CampbellCollaboration

Education Coordinating Group

Preschool predictors of later reading comprehension ability: a systematic review

Hanne Næss Hjetland, Ellen Irén Brinchmann, Ronny Scherer, Monica Melby-Lervåg

A Campbell Systematic Review
2017:14

Published: December 2017
Search executed: February 2016

CampbellCollaboration

The Campbell Library comprises:

- Systematic reviews (titles, protocols and reviews)
- Policies and Guidelines Series
- Methods Series

Go to the library to download these resources, at:

www.campbellcollaboration.org/library/

Better evidence for a better world

Colophon

Title	Preschool predictors of later reading comprehension ability: a systematic review
Authors	Hjetland, Hanne Næss Brinchmann, Ellen Irén Scherer, Ronny Melby-Lervåg, Monica
DOI	10.4073/csr.2017.14
No. of pages	155
Citation	Hjetland H. N., Brinchmann E. I., Scherer R., & Melby-Lervåg M. Preschool predictors of later reading comprehension ability: a systematic review. Campbell Systematic Reviews 2017:14 DOI: 10.4073/csr.2017.14
ISSN	1891-1803
Copyright	© Hjetland et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Roles and responsibilities	The members of the review team possess substantial expertise in terms of both content and methodology. All of the contributors to this review are working in the field of language and reading comprehension. Professor Monica Melby-Lervåg has extensive experience with conducting meta-analyses and has the required statistical analysis competence. The first and last authors have also completed a two-day course on meta-analysis with Michael Borenstein (October 2013), using Comprehensive Meta-Analysis version 3. All authors have knowledge of and experience with structural equation modeling, and they attended a two-day workshop on meta-analytic structural equation modeling with Associate Professor Mike Cheung (National University of Singapore) (Oslo, 6-7 October 2015). In addition, the review team has experience with electronic database retrieval and coding.
Editors for this review	Editor: Sandra Wilson Managing editor: Carlton J. Fong
Declarations of interest	The authors have no vested interest in the outcomes of this review, nor any incentive to represent findings in a biased manner.
Corresponding author	Hanne Næss Hjetland University of Oslo, Department of Special Needs Education P.O. Box 1140 Blindern Oslo N-0318 Norway E-mail: h.n.hjetland@isp.uio.no Full list of author information is available at the end of the article

Campbell Systematic Reviews

The Campbell Collaboration was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services. Campbell offers editorial and methodological support to review authors throughout the process of producing a systematic review. A number of Campbell's editors, librarians, methodologists and external peer reviewers contribute.

The Campbell Collaboration
P.O. Box 4404 Nydalen
0403 Oslo, Norway
www.campbellcollaboration.org

Table of contents

Plain language summary

Preschool language skills are associated with better reading comprehension at school

The evidence suggests that successful instruction for reading comprehension should target a broad set of language skills.

The review in brief

A variety of language skills related to both language comprehension (e.g., vocabulary and grammar) and code-related skills (e.g., phonological awareness and letter knowledge) is important for developing decoding skills and, in turn, reading comprehension in school. Thus, reading comprehension instruction is more likely to be successful if it focuses on a broad set of language skills.

What is this review about?

Determining how to provide the best instruction to support children's reading comprehension requires an understanding of how reading comprehension actually develops. To promote our understanding of this process, this review summarizes evidence from observations of the development of language and reading comprehension from the preschool years into school. The main outcome in this review is reading comprehension skills.

What is the aim of this review?
This Campbell systematic review examines the relationships between skills in preschool and later reading comprehension. The review summarizes evidence from 64 longitudinal studies that have observed these relationships.

Understanding the development of reading comprehension and its precursors can help us develop hypotheses about what effective instruction must comprise to facilitate well-functioning reading comprehension skills. These hypotheses can be tested in randomized controlled trials.

What studies are included?

This review includes studies that observe the relationship between preschool language and code-related skills and later reading comprehension. A total of 64 studies were identified, all of which were included in the analysis. However, several of them suffered from considerable attrition, used convenience sampling, included a selected sample and failed to report on important study and sample characteristics.

The studies spanned 1986 to 2016 and were mostly performed in the USA, Europe and Australia.

