THE CAMPBELL COLLABORATION

Protocol:
The Effect of Linguistic Comprehension Training on Language and Reading Comprehension: A Systematic Review

Kristin Rogde, Åste Mjelve Hagen, Monica Melby-Lervåg, Arne Lervåg

Submitted to the Coordinating Group of:

☐	Crime and Justice
☒	Education
	☐ Disability
☐	International Development
	☐ Nutrition
☐	Social Welfare
☐	Other:

Plans to co-register:

☒	No		
☐	Yes	☐ Cochrane	☐ Other
☐	Maybe		

Date Submitted:
Date Revision Submitted:
Approval Date:
Publication Date: 02 May 2016

Note: Campbell Collaboration Systematic Review Protocol Template version date: 24 February 2013

BACKGROUND

The Problem

The ability to understand and express language in both its oral and written forms is a crucial aspect of human development. Developing linguistic comprehension skills is fundamental to all higher-level cognitive activities, learning, and sets the stage for reading development (McNamara & Magliano, 2009). Unfortunately, there are large differences in language comprehension skills among children (Biemiller & Slonim, 2001; Hart & Risley, 1995; Melby-Lervåg & Lervåg, 2014), and difficulties in reading comprehension are relatively prevalent among students across countries. In the U.S., 31% of the students in fourth grade and 24% of the students in eight grade performed below the basic level on the NAEP [National Assessment of Educational Progress] reading test in 2015 (National Center for Educational Statistics [NCES], 2015). The proportion of children reading below the basic level is even higher among children from families with low socioeconomic status (SES) and from minority race/ethnicity groups like black and Hispanic children. The situation of low-level reading skills among students is similar in North America and several Europeans countries (Organisation for Economic Co-operation and Development [OECD], 2010ab). Linguistic comprehension and reading comprehension skills are necessary for content-area learning in all subjects and thus are influential factors for academic success. Children who lack a strong foundation of linguistic and reading comprehension skills are more likely to experience academic difficulties and drop-out from school. Developing effective instructional practices is therefore of the utmost importance to the field of education.

This review aims to improve our understanding of intervention studies targeting two core constructs: linguistic comprehension and reading comprehension. Linguistic comprehension is defined as the process by which lexical (i.e., word) information, sentences and discourses are interpreted (Gough & Tunmer, 1986;). It refers to the ability to understand and express oral language, often assessed by tests of vocabulary or listening comprehension (Bornstein, Hahn, Putnick, & Suwalsky, 2014; Foorman, Herrera, Petscher, Mitchell, & Truckenmiller, 2015b; Klem et al., 2014, Melby-Lervåg & Lervåg, 2014). Vocabulary is a core component in linguistic comprehension. Vocabulary has typically been divided into either expressive and receptive vocabulary or depth and breadth vocabulary (Ouellette, 2006). However, several more recent studies using latent variables have shown that these are highly related constructs that are difficult to differentiate between (Bornstein, Hahn, Putnick, & Suwalsky, 2014; Klem et al., 2014). Although vocabulary is a core component in linguistic comprehension, skills such as syntax (the ability to understand and formulate sentences) and morphology (how words are formed), which build directly on vocabulary knowledge, are also often considered to be a part of a broader linguistic comprehension construct (Klem et al., 2014).

Reading comprehension can be defined as the active extraction and construction of meaning from all kinds of text (Snow, 2001). Linguistic comprehension is a well-known precursor to

reading comprehension success that develops long before formal reading instruction begins (Snow, Burns & Griffin, 1998). A close relationship between linguistic comprehension skills and the development of reading comprehension has been demonstrated in several longitudinal studies (Lervåg & Aukrust, 2010; Muter, Hulme, Snowling & Stevenson, 2004; Storch & Whitehurst, 2002). According to the simple view of reading (SVR), linguistic comprehension is an important factor that underpins the development of reading comprehension beyond word-level reading (Gough & Tunmer, 1986). In later grades, when decoding skills are fully mastered and the contribution of decoding skills to reading comprehension has lessened, linguistic comprehension and reading comprehension are nearly isomorphic constructs (Muter, Hulme, Snowling, & Stevenson, 2004; Storch & Whitehurst, 2002). It is therefore worrying that large between-child differences in linguistic comprehension are observed from an early age (Hart & Risley, 1995). Hart and Risley (2003) found a large gap in vocabulary size between children from different socioeconomic groups as early as three years of age. Researchers also point to large differences in vocabulary knowledge between young second-language learners and first-language learners (Lervåg & Aukrust, 2010; Melby-Lervåg & Lervåg, 2014). Notably, between-child differences in the ability to learn words appear to be maintained throughout primary school (Biemiller & Boote, 2006; Lervåg & Aukrust, 2010; Melby-Lervåg et al., 2012). Due to these differences, intervention studies have aimed to boost development in linguistic comprehension skills in both preschool and school-age children and to examine the effects on linguistic comprehension and reading comprehension outcome measures (for reviews, see, e.g., Bus, van Ijzendorn & Pellegrini, 1995; Elleman, Lindo, Morphy, & Compton, 2009; Stahl & Fairbanks, 1986; Marulis & Neuman, 2010; Mol, Bus, de Jong, & Smeets, 2008).

As we will see, earlier meta-analyses have shown that transfer effects to global measures of linguistic comprehension and reading comprehension are typically much smaller and less impressive than effect sizes for custom measures (see e.g., Elleman et al., 2009; Marulis & Neuman, 2010). Whereas measures that refer to generalized language include test items that have not been explicitly trained in an intervention, customized tests are designed by researchers to show the effects of targeted training. Custom measures thus provide information on whether children have learned something that has been explicitly covered in an intervention (e.g., directly trained words in a training program). However, the ultimate goal of language-based interventions is to help children accelerate their further growth in linguistic comprehension and reading comprehension skills. If we are to narrow the gap between children with a low and a high vocabulary, we should focus on providing children with skills to continuously develop new knowledge of words. Unraveling the important factors that contribute to this generalization of knowledge thus becomes essential.

Despite the discrepancy in effects on global outcome measures vs. custom outcome measures, vocabulary training is recommended on a regular basis in schools. As noted by Coyne et al., (2010), researchers and practitioners often hold an implicit hypothesis that vocabulary instruction affords additional benefits beyond learning target words. However,

the subsequent conclusions we are able to draw from these different assessment practices (custom vs. global outcome measures) will be of different practical value.

The primary aim of this review is to provide an overview of studies on interventions targeting linguistic comprehension and their effects on measures of generalized linguistic comprehension and/or reading comprehension. By strengthening our knowledge of this subject, we can potentially obtain insight into how related deficits can be ameliorated. This information is critical in making policy decisions about whether such programs are suitable for implementation in early childhood education and later schooling. Reviewing training studies may also provide a more refined understanding of the underlying causal mechanisms through which interventions are effective. This aspect is vital for providing a sound theoretical foundation for constructing better and more targeted intervention programs. In the following sections, we go into detail about how we plan to conduct this systematic review and how this review will contribute information not present in other reviews in the field.

The Intervention

Our review will focus on interventions that aim to improve linguistic comprehension and reading comprehension through linguistic comprehension tasks. In line with our definition of linguistic comprehension (see page 2), related tasks involve vocabulary (such as defining words and pointing at pictures of specific words), syntax (such as narrative skills, listening comprehension and building sentences), morphology (tasks related to the process of forming new words through diversion or composition), and other grammar tasks.

Several meta-analyses have summarized the literature on studies aiming to improve linguistic comprehension. Table 1 shows an overview of these meta-analyses. As shown in Table 1, when reviewing the literature, earlier studies that have attempted to improve linguistic comprehension through language-related tasks represent different study characteristics. Foci of instruction, outcome measures, participants (e.g., at risk, SES), implementation features (e.g., who is doing the training, dosage of intervention) and research designs are some factors that are likely to differ among studies and hence are important to incorporate as moderator variables in our review. Below we describe in more detail what constitutes a linguistic comprehension intervention program in regard to instructional approaches, measurement outcomes and research design. A more detailed description about studies to be included and excluded is given at page 22. Moreover, study characteristics will be coded and serve as potential moderators in our study. Information about this is presented at page 29 and in appendix II.

Instructional approaches to improve children's linguistic comprehension skills

The majority of studies conducted with preschool- and kindergarten-age children that comprise linguistic comprehension training use some type of shared book reading (see Table 1; Marulis & Neuman, 2010; Mol, Bus, de Jong, & Smeets, 2008; Mol, Bus, & de Jong, 2009). Shared book reading has also been the focus in prior reviews. Mol, Bus, de Jong & Smeets (2008) examined the effect of dialogic parent-child reading and Mol, Bus and de Jong (2009) targeted studies in which an interactive reading intervention was implemented in preschool or kindergarten classrooms. Marulis and Neuman (2010) included book-reading studies in addition to all types of vocabulary interventions. Some of the intervention studies in Marulis and Neuman's meta-analysis used storybooks within more comprehensive programs and some studies were not related to storybook reading at all (e.g., computer-based interventions). However, the meta-analysis by Marulis and Neuman (2010) showed that most studies containing vocabulary intervention for young children included some type of storybook reading.

If we look at studies targeting linguistic comprehension skills through storybook reading, a large number of these studies use dialogic reading in which the child is actively involved in a book-reading situation. This stands in contrast to typical shared reading in which the adult reads and the child listens (Hargrave & Sénéchal, 2000; Whitehurst & Lonigan, 1998). Studies examining the effect of just reading aloud on word learning have showed less noteworthy results (Biemiller & Boote, 2006; Senechal, Thomas, & Monker, 1995).

A commonly used strategy to teach children words in intervention studies has been to provide children with direct instruction in word meanings through storybook reading. This direct vocabulary training has been practiced in various ways. One instructional approach is to provide the children with brief explanations of word meanings during reading. This embedded vocabulary instruction targets the breadth of vocabulary knowledge and has the benefit of being time efficient as it allows for the instruction of many word meanings during a training session (Coyne, McCoach, Loftus, Zipoli & Kapp, 2009).

Another instructional approach is to provide children with rich instructions of words following storybook reading (Beck & McKeown, 2007). This includes providing multiple explanations and examples related to multiple contexts, and to let the children actively engage in the explanations and discussions of word meanings. This technique is expected to foster a child's depth of vocabulary knowledge in contrast to increasing a child's number of word meanings (Beck & McKeown, 2007; Coyne et al., 2009).

The meta-analysis by Marulis and Neuman (2010) showed that studies containing storybook reading differed in the way vocabulary was taught. In order to code the type of vocabulary training that was used in the studies, Marulis and Neuman (2010) chose to differentiate between studies that contained an explicit instruction of words, studies where the focus was on the implicit learning of words (e.g., storybook reading without deliberation about word meanings), and studies where a combination of explicit and implicit instruction was

delivered. Larger effect sizes for vocabulary outcomes were found for explicit and combined explicit and implicit instruction than for implicit instruction alone, which was found to be less effective.

There are few studies in which instruction of linguistic comprehension has been conducted at young age and the transfer effect to later reading comprehension has been examined at follow-up time points. A large portion of the intervention studies targeting reading comprehension skills through linguistic comprehension training has thus been conducted with school-age children (Elleman et al., 2009). However, there are exceptions. For instance, Fricke et al. (2013) conducted a study where improvements in language skills in preschool children generalized to a later reading comprehension outcome measure. This study included a range of activities in the training program (e.g., direct vocabulary training, creating and acting out stories) and aimed to improve both linguistic comprehension skills (vocabulary, grammatical competence, narrative skills, listening comprehension) and reading comprehension.

Taken together, studies targeting linguistic comprehension in preschool and kindergarten age children are not solely related to vocabulary learning through the context of storybook reading. In order to get an overview of the effect of linguistic comprehension training, several types of instructional approaches should therefore be taken into account and included in a meta-analysis.

