



**Cochrane**  
**Library**

Cochrane Database of Systematic Reviews

## Vocabulary interventions for second language (L2) learners up to six years (Protocol)

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

Hjetland HN, Hofslundsengen H, Klem M, Karlsen J, Hagen ÅM, Engevik LI, Geva E, Norbury C, Monsrud M-B, Bottegaard Næss K-A.

Vocabulary interventions for second language (L2) learners up to six years (Protocol).

*Cochrane Database of Systematic Reviews* 2021, Issue 9. Art. No.: CD014890.

DOI: [10.1002/14651858.CD014890](https://doi.org/10.1002/14651858.CD014890).

[www.cochranelibrary.com](http://www.cochranelibrary.com)

**Vocabulary interventions for second language (L2) learners up to six years (Protocol)**

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

**WILEY**

---

**TABLE OF CONTENTS**

HEADER .....	1
ABSTRACT .....	1
BACKGROUND .....	2
OBJECTIVES .....	6
METHODS .....	6
ACKNOWLEDGEMENTS .....	11
REFERENCES .....	12
APPENDICES .....	18
CONTRIBUTIONS OF AUTHORS .....	32
DECLARATIONS OF INTEREST .....	33
SOURCES OF SUPPORT .....	33

---

[Intervention Protocol]

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

Hanne Næss Hjetland<sup>1</sup>, Hilde Hofslundsengen<sup>2</sup>, Marianne Klem<sup>3</sup>, Jannicke Karlsen<sup>3</sup>, Åste M Hagen<sup>1</sup>, Liv Inger Engevik<sup>4</sup>, Esther Geva<sup>5</sup>, Courtenay Norbury<sup>1,6</sup>, May-Britt Monsrud<sup>3</sup>, Kari-Anne Bottegaard Næss<sup>1</sup>

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

**Contact address:** Hanne Næss Hjetland, [h.n.hjetland@isp.uio.no](mailto:h.n.hjetland@isp.uio.no).

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

**Publication status and date:** New, published in Issue 9, 2021.

**Citation:** Hjetland HN, Hofslundsengen H, Klem M, Karlsen J, Hagen ÅM, Engevik LI, Geva E, Norbury C, Monsrud M-B, Bottegaard Næss K-A. Vocabulary interventions for second language (L2) learners up to six years (Protocol). *Cochrane Database of Systematic Reviews* 2021, Issue 9. Art. No.: CD014890. DOI: [10.1002/14651858.CD014890](https://doi.org/10.1002/14651858.CD014890).

Copyright © 2021 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

## ABSTRACT

### Objectives

This is a protocol for a Cochrane Review (intervention). The objectives are as follows:

The primary objective is 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. The secondary objectives are to examine associations between L2 vocabulary interventions and general characteristics of L2 learners (e.g. age, L2 exposure and L1 skills), as well as specific characteristics of L2 learners who do not appear to benefit from treatment.

## BACKGROUND

### Description of the condition

The vocabulary knowledge of second language (L2) learners includes words in both the first language (L1) and the L2, which is evident when their conceptual vocabulary is considered (Monsrud 2019). However, 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). With some exceptions, immigrant students have poorer outcomes at all levels of education than their L1, non-immigrant, monolingual peers, though the outcomes vary between countries (OECD 2019).

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). In general, 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). Moreover, it appears that less common, more academic vocabulary (i.e. words that are more abstract and potentially more ambiguous), which are essential for reading comprehension and learning, are particularly challenging to acquire for L2 learners (Biemiller 2005; Jean 2009; Lin 2012).

Persistently poor vocabulary skills have a critical impact on both the individual and society. Reduced vocabulary may have negative, long-term, cumulative effects on further oral language comprehension development, reading comprehension and academic achievement (Snow 1995; Stanovich 1986), and may lead to academic failure and dropping out of school (Lervåg 2018). Poor vocabulary also has a negative impact on mental health (Snow 2016; Toppelberg 2002). For example, inadequate vocabulary acquisition has been associated with oppositional disorder (Gremillion 2014), increased involvement in crime (Anderson 2016), and reduced occupational opportunities in adulthood (Johnson 2010). Moreover, vocabulary is important for social communication (Næss 2017); children with poor language skills are more likely to be rejected by their peers and are less likely to initiate interactions, participate in social interactions and play (Brekke Stangeland 2017; Rice 1991), although, it is important to note that these findings are from studies with monolingual students.

The number of children entering school for whom the instructional language is a L2 is increasing in various parts of the world. For example, about one in 10 people in OECD countries are foreign born (OECD 2020b), and in the USA, the number of English language learners in public schools increased by 1.2 million between 2000 and 2017 (to 10.1% or 5.0 million English language learners in 2017; NCES 2020).

It is difficult to estimate the proportion of these children who have poor L2 vocabulary skills in the societal language, but Melby-Lervåg 2014 found large differences (Cohen's  $d = 1.12$  for pooled effect sizes) in oral language (including vocabulary) when comparing language and reading skills between L1 and L2 learners. These results suggest that L2 learners as a group perform

considerably worse than their monolingual counterparts, and the large variation within the group of L2 learners (Melby-Lervåg 2014), suggests that some L2 learners may have a language impairment. Moreover, a large proportion of immigrant children with poorer L2 vocabulary have had less than optimal opportunities to develop the type of vocabulary needed for academic learning (Hoff 2013).

The poor L2 vocabulary knowledge of L2 learners relative to their monolingual peers is worrisome because the rank-ordering 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 for poor L2 vocabulary knowledge can be both external and internal to the individual (Paradis 2011). External factors refer to the amount and quality of language exposure; internal factors refer to the individual differences in the language learning potential of each child (e.g. language learning aptitude, phonological short-term memory), where the lower end of the continuum involves children with developmental language disorder (DLD). Notably, as the term DLD was endorsed in a consensus study involving a panel of experts only in 2017 (Catalise, see Bishop 2017), the relevant research literature uses different terminology for labelling children with unexplained language problems; the term DLD also includes children formerly identified with specific language impairment.

Furthermore, it is important to consider that immigrant students are not a homogeneous group (OECD 2019), and discussions focusing on the 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 the vast variability in the range of policies and opportunities designed to enhance the learning of the societal language by the children of immigrants and refugees.

The risk factors for poor L2 vocabulary are related to aspects of low socioeconomic status, such as parental education, and to the amount and quality of exposure to the societal language. In general, compared to their monolingual peers, L2 learners coming from lower socioeconomic status backgrounds, with more restricted access to high-quality education, are more at risk of having poor vocabulary skills (OECD 2020b). Thus, factors that are associated with low socioeconomic status are found to contribute to the difficulties of acquiring and mastering a L2. Children learn new words through their everyday experiences (Hart 1995), and for young children, social interactions with their parents are often the main source of language exposure (Hart 1995; Rowe 2012; Weizman 2001). Research on both monolingual children (Hart 1995), and second language learners (Hoff 2013), has shown that parents from more privileged socioeconomic backgrounds often talk more to their children; use a broader, more sophisticated and more precise range of vocabulary; and engage their children in context-independent conversation more often than parents from lower socioeconomic backgrounds.

For children who are exposed to different languages at home and at school on a daily basis, vocabulary learning is distributed between the two languages (Monsrud 2019; Oller 2007). These children have to develop a command of one language at home and of a different language in preschool settings and school. This may mean

that a larger percentage of their exposure to new words in the language used outside the home occurs at school than is the case for their monolingual peers (Bialystok 2010; Paradis 2009; Webb 2020); this is in addition to new vocabulary in the language at home (Monsrud 2019). Consequently, the time spent with a L2 in educational contexts, such as kindergarten, may be of particular importance for vocabulary learning in a L2. Previous studies have found that both the amount of time spent in preschool (Karlsen 2017), and the teaching quality provided by teachers (Bowers 2011; Grøver 2018; Rydland 2014), are related to children's L2 vocabulary development.

Poor language development that is associated with problematic policies, insufficient exposure to the language and less than optimal teacher preparation and instructional strategies should not be considered a language disorder (OECD 2019). At the same time, being a L2 learner does not preclude language disorders, although multilingualism does not increase the risk of having a DLD (Farnia 2019; Paradis 2016). Monolingual and multilingual children are 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); problems with word finding and semantics often occur in children with DLD, and this group of children might therefore also benefit from more targeted vocabulary interventions.

### Description of the intervention

Systematic and focused vocabulary interventions in preschool settings have been shown to be effective for enhancing vocabulary development in L2 learners (Leacox 2014; Lugo-Neris 2010; Rogde 2016). Although primary studies in this area have yielded promising results, a meta-analysis of 43 studies that included both L1 and L2 learners reported a small overall effect size ( $g = 0.16$ ) of 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 the developmental trajectories of vocabulary skills have been shown to be stable from a young age (Bornstein 2014; Klem 2015), even a slight early enhancement in vocabulary skills might be beneficial for a particular child. However, it remains challenging to determine the approaches to vocabulary interventions that are effective for different age groups, and the extent to which such interventions generalise to unfamiliar and novel vocabulary and literacy tasks.