What are the main findings of this review?

Code-related skills in preschool (e.g., phoneme awareness and letter knowledge) are indirectly related to reading comprehension via word decoding. Linguistic comprehension is directly related to reading comprehension skills. Code-related skills and linguistic comprehension were strongly related. Moreover, language comprehension was more important for reading comprehension in older readers than in younger readers.

What do the findings of this review mean?

These results show that a broad set of language skills is important in developing reading comprehension. The results also suggest that successful instruction for reading comprehension should target a broad set of language skills.

In future studies, the effectiveness of instruction that targets such a set must be tested in randomized controlled trials. Additionally, future longitudinal studies should address issues of reliability, missing data and representativeness.

How up-to-date is this review?

The review authors searched for studies up to 2016. This Campbell systematic review was published in December 2017.

Executive summary/Abstract

Background

Knowledge about preschool predictors of later reading comprehension is valuable for several reasons. On a general level, longitudinal studies can aid in generating understanding and causal hypotheses about language and literacy development, both of which are crucial processes in child development. A better understanding of these developmental processes may guide the establishment of effective instruction and interventions to teach reading comprehension that can later be tested in randomized controlled trials. Knowledge about preschool precursors for reading comprehension skills can also aid in developing tools to identify children at risk of reading difficulties.

Objectives

The primary objective for this systematic review is to summarize the available research on the correlation between reading-related preschool predictors and later reading comprehension skills.

Search methods

We developed a comprehensive search strategy in collaboration with a search information retrieval specialist at the university library. The electronic search was based on seven different databases. We also manually searched the table of contents of three key journals to find additional references. Finally, we checked the studies included in two previous systematic reviews.

Selection criteria

The included studies had to employ a longitudinal non-experimental/observational design. To avoid the overrepresentation of participants with special group affiliation (e.g., participants with learning disabilities or second language learner status), we chose studies that included either a sample of typically developing children or an unselected cohort.

Data collection and analysis

The search resulted in 3285 references. After the duplicates were removed, all remaining references were screened for inclusion and exclusion. A total of 64 studies met the eligibility criteria.

The analysis was conducted in two steps. First, the predictive relation between the abilities assessed in preschool and later reading comprehension skills was analyzed using Comprehensive Meta-analysis (CMA) software. Second, we used the correlation matrices in the included studies to further explore these relations by means of meta-analytic structural equation modeling.

Results

First, analyses of bivariate correlations showed that all the included predictors, except for non-word repetition, were moderately to strongly correlated with later reading comprehension in the bivariate analyses. Non-word repetition had only a weak to moderate contribution to later reading comprehension ability. To explain the between-study variation, we conducted a series of meta-regression analyses. Age at time of reading assessment could predict variations between studies in correlations related to the code-related predictors.

Second, meta-analytic structural equation modeling showed a significant indirect effect of code-related skills on reading comprehension via consecutive word recognition. Third, there was a strong relationship in preschool between language comprehension and code-related skills. Language comprehension had a moderate direct impact on reading comprehension. As hypothesized, this impact increased with age, and linguistic comprehension becomes more important for reading comprehension when children master decoding. Moreover, the overall individual variance in reading comprehension explained by the model was 59.5%; that of consecutive word recognition was 47.6%.

Authors' conclusions

Overall, our findings show that the foundation for reading comprehension is established in the preschool years through the development of language comprehension and code-related skills. Code-related skills and decoding are most important for reading comprehension in beginning readers, but linguistic comprehension gradually takes over as children become older. Taken together, these results suggest a need for a broad focus on language in preschool-age children.

Background

Development of reading comprehension

The ability to extract meaning from text is the core of reading comprehension. In today's information-driven society, the development of reading comprehension skills is of vital importance, both for academic performance and for participation in society and work-life (NELP, 2008).

