K" or "pre-primary" or preschool* or "pre-school*" or "elementary grade*" or "elementary school*") OR AB(baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or "pre-kindergarten*" or prekindergarten* or "pre-K" or "pre-primary" or preschool* or "pre-school*" or "elementary grade*" or "elementary school*")) OR TI ("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2") OR AB ("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2")) OR (MAINSUBJECT.EXACT("Babies") OR MAINSUBJECT.EXACT("Preschool children") OR MAINSUBJECT.EXACT("Preschool education") OR MAINSUBJECT.EXACT("Toddlers") OR MAINSUBJECT.EXACT("Children & youth")) AND (MAINSUBJECT.EXACT("Clinical trials") OR MAINSUBJECT.EXACT("Effectiveness studies") OR MAINSUBJECT.EXACT("Educational evaluation") OR MAINSUBJECT.EXACT("Experiments") OR TI (randomi*ed or randomly or "random allocation" or trial* or experiment* or prospective* OR longitudinal or blind* or control* or "follow up") OR AB (randomi*ed or randomly or "random allocation" or trial* or experiment* or prospective* OR longitudinal or blind* or control* or "follow up"))

PsycINFO Ovid (1806 onwards)

Searched 13 September 2021 (1018 records)

Searched 6 December 2022 (94 records)

- 1 Bilingual Education/
- 2 bilingualism/
- 3 multilingualism/
- 4 English as Second Language/
- 5 Foreign Language Education/
- 6 Foreign Language Learning/
- 7 native language/
- 8 (refugees/ or (immigrant\$ or migrant\$ or refugee\$).tw.) and (language\$ or read\$ or vocabular\$).tw.
- 9 (multilingual\$ or multi-lingual\$).tw.
- 10 (first language\$ or 1L or L1).tw.
- 11 (second\$ language\$ or 2L or L2).tw.
- 12 dual language\$.tw.
- 13 native language\$.tw.
- 14 (minority adj2 language\$).tw.
- 15 (bilingual\$ or bi-lingual\$).tw.
- 16 ((first adj2 second) and language\$).tw.
- 17 (both languages or two languages).tw.
- 18 (host adj1 language\$).tw.
- 19 or/1-18
- 20 (baby or babies or infant\$ or toddler\$ or child\$ or boy\$ or girl\$ or "day care" or "early years" or foundation stage\$ or key stage or kindergarten\$ or nurser\$ or "play group" or "play school" or pre-kindergarten\$ or prekindergarten\$ or pre-K or pre-primary or preschool \$ or pre-school\$ or elementary grade\$ or elementary school\$ or "1st year" or "2nd year" or "1st Grade\$" or "2nd Grade\$" or "First Grade \$" or "Second Grade\$" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2").tw.
- 21 19 and 20
- 22 limit 19 to (100 childhood <birth to age 12 yrs> or 120 neonatal <birth to age 1 mo> or 140 infancy <2 to 23 mo> or 160 preschool age <age 2 to 5 yrs> or 180 school age <age 6 to 12 yrs>)
- 23 21 or 22
- 24 clinical trials/
- 25 randomized clinical trials/
- 26 randomized controlled trials/
- 27 treatment effectiveness evaluation/
- 28 exp treatment outcomes/
- 29 followup studies/
- 30 longitudinal studies/

31 Placebo/
 32 Experiment Controls/
 33 exp program evaluation/
 34 (TAU or "treatment as usual" or "wait\$ list" or "business as usual").ab.
 35 (randomly or randomis\$ or randomiz\$).tw.
 36 ((control\$ or experimental) adj5 group\$).tw.
 37 or/24-36
 38 19 and 23 and 37

Scopus Elsevier (all available years)

Searched 17 September 2021 (1268 records)

Searched 6 December 2022 (131 records)