If we look at intervention studies targeting linguistic comprehension in school-age children with reading comprehension outcomes, Stahl and Fairbanks (1986) reported an effect size of 0.97 for reading comprehension measures developed by researchers and 0.30 for global measures of reading comprehension. Elleman et al. (2009) reported a mean effect of 0.50 for custom measures of reading comprehension and 0.10 for global measures of reading comprehension. Multiple components with elements related to reading and direct vocabulary training are often the case in these studies (for reviews see Stahl & Fairbanks; Stahl & Nagy, 1986). According to the National Reading Panel (NRP), studies vary greatly in types of instruction. Elleman et al. (2009) chose to exclude studies that only used repeated readings, read-aloud, or independent reading; some kind of instructional method for teaching vocabulary had to be delivered. In addition, if the studies included components to address comprehension, as well as vocabulary, the study was included (e.g., use of a multicomponent intervention with vocabulary instruction and other comprehension instruction). Intervention characteristics in Stahl and Fairbanks (1986) and Elleman et al.'s (2009) studies were coded based on depth of processing (coding relates to the amount of semantic processing and mental effort), definitional-contextual scale (e.g., the coding relates to if definitions or synonyms are provided without any use of context vs. understanding the word in context) and type of exposure (information related to repetitions and contexts were coded). Intervention characteristics like this may contribute to explaining why some interventions are more effective than others. These categories will therefore be included in the coding manual for this review (Appendix II).

In addition to vocabulary instruction, instructional approaches in studies to improve school-age children's linguistic comprehension skills have also included a wide range of activities in the training program to improve vocabulary and grammatical competence as well as narrative discourse and listening comprehension (Bowyer-Crane et al., 2008; Clarke, Snowling, Truelove & Hulme, 2010). These studies have been conducted with both school-aged children and younger children and have shown promising effects on both measures of linguistic comprehension and reading comprehension. These studies have not yet been assessed in a meta-analysis.

Measurement of outcomes

As shown in Table 1 (meta-analyses of educational interventions in the area of linguistic comprehension/vocabulary), there are large differences across the meta-analyses with regard to the mean effect size they report, which range from d = 0.29 to d = 1.21 for linguistic comprehension measures and from d = 0.10 to d = 0.70 for reading comprehension measures. These meta-analyses differed in their approaches to the diversity between the primary studies (i.e., they used different inclusion criteria and the mean effect sizes are based on different outcome measures). More specifically, as can be observed in Table 1, a particularly important aspect of study variation involves whether the study is designed to evaluate the impact of linguistic comprehension instruction on customized measures of content (e.g., target words) that have been the target for instruction or on distal measures of generalized language. When synthesizing earlier findings, it is therefore important to separate outcome measures in studies that target the effect of trained words vs. outcome measures in studies that are designed to improve children's global language skills.

However, the issue of assessment is not straightforward. One type of test that often is used to assess language or reading comprehension skills outside directly trained skills is a standardized test. When researchers aim to assess an intervention's impact on more generalized language and reading skills, they usually make sure not to include words in the vocabulary training program that are tested directly in standardized tests. Note also that global standardized measures and definition tasks on trained words can be seen as two ends of a continuum. It is therefore interesting to look at this more incrementally, and to see what characterizes effects for those that use measures that are in between these "extreme" ends of the continuum. This will be an important part of our moderator variable analyses.

Also, in addition to standardized tests, intervention studies designed to evaluate effects on distal measures of generalized language may also make use of outcome measures that are not standardized. For research purposes, the standardization of tests is in itself not a critical issue. Researchers may also make use of tests that are created to measure global language skills that are not standardized. However, standardized tests are more likely to be well proven and psychometrically tested and validated. Therefore, we will carefully code if the outcome measures of global language skills are standardized or not (see details of coding appendix II).

As already mentioned, the transfer effects to outcome measures of generalized language and reading comprehension are substantially smaller than the effects on customized measures of the targeted training (Table 1) (see for instance Elleman et al., 2009; Marulis & Neuman, 2010). However, some recent studies have shown that transferring the effects from linguistic comprehension training to standardized measures of language and reading comprehension is achievable (Bowyer-Crane et al., 2008; Clarke, Snowling, Truelove, & Hulme, 2010; Fricke et al., 2013).

Example of a study to be included in the review

The recent study by Fricke et al., (2013) is an example of a study that is to be included in this review. In a randomized controlled trial, 180 children from 15 UK nursery schools (12 children from each school) were randomly allocated to receive oral language training for 30 weeks (3 sessions every week) or to a waiting control group. Nursery school staff and teacher assistants delivered the intervention under training by the research staff. The training consisted of a comprehensive program to increase children's oral language skills. Activities targeted, for example, children's vocabulary, narrative skills, and active listening skills. In addition, components of training related to phonology and alphabetic language were incorporated into the training sessions the remaining 10 weeks of the program. Results showed that the intervention program had effects on standardized measures of language comprehension. Progress in oral language skills also generalized to a standardized measure of reading comprehension 6 months after the intervention ended.

Research design

Table 1 shows that the intervention studies also differ with regard to other potentially important variations between studies. One core aspect in which they differ is related to study design. This variation involves three main issues: (1) whether the studies have a control group, (2) whether the potential control group is active or passive, and finally, (3) whether participants have been randomized across conditions. Moreover, they also differ in other important aspects, such as the amount and duration of training provided. Some of the previous meta-analyses include studies with as little as a one-hour intervention (see for instance Elleman et al., 2009). Table 1 further shows that the studies also differ in regard to the type of training (e.g., book reading, vocabulary intervention), location (i.e., home or educational settings), context (classrooms or small groups), the individuals providing the training (i.e., parents, teachers or researchers) and the participants (children with or without learning disorders and in different age groups). Thus, studies focusing on linguistic comprehension training cannot be understood as a homogeneous group of studies, and different intervention characteristics should therefore be taken into account when conducting a systematic review of this topic.

Table 1. Meta-analyses of educational interventions in the area of linguistic comprehension/vocabulary

Study name	Results for main outcome measures	Design of the studies included	Meta-analytic method	Participant characteristics and intervention length	Intervention characteristics	Moderators (main findings)
Blok, 1999	1) Overall effect on oral language measures (vocabulary as well as phonology, grammar, etc.) d = 0.63** (k = 10)	Both experimental and quasi-experimental studies using a control group are included	Both random and fixed, unclear whether pretest data is controlled for	Includes studies of children up to 8 years, unclear whether studies of LD children are included	All studies include an intervention involving storybook reading to children in an educational setting.	No quantitative moderator analysis
Elleman et al. 2009	1) Vocabulary measured with standardized measures, d = 0.29** (k = 14) 2) Vocabulary measured with tests constructed for study, d = 0.79** (k = 18) 3) Reading comprehension measured using tests constructed for study, d = 0.50** (k = 23) 4) Reading comprehension measured with standardized measures, d = 0.10 (k = 16)	Studies must have employed either a pretest-posttest control group design, posttest control with randomization or pre-posttest within-subjects design using counterbalanced conditions. (% RCTs not reported)	Random effect models, effect sizes corrected for pretest differences when pretest data were available	Students from pre-K to grade 12 with English as the first language. Students with LD are also included. Treatment hours ranging from 1-37.5	Studied interventions with the goal of increasing word knowledge or comprehension that could be implemented in a classroom setting	When considering only custom measures, the effect size was correlated with control group strength and experiment vs quasi experiment design. High levels of discussion were associated with larger effect sizes
Fischel & Landry, 2008 (NELP report)	1) Oral language d = 0.63 (k = 19) 2) Phonological awareness d = 0.57 (k = 2)	Studies must have used a randomized experiment or a quasi-experiment with a control group (treated or untreated), 47% RCTs	Fixed and random effects, and with groups at a comparable level at pretest	Children from birth through the age of 5 both with and without language delays, studies with one single lesson were excluded, treatment ranged from less than two weeks to one year	Studies must have evaluated the effectiveness of interventions designed to explicitly and directly improve young children's language skills in terms of vocabulary, syntax or listening comprehension. Interventions primarily delivered by teachers (2 studies by parents)	No differences between children regarding age, ethnicity or population density. Studies of children younger than the age of 3 with higher effect sizes than studies of older children, no differences between play-related and non-play-related interventions, no significant difference between studies of children with and without language impairment
Fukkink & de Glopper, 1998	Mean effect d = 0.43	Included only studies with a control group	Data were analysed using the program META. Random effects models were used. d was corrected for small sample size	Participants were children and youth from midldle grade to 10th grade (with one exception). Average instruction was 5.5 hours.	Examined effects from an intervention where the treatment group was taught to derive word meaning from context. Outcomes were cloze tests, word definition tests or multiple choice vocabulary tests.	Clue instruction appears to be more effective than other instruction types or just practice (ß = 0.40). Effect size correlates negatively with class size (ß = -.03).

Study name	Results for main outcome measures	Design of the studies included	Meta-analytic method	Participant characteristics and intervention length	Intervention characteristics	Moderators (main findings)
Goodwin & Ahn, 2013	Morphological knowledge d = 0.44 Vocabulary d = 0.34 No effect on reading comprehension	Studies must have a control/comparison group compared to a morphological intervention group.	Random-effects model. Mixed effects models with moderator variables were applied to explain variations in effect sizes.	School-age children	Intervention effects of morphological instruction on language and literacy.	The mean effect of morphological intervention based on standardized tests was statistically lower than one from researcher-made tests. Intervention with an exclusive focus on morphological instruction was as effective as morphological intervention as a part of a more comprehensive instruction
Goodwin & Ahn, 2010	Morphological awareness d = 0.40 Vocabulary d = 0.40 Reading comprehension d = 0.24	Studies must have a control/comparison group compared to a morphological intervention group.	Variance-weighted analyses, fixed effect model.	Students with literacy difficulties	Investigating the effect of morphological interventions on literacy outcomes.	Interventions as a part of more comprehensive instruction was more effective at improving children's reading achievement than an intervention with an exclusive focus on morphological instruction.
Law, Garrett & Nye, 2004	Expressive vocabulary d = 0.98 (k = 2) Expressive syntax d = 0.70 (k = 5) Receptive syntax d = -0.04 (k = 1)	Studies must have used a randomized controlled trial	Data were analyzed using REVMAN software. Random effects models were used, not clear if d was corrected for small sample size	Participants were children or adolescents with primary speech and language difficulties	Interventions that aimed to improve expressive or receptive phonology, syntax or vocabulary were examined. Interventions were implemented by parents or clinicians. Outcomes must have been related to speech, receptive or expressive phonology, syntax or vocabulary	No difference between parent and clinician
Lonigan, C., Shanahan, T., & Cunningham, A. (2008), (NELP report)	1) Oral language d = 0.73 (d = 0.57 with outlier excluded, k = 16)	Studies must have used a randomized experiment or a quasi-experiment that was not seriously confounded (only one QE study included)	Fixed and random effects reported; groups at a comparable level at pretest	Children in pre-kindergarten or kindergarten, both at risk and not at risk. Number of instruction hours unclear	Studies must have evaluated the intervention effects from shared book reading (represented by change in frequency or in the style of shared reading activities), as implemented by either parents or in an educational setting	There were no differences between older and younger children, no differences between at-risk and not at-risk children, and no differences between parent- or teacher-delivered interventions