Vocabulary interventions for L2 learners between birth and six years of age typically aim to provide increased experiences with, and exposure to, words and meanings in the L2 in order to improve the understanding and use of targeted words in social communication. In turn, this may facilitate participation, language skills in general, listening comprehension, learning, access to the curriculum and narrative skills (Hagen 2017; Rogde 2016). New vocabulary knowledge is also expected to support the learning of related unfamiliar words, thus sustaining vocabulary growth over time (Rogde 2016).

Interventions can vary in terms of the intervention approaches, content words, activities, strategies, delivery mode, delivery settings, intervention providers, organisation of the intervention

delivery, the 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 incidentally. Explicit learning 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). Incidental learning includes the perception of an underlying structure without conscious awareness, and is often considered a 'naturalistic approach' in which word learning happens during interactions with children in naturalistic settings, such as during play activities. Caregivers may be taught or coached to engage in interactive behaviours that are thought to support vocabulary development – for example, listening to the child's initiation, naming and modelling of appropriate word labels 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 incidental approaches can also be complementary, and research has demonstrated successful interventions that use a combination of both approaches (Webb 2020).

### Content words and activities

#### Content words

When designing a vocabulary intervention, the starting point is usually selecting keywords to be learned in 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);
- existing books or educational materials, including sets of words that occur in the children's books 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 (morphology of words to transfer to other words with the same prefixes or suffixes; Torkildsen unpublished);
- 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).

If words are predefined, the choice of words is usually based on one or more of the above-mentioned aspects. If words are not predefined, the meaning may be inferred from the context or activity – for example, through a child’s personal story, homemade books or their existing books, rhymes or poems – in which the words may be described or focused upon (or both) during reading or storytelling (Bernhard 2006; Bernhard 2008; Boyce 2004). Note that the number of key words differs between intervention studies; for example, 20 key words were used in the study by Lugo-Neris 2010, 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 has been shown to benefit a diverse group of L2 children (Fitton 2018). However, single studies have produced contradictory results; large positive effects have been found in studies with non-randomised designs, while null results have been found in studies with randomised controlled designs (Fitton 2018). This large variation in results may, in addition to methodological issues, reflect differences in content, such as the choice of target words, or the use of activities or materials originally developed for other purposes and therefore only weakly related to the target words (Lawrence 2014). Interventions that were not adapted to a culturally and linguistically diverse population, such as those developed for a L1, could be another factor in the variations (Larson 2020).

In addition to shared picture book reading, which often includes scaffolding (Rogoff 1990), active listening and inference tasks (Hammer 2016; Van Kleeck 1994), 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 words and closed sentences and the choice of correct words or sentences;
- supporting the use of target words in a broader language context with including activities that target morphology, syntax and phonology (Hagen 2017; Næss unpublished; Stahl 1986); and
- gaming tasks, which are intended to give children comprehensible input and to encourage 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, which emphasised that depending on a single vocabulary instruction activity or strategy does not result in optimal learning. This conclusion was also supported in a meta-analysis by Marulis 2010.

### Delivery mode

Broadly, there are two common delivery modes in vocabulary interventions for young children: face-to-face and the real-time use of digital technology. 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), and there are indications that research-based digital educational interventions may be as effective for learning and retention as conventional delivery mode strategies (Chauhan 2017; Clark 2016). Relevant vocabulary learning strategies for apps include dictionary use or automatic translation (Wood 2018), phonological analysis (De Jong 2000), morphological analysis (Torkildsen unpublished), contextual analysis (Nagy 2001), picture book dialogues (Næss unpublished), and narratives and storytelling (Hur 2012). Apps also provide new, innovative and personal opportunities for vocabulary stimulation. Visual and audio exposure, interactive elements, direct feedback and the possibilities for individual adaptation may lead to both better memory of a word 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; Schuler 2012), but variations in gains in communication have been found, and 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 such digital tools remains unanswered, especially considering that there is 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 intervention providers

L2 vocabulary interventions for young children may be conducted in early education and care (nursery/preschool) settings, at home or as part of healthcare service provision. 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 (Jordan 2000).

### 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 a 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). Results from different clinical samples, however, 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 treatment as usual (business as usual 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 tease apart 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 2020a, business as usual, or standard care, in preschool settings varies greatly between countries in terms of enrolment rates, structures, investment and governance. For example, Nordic countries provide universal access to public sector preschool, while other countries use the private sector or a mixture of public and private. Vermeer 2016 investigating the quality and structural features in 23 countries using Environment Rating Scales found the mean caregiver-child ratio to be 8.6, with a range from 3 to 25 children per caregiver. The results also showed an overall higher average quality of child care in Australia, New Zealand and North America than in South America, Asia and Europe. Preschools in Australia, New Zealand and the USA seem to be more focused on educational outcomes than European preschool settings, and the greater use of quality rating systems, such as the Environment Rating Scales in the USA, might also raise awareness on the importance of caregiver sensitivity in interactions with children among preschool teachers (Vermeer 2016). In other words, treatment as usual will also vary between contexts. In addition, OECD 2020a noted children's experience can also vary within a country in terms of the preschool setting and the staff working in that setting. As book reading and talking about word meanings are activities that usually take place in preschools (OECD 2020a), information about business as usual and standard care is needed to identify the components in both the intervention and control groups that make the intervention different from daily practice.

### 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 of L2 learners' vocabulary. However, the general underlying strategies used in previous interventions can shed some light

on 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 or indirect vocabulary 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 incidentally (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). It can therefore be hypothesised that vocabulary develops continuously as a result of both implicit and explicit learning experiences; knowledge about a word 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 skills before the intervention as a baseline measure (pre-test) in order to compare them to the results after the intervention (post-test). The effects may be assessed immediately after the intervention or after a certain period of time in order to determine its longer-term impacts, or both. Longer-term impacts may also be tested with measures of children's reading comprehension (word-level, sentence-level or passage-level). Testing is typically done 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 the 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 upon 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 skills.

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 activities 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 take part 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 report such outcomes. Finally, a higher rate of attrition in the treatment group than in the control group may indicate negative reactions to the demands of the intervention, which may result in an overestimation of the treatment's benefits, especially in studies that do not employ an intention-to-treat design.

### Why it is important to do this review

As discussed, reduced vocabulary can impede learning in school, leading to academic failure and dropping out. It is therefore 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 will be applied in this systematic review and meta-analysis.

The primary objective of this review is to examine the effect of L2 vocabulary interventions on L2 learners when a rigorous randomised controlled trial (RCT) design is 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 2021a](#)); a review by [Rogde 2019](#) 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.

Previous reviews of L2 vocabulary interventions have included studies conducted exclusively in the English language ([Fitton 2018](#); [Hur 2020](#); [Larson 2020](#)). There is therefore 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, it will be possible to examine how different child characteristics are associated with an intervention's effect.

Previous reviews have variously considered a range of different language skills ([Larson 2020](#)), literacy alone ([Hur 2020](#)), or only shared book reading ([Fitton 2018](#)). The objective of this review is to include studies that were designed with the aim of improving L2 vocabulary skills and to examine how different approaches to vocabulary learning (e.g. explicit word learning, incidental learning in context) and different variables, such as dosage, setting (home versus school) and provider (teaching assistants, parents, teachers, speech and language pathologists), are 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 will look at 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 need to determine what happens to the children after the intervention has ended. By compiling information on different approaches, delivery agents, dosages and child characteristics associated with intervention success, we will 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 will provide crucial knowledge for policy makers who are planning for future resources and support needs.

## OBJECTIVES

The primary objective is 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. The secondary objectives are to examine associations between L2 vocabulary interventions and general characteristics of L2 learners (e.g. age, L2 exposure and L1 skills), as well as specific characteristics of L2 learners who do not appear to benefit from treatment.

## METHODS

### Criteria for considering studies for this review

#### Types of studies

We will include randomised controlled trials (RCTs). We will also include studies that use cluster randomisation as well as



randomisation at the individual level. We will exclude other study designs (e.g. quasi-experimental, within-subjects).

### Types of participants

Eligible studies must report on second language (L2) learners aged up to six years who have participated in L2 vocabulary interventions. The sample can include only children who are 5 years 11 months or younger at pre-test. We will not include studies involving children of six years or older at pre-test; we will determine this based on the author-reported age range at pre-test.

Eligible participants may also be reported by the study author(s) as being L2 learners with developmental language disorder (DLD), who thus have language deficits in both the first language (L1) and L2. We will not include samples including children diagnosed with severe learning or developmental disorders (e.g. autism, sensory impairments, intellectual disability).

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

### Types of interventions

#### Experimental intervention

- We will include any vocabulary intervention that aims 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 are acceptable, with the delivery agent being a teacher (preschool, kindergarten or school), teaching assistant, researcher, speech and language pathologist or caregiver (e.g. parent).
  - \* Interventions may 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) or explicit instruction on target vocabulary skills (Larson 2020)).
  - \* There will be no exclusion criteria based on dose, duration, intensity or different aspects of implementation quality because these aspects will be considered in the risk of bias analysis.

#### Control intervention

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

### Types of outcome measures

We will include studies that meet the above inclusion criteria regardless of whether they report on the primary and secondary outcomes listed below.