((TITLE (("first language" OR "second language")) OR TITLE-ABS-KEY ((("first language" OR "second language") W/5 (read* OR vocabular*))) OR TITLE-ABS-KEY (((l1 OR l2 OR l1 OR 2l) W/5 (language* OR read* OR vocabular*))) OR TITLE-ABS-KEY (((immigrant* OR migrant* OR refugee*) W/5 (language* OR read* OR vocabular*))) OR TITLE-ABS-KEY ((bilingual* OR "bi-lingual*" OR multilingual* OR "multi-lingual*" OR "both languages" OR "two languages" OR "dual language*" OR "native language" OR (minority W/2 language*)) W/5 (read* OR vocabular*))) AND (TITLE-ABS-KEY ((baby OR babies OR infant* OR toddler* OR child* OR boy* OR girl* OR "day care" OR "early years" OR "foundation stage*" OR "key stage" OR kindergarten* OR nurser* OR "play group" OR "play school" OR pre-kindergarten* OR prekindergarten* OR pre-k OR pre-primary OR preschool* OR pre-school* OR "elementary grade*" OR "elementary school*" OR "1st year" OR "2nd year" OR "1st Grade*" OR "2nd Grade*" OR "First Grade*" OR "Second Grade*" OR "Primary One" OR "Primary Two" OR "Primary 1" OR "Primary 2"))) AND (TITLE-ABS-KEY ((randomi*ed OR randomly OR "random allocation" OR trial* OR experiment* OR prospective* OR longitudinal OR blind* OR control* OR "follow up")))

Web of Science Core Collection, Clarivate; including Science Citation Index-Expanded Web of Science, Clarivate (1970 onwards); Social Sciences Citation Index Web of Science, Clarivate (1970 onwards); Conference Proceedings Citation Index-Science Web of Science, Clarivate (1990 onwards); Conference Proceedings Citation Index-Social Science and Humanities Web of Science, Clarivate (1990 onwards); Emerging Sources Citation Index Web of Science, Clarivate (2015 onwards)

Searched 15 September 2021 (2798)

Searched 7 December 2022 (381 records)

18 #12 AND #11

Indexes=ESCI Timespan=All years

17 #12 AND #11

Indexes=CPCI-SSH Timespan=All years

16 #12 AND #11

Indexes=CPCI-S Timespan=All years

15 #12 AND #11

Indexes=SSCI Timespan=All years

14 #12 AND #11

Indexes=SCI-EXPANDED Timespan=All years

13 #12 AND #11

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

12 TI=(randomi*ed or randomly or trial* or prospective* OR longitudinal or blind* or control or group*) OR AB=(randomi*ed or randomly or trial* or prospective* OR longitudinal or blind* or control or group*)

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

11 #10 AND #7

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

10 #9 OR #8

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

9 TI= ("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2") OR AB =("1st year" or "2nd year" or "1st Grade*" or "2nd Grade*" or "First Grade*" or "Second Grade*" or "Primary One" or "Primary Two" or "Primary 1" or "Primary 2")

Vocabulary interventions for second language (L2) learners up to six years of age (Review)

Copyright © 2023 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

8 TI=(baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or pre-kindergarten* or prekindergarten* or pre-K or pre-primary or preschool* or pre-school* or "elementary grade*" or "elementary school*") OR AB=(baby or babies or infant* or toddler* or child* or boy* or girl* or "day care" or "early years" or "foundation stage*" or "key stage" or kindergarten* or nurser* or "play group" or "play school" or pre-kindergarten* or prekindergarten* or pre-K or pre-primary or preschool* or pre-school* or "elementary grade*" or "elementary school*")
 Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

7 #6 OR #5 OR #4 OR #3 OR #2 OR #1

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

6 TI=((immigrant* or migrant* or refugee*) and (language* or read* or vocabular*)) or AB=((immigrant* or migrant* or refugee*) NEAR/3 (language* or read* or vocabular*))

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

5 TI(("dual language*" or "native language*" or "minority language*" or "both languages" or "two languages" or "host language*") AND (read* or vocabular*)) OR AB(("dual language*" or "native language*" or "minority language*" or "both languages" or "two languages" or "host language*") NEAR/3 (read* or vocabular*))

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

4 TI=((first NEAR/2 second) NEAR/3 language*) OR AB=((first NEAR/2 second) NEAR/3 language*)

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

3 TI=((bilingual* or "bi lingual*" OR multilingual* or "multi lingual*") AND (language* or read* or vocabular*)) or AB=((bilingual* or "bi lingual*" OR multilingual* or "multi lingual*") NEAR/3 (language* or read* or vocabular*))