Longitudinal studies that follow children's language and literacy skills over time can contribute to our knowledge of children's development of reading comprehension. However, the findings of such studies are not merely of theoretical interest; the knowledge gained from longitudinal studies is also of practical importance. For instance, by identifying precursors of reading comprehension, we may be able to recognize signs of delayed or divergent development. Thus, when a child shows early signs of poor language development, we can implement, with greater certainty, additional efforts to prevent later reading struggles. Moreover, by gaining insight into children's literacy development, we can develop causal hypotheses about how to enrich their learning environments and adapt instructional activities according to their individual needs. In summary, longitudinal studies of reading comprehension are important for at least two reasons: 1) to strengthen our ability to recognize and remedy early signs of reading comprehension difficulties and 2) to help us provide learning contexts that allow children to build a solid foundation for reading comprehension. Although longitudinal studies are an important means of generating causal hypotheses and theory to understand a phenomenon, to provide more conclusive knowledge about causality and the effectiveness of instruction, this must be tested in randomized controlled trials.

Over the past 15 years, longitudinal studies of reading comprehension have increased rapidly, but their results have been inconsistent. Studies vary greatly in the reported strength of early predictors of reading comprehension. For instance, some studies identify strong predictive relations between vocabulary and later reading comprehension (Roth, Speece, & Cooper, 2002), whereas others show a weak relation (Fricke, Szczerbinski, Fox-Boyer, & Stackhouse, 2016). This situation is problematic because divergent findings that are not clearly replicated limit the conclusions that we can draw from previous research.

Although this variation in the results of previous studies may be explained in various ways, some of the discrepancies most likely stem from measurement issues. For instance, different measures vary in their reliability. Because measurement error attenuates correlations in bivariate relations, the reliability of measures affects the strength of the relations between variables (Cole & Preacher, 2014). If the reliability of predictors differs, a predictor with good reliability is likely to supersede and explain variation beyond those with poor reliability (Cole & Preacher, 2014).

Additionally, we often regard abilities such as reading comprehension and language skills as constructs, even when they are measured using single indicators (usually a psychometric test). Because different tests capture various parts of a theoretical construct, results will vary depending on test properties. This variation is especially relevant for the many studies in this field that do not use latent variables to examine construct dimensionality and to control for measurement error (Bollen, 1989).

In conclusion, issues concerning measurement and construct validity may cause irrelevant variation in the results of single studies, and therefore limit conclusions based on previous research. Thus, the present study sought to meet the great need for a systematic review that summarizes longitudinal studies of reading comprehension while considering measurement issues. More specifically, we aimed to provide the best possible estimates of early predictors of reading comprehension by using a structural equation modeling approach (SEM) to meta-analysis (Cheung, 2015). We argue that the present study has both methodological and theoretical merits; it represents a promising avenue for summarizing research findings across studies and adds to our understanding of the development of reading comprehension.

Theories of reading comprehension that can inform the review

According to Gough and Tunmer's (1986) "simple view of reading", reading comprehension is the product of decoding and linguistic comprehension. Hoover and Gough (1990) define decoding as efficient word recognition: "[it is] the ability to rapidly derive a representation from printed input that allows access to the appropriate entry in the mental lexicon, and thus, the retrieval of semantic information on the word level" (p. 130). Linguistic comprehension is defined as "the ability to take lexical information (i.e., semantic information at the word level) and derive sentence and discourse interpretations" (Hoover & Gough, 1990, p. 131).

Notably, this "simple view" does not deny that capacities such as phonemic awareness, vocabulary knowledge, or orthographic awareness are important to reading; rather, it suggests that they are sub-skills or predictors of decoding and/or linguistic comprehension (Conners, 2009). Because the two components (decoding and linguistic comprehension) and their underlying skills simultaneously affect one another, fully disentangling these skills is difficult (Clarke, Truelove, Hulme, & Snowling, 2014).

Although there is support for the “the simple view of reading”, there are also researchers who argue that this model is too simple to explain the full complexity of reading comprehension (Chen & Vellutino, 1997; Conners, 2009; Hoover & Gough, 1990). For instance, some longitudinal studies provide support for models depicting an augmented simple view of reading (Geva & Farnia, 2012; Johnston & Kirby, 2006; Oakhill & Cain, 2012). In addition to decoding and linguistic comprehension, these augmented models typically include domain-general cognitive skills as part of the reading comprehension construct. Additionally, some augmented models depict language-related skills such as verbal working memory and inference skills as distinct components of reading comprehension (Cain, Oakhill, & Bryant, 2004). Some longitudinal studies show significant contributions to reading comprehension from cognitive skills, working memory and inference skills beyond word recognition and linguistic comprehension (Oakhill & Cain, 2012). However, the results of these studies are not consistent. For instance, a cross-sectional study did not find a predictive ability of verbal working memory beyond decoding, listening comprehension and vocabulary (Cutting & Scarborough, 2006).