Study name	Results for main outcome measures	Design of the studies included	Meta-analytic method	Participant characteristics and intervention length	Intervention characteristics	Moderators (main findings)
Marulis & Neuman, 2010	**Posttest:** 1) Vocabulary measured with standardized measures d = 0.71** (k = 36) 2) Vocabulary measured with tests constructed for study d = 1.21** (k = 19) **Follow up:** 1) Vocabulary overall d = 1.08** (k = 11)	A randomized controlled trial, a pretest-intervention-posttest with a control group, or a post intervention comparison between preexisting groups were included.	Random effects model, unclear whether effect sizes are corrected for posttest differences	Children within ages of birth through 6 included, range of training time from 1 day to 270 days	All studies include a training, intervention or specific technique to improve word learning. Interventions were implemented in a home, school or clinical setting	Studies in which trained adults implemented the intervention and studies in which explicit and implicit instruction were combined had larger effects; middle and upper income children at risk had better effects than those from poor SES backgrounds
Mol, Bus & de Jong, 2009	1) Expressive vocabulary d = 0.62** (k = 20) 2) Receptive vocabulary d = 0.45** (k = 23)	Both experimental and quasi-experimental studies using a control group were included (overall 61 % RCTs); studies were excluded if dialogic reading was a part of a larger intervention	Random effects models used, unclear whether pretest scores are controlled for	Includes children in preschool and kindergarten or first grade classrooms, intervention time ranging from 2-36 weeks; included both at-risk children and unselected	Includes studies in which teachers or graduate students were instructed to implement an interactive reading intervention following the main principles of dialogic reading	Studies that were highly controlled and executed by examiners had higher effect sizes than when interventions were implemented by teachers; individuals interacting with examiners had better effects than larger groups with teachers and experimenters; studies with high fidelity to treatment revealed higher effect sizes
Stahl & Fairbanks, 1986	1) Reading comprehension measured using tests constructed for study *d* = 0.97** (k = 41, control groups not exposed to target words) 2) Reading comprehension measured with standardized measures *d* = 0.30** (k = 16, control groups not exposed to target words)	Studies must have used a control group (treatment in which the children were given the target words and were told to study them as they liked or untreated[1])	Fixed effects, effect sizes calculated on the basis of posttest means (unclear whether pretest means were used as a correction)	The age of the children was from 2nd grade to college. The number of hours trained and characteristics of children included is unclear	The intervention must have focused on vocabulary instruction to learn word meanings	No moderator analysis for vocabulary outcomes
Swanson et al. 2011	1) Vocabulary d = 1.02** (k = 51) 2) Reading comprehension (overall) d = 0.70** (k = 22)	Studies using a treatment control group design, single group or single subject design were included	Fixed and random effects reported (HLM), unclear whether pretest scores are corrected for	Participants were 3-8 years old and at risk of reading disabilities (low achievement on language measures, low socioeconomic background, family risk, or	Examined the effect of storybook read aloud interventions in daycare, preschool or school settings (dialogic reading, repeated reading, story reading with questioning, computer-assisted story reading, story reading	Despite large variation between intervention types, only a small amount of variance was accounted for by intervention type

Study name	Results for main outcome measures	Design of the studies included	Meta-analytic method	Participant characteristics and intervention length	Intervention characteristics	Moderators (main findings)
				attending school with poor reading achievement); number of sessions ranging from 3-80	with extended vocabulary activities)	

How the intervention may work

In this section, we outline the theoretical foundation for the review and develop a model for the relations that are to be tested in the review.

Theoretical background

At least three important theoretical perspectives set the stage for our review. The first concerns the development of skills related to our primary outcome, linguistic comprehension. Two issues are particularly striking. The first is that different aspects of linguistic comprehension appear to develop with a high degree of interdependence. Several cross-sectional and longitudinal studies using observed variables have indicated that expressive and receptive vocabulary, grammar and syntax and verbal memory are related skills that reflect a common factor (Colledge, et al., 2002; Johnson et al., 1999; MacDonald, 2013; MacDonald & Christiansen, 2002; Pickering & Garrod, 2013). This hypothesis has recently gained more conclusive support in large-scale longitudinal studies using latent variables that correct for measurement errors: Bornstein, Hahn, Putnick, and Suwalsky (2014) found a unitary core language construct from early childhood to adolescence. Additionally, Klem et al. (2014) found a unidimensional latent language factor (defined by sentence repetition, vocabulary knowledge and grammatical skills) in a longitudinal study of children 4-6 years of age.

A second important issue is the robust longitudinal stability within the linguistic comprehension domain. A stable rank order of children's vocabulary knowledge is preserved during both preschool and later school years (Melby-Lervåg et al., 2012; Storch & Whitehurst, 2002). Studies by Klem et al. (2014) and Bornstein et al. (2014) showed a unitary core language construct and found that this construct is highly stable over time. All of these studies suggest that although all children's linguistic comprehension skills improve over time, the rank order between children is more or less preserved. This implies that altering children's language levels relative to other children is a complex and challenging endeavor. Nonetheless, as Bornstein et al. (2014) note, stability does not mean that it is impossible to change language skills through intervention. For example, Bornstein et al. (2014) estimated a latent correlation of 0.78 from 4 to 10 years; this leaves 39% of the variance in the 10-year core language skill unexplained by 4-year language.

The second theoretical issue involves the relationship between our primary outcomes of linguistic comprehension and reading comprehension. How could improvement in linguistic comprehension transfer to reading comprehension? The Simple View of Reading is a well-established theoretical model of reading comprehension (Gough & Tunmer, 1986). This model presents reading comprehension as the product of decoding and linguistic comprehension skills and is formalized as the equation "Decoding x Linguistic comprehension = Reading Comprehension". In this model, linguistic comprehension is an important underpinning in the development of reading comprehension beyond word-level reading (Gough & Tunmer, 1986). Whereas decoding is an important predictor of reading

skills in the early reading phase, linguistic comprehension is understood as an essential predictor for the further development of reading comprehension (Hoover & Gough, 1990; Muter et al., 2004; Storch & Whitehurst, 2002). The rationale for fostering linguistic comprehension skills to provide for later reading comprehension proficiency is related to the fact that linguistic comprehension is a well-known precursor to reading success that develops long before formal reading instruction begins (Snow, Burns, & Griffin, 1998; Whitehurst & Lonigan, 1998). However, notably, at an older age (when linguistic comprehension explains the majority of variation in reading comprehension), reading comprehension has also proven to be a highly stable construct (see Lervåg & Aukrust, 2010).

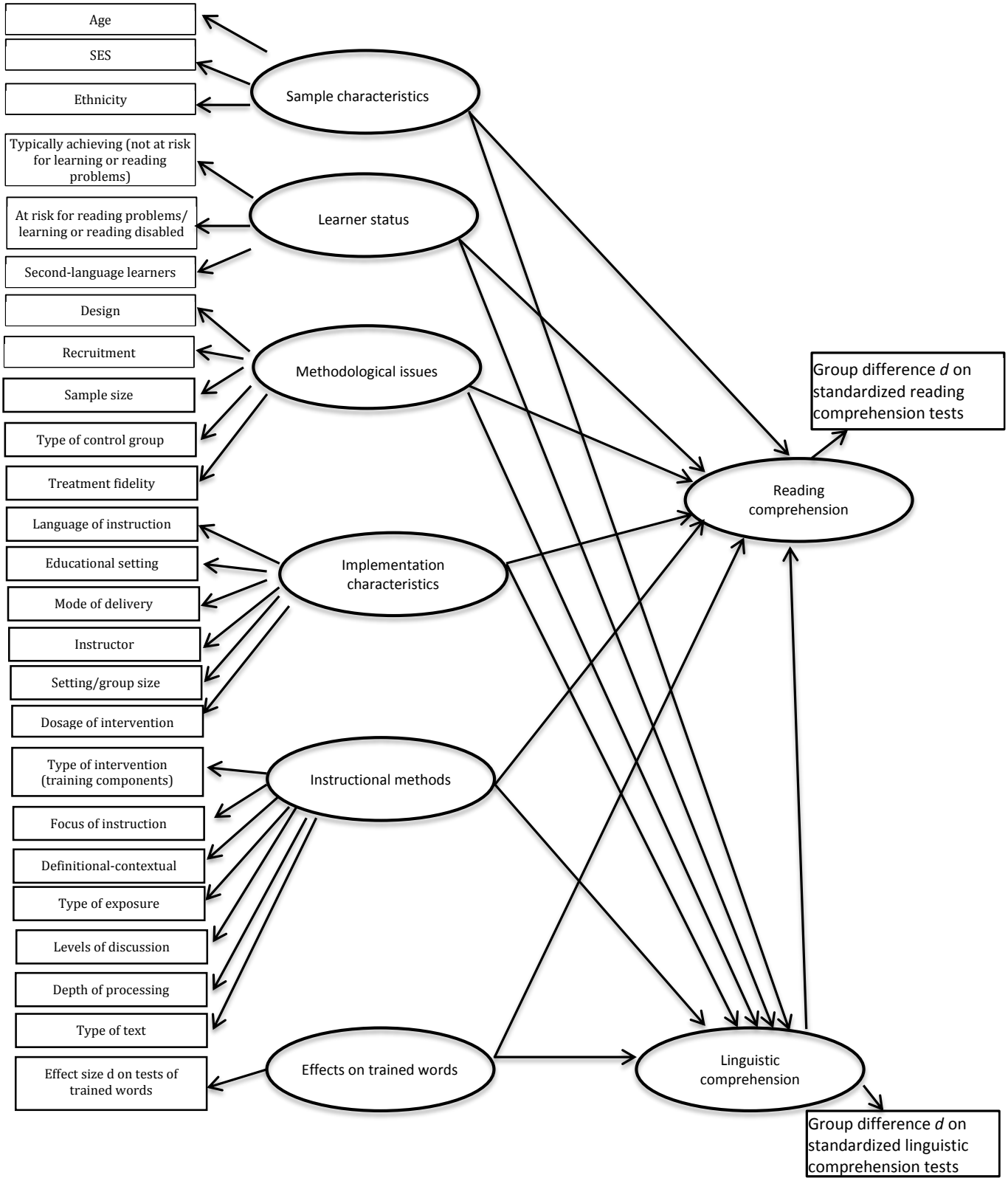
Third, theories on the nature of how and to what extent we are able to transfer what we learn will be an important issue of our study (see Barnett & Ceci, 2002; Bransford & Schwartz, 1999; Carraher & Schliemann, 2002). Two issues are at play: 1) the transfer of effects from criterion measures that contain the specific words that are used in the intervention to standardized tests of linguistic comprehension and 2) the transfer of effects on linguistic comprehension to reading comprehension. Numerous studies indicate that children can easily be taught the meaning of novel words with which they are presented in an intervention (see Elleman et al., 2009). This phenomenon is often referred to as "near transfer." However, in an intervention program, a child is typically presented with 3-6 novel words per week (Elleman et al., 2009). This amount is hardly sufficient to close the gap with children who have superior linguistic comprehension or the gap that exists between first- and second-language learners because the comparison children also continuously develop their language skills. For instance, in the studies that provide direct vocabulary training either embedded in story book reading or as a separate component, it is important to note that there have been no intervention studies that have taught over 150 words or that have lasted more than 104 hours (at least up until 2009) (Elleman 2009). Biemiller (2005) has estimated that it would require 1,000 word roots to have a chance to affect general reading comprehension. A vocabulary training of this magnitude would not be realistic to implement, and has not been done up until now. Thus, for the studies that do show positive effects on global measures (e.g., Bowyer-Crane et al., 2008; Clarke, Snowling, Truelove & Hulme, 2010; Fricke et al., 2013), it is not likely that training specific definitions of words is the causal factor that underpins this improvement. Most likely there are other factors in the instruction that have led to the gains on standardized measures. Language interventions need to teach children skills that are transferrable so that they can use them in their general language development. These strategies can then be used when they encounter new words and unfamiliar sentences and not merely for the specific words taught in the intervention. As Taatgen (2013) states, “[T]ransfer in education is not necessarily based on content and semantics but also on the underlying structure of skills” (p. 469). Thus, to achieve long-reaching transfer in language interventions (i.e., transfer beyond the specific words on which children are trained to more global language skills), an intervention also needs to focus on strategies that can be used in general language learning.