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

2 TI=((L1 or L2 or 1L or 2L) AND (language* or read* or vocabular*)) OR AB=((L1 or L2 or 1L or 2L) NEAR/5 (language* or read* or vocabular*))

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

1 TI=("first language*" or "second language*") OR AB=("first language*" or "second language*")

Indexes=SCI-EXPANDED, SSCI, CPCI-S, CPCI-SSH, ESCI Timespan=All years

Sociological Abstracts ProQuest (1952 onwards)

Searched 16 September 2021 (180 records)

Searched 8 December 2022 (56 records)

((MAINSUBJECT.EXACT("Second Language Learning") OR MAINSUBJECT.EXACT("Bilingualism") OR MAINSUBJECT.EXACT("Bilingual Education") OR MAINSUBJECT.EXACT("Multilingualism")) OR TI((immigrant* OR migrant* OR refugee*) AND (language* OR read* OR vocabular*)) OR AB ((immigrant* OR migrant* OR refugee*) NEAR/5 (language* OR read* OR vocabular*))) OR TI(("first language" OR "first languages") OR ("second language" OR "second languages")) OR (noft((L1 OR L2 OR 1L OR 2L) AND (language* OR read* OR vocabular*))) OR TI(("dual language" OR ("native language" OR "minority language" OR "minority languages") OR "both languages" OR "two languages" OR ("host language" OR "host languages")) OR TI(((bilingual* OR "bi lingual*" OR multilingual* or "multi lingual*") AND (language* OR read* OR vocabular*))) OR AB(((bilingual* OR "bi lingual*" OR multilingual* or "multi lingual*") AND (language* OR read* OR vocabular*))) OR noft(((first NEAR/2 second) NEAR/5 language*)) AND noft((randomi*ed OR randomly OR trial* OR experiment* OR PROSPECTIVE* OR longitudinal OR BLIND* OR CONTROL*)) AND ((ti((baby OR babies OR infant* OR toddler* OR child* OR boy* OR girl* OR "day care" OR "early years" OR "foundation stage*" OR "key stage" OR kindergarten* OR nurser* OR "play group" OR "play school" OR pre-kindergarten* OR prekindergarten* OR pre-K OR pre-primary OR preschool* OR pre-school* OR "elementary grade*" OR "elementary school*")) OR ab((baby OR babies OR infant* OR toddler* OR child* OR boy* OR girl* OR "day care" OR "early years" OR "foundation stage*" OR "key stage" OR kindergarten* OR nurser* OR "play group" OR "play school" OR pre-kindergarten* OR prekindergarten* OR pre-K OR pre-primary OR preschool* OR pre-school* OR "elementary grade*" OR "elementary school*")) OR ti(("1st year" OR "2nd year" OR "1st Grade*" OR "2nd Grade*" OR "First Grade*" OR "Second Grade*" OR "Primary One" OR "Primary Two" OR "Primary 1" OR "Primary 2")) OR ab(("1st year" OR "2nd year" OR "1st Grade*" OR "2nd Grade*" OR "First Grade*" OR "Second Grade*" OR "Primary One" OR "Primary Two" OR "Primary 1" OR "Primary 2")) OR (MAINSUBJECT.EXACT("Infants") OR MAINSUBJECT.EXACT("Children") OR MAINSUBJECT.EXACT("Preschool Children")))

Cochrane Database of Systematic Reviews (CDSR; current issue), in the Cochrane Library

Searched 14 September 2021 (4 records)

Searched 8 December 2022 (2 records)

Vocabulary interventions for second language (L2) learners up to six years of age (Review)