#### Primary outcomes

- Receptive L2 vocabulary (both proximal and distal)
- Expressive L2 vocabulary (both proximal and distal)
- Mean length of utterance (potential adverse effects)\*

\*Mean length of utterance is 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.

#### Measurement of outcomes

*Proximal measures* will include taught L2 vocabulary that is measured in terms of either:

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

*Distal measures* will assess L2 vocabulary that is not included among the directly trained words in the intervention. Eligible outcomes will include standardised tests such as:

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

Data should be presented 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 is included as a measure of social and emotional behavior. As this is a wide concept, we will only include this indicator of this theoretical concept.

#### Timing of outcome measures

If available, we will extract and analyse outcome assessments at the first post-test (assessed immediately after the intervention programme) and over a longer term (assessed at least one month after the intervention programme ends). If studies report on more than one long-term follow-up, we will use the last reported time point.

#### Hierarchy of outcome measures

If a study reports on more than one measure for an outcome, we will select the most commonly used measure.

### Search methods for identification of studies

The electronic searches for candidate studies will be shared between the Cochrane Information Specialist for Cochrane Developmental, Psychosocial and Learning Problems (CDPLP) and

the review team. The Information Specialist will conduct searches in all databases that are listed under Electronic searches, except for the search in Linguistics and Language Behavior Abstracts, which will be conducted by the review team. The review team will also conduct the searches outlined in the section [Searching other resources](#).

### Electronic searches

The search strategy for PsycINFO is provided in [Appendix 2](#). This strategy will be adapted for the databases listed below.

- Cochrane Central Register of Controlled Trials (CENTRAL; current issue) in the Cochrane Library, which includes the Cochrane Developmental, Psychosocial and Learning Problems Specialised Register
- MEDLINE Ovid (1946 onwards)
- MEDLINE In-Process and Other Non-Indexed Citations Ovid (current issue)
- MEDLINE Epub Ahead of Print Ovid (current issue)
- Embase Ovid (1974 onwards)
- ERIC EBSCOhost (1966 onwards)
- Education Abstracts (H.W. Wilson) EBSCOhost (1983 onwards)
- Education Database Proquest (1988 onwards)
- Linguistics and Language Behavior Abstracts ProQuest (LLBA; 1973 onwards)
- PsycINFO Ovid (1806 onwards)
- Scopus Elsevier (all available years)
- 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)
- ProQuest Dissertations & Theses Global (1743 onwards)
- Sociological Abstracts ProQuest (1952 onwards)
- *Cochrane Database of Systematic Reviews* (CDSR; current issue), in the Cochrane Library
- Epistemonikos ([www.epistemonikos.org](http://www.epistemonikos.org))
- ClinicalTrials.gov ([www.clinicaltrials.gov](http://www.clinicaltrials.gov))
- WHO International Clinical Trials Registry Platform (WHO ICTRP; [trialsearch.who.int](http://trialsearch.who.int))

We will not limit the searches by year of publication, language of publication or publication type. For publications published in languages other than English we will contact people with knowledge of that specific language in our network or use a language service to get the pertinent information translated.

### Searching other resources

In addition to the electronic searches noted above, we will identify 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, we will handsearch the following journals.

- International Journal of Bilingual Education and Bilingualism
- Bilingualism Research Journal
- Early Education and Development
- Journal Early Childhood Research Quarterly
- Journal of Speech, Language, and Hearing Research

We will search the following grey literature databases.

- OpenGrey ([www.opengrey.eu](http://www.opengrey.eu))
- Google Scholar ([scholar.google.com/](http://scholar.google.com/)). (As Google Scholar does not have a limit on the number of hits, we will screen the first 500 references).

The first review author (HNN) will contact 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. We will run a new search before publication to find out if any of our included studies have been retracted or corrected.

The first review author (HNN) will be responsible for searching these other resources.

### Data collection and analysis

#### Selection of studies

We will import all records yielded by the searches into EndNote and remove any duplicates. We will then export all records to Covidence ([Covidence 2020](#)), where we will remove any remaining duplicates before adopting a two-stage approach to screening; a form will be developed in Covidence to facilitate the screening process based on the inclusion criteria. The first stage will involve screening the titles and abstracts of all records against the eligibility criteria (see [Criteria for considering studies for this review](#)). This will be done by two review authors (HNN, HH) independently to ensure reliability. Records deemed potentially eligible, or those that do not provide sufficient information to evaluate eligibility based on the inclusion criteria will progress to the second stage: screening of full texts. We will import the full texts into Covidence where again, to ensure reliability, two review authors (HNN, HH) will independently screen the texts for inclusion based on the selection criteria. We will record and report the main reasons for any exclusions. At both stages, we will report the inter-rater agreement between the two screeners using the Kappa statistic, and any conflicts between the two screeners will be resolved by consulting a third review author (K-ABN). We will record decisions made throughout the selection process and present these in a PRISMA flow diagram, which will include references to both included and excluded studies and the number of studies assessed at each stage.

#### Data extraction and management

Two review authors (HNN, JK) will independently extract the following information and data from the included studies using Covidence ([Covidence 2020](#)).

- 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, masking, 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, socioeconomic status and other reported risk factors)
- Intervention details (activities, instructional approach, intervention protocols, language of instruction, intervention provider, method of delivery, dosage (frequency and duration), staff qualifications, fidelity, description of business as usual 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.)

Any disagreements in coding will be resolved by consulting a third review author (HCH). We will use the data extraction form in [Appendix 3](#).

### Assessment of risk of bias in included studies

We will assess risk of bias using Cochrane's revised risk of bias tool for randomised trials (RoB 2; [Sterne 2019](#)). Two review authors (HNH, HH) will independently assess each individual primary outcome and the secondary outcomes L2 listening comprehension and narrative skills in the included studies. Both review authors will resolve any conflicts by discussion; a third review author (KABN) will arbitrate, if necessary. We are interested in effects of assignment to intervention, estimated using intention-to-treat (ITT) analyses.

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 will judge the risk of bias in each of the five domains using the RoB 2 signalling questions. We will assess risk of bias in the included primary outcomes at immediate post-test L2 vocabulary (receptive and expressive, both near and distal) and mean length of utterance (potential adverse effects). In addition to the primary outcomes, we will include L2 listening comprehension and narrative skills. We will use templates for randomised parallel-group trials and cluster-randomised parallel-group trials ([Sterne 2019](#)). We will use the Excel tool to make decisions for parallel-group trials ([www.riskofbias.info/welcome/rob-2-0-tool/current-version-of-rob-2](http://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](http://www.riskofbias.info/welcome/rob-2-0-tool/rob-2-for-cluster-randomized-trials)). The Excel spreadsheets with responses to signalling questions will be available as supplementary information.

For cluster-randomised parallel-group trials, there are 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 will be aware of some differences. For example, in domain (1), randomisation could be based on geography, there could be imbalance in cluster or participant characteristics. Also, there is a risk of bias if recruitment of individual eligible participants is done after randomisation of clusters. In domain (2), we will assess if clusters and participants are analysed in their assigned group.

Based on the domain-level judgement of risk of bias, we will reach an overall judgement of risk of bias for the outcome in each included study: we will use '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. We will complete a risk of bias table, with a justification for the judgement, and present it in the published review.

### Measures of treatment effect

#### Continuous data

We expect that outcome measures (see [Types of outcome measures](#)) from L2 vocabulary intervention studies will be reported as continuous variables. We will use the standardised mean difference (SMD) (i.e. Cohen's d or Hedges' g) 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, where possible. If the same measures are used then we will use mean, SD and sample size to compute a mean difference (MD). If different measures are used to explore the same construct we will analyse using SMD. We will use RevMan Web to conduct the meta-analysis of the treatment effect ([RevMan Web 2020](#)), and will present SMD and MD alongside 95% CIs. We will use random-effects meta-analysis.

#### Dichotomous data

We do not expect our chosen outcomes to be presented using dichotomous data.

### Unit of analysis issues

#### Cluster-randomised trials

Randomisation in vocabulary intervention research involving preschool-aged children can be conducted at different levels, such as individuals, department in preschool, preschool centre and geographical level/district levels, depending on the implementation of the intervention. Another type of clustering may relate to pre-existing conditions due to the diversity in the L2 learner group, such as their L1, time since arrival/experience with L2 and experiences from previous interventions. We will include all types of clustered-RCTs.

If corrected data for cluster-RCTs are reported (i.e. intra-cluster correlation coefficient (ICC)), we will use these data in meta-analyses. Where ICC data or values have not been provided, we will contact the authors for further information. If these corrected data are not provided but values that can be used to calculate the intra-cluster correlation are reported, we will estimate corrected data to be used in meta-analyses ([Higgins 2021a](#)).

If a cluster-RCT has not adjusted for clustering in their analysis, we will make the adjustment by multiplying the standard errors of the estimates by the square root of the design effect, where the design effect is calculated as  $1 + (\text{average cluster size} - 1) \times \text{ICC}$  (Higgins 2021b).