After undertaking this review, we may be able to answer critical policy questions regarding whether or how we, based on the evidence to date, can most efficiently create robust long-term improvements in children's linguistic comprehension and reading comprehension skills by targeting their linguistic comprehension. Because linguistic comprehension and reading comprehension skills are crucial for both academic performance and societal participation, the development of methods to improve these skills in children is of great importance both for the individual and for society. Moreover, our review sheds light on important theoretical issues related to the nature of language learning, such as to what extent we, despite the high stability of linguistic comprehension and reading comprehension, are able to alter these skills and whether skills transfer from specific tasks integrated in the intervention to more generalized tasks in standardized tests.

Model of the relations between intervention, moderator variables and outcomes

Figure 1 illustrates the causal mechanisms behind the intervention. As can be observed, linguistic comprehension is hypothesized to be a mediator that affects the strength of the effects on reading comprehension skills. However, we also predict important direct effects on linguistic comprehension. Moderators in the model are hypothesized to have an impact on the effects on both linguistic comprehension and reading comprehension. Notably, the indicators that are considered representations for the latent constructs in the model (and coded in the systematic review) are clearly defined in the next section.

Figure 1. Model of the relations studied in the review



Why it is Important to do the Review

Table 1 shows that our literature search found 12 previous reviews on this topic. Thus, it may be asked why there is a need to perform an additional review. As we demonstrate in this section, although several reviews have been published on this matter, they do not fully test the model that we aim to examine in this systematic review. Additionally, several of the previous meta-analyses are now outdated; many new studies are not included. The incorporation of these new studies makes our review substantially different from and a valuable extension of recent reviews.

As shown in Table 1, earlier meta-analyses have included studies that use both customized tests for the targeted training and standardized tests in their examination of training effects. However, in many of the previous meta-analyses, a mean effect size was calculated to combine these two test types (see Table 1, Mol, Bus, & de Jong, 2009; Swanson et al. 2011). This procedure yields a biased result for the effects of linguistic comprehension training. In contrast, the planned review will exclusively examine studies that report standardized measures in addition to measures of the words that are trained. For the overall mean effect size, the effects on standardized tests and custom measures are calculated separately. Importantly, we plan to examine whether the effects on standardized measures are mediated by the effects on custom measures. This approach enables the review to make a unique contribution to the study of the effectiveness of linguistic comprehension intervention in educational settings.

Several meta-analyses on the topic have exclusively examined the value of shared book reading (e.g., Blok, 1999; Bus, van Ijzendoorn, & Pellegrini, 1995; Mol, Bus, de Jong, & Smeets, 2008), whereas others have included several types of vocabulary interventions in addition to print-based training (Elleman et al., 2009; Marulis & Neuman, 2010). Similar to Marulis and Neuman (2010) and Elleman et al. (2009), we want to include training studies that focus on both shared book reading and other types of vocabulary instruction. We will include studies with children older than those in Marulis and Neuman (2010), who examined children from birth to age nine. We will also include studies that contain a broader view of oral language training (e.g., training focusing on listening comprehension, narrative skills, and morphology/grammatical skills). Such studies have been published more recently and have not been included in any of the previous meta-analyses (e.g., Clarke et al. 2010, Fricke et al., 2013).

Another reason for conducting this review is to examine the question of how dosage (duration and amount of training) of training is related to intervention effects on global language comprehension skills. Earlier meta-analyses do not give an answer to this question. The previous work by Elleman et al. (2009) and Marulis & Neuman (2010) found that longer duration of treatment is not related to higher effect sizes in comprehension. However, because Elleman et al. (2009) and Marulis & Neuman (2010) include studies with both global and custom outcome measures in their analyses, this finding is not applicable to the

question of how dosage of training is related to increasing global comprehension skills (i.e., outcome measures of generalized language). As Elleman et al. (2009) note from their study, studies using custom measures of vocabulary were short in duration (more than half of the studies were conducted in less than 10 hours) and studies using a standardized measure, in general, had longer interventions. The finding that duration is not an important factor to increase vocabulary skills is not surprising considering the included studies and method of analysis in these reviews. In contrast, we can hypothesize that the picture will look different when it comes to interventions designed to increase global language skills.

Regarding settings, the above-mentioned meta-analyses vary in their inclusion criteria. Bus et al. (1995) and Mol et al. (2008) studied book reading in parent-child settings and excluded interventions implemented in educational settings. Blok (1999) and Elleman et al. (2009) included only training studies in educational settings, whereas Marulis and Neuman (2010) included training studies implemented in both home and educational settings. Our aim is to focus on language training conducted exclusively in educational settings because these studies have the most relevance for educational policy and practice. Thus, we want to exclude interventions implemented by parents or in the child's home environment. An additional reason for this is that this review will focus on how linguistic comprehension and reading comprehension can be improved in an educational setting.

Also, this review will be the first to expand the current literature in incorporating training studies from both an early age in preschool and school-age children. We include studies conducted in preschool and later educational settings up to the end of secondary school. Notably, the National Early Literacy Panel (2008) studied shared-reading interventions in children aged zero to five, and no studies examined the impact of intervention on reading as an outcome variable. Similarly, Marulis and Neuman (2010) targeted only the very early years of vocabulary development (birth through age 6) and did not include measures of reading comprehension. Elleman et al. (2009) examined the impact of vocabulary instruction on reading comprehension in school-age children where the majority of the studies include instruction conducted in grades 3 to 5.

Further, many of the previous meta-analyses included studies without an appropriate control group, e.g., within-subjects designs (see for instance Table 1, Elleman et al. 2009) or comparisons between pre-existing groups (Table 1, Marulis & Neuman, 2010). Our review will only include information from randomized controlled trials and quasi-experiments with a control group and measures of baseline differences. We will also focus on the type of control group that is used in the study (more details about this is given in the coding manual, appendix II). Several studies have shown that this can be an important factor in explaining differences between studies (see Boot et al. 2013). Further, in contrast with the majority of previous reviews (see Table 1), we will also code measures of follow-up effects if this is reported because the practical value of such interventions also depends on the extent to which intervention effects are lasting. Altogether, this emphasis on methodology in our

review generates the best data for drawing conclusions on the effectiveness of intervention programs.

In summary, our planned review will take a different approach in summarizing linguistic comprehension training studies than earlier reviews on the topic. The primary interest and need for this review are linked to the question of whether linguistic comprehension training effects can be transferred to standardized measures. To obtain more knowledge on this point, this review will take a different approach to the type of training studies that will be included compared to prior reviews (e.g., types of research designs and outcome measures). The planned review will expand the current literature by including recent research from the past several years. For instance, Elleman et al. (2009) included studies from 1950 to 2006, and Marulis and Neuman (2010) conducted their search through September 2008. In addition, we will conduct a thorough search for grey literature in collaboration with an information retrieval expert. Altogether, this approach will make our study a valuable addition to previous reviews. Based on this most current knowledge base, the review may be able to answer important policy questions related to the extent to which children's global linguistic comprehension and reading comprehension skills can be altered through interventions and, if so, how interventions that promote global linguistic comprehension and reading comprehension can be best constructed and implemented.

OBJECTIVES

To summarize the discussion from the previous sections, this review has several different objectives. As stated in the previous sections, there is a large variation between studies with regard to how effective linguistic comprehension training appears to be. One objective is to understand this conflicting evidence. Additionally, there are a number of previous reviews in the field with highly conflicting evidence. Another objective of the present review is to produce more general statements on relations and treatment effects. This will be done through the synthesis of results from studies that use the most valid methodology to draw conclusions on intervention effects. We also aim to answer questions on a meta level regarding the mediation and moderation of effects that are difficult to address in the individual component studies. Finally, we aim to build connections between related areas of research in disparate intervention types that previous studies have treated separately (i.e., shared book reading, vocabulary interventions, broader language interventions) in one review.

The aim of this systematic review is to summarize the effectiveness of linguistic comprehension training on standardized linguistic comprehension and reading comprehension measures.

This review aims to answer:

1. To what extent can linguistic comprehension training be generalized to global tests of linguistic comprehension and reading comprehension?

2. Which features of linguistic comprehension training (e.g., shared book reading, vocabulary instruction, instruction targeting multiple linguistic dimensions) are associated with generalized effects on language and/or reading comprehension in a given age group?

3. Are the effects on reading comprehension mediated through linguistic comprehension gains?

METHODOLOGY

Inclusion and exclusion criteria for the studies in the systematic review

Study designs

The review will include quantitative evidence from randomized controlled trials and quasi-experimental studies. Quasi-experimental studies must include a pre-post design with a control group that report baseline differences. Randomized controlled trials (true experiments) will be eligible without pre-test data. Within-subject, single subject, or pre-post designs with no control design will not be eligible. Although within-subject designs can be reasonably to apply in case one wants to examine differential response to different types of vocabulary instruction methods, this is not within the scope of our review. When our focus is on experimental research trying to increase linguistic comprehension skills in general (i.e. effects on global outcomes), within-subject designs may result in biased findings because of carryover effects (e.g., practice effect). Control conditions can include no treatment, waiting list treatment, treatment as usual or alternative treatment. However, if the control condition in a study includes some type of alternative treatment, this will be carefully considered before inclusion. For instance if the control condition include a different type of instructional method for linguistic comprehension training or a type of training that targets other language constructs (e.g. phonological awareness training or morphological awareness training) that may relate to the outcome measures we include in the review, this study will be carefully coded related to this. Moderator analyses to examine potential differences between control conditions will be conducted (see coding details Appendix II).

Language of publication

Studies must be reported in English to be reported in the review. Beyond this, any language of instruction will be considered to be included as long as a full-text translation of the study is provided.

Years of publication

Studies between 1986 and up until today will be eligible for inclusion. The reason for why we focus on the last 30 years is that it is important that the educational settings in which the studies are conducted are comparable over time.

Types of participants

The review will include studies conducted in preschool and educational settings up to the end of secondary school, corresponding approximately to ages 4-16. Groups of unselected typically achieving children, second-language learners, children with language problems, or children at risk for language and reading problems are included in the study. Children with special diagnosis like autism or other physical, mental or sensory disabilities will not be included in the review.

In the analysis, sample characteristics will be used as moderator variables (see coding details Appendix II.

Types of interventions

Studies must include an instructional method targeting linguistic comprehension skills. Vocabulary training studies and studies incorporating vocabulary training within a more extensive approach to linguistic comprehension training (e.g., activities fostering grammatical knowledge, listening comprehension and narrative skills) will be eligible. Studies that only train grammatical skills (e.g., morphology or syntax) with the aim of improving decoding skills will not be included. The aim of the instruction program must be to improve linguistic comprehension and include outcome measures of comprehension.

Some types of studies that include outcome measures of linguistic comprehension will be excluded from this review. Studies targeting linguistic comprehension through the use of sign language, as was included in the meta-analysis of vocabulary training by Marulis & Neuman (2010), will not be eligible (e.g., Daniels, 2004). Language interventions with a main focus on phonology are also omitted from this review. The reason for this is that phonological awareness tasks focus more on the sound system in language (phonology) than on the meaning-based (semantic) aspects of language. Thus, the literature on interventions focusing on phonological awareness training is beyond the scope of this review. Additionally, there is a large set of studies that focus on the improvement of linguistic comprehension by targeting broader cognitive skills such as working memory or auditory processing. Those studies are also beyond the scope of this review (e.g., Melby-Lervåg & Hulme, 2012; Strong, Torgerson, Torgerson, & Hulme, 2011).

Studies that target linguistic comprehension skills prior to conventional reading in preschool and kindergarten are likely to include vocabulary training within the context of storybook reading. Studies that only include traditional shared reading instruction (e.g., just reading to

children), in contrast to studies incorporating strategies like dialogic reading, will not be eligible. There are two reasons for this exclusion. Firstly, studies targeting the effect of conventional reading have shown less impressive results for targeting vocabulary knowledge than studies of dialogic reading. Secondly, studies of dialogic reading typically include a comparison group who receive traditional shared reading, while studies targeting the effect of just traditional shared reading are likely to compare the results in the intervention group to a control group who have received no additional activities. This mix-up in comparison groups should be avoided. In addition, short experiments to test the response to different instructional methods of vocabulary knowledge will be excluded.