Copyright © 2023 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ID Search

- #1 [mh multilingualism]
- #2 ([mh refugees] AND (language* or read* or vocabular*)):TI,AB or ((Immigrant* or migrant* or refugee*) AND (language* or read* or vocabular*)):TI,AB,KW
- #3 (multilingual* or multi next lingual*):TI
- #4 (first next language*):TI
- #5 (second* next language*):TI
- #6 ((L1 or L2 or 1L or 2L) near/10 (language* or read* or vocabular*)):TI,AB,KW
- #7 (dual next language*):TI
- #8 (native next language*):TI
- #9 (minority near/2 language*):TI
- #10 (bilingual* or bi next lingual*):TI
- #11 ((first near/2 second) near/5 language*):TI
- #12 (both next languages or two next languages):TI
- #13 (host near/1 language*):TI
- #14 {or #1-#13}
- #15 [mh infant] or [mh child]
- #16 (baby or babies or infant* or toddler* or child* or boy* or girl* or day care or early next years or foundation next stage* or key next stage or kindergarten* or nurser* or play next group or play next school or pre next kindergarten* or prekindergarten* or pre next K or pre next primary or preschool* or pre next school* or elementary next grade* or elementary next school*):TI
- #17 (1st next year* or 2nd next year* or 1st next Grade* or 2nd next Grade* or First next Grade* or Second next Grade* or Primary next One or Primary next Two or Primary next "1" or Primary next "2"):TI
- #18 {or #15-#17}
- #19 #14 and #18 in Cochrane Reviews, Cochrane Protocols

Epistemonikos

Searched 13 September 2021 (49 records)

Searched 8 December 2022 (6 records)

title:("first language" OR "second language" OR "dual language" OR "native language" OR "minority language" OR "both languages" OR "two languages" OR "host language" OR L1 OR L2 OR 1L OR 2L)

Limited to Publication type Systematic reviews and Systematic review questions: Interventions

ClinicalTrials.gov

Searched 13 September 2021 (16 records)

Searched 8 December 2022 (3 records)

Interventional Studies CONDITION| | SECOND LANGUAGE or FIRST LANGUAGE OR DUAL LANGUAGE OR BILINGUAL OR BI-LINGUAL OR MULTILINGUAL OR MULTI-LINGUAL | Child [4] ct.gov_2

Interventional Studies TITLE | SECOND LANGUAGE or FIRST LANGUAGE OR DUAL LANGUAGE OR BILINGUAL OR BI-LINGUAL OR MULTILINGUAL OR MULTI-LINGUAL | Child [2] ct.gov_3

Interventional Studies CONDITION| native language or minority language or both languages or two languages or host language | Child [0]

Interventional Studies TITLE | native language or minority language or both languages or two languages or host language | Child [0]

Interventional Studies intervention | VOCABULARY or READING or LANGUAGE | Child [10] ct.gov_1

World Health Organization International Clinical Trials Registry Platform (WHO ICTRP; trialsearch.who.int)

Searched 13 September 2021 (26 records)

Searched 8 December 2022 (42 records)

TITLE| SECOND LANGUAGE or FIRST LANGUAGE OR DUAL LANGUAGE OR BILINGUAL OR BI-LINGUAL OR MULTILINGUAL OR MULTI-LINGUAL [10 records] limited to children

CONDITION| SECOND LANGUAGE or FIRST LANGUAGE OR DUAL LANGUAGE OR BILINGUAL OR BI-LINGUAL OR MULTILINGUAL OR MULTI-LINGUAL [2 records] limited to children

TITLE | native language or minority language or both languages or two languages or host language | [1] limited to children

Vocabulary interventions for second language (L2) learners up to six years of age (Review)

Copyright © 2023 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

CONDITION | native language or minority language or both languages or two languages or host language | [0] limited to children

intervention | VOCABULARY or READING or LANGUAGE | [18] limited to children

Google Scholar

Searched 23 September 2021 (the 500 first records)

Searched 14 December 2022 (the 100 first records)

("second language" OR immigrant* OR bilingual* OR multilingual* OR "dual language*") AND (child* OR "day care" OR kindergarten* OR nurser* OR pre-primary OR preschool* OR pre-school*) AND (randomi*ed OR randomly OR trial* OR experiment* OR control* OR "follow up")

Open Grey

Searched 23 September 2021 (640 records)

Searched 14 December 2022 (30 records)