In summary, the simple view of reading defines reading comprehension as a product of decoding and linguistic comprehension, whereas the augmented view of reading advocates a wider perspective on the linguistic and cognitive processes involved in reading comprehension.

The simple view of reading has been, and still is, an influential framework for explaining the abilities necessary for reading with understanding in children in primary and early secondary school (which is the main focus of this review). However, notably, other theoretical models exist that have commonly been used to understand the development of reading comprehension. As mentioned above, some of the alternative models posit that there is a need to modify or augment the simple view of reading by adding other components or redefining the definition of reading. With the component model of reading, Joshi and Aaron (2000) proposed adding speed of processing as an additional component in the simple view of reading. Speed of processing explained 10% of additional variance beyond that of decoding and listening comprehension; thus, a modified (augmented) model of reading is proposed ($R = D \times C + S$). In the Reading Efficiency Model, reading is defined as the ability to comprehend text and the ability to read text fluently (Høien-Tengesdal & Høien, 2012). This proposed model is expressed as $R_E = D_E \times LC$. Rather than traditional reading comprehension, R_E is a composite score that combines reading comprehension and oral text reading fluency.

In addition, for children older than primary or secondary school age, a number of models have been used to describe how reading comprehension evolves (Cromley & Azevedo, 2007; Kintsch, 1988; McNamara & Kintsch, 1996; Perfetti & Stafura, 2014). However, despite the variety of different models used to explain the development of reading comprehension, it seems fair to conclude that in elementary school children, the simple view of reading is the framework with the strongest empirical support.

Word decoding is an important part of reading comprehension, and for word decoding, there are a number of different models. For instance, connectionist models (Seidenberg, 2005) explain the underlying mechanisms of word reading. Another important framework for understanding word decoding is the dual route theory. This theory is based on the notion that there are two routes from print to speech (i.e., reading aloud) – one that consults the mental lexicon and one that does not (Coltheart, 2006). By breaking the reading construct down into simpler components that are more immediately amenable to examination, the hope is that the greater our understanding is of the components, the closer we are to understanding reading. To understand how we comprehend sentences, it is necessary to know how we recognize whole words. Thus, if we have knowledge about how we recognize whole words in a text, we have a better chance of understanding how people comprehend sentences.

Thus, a primary goal of our review is to gain understanding about how reading comprehension and its precursors develop from preschool and into elementary school. We know that many studies have examined this process; thus, our main aim is to ascertain which findings in the previous studies are robust and replicated across studies and which need further support and investigation. As mentioned, one theoretical issue that has been debated is whether previous research supports the "simple" two-factor model of reading comprehension or whether we need to broaden our understanding of the reading comprehension construct and the skills underlying children's ability to understand written text. Thus, in the following sections, we will further discuss the main predictors of reading comprehension that have been used in previous studies, namely, decoding, linguistic comprehension and domain-general cognitive skills.

Preschool predictors of decoding

Concerning the decoding component in the simple view of reading, previous studies have consistently demonstrated that phonological awareness, letter knowledge and rapid automatized naming (RAN) play a key role in its development. These variables are thus, central to the process of learning, and later automatizing, letter-sound correspondences.

This central role was demonstrated in a study by Lervåg, Bråten, and Hulme (2009), who conducted a two-year longitudinal study in which phoneme awareness, letter-sound knowledge, and non-alphanumeric RAN were measured four times, beginning 10 months before the onset of reading instruction. The results showed unique contributions from the three predictor variables to the growth of decoding skills in the early stages of development.