Studies of school-age children are likely to include linguistic comprehension instruction in the context of text reading. Similar to Elleman et al. (2009), studies that only use repeated reading, read-aloud or independent reading without any instructional method for teaching linguistic comprehension skills will be excluded from the review. Moreover, studies in which the instruction program includes components of e.g. cognitive, social or number skills in addition to linguistic comprehension instruction will be excluded.

Settings

Studies that include training provided in educational settings will be eligible. An intervention has to be conducted within a day-care center, preschool, kindergarten or school setting. The intervention can be delivered by a teacher, assistant or a researcher. It can be provided within a classroom setting, in groups outside the classroom setting or individually. Interventions implemented by parents or in children's home environments will not be included. Interventions implemented in an educational setting plus home condition will not be eligible. This exclusion of parent-child studies is primarily because we want to be able to provide information about how intervention programs should be constructed in educational settings. Hence, we cannot include information from studies that are based on training in different contexts outside this (e.g., home settings). As a group, parents do not have the pedagogical education or experiences that are likely to be present for the persons giving the training in educational settings. In addition, differences in educational background between parents are likely to influence how the children will benefit. Moreover, according to Mol et al. (2008), many studies of parent-child book reading (which is the most common method of home-based linguistic comprehension instruction) lack control of what actually happens in the control and experimental groups, and the instruction programs are in many cases based on voluntary participation.

Training dosage

We will not operate with a prescribed cut off in relation to training dosage. However, because of other defined inclusion and exclusion criteria for this review (e.g., only studies that include global outcome measures are eligible; only RCT and quasi-experiments will be eligible) it is possible that we won't end up with very small-scale studies of a very limited time-frame or just a few training sessions. Moreover, the amount of training that is delivered

to participants will be coded in line with how Marulis and Neuman (2010) coded the concept in relation to duration of training, frequency and intensity (for coding specifications see appendix II).

Types of outcome measures

For studies to be included in the review, an intervention effect on at least one of the two following primary outcome variables must be reported:

1. *Linguistic comprehension:* Reported outcomes must be measured using tests that include items that have not been explicitly trained for in an intervention (e.g., standardized tests; tests created for research purposes that include items not trained in the intervention). Eligible outcomes can include both expressive and receptive tests of linguistic comprehension (e.g. tests of listening comprehension, grammar, morphology, vocabulary and narrative skills).

2. *Reading comprehension:* Reported outcomes must be measured using tests that include test items that have not been explicitly trained in an intervention (e.g., standardized tests; tests created for research purposes that include items not trained in the intervention). Studies to be included with a reading comprehension outcome must report a test in which the child reads a text and answers questions related to the content. Both reported immediate outcomes and long-term outcome effects will be coded. See more about coding details in appendix II.

Studies that only report custom measures that include test items that have been explicitly trained for in an intervention will not be included in the review. However, if custom tests that include test items that have been directly covered in an intervention are reported in addition to measures of generalized language and reading comprehension, this information will be coded (more about the background and rational for excluding studies that solely report custom measures is provided on page 7. A list of the outcomes of interest to be coded is provided in Appendix II.

Search Strategy

The search for relevant literature will include studies published in English since 1986. The search will use multiple sources of information and will be supervised and conducted in collaboration with an information retrieval specialist at the Library of Humanities and Social Sciences, University of Oslo. Details of the search strategy are included in Appendix I.

SEARCH TOOLS AND STRATEGIES

1. **Database selection and strategies.**

- Databases: Studies will be located using electronic searches in the following: PsycINFO, ERIC, ISI Web of Science, ProQuest Digital Dissertations, Linguistics and Language Behavior Abstracts (LLBA), Scopus Science Direct, BASE (Bielefeld Academic Search Engine) and Google Scholar.

➢ Strategy: Keywords and descriptors: A list of search terms that will be used to identify articles is presented in Table 1, Appendix I).

2. **Manual searches**

- Hand search of journals relevant to the topic. The most recent issues of identified key journals will be manually searched towards the end of the information retrieval stage in order to include studies that may not yet have been indexed by the databases. At this point, we will check essential journals that have been valuable in providing included studies up to this point. Table 2 in Appendix I shows a list over relevant key journals.

- Scanning reference lists in earlier meta-analyses and studies. Table 1 in Appendix I shows an overview of earlier meta-analyses. The reference lists of these reviews will be examined for relevant studies. In addition, the authors will consult the reference lists of previous studies.

4. Locating Grey literature. Unpublished reports, such as dissertations, technical reports, and conference presentations, will be located via searches in

- OpenGrey.eu

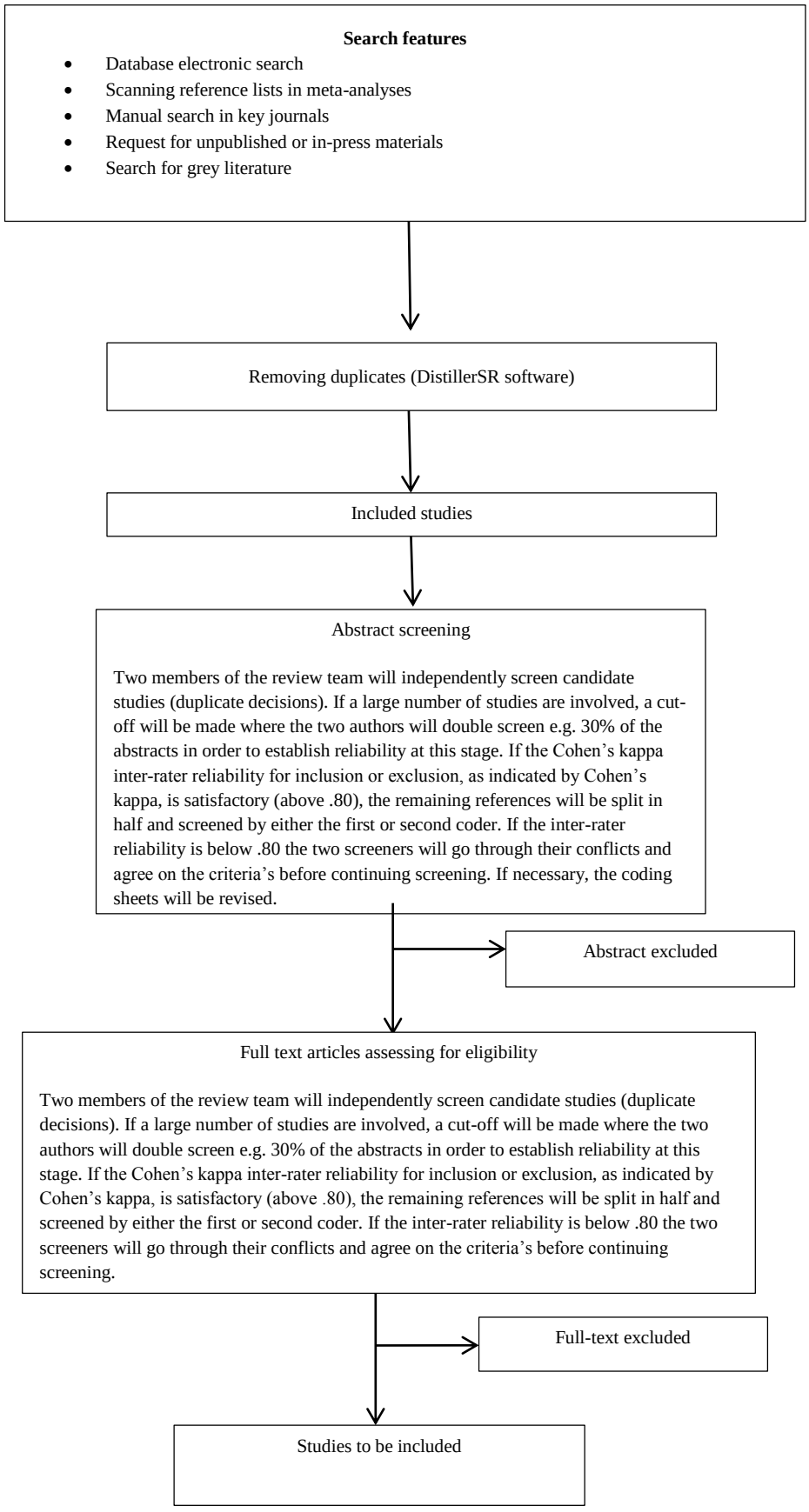
- Proquest Dissertations and Theses

- Search for literature at http://earlyliteracylearning.org/productscr.php

Personal contacts. The listserve of the Society for Research on Educational Effectiveness and Society for the Scientific Study of Reading will be used to ask researchers for in-press or unpublished material.

Procedures for screening and inclusion of studies

Figure 2 below shows the procedure for screening studies to be included in the review.


Search features
Database electronic search
Scanning reference lists in meta-analyses
Manual search in key journals
Request for unpublished or in-press materials
Search for grey literature
Removing duplicates (DistillerSR software)
Included studies
Abstract screening
Two members of the review team will independently screen candidate studies (duplicate decisions). If a large number of studies are involved, a cut-off will be made where the two authors will double screen e.g. 30% of the abstracts in order to establish reliability at this stage. If the Cohen's kappa inter-rater reliability for inclusion or exclusion, as indicated by Cohen's kappa, is satisfactory (above .80), the remaining references will be split in half and screened by either the first or second coder. If the inter-rater reliability is below .80 the two screeners will go through their conflicts and agree on the criteria's before continuing screening. If necessary, the coding sheets will be revised.
Abstract excluded
Full text articles assessing for eligibility
Two members of the review team will independently screen candidate studies (duplicate decisions). If a large number of studies are involved, a cut-off will be made where the two authors will double screen e.g. 30% of the abstracts in order to establish reliability at this stage. If the Cohen's kappa inter-rater reliability for inclusion or exclusion, as indicated by Cohen's kappa, is satisfactory (above .80), the remaining references will be split in half and screened by either the first or second coder. If the inter-rater reliability is below .80 the two screeners will go through their conflicts and agree on the criteria's before continuing screening.
Full-text excluded
Studies to be included

Description of methods used in primary research

Randomized controlled trials and quasi-experimental studies with pre-post designs and control/comparison groups will be included in the review.

The use of pretests in the studies' designs is required to determine whether groups are equivalent and comparable prior to a given intervention program. Posttest measurements in the studies enable us to compute group differences and changes over time. Most studies include a post-test measure conducted immediately after the intervention program has ended. Some studies may also include data from follow-up measurements at later points in time, which we will include in the review.

Criteria for the determination of independent findings

Violating the assumption of independence by computing an overall effect size based on information from the same sample more than once can lead to biased estimates (Borenstein et al. 2007). Thus, several considerations will be made before the studies were coded.

Multiple reports of the same study. It is possible to come across multiple reports from one single study. In such cases the coders will make sure to have the total number of reports before the coding process starts. All the different reports may contribute with additional information to the coding manual. The review authors will choose to extract information from the report that best fit the information needed for each item in the coding manual. Multiple publications within our sample of studies can be detected by finding characteristics such as identical sample sizes, authors, intervention programs or outcome reports. Because multiple publications can lead to an incorrect weighting of study results, authors will be contacted if there are uncertainties regarding the multiple publication of original research.

Multiple studies in single reports. Several studies may be described in one single report. In such cases, each study will be coded separately.