### Studies with multiple treatment groups

For studies that compare multiple treatment groups to a control group, we will select the treatment that has received the highest dose or that has the most vocabulary-based intervention before comparing it with the control group. If there is more than one control group (e.g. one of the control groups includes both L1 and L2 learners), we will select the control group with only L2 learners.

### Dealing with missing data

Because incomplete outcome data can introduce bias, we will take steps to collect missing data to follow the ITT principle (Higgins 2021a). In situations with missing outcome data and information, the first review author (HMH) will contact the study authors. We will report missing data in the data extraction form and in the risk of bias tables.

### 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 will test the heterogeneity of effect sizes using the  $\text{Chi}^2$  statistic. This will establish the degree to which the variation in effect size is caused by true heterogeneity and not due to chance (Borenstein 2011). The  $\text{Chi}^2$  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 will report  $\text{Tau}^2$  as an indicator of the magnitude of variation in effect sizes between studies. We will also assess 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 will adhere to the recommended rules of thumb in Section 10.10.2 of the *Cochrane Handbook for Systematic Reviews of Interventions* (Deeks 2021), 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 will interpret statistically significant, unexplained heterogeneity in the results with caution.

### Assessment of reporting biases

To statistically estimate the impact of publication bias, researchers have commonly used funnel plots in combination with a trim-and-fill analysis. If there are a sufficient number of studies (i.e. 10 studies or more) we will create funnel plots using RevMan Web (RevMan Web 2020). Notably, there are several problems associated

with the validity of the funnel plot/trim-and-fill method (Lau 2006), especially when it is used in the presence of large between-study variation (Terrin 2003). Therefore, the results from the funnel plot/trim-and-fill analysis must be interpreted with caution. When interpreting the results, we will consider possible explanations for funnel plot asymmetry, including true heterogeneity of the effect with respect to sample size, bias of small trials and publication bias (Deeks 2021).

### Data synthesis

We will conduct meta-analyses using RevMan Web if outcome data are available from at least two RCTs (RevMan Web 2020). Because we expect that the effect size will vary between studies, we will use random-effect models, as these 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 will conduct a meta-analysis of treatment effects for effect sizes obtained at both immediately post-intervention and after a longer period follow-up (e.g. 6 months after the intervention or 1 year after the intervention) to examine any long-term maintenance of the treatment effect.

In the event that we are unable to perform a meta-analysis, we will provide a narrative summary of the available data instead.

### Subgroup analysis and investigation of heterogeneity

We will conduct subgroup (moderator) analyses with the variables noted below when there are at least 10 studies in the meta-analysis that reported on the moderator variable (Deeks 2021).

- Characteristics of L2 vocabulary interventions
  - \* Instruction: interventions providing explicit word definitions versus interventions not providing explicit word definitions
  - \* Dosage of intervention: duration and amount of training (i.e. number of hours with intervention)
- Sample characteristics
  - \* By age group (samples under three years of age and three years of age and over)
  - \* Samples identified with developmental language disorder (DLD) compared with those who are not identified with a DLD.
  - \* Amount of L2 exposure (number of years in preschool)
  - \* Level of L1 skills (for this analysis we will ask authors of studies that include data in L1 vocabulary skills for the raw data in order to examine if level of L1 skills predicts effect of intervention effect)

### Sensitivity analysis

We plan to perform the following sensitivity analyses to assess the robustness of the results to decisions made throughout the review process.

- We will repeat the analysis after excluding studies in which the overall risk of bias was high.
- We will conduct alternative meta-analyses using the fixed-effect model.

Because we may identify a need for other sensitivity analyses during the review process (Deeks 2021), we will consider conducting additional sensitivity analyses, if necessary.

### Summary of findings and assessment of the certainty of the evidence

We will create a summary of findings table that includes the primary outcomes at immediate post-test L2 vocabulary (both receptive and expressive vocabulary and proximal and distal) and mean length of utterance. In addition to the primary outcomes, we will include L2 listening comprehension and narrative skills in the summary of findings table. We have chosen these secondary outcomes as they are broader language measures that measure mastering of language ability in a more real-life context.

We will assess the certainty of the evidence by using GRADEprofiler (GRADEpro GDT 2020). Two review authors (HNN, HH) will independently assess the certainty of the evidence as high, moderate, low or very low and a third review author (MK) will assist to settle any disagreements. The rating will be based on five domains (inconsistency, indirectness, imprecision, publication bias and the overall risk of bias) to determine how confident we are that the estimated effect reflects the true effect (GRADEpro GDT 2020). We will present the GRADE ratings in the tables and provide reasons for downgrading the certainty in the footnotes of the table.

High-certainty evidence would mean that the true effect would be close to the estimated effect. Moderate-certainty evidence would imply that the true effect is likely close to the estimate but with the possibility that the effect could be substantially different. Low-certainty evidence would reflect that the true effect could be different than the estimated effect. Very low-certainty evidence would imply that the true effect is likely to be different from the estimated effect.

### ACKNOWLEDGEMENTS

The authors would like to give a big thank you to Dr Joanne Duffield (Managing Editor), Dr Sarah Davies (Deputy Managing Editor) and Margaret Anderson (Information Specialist), and Cochrane Developmental, Psychosocial and Learning Problems (CDPLP), for all their expert help and support throughout the preparation of the protocol.

CDPLP editorial team would like to acknowledge the following peer reviewers for their time and comments: Dr Jin Hee Hur, California State University, Bakersfield (CA), USA; Dr Brook Sawyer, Lehigh University, Bethlehem (PA), USA; and Genna White, UK.

## REFERENCES

### Additional references

#### Anderson 2016

Anderson SA, Hawes DJ, Snow PC. Language impairments among youth offenders: a systematic review. *Children and Youth Services Review* 2016;**65**:195-203. [DOI: [10.1016/j.chilcyouth.2016.04.004](https://doi.org/10.1016/j.chilcyouth.2016.04.004)]

#### 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.

#### Beck 2013

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

#### Bernhard 2006

Bernhard JK, Cummins J, Campoy FI, Ada AF, Winsler A, Bleiker C. Identity texts and literacy development among preschool English language learners: enhancing learning opportunities for children at risk for learning disabilities. *Teachers College Record* 2006;**108**(11):2380-405. [DOI: [10.1111/j.1467-9620.2006.00786.x](https://doi.org/10.1111/j.1467-9620.2006.00786.x)] [ERIC #: EJ746180]

#### Bernhard 2008

Bernhard, JK, Winsler A, Bleiker C, Ginieniewicz J, Madigan AL. "Read my story!" Using the Early Authors Program to promote early literacy among diverse, urban preschool children in poverty. *Journal of Education for Students Placed at Risk* 2008;**13**(1):76-105. [DOI: [10.1080/10824660701860458](https://doi.org/10.1080/10824660701860458)]

#### 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). Washington (DC): AERA, 2005:183. [WEB PAGE: Available at [www.aera.net/Portals/38/docs/Annual\\_Meeting/AM\\_2005\\_000\\_Full%20Program.pdf](http://www.aera.net/Portals/38/docs/Annual_Meeting/AM_2005_000_Full%20Program.pdf)]

#### 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 2004

Boyce LK, Cook GA, Roggman LA, Innocenti MS, Jump VK, Akers JF. Sharing books and learning language: what do Latina mothers and their children do? *Early Education and Development* 2004;**15**(4):371-86. [DOI: [10.1207/s15566935eed1504\\_2](https://doi.org/10.1207/s15566935eed1504_2)]

#### 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)]

#### Brekke Stangeland 2017

Brekke Stangeland E. 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]

#### 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)]

#### 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]

#### 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]

Veritas Health Innovation Covidence. Version accessed 15 January 2020. Melbourne, Australia: Veritas Health Innovation, 2020. Available at [covidence.org](http://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)]

#### 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]

#### Dakun 2000

Dakun W. Vocabulary acquisition: implicit learning and explicit teaching. *REACT* 2000;**2**:15-22. [WEB PAGE: Available at [repository.nie.edu.sg/bitstream/10497/3825/3/REACT-2000-2-15.pdf](http://repository.nie.edu.sg/bitstream/10497/3825/3/REACT-2000-2-15.pdf)]

#### 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](http://www.jstor.org/stable/41385916)]

#### Deeks 2021

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 6.2 (updated February 2021). Cochrane, 2021. Available from [www.training.cochrane.org/handbook](http://www.training.cochrane.org/handbook).

#### 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]

#### 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 2009

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

#### Dunn 2018

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

#### 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](http://www.sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool).