("bilingual education" OR bilingualism OR multilingualism OR "english as second language" OR "foreign language education" OR "foreign language learning" OR "native language" OR refugees OR immigrant* OR migrant* OR refugee* OR multilingual* or multi-lingual* OR "first language*" OR 1L OR "second* language*" OR 2L OR "dual language*" OR "native language*" OR "minority adj2 language*" OR bilingual* OR bi-lingual* OR "first adj2 second language*" OR "both languages" OR "two languages" OR "host adj1 language*") AND (baby OR babies OR infant* OR toddler* OR child* OR boy* OR girl* OR "day care" OR "early years" OR "foundation stage*" OR "key stage" OR kindergarten* OR nurser* OR "play group" OR "play school" OR pre-kindergarten* OR prekindergarten* OR pre-K OR pre-primary OR preschool* OR pre-school* OR "elementary grade*" OR elementary school* OR "1st year" OR "2nd year" OR "1st Grade*" OR "2nd Grade*" OR "First Grade*" OR "Second Grade*" OR "Primary One" OR "Primary Two" OR "Primary 1" OR "Primary 2") AND ("clinical trials" OR "randomized clinical trials" OR "randomized controlled trials" OR "treatment effectiveness evaluation" OR "treatment outcomes" OR "followup studies" OR "longitudinal studies" OR placebo OR "experiment controls" OR "program evaluation" OR TAU OR "treatment as usual" OR "wait* list" OR "business as usual" OR randomly OR randomis* OR randomize* OR control* or experimental adj5 group*)

Appendix 3. Methods included in the protocol but not used in review

Methods included in the protocol, but not used in review (Hjetland 2021)

Data collection and analysis

Because the meta-analyses included < 10 studies, we were unable to create and examine a funnel plot to assess reporting bias.

We were unable to conduct subgroup (moderator) analyses, as planned, because there were < 10 studies in the meta-analyses that reported on the moderator (Deeks 2022).

Because none of the studies included in the meta-analyses had an overall high risk of bias, we did not conduct the planned sensitivity analysis excluding studies with a high risk of bias.

HISTORY

Protocol first published: Issue 9, 2021

CONTRIBUTIONS OF AUTHORS

For contributions that were developed and written at the protocol stage (background and methods), see Hjetland 2021.

HNH led the development of the review.

HNH conducted the searches in Open Grey, Linguistics and Language Behavior Abstracts ProQuest (LLBA), and Google Scholar.

HNH, ÅMH, M-BM, KAN conducted the handsearches.

HNH prepared the [Characteristics of included studies](#) table and ÅMH, LIE, M-BM, CN each contributed by describing the content of the respective interventions.

HNH and HH screened all studies at both abstract and full-text stage. JK resolved any disagreements.

HNH and MK extracted data from the included articles. HH resolved coding disagreements.

HNH entered the data into Review Manager Web.

HNH and HH assessed risk of bias and the certainty of the evidence in the included studies, and constructed [Summary of findings 1](#).

KAN resolved any disagreements in risk of bias ratings and MK resolved disagreements for GRADE ratings.

HNH conducted the analyses and wrote the [Results](#) section.

HH and HNH wrote the first draft of [Summary of main results](#) and [Potential biases in the review process](#).

KAN and ÅMH wrote the first draft of [Overall completeness and applicability of evidence](#).

HNH wrote the first draft of [Quality of the evidence](#).

JK and HNH wrote the first draft of [Agreements and disagreements with other studies or reviews](#).

CN wrote the first draft of [Authors' conclusions](#).

EG contributed to conceptual discussions, reviewed and contributed to revisions.

All review authors were involved in the conception of the review, design of the review, interpretation of the data, revisions, discussions, and provided comments on drafts.

Both HNH and KAN are guarantors for the review.

DECLARATIONS OF INTEREST

HNH: none. Her postdoctoral fellow position (from January 2020 to July 2022) was part financed (75%) by a research grant (#299197) from Norges Forskningsråd (Research Council of Norway) for the project 'Second Language Learner Plus (SL+)'; paid to University of Oslo.