Further studies and meta-analyses yielded similar findings (Hatcher, Hulme & Snowling, 2004; Høien & Lundberg, 2000; Lundberg, Frost & Petersen, 1988; Melby-Lervåg & Lervåg 2011). Despite the strong evidence supporting the predictive powers of phonological awareness, letter knowledge and RAN, there is still some uncertainty as to how these variables are related to one another and to the development of decoding. Indeed, the role of phonological awareness and letter knowledge is easy to understand. After all, in alphabetic languages decoding can be defined as the very process of linking letters and phonemes. This central role of letter knowledge and phonological awareness is also reflected on an empirical level.

For instance, Muter, Hulme, Snowling, and Stevenson (2004) reported that letter knowledge measured upon school entry is a powerful predictor of early decoding ability. Likewise, Melby-Lervåg, Lyster, and Hulme (2012b) noted that phoneme-level awareness is especially crucial for the development of decoding skills. In the latter study, phonological awareness and letter knowledge assessed upon school entry explained 54% of the variance in decoding ability one year later (Melby-Lervåg, Lyster, & Hulme, 2012b).

Although it is easy to explain the importance of letter knowledge and phoneme awareness, the role of RAN in the development of decoding is less intuitive. RAN refers to the speed at which one can identify known symbols, numbers or letters, but why this ability explains unique variation in decoding is not immediately clear. However, several explanations have been suggested. In one view, naming speed represents a demanding combination of attentional, perceptual, conceptual, memory, lexical, and articulatory processes that in turn enhances or constrains one's ability to recognize orthographic patterns in a text (Wolf, Bowers, & Biddle, 2000). Additionally, several studies have shown that particularly in transparent orthographies RAN, together with phoneme awareness and letter knowledge, is a strong predictor of growth in reading fluency (Lervåg & Hulme, 2009).

Concerning the relationship with reading comprehension, most studies find that rapid naming operates indirectly to influence reading comprehension through word decoding. For instance, Johnston and Kirby (2006) observed that the unique contribution of naming speed was relatively small and that naming speed contributed primarily in terms of word recognition. They also acknowledged that when the word recognition component is included, naming speed does not uniquely contribute to reading comprehension.

Because previous studies have identified phonological awareness, letter knowledge and RAN as unique predictors of decoding, these three variables were included in the present meta-analysis. The aim of this meta-analysis is to investigate how the variables are related to one another and how they contribute to the development of decoding and reading comprehension.

Preschool predictors of linguistic comprehension

In contrast to decoding, which is a constrained skill and a more unitary construct, several recent studies show that linguistic comprehension comprises a broad set of language skills that are imperative to the ability to understand spoken language (Bornstein, Hahn, Putnick, & Suwalsky, 2014; Hjetland et al., under review; Klem, Melby-Lervåg, Hagtvet, Lyster, Gustafsson, & Hulme, 2015; Lervåg, Hulme, & Melby-Lervåg, 2017). These studies show that the linguistic comprehension construct consists of both receptive and expressive vocabulary, grammatical skills, and narrative skill. This core language comprehension construct is highly stable; the rank order between children in linguistic comprehension remains almost unchanged over time (Melby-Lervåg et al., 2012a).

Although the simple view of reading postulates that reading comprehension is the product of decoding and linguistic comprehension, the relative impact of these two components changes over time. In a meta-analysis, García and Cain (2014) showed that the contribution of decoding to reading comprehension decreases with age, whereas the contribution of listening comprehension increases. In other words, as children progress in their development, linguistic comprehension becomes paramount for good reading comprehension (Carroll, 2011).

The relation between reading and linguistic comprehension is not difficult to explain; to comprehend what one reads, one must understand language in its spoken form (Cain & Oakhill, 2007). However, as mentioned, linguistic comprehension is a complex ability that consists of several sub-skills, such as vocabulary, grammar, and inference skills (Kim, Oines, & Sikos, 2015; Lervåg, Hulme, & Melby-Lervåg, 2017). Among these skills, vocabulary and grammar are emphasized as particularly important aspects of language that are likely to influence reading development (Brimo, Apel, & Fountain, 2017). For instance, vocabulary knowledge is believed to have an impact both in learning to recognize individual words and in developing text comprehension skills (Cain & Oakhill, 2007). Similarly, some researchers have suggested that grammatical abilities such as syntactic and morphological knowledge may contribute to reading comprehension by helping students detect and correct word recognition errors and infer the meanings of unknown words (Cain & Oakhill, 2007).