#### Ellis 1994

Ellis NC. Vocabulary acquisition: the implicit ins and outs of explicit cognitive mediation. In: Ellis NC, editor(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)]

**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]

**GRADEpro GDT 2020 [Computer program]**

McMaster University (developed by Evidence Prime) GRADEpro GDT. Version accessed 15 January 2020. Hamilton (ON): McMaster University (developed by Evidence Prime), 2020. Available from [gradepro.org](http://gradepro.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)]

**Gremillion 2014**

Gremillion ML, Martel MM. Merely misunderstood? Receptive, expressive, and pragmatic language in young children with disruptive behavior disorders. *Journal of Clinical Child & Adolescent Psychology* 2014;**43**(5):765-76. [DOI: [10.1080/15374416.2013.822306](https://doi.org/10.1080/15374416.2013.822306)] [PMCID: PMC4311524] [PMID: 23924073]

**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 2018**

Grøver V, Lawrence J, Rydland V. Bilingual preschool children's second-language vocabulary development: the role of first-language vocabulary skills and second-language talk input. *International Journal of Bilingualism* 2018;**22**(2):234-50. [DOI: [10.1177/1367006916666389](https://doi.org/10.1177/1367006916666389)]

**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](http://journals.uncc.edu/dialog/article/view/421)]

**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.

**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 2021a**

Higgins JP, Lasserson T, Chandler J, Tovey D, Thomas, J, Flemyng E, et al. Methodological Expectations of Cochrane Intervention Reviews. London (UK): Cochrane, Version February 2021.

**Higgins 2021b**

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 6.1 (updated February 2021). Cochrane, 2021. Available from [training.cochrane.org/handbook](http://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: [10.1177/0271121419894623](https://doi.org/10.1177/0271121419894623)]

#### 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]

#### Jordan 2000

Jordan GE, Snow CE, Porche MV. Project EASE: the effect of a family literacy project on kindergarten students' early literacy skills. *Reading Research Quarterly* 2000;**35**(4):524-46. [DOI: [10.1598/RRQ.35.4.5](https://doi.org/10.1598/RRQ.35.4.5)] [WEB PAGE: Available at [www.jstor.org/stable/748097](http://www.jstor.org/stable/748097)]

#### 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 S-AH, 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]

#### Larson 2020

Larson AL, Cycyk 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)]

#### Lau 2006

Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *BMJ* 2006;**333**(7568):597-600. [DOI: [10.1136/bmj.333.7568.597](https://doi.org/10.1136/bmj.333.7568.597)] [PMCID: PMC1570006] [PMID: 16974018]

#### 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-Y, 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]

#### 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]

#### 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]

**Monsrud 2019**

Monsrud M-B, Rydland V, Geva E, Thurmann-Moe AC, Lyster S-AH. The advantages of jointly considering first and second language vocabulary skills among emergent bilingual children. *International Journal of Bilingual Education and Bilingualism* 2019 Jun 18 [Epub ahead of print]. [DOI: [10.1080/13670050.2019.1624685](https://doi.org/10.1080/13670050.2019.1624685)]

**Næss 2017**

Næss K-AB, Nygaard E, Ostad J, Dolva A-S, Lyster S-AH. The profile of social functioning in children with Down syndrome. *Disability and Rehabilitation* 2017;**39**(13):1320-31. [DOI: [10.1080/09638288.2016.1194901](https://doi.org/10.1080/09638288.2016.1194901)] [PMID: 27442783]

**Næss unpublished**

Næss K-AB, Hokstad S, Engevik LI, Lervåg A, Smith E. A randomized trial of the Digital Down Syndrome Language Plus (DSL+) vocabulary intervention programme. Unpublished manuscript.

**Nagy 2001**

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.

**NCES 2020**

National Center for Education Statistics (NCES). The Condition of Education 2020. English Language Learners in Public Schools; May 2020. Available at [nces.ed.gov/programs/coe/indicator\\_cgf.asp](https://nces.ed.gov/programs/coe/indicator_cgf.asp).

**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](https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf).

**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**(11):1247-57. [DOI: [10.1111/jcpp.12573](https://doi.org/10.1111/jcpp.12573)] [PMCID: PMC5082564] [PMID: 27184709]

**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. 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)]

**OECD 2020b**

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

**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)]

**Paradis 2016**

Paradis J. An agenda for knowledge-oriented research on bilingualism in children with developmental disorders. *Journal of Communication Disorders* 2016;**63**:79-84. [DOI: [10.1016/j.jcomdis.2016.08.002](https://doi.org/10.1016/j.jcomdis.2016.08.002)] [PMID: 27663904]

**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)]

**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, Haring 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/s11145-011-9351-6](https://doi.org/10.1007/s11145-011-9351-6). [DOI: [10.1007/s11145-011-9336-5](https://doi.org/10.1007/s11145-011-9336-5)]

#### 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]

The Cochrane Collaboration Review Manager Web (RevMan Web). Version 1.22.0. The Cochrane Collaboration, 2020. Available at [revman.cochrane.org](http://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 2014

Rydland V, Grøver V, Lawrence J. The second-language vocabulary trajectories of Turkish immigrant children in Norway from ages five to ten: the role of preschool talk exposure, maternal education, and co-ethnic concentration in the neighborhood. *Journal of Child Language* 2014;**41**(2):352-81. [DOI: [10.1017/S0305000912000712](https://doi.org/10.1017/S0305000912000712)] [PMID: 23442820]

#### Schuler 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]

#### 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.

#### 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](http://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]

#### Terrin 2003

Terrin N, Schmid CH, Lau J, Olkin I. Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine* 2003;**22**(13):2113-26. Erratum in: *Statistics in Medicine* 2005;**24**(5):825-6. DOI: [10.1002/sim.2049](https://doi.org/10.1002/sim.2049). [DOI: [10.1002/sim.1461](https://doi.org/10.1002/sim.1461)] [PMID: 12820277]

#### 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 unpublished

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. Unpublished manuscript.

### 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 2016

Vermer 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**(1):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.

### 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]

### Williams 2007

Williams KT. *Expressive Vocabulary Test*. 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]

### 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)]

## 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

(Continued)

Fitton 2018	Meta-analysis	"The purpose of the present meta-analysis is 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)	<ul style="list-style-type: none"> <li>• PsycINFO, Education Resource Information Center (ERIC), MEDLINE, Academic Search Premier, and ProQuest Social Sciences</li> <li>• Studies published in English between 1 January 1981 and 30 April 2017</li> </ul>	Empirical studies that used experimental design (quantitative)	Shared book reading	<ul style="list-style-type: none"> <li>• Participant sample comprising at least 80% ELs</li> <li>• Study was conducted in the USA</li> <li>• Age: 12 years old or younger</li> </ul>
Hur 2020	Systematic review	"The purposes of this systematic review were to describe key features of English early literacy interventions provided to children who were DLLs and their effects on English early literacy skills." (quote)	<ul style="list-style-type: none"> <li>• PsycINFO, Education Resource Information Center (ERIC), MEDLINE, Academic Search Premier, and ProQuest Social Sciences</li> <li>• Peer-reviewed journals</li> <li>• Published in English, through to August 2016</li> </ul>	<ul style="list-style-type: none"> <li>• Group experimental research designs in which participants were randomly assigned to conditions</li> <li>• Quasi-experimental research designs</li> <li>• SCEDs</li> </ul>	English early literacy interventions	<ul style="list-style-type: none"> <li>• DLLs (English learners)</li> <li>• Children with or without disability</li> <li>• Age: birth to age 5 years</li> </ul>
Larson 2020	Systematic review	"The purpose was to discuss how cultural and linguistic factors were addressed in the interventions, examine the methodological rigor of the studies, 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)	<ul style="list-style-type: none"> <li>• PsycINFO, ERIC, MEDLINE, PubMed, Academic Search Complete, Web of Science, and Google Scholar</li> <li>• Peer-reviewed articles</li> <li>• Published in English between 1975 and 2015</li> </ul>	<ul style="list-style-type: none"> <li>• Randomised or non-randomised experimental designs, including single-case experimental designs with at least two participants and adequate experimental control</li> </ul>	Interventions focused on four areas: <ul style="list-style-type: none"> <li>• explicit instruction on targeted skills;</li> <li>• classroom curriculum interventions;</li> <li>• interactive book reading and/or book making interventions; and</li> </ul>	<ul style="list-style-type: none"> <li>• Young children from CLD backgrounds</li> <li>• Age: birth to age 5</li> </ul>

(Continued)

- natural-istic, routines-based inter-ventions

---

### Footnotes

**CLD:** culturally and linguistically diverse; **DLLs:** dual language learners; **ELs:** English learners; **SCEDs:** single case experimental designs

---

## Appendix 2. Search strategy for PsycINFO Ovid

- 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).tw.
- 11 (second\$ language\$ or 2L).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 randomiz\$ or randomiz\$).tw.
- 36 ((control\$ or experimental) adj5 group\$).tw.
- 37 or/24-36
- 38 23 and 37

## Appendix 3. Data extraction form

Variable information	Description, as stated in paper/report	Location in text or source (page number & figure/table/other)
<b>Reference details</b>		
<b>Review title or ID (identifier):</b>		
<b>Study ID</b> (surname of first author and year first full report of study was published, e.g. Smith 2001):		
<b>Report ID:</b>		
<b>Report ID of other reports of this study, including errata or retractions:</b>		
<b>Notes:</b>		
<b>Methods</b>		
Aim of study		
Design (e.g. randomised controlled trial (RCT) or quasi-RCT)		
Unit of allocation (by individuals, cluster/groups)		
Start date		
End date		
Duration of participation (from recruitment to last follow-up)		
Ethical approval needed/obtained for study		Specify: Yes/No/Unclear
Notes		
<b>Participants</b>		
Participants characteristics (e.g. L2 learners with DLD)		
Method of recruitment of participants (e.g. phone, mail, social media, municipality, clinic)		
Informed consent obtained		Specify: Yes/ No/Unclear
Total number randomised		
Clusters (if applicable, number, type, number of people per cluster)		
Withdrawals and exclusions (if not provided below by outcome)		
Age at recruitment (M and SD)		
Sex		
Language spoken: L1		