Multiple comparison groups and multiple interventions. It is likely to come across studies that make use of several control groups. Only the intervention and control groups that meet the eligibility criteria will be included. Also, some of the studies are likely compare the same control to different experimental groups and are included in the same analysis of a mean effect size for treated and untreated controls. Since correlations between the multiple comparisons and their outcomes are likely to be rarely reported in the original studies, we will include these studies in the analyses assuming zero correlation between the outcomes.

Multiple outcomes. When a study has multiple indicators for the same construct (for instance more than one global measure of reading comprehension) the mean of the indicators will be used.

Multiple time-point. When a study has multiple time-points, we will use pretest, immediate posttest and the follow up measure that is closes to 6 months after the intervention.

Details of study coding categories

Differential effects from the different studies may be influenced by systematic differences between the studies. Thus, the coding and examination of moderator variables will account for differences between the studies. Studies will be coded on variables related to:

- Contextual information (e.g., year, type of publication)
- Sample characteristics (age (if only Grade is reported age will be recoded to the mean age in years of the grade); SES; Ethnicity; language status; at risk status; reading level status, IQ)
- Methodological information: (e.g., design; recruitment; sample size; type of control group; attrition), treatment/implementation characteristics (setting, mode of delivery; instructor; group size; dosage of intervention)
- Risk of bias:

  Examining the Strength of Evidence using GRADE systems are available for rating research-based evidence (Sanderson, Tatt, Higgins, 2007). In this review, to assess the strength of the evidence, we will adapt the Grading of Recommendations, Assessment, Development and Evaluation system (GRADE; Guyatt et al., 2008; 2011). The GRADE system rates the quality of evidence and is used widely by, for instance, Cochrane and the World Health Organization (Guyatt et al., 2008; 2011). The quality reflects the confidence in the results based on the study limitations and risk of bias. The Cochrane risk of bias tool will be used to assess risk of bias (Higgins et al., 2011). There are three categories: low risk of bias, unclear risk of bias and high risk of bias. Each study will be rated in the "Risk of bias table" (see appendix 2) where judgement of categories (low risk of bias, unclear risk of bias and high risk of bias) will be made according to the domains; sequence generation, allocation concealment, blinding, incomplete outcome data; selective outcome reporting; other sources of bias. Dual assessment of risk of bias will be conducted. In addition to the risk of bias tool, more information related to methodological information will be coded to assess information about study quality (see appendix II).

- Instructional methods (type of training; definitional-contextual scale; type of exposure; level of discussion; type of text involved in the training; depth of processing). These characteristics are included in order to examine why some interventions may be more effective than others. Some of the categories are adapted from Stahl & Fairbanks (1986) and was also used in the study by Elleman et al., 2009).
- Session duration (Hours of training. If only number of sessions are reported, a session is considered to be 30 minutes)

- Outcome (name of test; type of test; global vs custom measure; linguistic comprehension test format; reading comprehension test format).

- Effect size coding

More about the coding procedures is given in Appendix II.

Procedures For Making Inclusion / Exclusion Decisions

The coding of studies will be conducted by two of the authors. Preferably, extracting study characteristics will be made in duplicate.

If there is missing or unclear information on key variables, the authors will be contacted to obtain the necessary coding information. If any key information is unavailable, the variable will be coded as missing.

Statistical procedures and conventions

The analyses will be conducted using mainly the "Comprehensive meta-analysis" program (Borenstein, Hedges, Higgins, & Rothstein, 2005). We will calculate effect sizes by dividing the differences in gain between pre- and post-test in the treatment and the control group by the pooled standard deviation for each group at pretest; this method of effect size calculation for pretest-posttest control group designs is recommended (Morris, 2008). Thus, when the effect size is positive, the group receiving training has made greater pretest-posttest gains than the control group. We will adjust the effect size for small samples using Hedges *g* (Hedges & Olkin, 1985). Effect sizes for follow-up tests will be calculated in an analogous way (pretest to follow-up).

The mean effect sizes will be calculated by a weighted average of individual effect sizes using a random-effects model. Outliers larger than 3 SDs from the mean will be excluded. Notably, since we examine effects from interventions that are rather different when it comes to intervention content and age group they are targeted towards, depending on the number of studies it will be preferably to do separate analyses of the main categories of interventions (i.e. shared book reading, vocabulary training, general language training).

For moderator analyses, we would prefer to the whole model indicated in Figure 1 using a structural equation procedure (MASEM). However, given the potential number of studies and issues with missing data (and that very few experiments tend to report correlations), this is not likely to be realistic because of problems associated with the number of parameters versus the number of studies. However, we will use meta-regression procedures to test aspects of the models in different analyses. Notably, the moderators in Model 1 are restricted to the primary issues that, based on prior studies, have demonstrated clear importance. Increasing the number of moderators is bound to result in Type 2 errors in the analysis that we plan to perform. Thus, we plan to test only the moderators for which there are clear theoretical motivations. Rather than using MASEM it is likely that we will use method of

moments meta-regression for continuous variables (e.g. age, training duration etc). To examine whether effects on reading comprehension are mediated through linguistic comprehension gains, we will set up mediation models using meta-regression. For categorical moderator variables, studies will be separated into subsets based on the categories in the moderator variables (e.g., experiments vs. quasi experiments). A *Q*-test will be used to examine whether the effect sizes differed between subsets. When there are fewer than four studies in a subset ($k < 4$), this analysis will not be conducted. The overlap between confidence intervals will be used to examine the size of the difference between subsets of studies.

Publication bias refers to the notion that a mean effect size can be upwardly biased because only studies with large or significant effects get published (i.e., file-drawer problem with entire studies), or that authors only report data on variables that show effects (often referred to as *p*-hacking, or the file-drawer problem for parts of studies; see Simmons, Nelson, & Simonsohn, 2011). In line with recommendations for meta-analyses, we will make special efforts to retrieve studies from the grey literature and used this as a moderator when possible (Higgins & Green, 2011). To estimate the impact from publication bias statistically, commonly funnel plots have been used in combination with a trim-and-fill analysis. However, there are several problems with the funnel plot/trim-and-fill method (Lau, Ioannidis, Terring, Schmid, & Olkin, 2006). *P*-curve is a recently developed method that deals with the weaknesses in the funnel plot/trim- and-fill analysis (Simonsohn, Nelson & Simmons, 2014). A *p*-curve plots the distribution of statistically significant *p* values ($p < 0.05$) in published studies, and the shape of the *p*-curve is a function only of the effect size and sample size, when the power level is taken into account. If there are true effects, one expects the distribution of published *p*-values to be right-skewed with more low (.01s) than high (.04s) *p*-values. However, if a set of studies is affected by publication bias (because researchers discard entire studies or discard analyses or parts of studies), the *p*-curve becomes left-skewed or flat. Such a form of *p*-curve is said to provide "no evidential value" (i.e., no support for an appreciable effect size). For studies that have published significant findings we will use p-curve to examine bias in these findings.

When coding articles, it is expected that there will be numerous instances of missing data. If data is critical to calculate an effect size, articles with missing data will be excluded if authors do not respond to an email request to provide the data. In cases where an effect size cannot be computed on one outcome but data were missing on other outcomes or moderator variables, the study will be included in all the analyses for which sufficient data were provided.

Treatment of qualitative research

Qualitative research will not be included in the review.

REFERENCES

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin, 128,* 612–637. doi:10.1037/0033-2909.128.4.612.

Beck, I. L., & McKeown, M. G. (2007). Increasing young low- income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal*, *107*(3), 251-271. doi: 10.1086/511706

Biemiller, A. & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98(1)*, 44-62. doi: 10.1037/0022-0663.98.1.44

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology*, *93*, 498–520.

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science,* 8, 445-454. Doi: 10.1177/1745691613491271

Blok, H. (1999). Reading to young children in educational settings: A meta-analysis of recent research. *Language Learning, 49*(2), 343-371. doi: 10.1111/0023-8333.00091

Bornstein, M. H., Hahn, C. S., Putnick, D. L., & Suwalsky, J. T. (2014). Stability of Core language skill from early childhood to adolescence: A latent variable approach. *Child Development*, *5*, 1346–1356. doi:10.1111/cdev.12192.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. (2005). Comprehensive meta-analysis 2.0. *Biostatistics*.

Bowyer-Crane, C., Snowling, M., Duff, F., Fieldsend, E., Carroll, J., M., Miles, J., . . . Hulme, C. (2008). Improving early language and literacy skills: Differential effects of an oral language versus a phonology with reading intervention. *Journal of Child Psychology and Psychiatry, 49*, 422-432. doi: 10.1111/j.1469-7610.2007.01849.x

Bus, A. G., van Ijzendorn, M. H. & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read. A meta- analysis on intergenerational transmission of literacy. *Review of Educational Research, 65*, 1-21. doi: 10.3102/00346543065001001

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 61–100). Washington, DC: American Educational Research Association.

Carraher, D., & Schliemann, A. (2002). The transfer dilemma. *Journal of the Learning Sciences, 11*, 1-24. doi:10.1207/S15327809JLS1101_1.

Clarke, P. J., Snowling, M. J., Truelove E. & Hulme, C. (2010). Ameliorating Children's Reading-Comprehension Difficulties. A Randomized Controlled Trial. *Psychological Science, 21 (8)*, 1106-1116. doi: 10.1177/0956797610375449

Colledge, E., Bishop, D. V., Koeppen-Schomerus, G., Price, T. S., Happé, F. G., Eley, T. C., Dale, P. S., & Plomin, R. (2002). The structure of language abilities at 4 years: A twin study. *Developmental Psychology, 38,* 749–757. doi:10.1037/0012-1649.38.5.749.

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli Jr, R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal*, *110*(1), 1-18. doi: 10.1086/598840

Coyne, M. D., McCoach, D. B., Loftus, S., Zipoli Jr, R., Ruby, M., Crevecoeur, Y. C., & Kapp, S. (2010). Direct and extended vocabulary instruction in kindergarten: Investigating transfer effects. *Journal of Research on Educational Effectiveness*, *3*(2), 93-120. doi: 10.1080/19345741003592410

Daniels, M. (2004). Happy hands: The effect of ASL on hearing childrens literacy. *Reading Research and Instruction, 44*, 86-100. doi: 10.1080/19388070409558422

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*(1), 1-44. doi: 10.1080/19345740802539200

Fischel, J., & Landry, S. H. (2008). Impact of language enhancement interventions on young children's early literacy skills. In *Developing early literacy: Report of the National Early Literacy Panel*, 211-231. Washington, DC: National Institute for Literacy.

Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehension in Kindergarten through Grade 2. *Reading and Writing*, *28*(5), 655-681.

Fricke, S., Bowywer-Crane, C., Haley, A.j., Hulme, C. & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54 (3)*, 280-290. doi: 10.1111/jcpp.12010

Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. *Review of Educational Research*, 68(4), 450-469. doi: 10.1023/A:1015559702157

Goodwin, A. & Ahn, S. (2010). A meta-analysis of morphological interventions: effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia*, 60 (2), 183-208. doi: 10.1007/s11881-010-0041-x

Goodwin, A. & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. Scientific Studies of Reading, 17(4), 257-285. doi: 10.1080/10888438.2012.689791

Gough, P. and Tunmer, W. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6–10. doi: 10.1177/074193258600700104

Guyatt, G. H., Oxman, A. D., Kunz, R., Falck-Ytter, Y., Vist, G. E., Liberati, A., & Schünemann, H. J. (2008). Rating quality of evidence and strength of recommendations: Going from evidence to recommendations. *BMJ: British Medical Journal*, *336*(7652), 1049.

Guyatt, G. H., Oxman, A. D., Schünemann, H. J., Tugwell, P., & Knottnerus, A. (2011). GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *Journal of clinical epidemiology*, *64*(4), 380-382.

Hargrave, A. C. & Sénéchal, M. (2000). Book reading interventions with language-delayed preschool children: The benefits of regular reading and dialogic reading. *Early Childhood Research Quarterly, 15,* 75-90. doi:10.1016/S0885-2006(99)00038-1

Hart, B. & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Paul H. Brookes Publishing.