Although Cain and Oakhill (2007) emphasized the role of vocabulary and grammatical abilities, knowing the meanings of words and sentences is not always sufficient to understand written materials. In text, information is often implied, and readers must use their background knowledge and reasoning skills to discover what is not directly stated. Accordingly, studies have demonstrated that higher-order linguistic processes explain variance in reading comprehension beyond vocabulary and grammar (Oakhill & Cain, 2012). However, although it can be argued that background knowledge and inference skills represent important aspects of the linguistic comprehension construct, these types of traits are usually not measured at early developmental levels. Because the present study is

concerned with preschool predictors of reading comprehension, background knowledge and inference skills were not included as variables in our review.
Concerning the relationships between linguistic comprehension and reading comprehension, both Lervåg et al. (2017) and Foorman, Koon, Petscher, Mitchell, and Truckenmiller (2015) demonstrated that measures of vocabulary and grammatical abilities were not unique predictors of reading comprehension. Instead, the vast amount of variance in reading comprehension in fourth-10^{th} grades was accounted for by a general oral language factor comprising both vocabulary and grammar. Thus, in the present review, we used a structural equation modeling approach to investigate the shared contribution of vocabulary and grammatical abilities to reading comprehension. However, the bivariate correlation between each of these predictors and reading comprehension is also reported.

Preschool domain-general cognitive skills as predictors of later reading comprehension

As mentioned earlier, in addition to linguistic comprehension and decoding, research has suggested that cognitive abilities such as memory are an integral part of the reading comprehension construct. Two different memory functions are often considered: (1) short-term memory, that is, "the capacity to store material over time in situations that do not impose other competing cognitive demands" (Florit, Roch, Altoè, & Levorato, 2009, p. 936), and (2) working memory, that is, "the capacity to store information while engaging in other cognitively demanding activities" (Florit et al., 2009, p.936). In a longitudinal study conducted by Cain et al. (2004), working memory capacity and component skills of comprehension predicted unique variance in reading comprehension. Florit et al. (2009) referred to previous studies that suggest that reading comprehension partly depends on the capacity of working memory to maintain and manipulate information. Cain et al. (2004) noted that working memory capacity appears to be directly related to reading comprehension over and above short-term memory, word reading, and vocabulary knowledge. In addition to linguistic comprehension and decoding, this review also aims to explore the contribution of memory skills (i.e., short-term memory and working memory) to reading comprehension. We must consider, however, that many memory tasks are language based. In some studies, these tasks have been found to load on a linguistic comprehension factor rather than a separate memory factor or domain-general cognitive skills (Klem et al., 2015; Lervåg et al., 2017; Melby-Lervåg et al., 2012a). Thus, this consideration is important as we examine and interpret the relationship between memory and reading comprehension.

In addition to working memory and other memory skills, studies have also found that other domain-general cognitive skills, such as nonverbal IQ, uniquely explain variation in reading comprehension skills. The present review therefore includes components of domain-general cognitive skills.