(Continued)

Language treatment given delivered in (L2)

Socioeconomic status: indicator and M, SD

Subgroups measured

Subgroups reported

Notes

**Intervention programme details**

Description of treatment and name of treatment programme

Description of the vocabulary activities in the intervention

Description of criteria for selecting focus words

Number randomised to group (specify whether number of people or clusters)

Duration of treatment period (number of weeks)

Duration of treatment period (hours per week)

Delivery agent (e.g. preschool teacher, teacher, parent, SLP)

Resource requirements (e.g. staff numbers)

Compliance

Description of content of control condition

Notes

**Outcomes** (copy and paste the following rows for each outcome)

Name of outcome

Expressive or receptive measure

Standardised or researcher made

Time points measured (pre, post, follow-up)

Scales: upper and lower limits (indicate whether high or low score is good)

Imputation of missing data (e.g. assumptions made for intention-to-treat analysis)

Assumed risk estimate (e.g. baseline or population risk noted in Background)

Power (e.g. power and sample size calculation, level of power achieved)

Notes

**Data and analysis** (copy and paste the appropriate rows for each outcome, including additional rows for each time point and subgroup)



(Continued)

Outcome (name of variable)

Intervention group: mean

Intervention group: standard deviation (SD) (or other variance, specify)

Intervention group: number of participants

Control group: mean

Control group: standard deviation (SD) (or other variance, specify)

Control group: number of participants

Any other results reported (e.g. mean difference, confidence interval, P value)

Intervention group: number of missing participants

Intervention group: reasons missing

Control group: number of missing participants

Control group: reasons missing

Unit of analysis (individuals, cluster/groups)

Statistical methods used and appropriateness of these (e.g. adjustment for correlation)

Reanalysis required?

Specify: Yes/No/Unclear

Reanalysis possible?

Specify: Yes/No/Unclear

Reanalysis results

Notes

**Conclusions and supplementary information**

Key conclusions of study authors

References to other relevant studies

Correspondence required for further study information (from and to whom, for what and when)

Notes

*Footnotes*

DLD: developmental language disorder; ID: identifier; L1: first language; L2: second language; M: mean; RCT: randomised controlled trial; SD: standard deviation; SLP: speech-language pathologists.

## Appendix 4. Revised Cochrane risk of bias tool for randomised trials (RoB 2)

### Domain 1. Risk of bias arising from the randomisation process

Signalling questions	Elaboration	Response options
<b>1.1 Was the allocation sequence random?</b>	<p>Answer 'Yes' if a random component was used in the sequence generation process. Examples include computer-generated random numbers; reference to a random number table; coin tossing; shuffling cards or envelopes; throwing dice; or drawing lots. Minimisation is generally implemented with a random element (at least when the scores are equal), so an allocation sequence that is generated using minimisation should generally be considered to be random.</p> <p>Answer 'No' if no random element was used in generating the allocation sequence or the sequence is predictable. Examples include alternation; methods based on dates (of birth or admission); patient record numbers; allocation decisions made by clinicians or participants; allocation based on the availability of the intervention; or any other systematic or haphazard method.</p> <p>Answer 'No information' if the only information about randomisation methods is a statement that the study is randomised.</p> <p>In some situations a judgement may be made to answer 'Probably no' or 'Probably yes'. For example, in the context of a large trial run by an experienced clinical trials unit, absence of specific information about generation of the randomisation sequence, in a paper published in a journal with rigorously enforced word count limits, is likely to result in a response of 'Probably yes' rather than 'No information'. Alternatively, if other (contemporary) trials by the same investigator team have clearly used non-random sequences, it might be reasonable to assume that the current study was done using similar methods.</p>	Y/PY/PN/N/NI
<b>1.2 Was the allocation sequence concealed until participants were enrolled and assigned to interventions?</b>	<p>Answer 'Yes' if the trial used any form of remote or centrally administered method to allocate interventions to participants, where the process of allocation is controlled by an external unit or organisation, independent of the enrolment personnel (e.g. independent central pharmacy, telephone or internet-based randomisation service providers).</p> <p>Answer 'Yes' if envelopes or drug containers were used appropriately. Envelopes should be opaque, sequentially numbered, sealed with a tamper-proof seal and opened only after the envelope has been irreversibly assigned to the participant. Drug containers should be sequentially numbered and of identical appearance, and dispensed or administered only after they have been irreversibly assigned to the participant. This level of detail is rarely provided in reports, and a judgement may be required to justify an answer of 'Probably yes' or 'Probably no'.</p> <p>Answer 'No' if there is reason to suspect that the enrolling investigator or the participant had knowledge of the forthcoming allocation.</p>	Y/PY/PN/N/NI
<b>1.3 Did baseline differences between intervention groups suggest a problem with the randomisation process?</b>	<p>Note that differences that are compatible with chance do not lead to a risk of bias. A small number of differences identified as 'statistically significant' at the conventional 0.05 threshold should usually be considered to be compatible with chance.</p> <p>Answer 'No' if no imbalances are apparent or if any observed imbalances are compatible with chance.</p> <p>Answer 'Yes' if there are imbalances that indicate problems with the randomisation process, including:</p>	Y/PY/PN/N/NI

(Continued)

(1) substantial differences between intervention group sizes, compared with the intended allocation ratio;

OR

(2) a substantial excess in statistically significant differences in baseline characteristics between intervention groups, beyond that expected by chance;

OR

(3) imbalance in one or more key prognostic factors, or baseline measures of outcome variables, that is very unlikely to be due to chance and for which the between-group difference is big enough to result in bias in the intervention effect estimate.

Also answer 'Yes' if there are other reasons to suspect that the randomisation process was problematic:

(4) excessive similarity in baseline characteristics that is not compatible with chance.

Answer 'No information' when there is no useful baseline information available (e.g. abstracts, or studies that reported only baseline characteristics of participants in the final analysis).

The answer to this question should not influence answers to questions 1.1 or 1.2. For example, if the trial has large baseline imbalances, but study authors report adequate randomisation methods, questions 1.1 and 1.2 should still be answered on the basis of the reported adequate methods, and any concerns about the imbalance should be raised in the answer to question 1.3 and reflected in the domain-level 'Risk of bias' judgement.

Trialists may undertake analyses that attempt to deal with flawed randomisation by controlling for imbalances in prognostic factors at baseline. To remove the risk of bias caused by problems in the randomisation process, it would be necessary to know, and measure, all the prognostic factors that were imbalanced at baseline. It is unlikely that all important prognostic factors are known and measured, so such analyses will, at best, reduce the risk of bias. If review authors wish to assess the risk of bias in a trial that controlled for baseline imbalances in order to mitigate failures of randomisation, the study should be assessed using the ROBINS-I tool.

<b>Risk of bias judgement</b>	See algorithm provided in <a href="#">Sterne 2019</a> , p 6	Low/High/Some concerns
-------------------------------	---	------------------------

## Domain 2. Risk of bias due to deviations from the intended interventions (effect of assignment to intervention)

Signalling questions	Elaboration	Response options
<b>2.1. Were participants aware of their assigned intervention during the trial?</b>	If participants are aware of their assigned intervention it is more likely that health-related behaviours will differ between the intervention groups. Blinding participants, most commonly through use of a placebo or sham intervention, may prevent such differences. If participants experienced side effects or toxicities that they knew to be specific to one of the interventions, answer this question 'Yes' or 'Probably yes'.	Y/PY/PN/N/NI

(Continued)