HH: declares a research grant (#299197) from Norges Forskningsråd (Research Council of Norway) for the project 'Second Language Learner Plus (SL+)'; paid to Western Norway University of Applied Sciences.

MK: none. Her current position is part financed by a research grant (#299197) from Norges Forskningsråd (Research Council of Norway) for the project 'Second Language Learner Plus (SL+)'; paid to Statped.

JK: none. Her current position is part financed by a research grant (#299197) from Norges Forskningsråd (Research Council of Norway) for the project 'The Second Language Learner Plus (SL+)'; paid to Statped.

ÅMH: none. Her current position is part financed by a research grant (#299197) from Norges Forskningsråd (Research Council of Norway) for the project 'Second Language Learner Plus (SL+)'; paid to the University of Oslo.

LIE: none. Her current position is part financed by a research grant (#299197) from Norges Forskningsråd (Research Council of Norway) for the project 'Second Language Learner Plus (SL+)'; paid to Western Norway University of Applied Sciences.

EG: none.

CN: none.

MBM: none. Her current position is part financed by a research grant (#299197) from Norges Forskningsråd (Research Council of Norway) for the project 'Second Language Learner Plus (SL+)'; paid to Statped.

KAN: reports a grant (#299197) from Norges Forskningsråd (Research Council of Norway) for the project 'Second Language Learner Plus (SL+)', which seeks to develop and test the effects of a vocabulary intervention program and on which she is the Principal Investigator; paid to University of Oslo.

SOURCES OF SUPPORT

Internal sources

- Department of Special Needs Education, University of Oslo, Norway

Support for the preparation of the protocol and review to be carried out during office hours for HNH, KAN, ÅMH, and CN.

- Department of Language, Literature, Mathematics and Interpreting, Western Norway University of Applied Sciences, Norway

Support for the preparation of the protocol and review to be carried out during office hours for HCH.

- Applied Psychology and Human Development, University of Toronto, Canada

Support for the preparation of the protocol and review to be carried out during office hours for EG.

- Statped, Norway

Support for the preparation of the protocol and review to be carried out during office hours for JK, MK, and MBM.

- Department of Pedagogy, Religion and Social Studies, Western Norway University of Applied Sciences, Norway

Support for the preparation of the protocol and review to be carried out during office hours for LIE.

External sources

- The Research Council of Norway, Norway

This review is part of a project, SL+, which is funded by the Research Council of Norway (grant number 299197). The funder had no role in the design, conduct, or publication of this review.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Due to workload pressures, JK replaced KAN in resolving conflicts between HNH and HH after screening full texts for eligibility, and MK replaced JK in extracting data.

In [Objectives](#), we replaced the word "treatment" with "intervention" because the latter is more in keeping with the literature on which the review is based. In addition, we deleted the phrase "characteristics of L2 learners who do not appear to benefit from treatment" from the secondary objective because we had not specified how this would be examined as an outcome.

Distal outcome measures were prioritised for inclusion in the meta-analysis as we were interested in the broad impacts of early vocabulary intervention. Distal is defined as vocabulary that was not explicitly taught in the intervention itself, and broader measures of language, for instance, listening comprehension, and narrative skills. The latter were often vehicles for vocabulary instruction and, therefore, it was reasonable to expect children may have benefitted from repeated exposure. For listening comprehension and narrative skills, we included only distal measures. This had not been specified in the protocol ([Hjetland 2021](#)).

In the protocol, we stated that other sensitivity analyses than the one listed could be considered ([Hjetland 2021](#)). Therefore, we decided to run a sensitivity analysis where we excluded the one study that included children with clinically identified developmental language disorder. This is because recent terminology distinguishes low language proficiency due to reduced exposure (i.e. L2 learners) and persistent poor language despite adequate exposure characteristic of DLD ([Bishop 2017](#)).

We planned to conduct a search in ProQuest Dissertations & Theses Global (PQDT). However, as educational theses records from this database are included in Proquest Education Database, we did not conduct a separate search in PQDT.

INDEX TERMS

Medical Subject Headings (MeSH)

Cognition; *Communication; Emotions; Language; *Vocabulary

MeSH check words

Child; Child, Preschool; Humans