Hart, B. & Risley, T. R. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator*, *27* (1), 4-9.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Higgins JPT, Altman DG, Sterne JAC (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.

Hoover, W. and Gough, P. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal,* 2, 127–160. doi:10.1007/BF00401799

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., Brownlie, E. B., Douglas, L., Taback, N., Lam, I., & Wang, M. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language and Hearing Research,* 42, 744-760. doi:10.1044/jslhr.4203.744.

Klem, M., Melby-Lervåg, M., Hagtvet, B. , Lyster, S. A. H., Hulme, C., & Gustafsson, J. E. (2014). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*. doi:10.1111/desc.12202

Lau, J., Ioannidis, J. P. A., Terring, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *British Medical Journal, 333,* 597–600. doi:10.1136/bmj.333.7568.597

Law, J., Garrett, Z., Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorders: A meta-analysis. *Journal of Speech, Language and Hearing research, 47, 924-943*.

Lervåg, A. & Aukrust, V. G. (2010). Vocabulary knowledge is a critical determinant of the difference in reading comprehension growth between first and second language learners. *The Journal of Child Psychology and Psychiatry, 51 (5),* 612-620. doi:10.1111/j.1469-7610.2009.02185.x

Lonigan, C., Shanahan, T., & Cunningham, A. (2008). Impact of shared-reading interventions on young children`s early literacy skills. In *Developing early literacy: Report of the National Early Literacy Panel.,* 153-171. Washington, DC: National Institute for Literacy.

MacDonald, M. C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology, 4*, 1-16. doi:10.3389/fpsyg.2013.00226.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review, 109*, 35-54. doi:10.1037/0033-295X. 109.1.35

McNamara. D., S. & Magliano, J. (2009). Toward a comprehensive model of comprehension. Psychology of Learning and Motivation, 51, 297-384. doi: 10.1016/s0079-7421(09)51009-2

Marulis, L. M. & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300-335. doi:10.3102/0034654310377087

Melby-Lervåg, M. & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology,* 49(2), 270-291. doi:10.1037/a0028228.

Melby-Lervåg, Monica & Lervåg, Arne (2014). Reading Comprehension and its Underlying Components in Second Language Learners: A Meta-analysis of Studies Comparing First and Second Language Learners. *Psychological bulletin*. ISSN 0033-2909. *140*(2), s 409- 433. doi:10.1037/a0033890

Mol S. E., Bus A. G., de Jong, M. T. & Smeets, D. J. H. (2008) Added Value of Dialogic Parent–Child Book Readings: A Meta-Analysis, Early Education and Development, 19:1, 7-26. doi:10.1080/10409280701838603

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*(2), 979-1007. doi: 10.3102/0034654309332561

Morris, S. B. (2008). Estimating effect sizes from pretest-posttest-control group designs. *Organizational Research Methods, 11*, 364-386. doi:10.1177/1094428106291059

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes and language skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40,* 665–681. doi:10.1037/0012-1649.40.5.665

National Early Literacy Panel (NELP). (2008). *Developing early literacy: Report of the National Early Literacy Panel.* Washington, DC: National Institute for literacy. Retrieved from https://www.nichd.nih.gov/publications/pubs/documents/NELPReport09.pdf

OECD. (2010a). *Reading for change: performance and engagement across countries. Results from PISA 2000.* Retrieved from http://www.oecd.org/edu/school/programmeforinternationalstudentassessmentpisa/33690904.pdf

OECD. (2010b). *Equal opportunities?: The labour market integration of the children of immigrants*. Retrived from http://www.oecd-ilibrary.org/social-issues-migration-health/equal-opportunities_9789264086395-en

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554. doi:10.1037/0022-0663.98.3.554

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36* (4), 329-347. doi:10.1017/S0140525X12001495.

Sanderson, S., Tatt, I. D., & Higgins, J. P. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: a systematic review and annotated bibliography. *International journal of epidemiology*, *36*(3), 666-676. doi: 10.1093/ije/dym018

Sénéchal, M., Thomas, E., & Monker, J. A. (1995). Individual differences in 4-year-old children's acquisition of vocabulary during storybook reading. *Journal of Educational Psychology*, *87*(2), 218. doi: 10.1037/0022-0663.87.2.218

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*, 1359-1366.

Snow, C. (2001). *Reading for understanding*. Santa Monica, CA: RAND Education & the Science and Technology Police Institute.

Snow, C., Burns, M. S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Strong, K. G., Torgerson, C.J., Torgerson, D., Hulme, C. (2011). A systematic meta-analytic review of evidence for the effectiveness of the FastForWord language intervention program. *Journal of child psychology and psychiatry,* 52, 224-235.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*, 72-110. doi:10.3102/00346543056001072

Storch, S. A. & Whitehurst, G.J. (2002) Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38, 934-937. doi:10.1037/0012-1649.38.6.934

Swanson, E., Vaughn, S., Wanzek, J., Petscher, Y., Heckert, J., Cavanaugh, C., ... & Tackett, K. (2011). A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. *Journal of learning disabilities*, *44*(3), 258-275. doi:10.1177/0022219410378444

Taatgen, N. A. (2013). The nature and transfer of cognitive skills. Psychological Review, 120, 439-471. doi:10.1037/a0033138.

Whitehurst, G. J., & Lonigan, C. J. (1998). Child development and emergent literacy. *Child Development, 69*, 848-872. doi:10.1111/j.1467-8624.1998.tb06247.x.

SOURCES OF SUPPORT

The project is funded by the Norwegian Research Council's Research and Innovation in the Educational Sector (FINNUT) program.

DECLARATIONS OF INTEREST

The authors declare that there are no conflicts of interest.

REVIEW AUTHORS

Lead review author:

Name:	**Kristin Rogde**
Title:	PhD Candidate
Affiliation:	University of Oslo, Department of Education
Address:	Postboks 1092 Blindern
City, State, Province or County:	Oslo
Postal Code:	0317
Country:	Norway
Phone:	0047-22857802
Mobile:	0047-93636424
Email:	kristin.rogde@hotmail.com

Co-authors:

Name:	**Åste Mjelve Hagen**
Title:	Postdoc
Affiliation:	University of Oslo, Department of Education
Address:	Postboks 1092 Blindern
City, State, Province or County:	Oslo
Postal Code:	0317
Country:	Norway
Phone:	0047-22857025
Email:	a.m.m.hagen@iped.uio.no

Name:	**Monica Melby-Lervåg**
Title:	Professor

Affiliation:	University of Oslo, Department of Special Needs Education
Address:	Postboks 1140 Blindern
City, State, Province or County:	Oslo
Postal Code:	0318
Country:	Norway
Phone:	0047-22858138
Email:	monica.melby-lervag@isp.uio.no

Name:	**Arne Lervåg**
Title:	Professor
Affiliation:	University of Oslo, Department of Education
Address:	Postboks 1092 Blindern
City, State, Province or County:	Oslo
Postal Code:	0317
Country:	Norway
Phone:	0047-22844701
Email:	a.o.lervag@iped.uio.no

ROLES AND RESPONSIBILITIES

There is both content and methodological expertise within the review team. The authors are all working on related topics within the field of language, reading development, and intervention. Professor Monica Melby-Lervåg and Professor Arne Lervåg have conducted several meta-analyses and have expertise in statistical analysis. Furthermore, the review team has experience with electronic database retrieval and has access to library support staff when needed.

- Content: Kristin Rogde & Åste M. Hagen
- Systematic review methods: Kristin Rogde, Åste M. Hagen & Monica Melby-Lervåg
- Statistical analysis: Kristin Rogde, Monica Melby-Lervåg & Arne Lervåg
- Information retrieval: Kristin Rogde & Åste M. Hagen

PRELIMINARY TIMEFRAME

Approximate date for submission of the systematic review – June 15th, 2016

PLANS FOR UPDATING THE REVIEW

A new search will be conducted every other year.

AUTHORS' RESPONSIBILITIES

By completing this form, you accept responsibility for preparing, maintaining and updating the review in accordance with Campbell Collaboration policy. The Campbell Collaboration will provide as much support as possible to assist with the preparation of the review.

A draft review must be submitted to the relevant Coordinating Group within two years of protocol publication. If drafts are not submitted before the agreed deadlines, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to de-register the title or transfer the title to alternative authors. The Coordinating Group also has the right to de-register or transfer the title if it does not meet the standards of the Coordinating Group and/or the Campbell Collaboration.

You accept responsibility for maintaining the review in light of new evidence, comments and criticisms, and other developments, and updating the review at least once every five years, or, if requested, transferring responsibility for maintaining the review to others as agreed with the Coordinating Group.

Publication in the Campbell Library

The support of the Coordinating Group in preparing your review is conditional upon your agreement to publish the protocol, finished review, and subsequent updates in the Campbell Library. The Campbell Collaboration places no restrictions on publication of the findings of a Campbell systematic review in a more abbreviated form as a journal article either before or after the publication of the monograph version in *Campbell Systematic Reviews*. Some journals, however, have restrictions that preclude publication of findings that have been, or will be, reported elsewhere and authors considering publication in such a journal should be aware of possible conflict with publication of the monograph version in *Campbell Systematic Reviews*. Publication in a journal after publication or in press status in *Campbell Systematic Reviews* should acknowledge the Campbell version and include a citation to it. Note that systematic reviews published in *Campbell Systematic Reviews* and co-registered with the Cochrane Collaboration may have additional requirements or restrictions for co-publication. Review authors accept responsibility for meeting any co-publication requirements.

I understand the commitment required to undertake a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by: Kristin Rogde

Date: 10.03.2015

Appendix I. Details of search strategy for the identification of eligible studies.

TABLE 1. DATABASE SEARCH TERMS
TABLE 2. SEARCH IN PRIOR REVIEWS/META-ANALYSES
TABLE 3. SEARCH IN JOURNALS

TABLE 1. DATABASE SEARCH TERMS (PUBLISHED 1986-)

Search Terms:	
(OR between all terms) Intervention training instruct* learn* teach* AND (OR between all terms) experiment* "quasi-experiment*" quasiexperiment* RCT "randomised controlled trial" "randomized controlled trial" AND (OR between all terms) Vocabulary "Word knowledge" "Word learning" "Linguistic comprehension" "Oral language" "Listening Comprehension" "Language proficiency" "Language comprehension" "Language skills" "Language abilities" "Language acquisition" Gramma* Syntax Syntactic* Semantic* Morph* "Narrative skills" "Narrative comprehension" Reading "Text comprehension" "sentence comprehension" "passage comprehension"	Search filters: ➢ Date range: 1986- current ➢ Languages: English ➢ Age, school age

TABLE 2. SEARCH IN PRIOR REVIEWS/META-ANALYSES

List of studies:
Blok, H. (1999). Reading to young children in educational settings: A meta-analysis of recent research. *Language Learning, 49*(2), 343-371. doi: 10.1111/0023-8333.00091
Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*(1), 1-44. doi: 10.1080/19345740802539200
Fischel, J., & Landry, S. H. (2008). Impact of language enhancement interventions on young children's early literacy skills. In *Developing early literacy: Report of the National Early Literacy Panel*, 211-231. Washington, DC: National Institute for Literacy.
Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. *Review of Educational Research*, 68(4), 450-469. doi: 10.1023/A:1015559702157
Garrett, Z., Nye, C. (2004). The efficacy of treatment for children with developmental speech and language delay/disorders: A meta-analysis. *Journal of Speech, Language and Hearing research, 47, 924-943.*
Goodwin, A. & Ahn, S. (2010). A meta-analysis of morphological interventions: effects on literacy achievement of children with literacy difficulties. *Annals of Dyslexia*, 60 (2), 183-208. doi: 10.1007/s11881-010-0041-x
Goodwin, A. & Ahn, S. (2013). A meta-analysis of morphological interventions in English: Effects on literacy outcomes for school-age children. Scientific Studies of Reading, 17(4), 257-285. doi: 10.1080/10888438.2012.689791
Lonigan, C., Shanahan, T., & Cunningham, A. (2008). Impact of shared-reading interventions on young children`s early literacy skills. In *Developing early literacy: Report of the National Early Literacy Panel.,* 153-171. Washington, DC: National Institute for Literacy.
Marulis, L. M. & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300-335. doi: 10.3102/0034654310377087
Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: tool to stimulate print knowledge as well as oral language. *Review of Educational Research*, *79*(2), 979-1007. doi: 10.3102/0034654309332561
Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*, 72-110. doi: 10.3102/00346543056001072