<p><b>2.2. Were carers and people delivering the interventions aware of participants' assigned intervention during the trial?</b></p>	<p>If carers or people delivering the interventions are aware of the assigned intervention then its implementation, or administration of non-protocol interventions, may differ between the intervention groups. Blinding may prevent such differences. If participants experienced side effects or toxicities that carers or people delivering the interventions knew to be specific to one of the interventions, answer question 'Yes' or 'Probably yes'. If randomised allocation was not concealed, then it is likely that carers and people delivering the interventions were aware of participants' assigned intervention during the trial.</p>	<p>Y/PY/PN/N/NI</p>
<p><b>2.3. If Y/PY/NI to 2.1 or 2.2: were there deviations from the intended intervention that arose because of the trial context?</b></p>	<p>For the effect of assignment to intervention, this domain assesses problems that arise when changes from assigned intervention that are inconsistent with the trial protocol arose because of the trial context. We use the term trial context to refer to effects of recruitment and engagement activities on trial participants and when trial personnel (carers or people delivering the interventions) undermine the implementation of the trial protocol in ways that would not happen outside the trial. For example, the process of securing informed consent may lead participants subsequently assigned to the comparator group to feel unlucky and therefore seek the experimental intervention, or other interventions that improve their prognosis.</p> <p>Answer 'Yes' or 'Probably yes' only if there is evidence, or strong reason to believe, that the trial context led to failure to implement the protocol interventions or to implementation of interventions not allowed by the protocol.</p> <p>Answer 'No' or 'Probably no' if there were changes from assigned intervention that are inconsistent with the trial protocol, such as non-adherence to intervention, but these are consistent with what could occur outside the trial context.</p> <p>Answer 'No' or 'Probably no' for changes to intervention that are consistent with the trial protocol, for example, cessation of a drug intervention because of acute toxicity or use of additional interventions whose aim is to treat consequences of one of the intended interventions.</p> <p>If blinding is compromised because participants report side effects or toxicities that are specific to one of the interventions, answer 'Yes' or 'Probably yes' only if there were changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context.</p> <p>The answer 'No information' may be appropriate, because trialists do not always report whether deviations arose because of the trial context.</p>	<p>NA/Y/PY/PN/N/NI</p>
<p><b>2.4 If Y/PY to 2.3: were these deviations likely to have affected the outcome?</b></p>	<p>Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context will impact on the intervention effect estimate if they affect the outcome, but not otherwise.</p>	<p>NA/Y/PY/PN/N/NI</p>
<p><b>2.5. If Y/PY/NI to 2.4: were these deviations from intended intervention balanced between groups?</b></p>	<p>Changes from assigned intervention that are inconsistent with the trial protocol and arose because of the trial context are more likely to impact on the intervention effect estimate if they are not balanced between the intervention groups.</p>	<p>NA/Y/PY/PN/N/NI</p>
<p><b>2.6 Was an appropriate analysis used to estimate the effect of assignment to intervention?</b></p>	<p>Both intention-to-treat (ITT) analyses and modified intention-to-treat (mITT) analyses excluding participants with missing outcome data should be considered appropriate. Both naive 'per-protocol' analyses (excluding trial participants who did not receive their assigned intervention) and 'as treated' analyses (in which trial participants are grouped according to the intervention that they received, rather than according to their assigned intervention) should be considered inappropriate. Analyses excluding eligible trial participants post-randomisation should also be considered inappropriate, but post-randomisa-</p>	<p>Y/PY/PN/N/NI</p>

(Continued)

tion exclusions of ineligible participants (when eligibility was not confirmed until after randomisation, and could not have been influenced by intervention group assignment) can be considered appropriate.

<b>2.7 If N/PN/NI to 2.6: was there potential for a substantial impact (on the result) of the failure to analyse participants in the group to which they were randomised?</b>	This question addresses whether the number of participants who were analysed in the wrong intervention group, or excluded from the analysis, was sufficient that there could have been a substantial impact on the result. It is not possible to specify a precise rule: there may be potential for substantial impact even if fewer than 5% of participants were analysed in the wrong group or excluded, if the outcome is rare or if exclusions are strongly related to prognostic factors.	NA/Y/PY/PN/N/NI
<b>Risk of bias judgement</b>	See algorithm provided in <a href="#">Sterne 2019</a> , p10	Low/High/Some concerns
<b>Optional: what is the predicted direction of bias due to deviations from intended interventions?</b>	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterised either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA/Favours experimental/Favours comparator/Towards null/Away from null/Unpredictable

### Domain 3. Missing outcome data

Signalling questions	Elaboration	Response options
<b>3.1 Were data for this outcome available for all, or nearly all, participants randomised?</b>	<p>The appropriate study population for an analysis of the ITT effect is all randomised participants.</p> <p>'Nearly all' should be interpreted that as the number of participants with missing outcome data is sufficiently small that their outcomes, whatever they were, could have made no important difference to the estimated effect of intervention.</p> <p>For continuous outcomes, availability of data from 95% of the participants will often be sufficient. For dichotomous outcomes, the proportion required is directly linked to the risk of the event. If the observed number of events is much greater than the number of participants with missing outcome data, the bias would necessarily be small.</p> <p>Only answer 'No information' if the trial report provides no information about the extent of missing outcome data. This situation will usually lead to a judgement that there is a high risk of bias due to missing outcome data.</p> <p>Note that imputed data should be regarded as missing data, and not considered as 'outcome data' in the context of this question.</p>	Y/PY/PN/N/NI
<b>3.2 If N/PN/NI to 3.1: is there evidence that the result was not biased by missing outcome data?</b>	Evidence that the result was not biased by missing outcome data may come from: (1) analysis methods that correct for bias; or (2) sensitivity analyses showing that results are little changed under a range of plausible assumptions about the relationship between missingness in the outcome and its true value. However, imputing the outcome variable, either through methods such as 'last observation carried forward' or via multiple imputation based only on inter-	NA/Y/PY/PN/N

(Continued)

vention group, should not be assumed to correct for bias due to missing outcome data.

**3.3 If N/PN to 3.2: could missingness in the outcome depend on its true value?**

If loss to follow-up, or withdrawal from the study, could be related to participants' health status, then it is possible that missingness in the outcome was influenced by its true value. However, if all missing outcome data occurred for documented reasons that are unrelated to the outcome then the risk of bias due to missing outcome data will be low (for example, failure of a measuring device or interruptions to routine data collection).

NA/Y/PY/PN/N/NI

In time-to-event analyses, participants censored during trial follow-up, for example, because they withdrew from the study, should be regarded as having missing outcome data, even though some of their follow-up is included in the analysis. Note that such participants may be shown as included in analyses in CONSORT flow diagrams.

**3.4 If Y/PY/NI to 3.3: is it likely that missingness in the outcome depended on its true value?**

This question distinguishes between situations in which (i) missingness in the outcome could depend on its true value (assessed as 'Some concerns') from those in which (ii) it is likely that missingness in the outcome depended on its true value (assessed as 'High risk of bias'). Five reasons for answering 'Yes' are:

NA/Y/PY/PN/N/NI

1. Differences between intervention groups in the proportions of missing outcome data. If there is a difference between the effects of the experimental and comparator interventions on the outcome, and the missingness in the outcome is influenced by its true value, then the proportions of missing outcome data are likely to differ between intervention groups. Such a difference suggests a risk of bias due to missing outcome data, because the trial result will be sensitive to missingness in the outcome being related to its true value. For time-to-event data, the analogue is that rates of censoring (loss to follow-up) differ between the intervention groups.

2. Reported reasons for missing outcome data provide evidence that missingness in the outcome depends on its true value.

3. Reported reasons for missing outcome data differ between the intervention groups.

4. The circumstances of the trial make it likely that missingness in the outcome depends on its true value. For example, in trials of interventions to treat schizophrenia it is widely understood that continuing symptoms make drop out more likely.

5. In time-to-event analyses, participants' follow-up is censored when they stop or change their assigned intervention, for example, because of drug toxicity or, in cancer trials, when participants switch to second-line chemotherapy.

Answer 'No' if the analysis accounted for participant characteristics that are likely to explain the relationship between missingness in the outcome and its true value.

**Risk of bias judgement**

See algorithm provided in [Sterne 2019](#), p 16

Low/High/Some concerns

**Optional: what is the predicted direction of bias due to missing outcome data?**

If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterised either as being towards (or away from) the null, or as being in favour of one of the interventions.

NA/Favours experimental/Favours comparator/Towards null/Away from null/Unpredictable

**Domain 4. Risk of bias in measurement of the outcome**

Signalling questions	Elaboration	Response options
<b>4.1 Was the method of measuring the outcome inappropriate?</b>	<p>This question aims to identify methods of outcome measurement (data collection) that are unsuitable for the outcome they are intended to evaluate. The question <i>does not</i> aim to assess whether the choice of outcome being evaluated was sensible (e.g. because it is a surrogate or proxy for the main outcome of interest). In most circumstances, for prespecified outcomes, the answer to this question will be 'No' or 'Probably no'.</p> <p>Answer 'Yes' or 'Probably yes' if the method of measuring the outcome is inappropriate, for example because:</p> <p>(1) it is unlikely to be sensitive to plausible intervention effects (e.g. important ranges of outcome values fall outside levels that are detectable using the measurement method); or</p> <p>(2) the measurement instrument has been demonstrated to have poor validity.</p>	Y/PY/PN/N/NI
<b>4.2 Could measurement or ascertainment of the outcome have differed between intervention groups?</b>	<p>Comparable methods of outcome measurement (data collection) involve the same measurement methods and thresholds, used at comparable time points. Differences between intervention groups may arise because of 'diagnostic detection bias' in the context of passive collection of outcome data, or if an intervention involves additional visits to a healthcare provider, leading to additional opportunities for outcome events to be identified.</p>	Y/PY/PN/N/NI
<b>4.3 If N/PN/NI to 4.1 and 4.2: were outcome assessors aware of the intervention received by study participants?</b>	<p>Answer 'No' if outcome assessors were blinded to intervention status. For participant-reported outcomes, the outcome assessor is the study participant.</p>	NA/Y/PY/PN/N/NI
<b>4.4 If Y/PY/NI to 4.3: could assessment of the outcome have been influenced by knowledge of intervention received?</b>	<p>Knowledge of the assigned intervention could influence participant-reported outcomes (such as level of pain), observer-reported outcomes involving some judgement, and intervention provider decision outcomes. They are unlikely to influence observer - outcomes that do not involve judgement; for example, all-cause mortality.</p>	NA/Y/PY/PN/N/NI
<b>4.5 If Y/PY/NI to 4.4: is it likely that assessment of the outcome was influenced by knowledge of intervention received?</b>	<p>This question distinguishes between situations in which (i) knowledge of intervention status could have influenced outcome assessment but there is no reason to believe that it did (assessed as 'Some concerns') from those in which (ii) knowledge of intervention status was likely to influence outcome assessment (assessed as 'High'). When there are strong levels of belief in either beneficial or harmful effects of the intervention, it is more likely that the outcome was influenced by knowledge of the intervention received. Examples may include patient-reported symptoms in trials of homeopathy, or assessments of recovery of function by a physiotherapist who delivered the intervention.</p>	NA/Y/PY/PN/N/NI
<b>'Risk of bias' judgement</b>	<p>See algorithm provided in <a href="#">Sterne 2019</a>, p 18</p>	Low/High/Some concerns
<b>Optional: what is the predicted direction of bias in measurement of the outcome?</b>	<p>If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterised either as being towards (or away from) the null, or as being in favour of one of the interventions.</p>	NA/Favours experimental/Favours comparator/Towards null/Away from null/Unpredictable

**Domain 5. Risk of bias in selection of the reported result**

Signalling questions	Elaboration	Response options
<b>5.1 Were the data that produced this result analysed in accordance with a prespecified analysis plan that was finalised before unblinded outcome data were available for analysis?</b>	<p>If the researchers' prespecified intentions are available in sufficient detail, then planned outcome measurements and analyses can be compared with those presented in the published report(s). To avoid the possibility of selection of the reported result, finalisation of the analysis intentions must precede availability of unblinded outcome data to the trial investigators.</p> <p>Changes to analysis plans that were made before unblinded outcome data were available, or that were clearly unrelated to the results (e.g. due to a broken machine making data collection impossible) do not raise concerns about bias in selection of the reported result.</p>	Y/PY/PN/N/NI
<b>Is the numerical result being assessed likely to have been selected, on the basis of the results, from...</b>		
<b>5.2. ... multiple eligible outcome measurements (e.g. scales, definitions, time points) within the outcome domain?</b>	<p>A particular outcome domain (i.e. a true state or endpoint of interest) may be measured in multiple ways. For example, the domain pain may be measured using multiple scales (e.g. a visual analogue scale and the McGill Pain Questionnaire), each at multiple time points (e.g. 3, 6 and 12 weeks post-treatment). If multiple measurements were made, but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to outcome measurements that are eligible for consideration by the RoB 2 tool user. For example, if only a result using a specific measurement scale is eligible for inclusion in a meta-analysis (e.g. Hamilton Depression Rating Scale), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from a different measurement scale (e.g. Beck Depression Inventory).</p> <p>Answer 'Yes' or 'Probably yes' if:</p> <p>There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a domain was measured in multiple eligible ways, but data for only one or a subset of measures are fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results can arise from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a preconception, or vested interest in showing, that an experimental intervention is beneficial may be inclined to report outcome measurements selectively that are favourable to the experimental intervention.</p> <p>Answer 'No' or 'Probably no' if:</p> <p>There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome domain correspond to all intended outcome measurements.</p> <p>OR</p> <p>There is only one possible way in which the outcome domain can be measured (hence there is no opportunity to select from multiple measures).</p> <p>OR</p>	Y/PY/PN/N/NI



(Continued)

Outcome measurements are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results.

Answer 'No information' if:

Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment, and there is more than one way in which the outcome domain could have been measured.

### 5.3 ... multiple eligible analyses of the data?

A particular outcome measurement may be analysed in multiple ways. Examples include: unadjusted and adjusted models; final value versus change from baseline versus analysis of covariance; transformations of variables; different definitions of composite outcomes (e.g. 'major adverse effect'); conversion of continuously scaled outcome to categorical data with different cut-off points; different sets of covariates for adjustment; and different strategies for dealing with missing data. Application of multiple methods generates multiple effect estimates for a specific outcome measurement. If multiple estimates are generated but only one or a subset is reported on the basis of the results (e.g. statistical significance), there is a high risk of bias in the fully reported result. Attention should be restricted to analyses that are eligible for consideration by the RoB 2 tool user. For example, if only the result from an analysis of post-intervention values is eligible for inclusion in a meta-analysis (e.g. at 12 weeks after randomisation), and this is reported by the trial, then there would not be an issue of selection even if this result was reported (on the basis of the results) in preference to the result from an analysis of changes from baseline.

Y/PY/PN/N/NI

Answer 'Yes' or 'Probably yes' if:

There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that a measurement was analysed in multiple eligible ways, but data for only one or a subset of analyses are fully reported (without justification), and the fully reported result is likely to have been selected on the basis of the results. Selection on the basis of the results arises from a desire for findings to be newsworthy, sufficiently noteworthy to merit publication, or to confirm a prior hypothesis. For example, trialists who have a pre-conception or vested interest in showing that an experimental intervention is beneficial may be inclined to selectively report analyses that are favourable to the experimental intervention.

Answer 'No' or 'Probably no' if:

There is clear evidence (usually through examination of a trial protocol or statistical analysis plan) that all eligible reported results for the outcome measurement correspond to all intended analyses.

OR

There is only one possible way in which the outcome measurement can be analysed (hence there is no opportunity to select from multiple analyses).

OR

Analyses are inconsistent across different reports on the same trial, but the trialists have provided the reason for the inconsistency and it is not related to the nature of the results.

Answer 'No information' if:

Analysis intentions are not available, or the analysis intentions are not reported in sufficient detail to enable an assessment, and there is more than one way in which the outcome measurement could have been analysed.

(Continued)

<b>Risk of bias judgement</b>	See algorithm provided in <a href="#">Sterne 2019</a> , p 23	Low/High/Some concerns
<b>Optional: what is the predicted direction of bias due to selection of the reported result?</b>	If the likely direction of bias can be predicted, it is helpful to state this. The direction might be characterised either as being towards (or away from) the null, or as being in favour of one of the interventions.	NA/Favours experimental/Favours comparator/Towards null/Away from null/Unpredictable

## Overall risk of bias

Overall risk of bias judgement	Low/High/Some concerns
Optional: what is the overall predicted direction of bias for this outcome?	Favours experimental/Favours comparator/Towards null/Away from null/Unpredictable/Not applicable

Overall risk of bias judgement	Criteria
Low risk of bias	The study is judged to be at <b>low risk of bias for all domains</b> for this result.
Some concerns	The study is judged to raise <b>some concerns</b> in at least one domain for this result, but not to be at high risk of bias for any domain.
High risk of bias	The study is judged to be at <b>high risk of bias</b> in at least one domain for this result.  Or  The study is judged to have <b>some concerns</b> for <b>multiple domains</b> in a way that substantially lowers confidence in the result.

### Footnotes

Content retrieved from [Sterne 2019](#).

N: no; NA: not applicable; NI: no information; PY: probably yes; PN: probably no; ROBINS-I: Risk Of Bias In Non-randomized Studies of Interventions; Y: yes.

## CONTRIBUTIONS OF AUTHORS

HNH wrote all remaining sections (to those listed below), commented on and revised the protocol, and led the development and preparation of the entire Cochrane protocol.

HH wrote the Assessment of risk of bias in included studies and the 'Summary of findings and assessment of the certainty of the evidence' sections under Data collection and analysis.

MK wrote the [Types of outcome measures](#) section under [Criteria for considering studies for this review](#).

JK wrote the Description of the condition section in the Background.

ÅMH wrote the section on 'Assessing the impact of the intervention' under [How the intervention might work](#) in the Background.

LE wrote the Description of the intervention and the How the intervention might work sections in the Background.

EG commented on and helped revise the [Description of the condition](#) section in the Background.

CN wrote the section on 'Adverse effects' under [How the intervention might work](#) in the Background.

MBM commented on and helped revise the [Description of the condition](#) section in the Background.

K-ABN wrote the [Description of the intervention](#) and [How the intervention might work](#) sections in the Background, and wrote the [Unit of analysis issues](#) under [Data collection and analysis](#).

All co-authors were involved in revisions, discussions, and provided comments on all drafts.

Both HNH and K-ABN are the guarantors for the review.

## DECLARATIONS OF INTEREST

Hanne Næss Hjetland (HNH): none known. However, in the interests of transparency, HNH declares that her current position as Postdoctoral Fellow is part financed (75%) by the The Research Council of Norway (grant 299197).

Hilde Hofslundsengen (HH): none known

Marianne Klem (MK): none known

Jannicke Karlsen (JK): none known

Åste M Hagen (ÅMH): none known

Liv Inger Engevik (LIE): none known

Esther Geva (EG): none known

Courtenay Norbury (CN): none known

May-Britt Monsrud (M-BM): none known

Kari-Anne Bottegaard Næss (K-ABN): none known

## 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, K-ABN, 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 has had no role in the design, conduct, or publication of this protocol for the review.