List of studies:
Swanson, E., Vaughn, S., Wanzek, J., Petscher, Y., Heckert, J., Cavanaugh, C., ... & Tackett, K. (2011). A synthesis of read-aloud interventions on early reading outcomes among preschool through third graders at risk for reading difficulties. *Journal of learning disabilities*, *44*(3), 258-275. doi:10.1177/0022219410378444

TABLE 3. MANUAL SEARCH IN JOURNALS

Journals
Journal of Research in Reading
Scientific Studies in Reading
Reading and writing
Reading research quarterly
Journal of Research on Educational Effectiveness
Journal of Child Psychology and Psychiatry?
Early Childhood Research Quarterly
Learning and Instruction
Cognition and Instruction
Journal of Literacy Research,
International Journal of Language and Linguistics,
International Journal of Applied Linguistics and English Literature,
Journal of Language Teaching and Research
Language Learning
Journal of Learning Disabilities

Journal of Learning Disabilities
Reading & Writing Quarterly
Journal of Research in Reading
Reading and Writing

APPENDIX II: CODING MANUAL

STUDY IDENTIFICATION

[STUDY ID] Identification number of the study
[CODE DATE] Coder's initials
[CODER] Date of coding

There can be several articles/or reports based on one single study. If this is the case make sure to obtain all articles/or reports before entering information to the coding manual. If one article/or report include several studies – each study must be coded separately.

REPORT IDENTIFICATION

[1] Names of all authors
(Be aware that studies from the same author don't lead to duplicate findings)

[2] Year of publication:xxxx. If there are more than one report, choose the first date.

[3] Country where the study has been conducted.

1. Scandinavia: Norway, Sweden, Finland, Denmark
2. Great Britain
3. Other European country
4. USA
5. Canada
6. Australia/New Zealand
7. Other

[4] Publication type

1. Journal article (published)
2. Book (published)
3. Chapter in a book (published)
4. Dissertation (a: published data/b: unpublished data)
5. Report (a: published data/b: unpublished data)
6. Conference paper (a: published data/b: unpublished data)
7. Other unpublished study/data

If several publication types relate to one study, choose the one from where the effect sizes have been noted from; choose a published version over an unpublished.

SAMPLE CHARACTERISTICS

[5] Mean age of children receiving treatment (age at first time point) (months)
[6] Mean age of children in control group (age at first time point) (months)
[7] Mean age of children receiving treatment (age at first posttest point) (months)
[8] Mean age of children in control group (age at first posttest point) (months)
[9] Mean age of children receiving treatment (age at follow-up test point) (months)
[10] Mean age of children in control group (age at follow-up test point) (months)

If grade is given but not age, grade should be recoded to the mean age in years of the grade.

[11] Socioeconomic status

1. Low
2. Middle to high
4. Mixed

If SES status is not reported in line with the categories, use indicators like free/reduced lunches, parental educational level/income level or neighborhood income/poverty in the community to decide which category to code. If it is uncertain which category to choose, type in a brief description of relevant SES variables that relates to the sample (e.g. suburban; rural).

[12] Ethnicity_ describe

[13] Sample language learner status

1. Monolingual. The total sample has the instructional language as the first language
2. Mixed (parts of the sample have the instructional language as a second language_describe distribution)
3. Second-language learners

[] If the sample consists of second-language learners, describe sample characteristics related to the type of first languages.

[14] At risk status

1. Children not reported to be at risk (typically achieving)
2. Children reported to be at risk for language/reading problems or described as learning or reading disabled
3. Second-language learners (L2)

4. Mixed_describe

[15] Reading level of participants during the intervention period
1. prereading - before reading instruction
2. after reading instruction

[16] IQ: Information from nonverbal tests (e.g. scores from the subtest of the Wechsler Preschool and Primary Scales of Intelligence or scores from the Standford-Binet Intelligence Scale).

METHODOLOGICAL INFORMATION

[17] Design
1. Randomized controlled trial
2. Quasi-experimental design

Number of schools and classrooms – answer to the applicable questions:

[18] Number of schools involved in the study_

[19] Number of schools that is represented by the intervention group
[20] Number of schools that is represented by the control group

[21] Number of classrooms/groups represented in the intervention group
[22] Number of classrooms/groups represented in the control group

[23] Recruitment
1. Describe how sample was obtained_
2. Not reported

[24] Sample size
1. Sample size intervention group
2. Sample size control group

[25] Type of control group
1. Untreated/treatment as usual or waiting control group
2. Alternative treatment [also describe what? E.g. comprehension strategies, nonlinguistic related training]
3. Alternative treatment with linguistic aspects [make a description of what] (e.g. phonological awareness training, letter knowledge, orthographic knowledge)

[26] How has the control group been selected?

1. Describe_

[27] Details about the intervention: Has the intervention program been described in detail?

1. Yes
2. No (It is unclear what kind of instruction has been given or how the training sessions have been conducted and implemented. Essential information is omitted)

Attrition

[28] Number of participants lost to intervention group

[29] Number of participants lost to control group

[30] Has the person given the training been given instruction/training?

1. Yes
2. No
3. No information

If so, how much?

[31] Number of training meetings

[32] Hours of training sessions

[33] Does the study report strategies for monitoring treatment fidelity during the intervention? (Has there been any kind of observation/meetings/report system/video records/audio records to monitor the implementation of the intervention?)

1. Yes
2. No
3. No information

[34] Did the instruction program include concrete instructions for the training lessons (for the instructors?)? (This may be provided in an appendix or reported)

1. Yes
2. No
3. No information

[35] Any treatment fidelity assessment? Describe method and outcome

RISK OF BIAS – The Cochrane Collaboration`s tool for assessing risk of bias

Each study will be rated in the "Risk of bias table" below.

RISK OF BIAS TABLE

ENTRY	Judgement	Support for judgement
Sequence generation		
Allocation concealment		
Blinding		
Incomplete outcome reporting		
Other sources of bias		

Enter the judgement "Low risk of bias; unclear risk of bias or high risk of bias".

[36] Sequence generation

[34] Allocation concealment

[38] Blinding

[39] Incomplete outcome reporting

[40] Other sources of bias

TREATMENT/IMPLEMENTATION CHARACTERISTICS

[41] Language of instruction

_ Report the language in which the intervention program is given

[42] Educational setting

1. Day-Care Center
2. Preschool
3. Kindergarten
4. School

[43] Mode of delivery

1. Computerized
2. Tutor led

3. Individual practice
4. Mixed [what?]

[44] Instructor

1. Teacher (certified)
2. Teacher (assistant/not certified)
3. Researcher
4. Other (specify)

[45] Setting/group size

1. Classroom instruction (10-or more)
2. Larger groups 6-10
3. Small group instruction 5 or less
4. Individual instruction
5. Mixed small group and individual instruction

DOSAGE OF INTERVENTION

Duration of training

[46] Total number of weeks

Frequency

[47] A) Total number of training sessions

[48] B) Total number of training hours

Intensity

[49] Minutes per training session

If sessions are given but not the amount of time, consider a sessions to be 30 minutes (+ add a note if this is the case).

INSTRUCTIONAL METHODS

[50] Type of intervention (training components) in studies **before** formal reading instruction has begun

1. Interactive reading/dialogic reading (not traditional book reading, some interactive elements have to be included)
2. Interactive storybook reading/dialogic reading plus additional linguistic comprehension instruction in other settings
3. General linguistic comprehension instruction without the interactive storybook reading element

[51] Type of intervention (training components) in studies **after** formal reading instruction has begun

1. The intervention is delivered in the context of reading texts – strategier?

[52] Focus of instruction

1. The training focus on one specific dimension within linguistic comprehension [Describe which?] (e.g. vocabulary, morphology, syntax)

2. The training relates to several linguistic dimensions
3. In addition to 1. Or 2., the study also focus on meta-linguistic…inference skills

[53] DEFINITIONAL-CONTEXTUAL SCALE (relevant for studies of vocabulary training)

1. Definitional only. Definitions or synonyms are provided without any use of context (no examples of the word used in context).
2. Definitional emphasis. A limited amount of context is provided, but an emphasis is placed on learning the definition
3. Balanced. The instruction contains nearly equal emphasis on definitional and contextual information.
4. Contextual emphasis. Although definitions are provided or derived from context, the focus of instruction is on understanding the words in context
5. Context only. The focus of the instruction is to expose the child to words in context with no definitions provided.

[54] TYPE OF EXPOSURE

1. less than three repetitions in a single context
2. More than three repetitions in a single context
3. More than three repetitions in multiple contexts

[55] LEVELS OF DISCUSSION

1. Little to no discussion
2. High levels of discussion

[56] Word specific versus generative instruction

1. Teaching of specific words
2. Teaching of specific words and strategies to increase student's generative word knowledge.

[57] DEPTH OF PROCESSING

1. Association. The child learns an association between a new word and a definition (or a single context).

2. Comprehension. The child demonstrates comprehension by doing something with the definitional information such as classifying words or providing antonyms.
3. Generation. The student produce a novel response to the word (e.g., an original sentence, a restatement of the definition in the child's own words (oral or written)).

[58] TYPE OF TEXT INVOLVED IN THE TRAINING

1. Expository texts
2. Narrative texts
3. Other

OUTCOME

All type of linguistic comprehension and reading comprehension outcomes must be coded following these guidelines:

[59] Name of test: _

[60] Type of test: _ *(e.g., vocabulary; narrative skills; listening measures reading comprehension)*

[61] Is the test designed to capture effects of near or far transfer?

1. Global measure (standardized) (far transfer)
2. Global measure (researcher created) (far transfer)
3. Custom measure (near transfer – e.g., trained words)

[62] Internal consistency reliability for the outcome measure

1. Cronbach's alfa score_
2. Not reported

LINGUISTIC COMPREHENSION TESTS

[63] How is the test format?

1. Multiple choice
2. Comprehension measured by retelling / recall
3. Other – describe: _

READING COMPREHENSION TESTS

[64] Reading comprehension outcome

Is the child corrected on incorrect decoding of words during the reading test?

1. Yes
2. No
3. No information

[65] Is the test a read aloud test or a silent reading test?

1. Read aloud
2. Silent reading
3. No information

[66] How is the test format?

1. Multiple choice
2. Comprehension measured by retelling / recall
3. Control questions after reading (silent or aloud)

[67] Type of text to measure reading comprehension

1. Describe the type of text

EFFECT SIZE CODING

Two different types of effect sizes can be coded:

[68] Pretest effect size. The difference between the treatment and control group **before** treatment on the same variable as an outcome variable.

[69] Posttest effect size. The difference between the treatment and the control group **after** treatment on an outcome variable.

There may be several pretest effect sizes and posttest effect sizes if the study reports several outcome variables of interest. In addition, studies that include follow-up measurement will also provide several effect sizes for one outcome variable. It is important to discriminate between e.g., first- and second post-tests by adding information about this if necessary.

[70] Information about time of post-test

Enter information about time point for the post-test after treatment ended.

[71] Information about post-test follow-up time points.

If the study provides multiple post-tests, enter information about time after treatment.