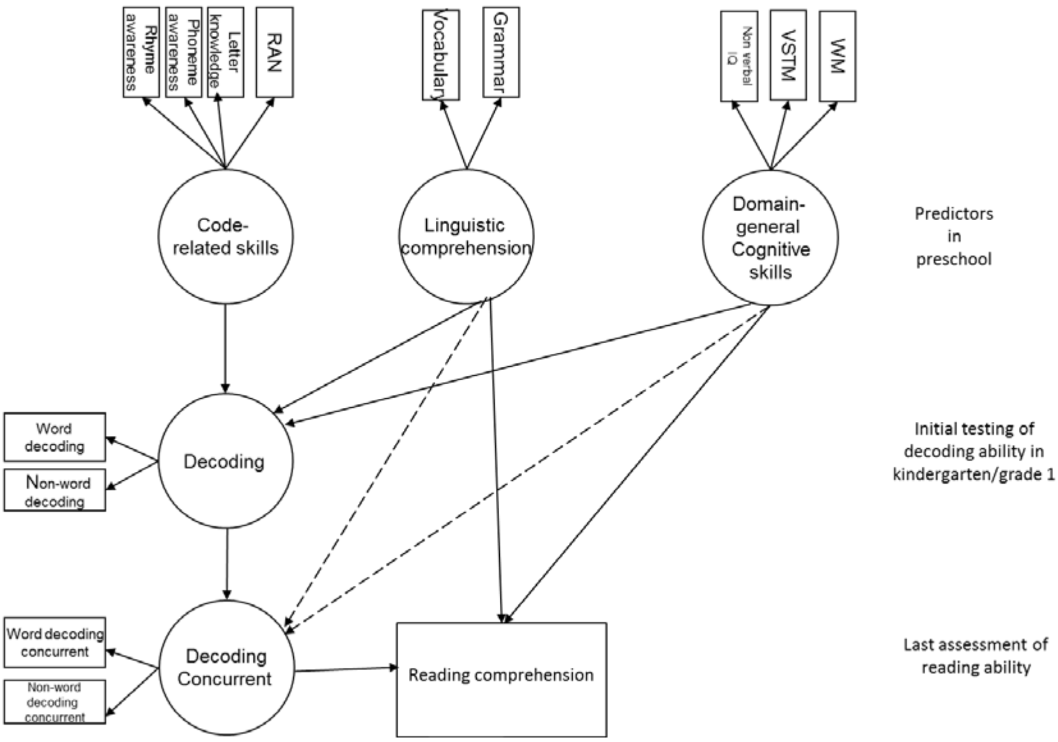
Model

Figure 1 summarizes the model of reading comprehension and its underlying cognitive and language-related skills that we examine in the current review. This model is based on theory and research findings that have been discussed in previous sections.

Notably, there are some revisions in the model that distinguish it from the protocol. The reasons for these revisions are twofold. First, we wanted to show the different predictive relations that we aim to include in the analyses. Second, we changed the model to better show the longitudinal aspect in this review.

Because the predictors are highly interrelated, examining the predictors of these three dimensions separately is somewhat problematic. For instance, some predictors may influence more than one factor related to later reading. This simple structure also works best for analyzing these important relations empirically. Examples of indicators are listed on the left side in the figure. To summarize the model, we predict that the code-related predictors (rhyme awareness, phoneme awareness, letter knowledge and RAN) have a large contribution in the early stages of learning to read and that vocabulary and grammar will have a larger contribution when children have become more-experienced readers. This model is what we aim to test. Our ability to do so depends on the extent of missing data in the primary studies.

Figure 1: Predictors of reading comprehension



Note: RAN = rapid automatized naming, WM = working memory, and VSTM = verbal short-term memory.

Definitions

To clarify the terminology that we will use throughout this review, we provide a description of the predictor terms included in the model shown in Figure 1.

Predictors of decoding:
Phonological awareness: "the ability to detect, manipulate, or analyze the auditory aspects of spoken language (including the ability to distinguish or segment words, syllables, or phonemes), independent of meaning" (NELP, 2008, p. vii).
Letter knowledge: "knowledge of the names and sounds associated with printed letters" (NELP, 2008, p. vii).
Rapid automatized naming (RAN): "the ability to rapidly name a sequence of repeating random sets of pictures of objects (e.g., 'car,' 'tree,' 'house,' 'man') or colors, letters, or digits" (NELP, 2008, p. vii).

Predictors of linguistic comprehension:
Vocabulary: the words with which one is familiar in a given language.
Grammar: knowledge about how words and their component parts are combined to form coherent sentences (i.e., morphology and syntax).

Domain-general cognitive skills:
Short-term memory: "the capacity to store material over time in situations that do not impose other competing cognitive demands" (Florit, Roch, Altoè, & Levorato, 2009, p. 936).
Working memory: "the capacity to store information while engaging in other cognitively demanding activities" (Florit et al., 2009, p. 936).
Nonverbal ability: the ability to analyze information and solve problems without using language-based reasoning.

Previous systematic reviews

Two other reviews share similarities with our review:

Similar to our review, the **National Early Literacy Panel (NELP, 2008)** review summarized longitudinal studies of reading comprehension. More specifically, the authors examined whether decoding, spelling, and reading comprehension could be predicted by a wide range of variables, including alphabetic knowledge, phonological awareness, rapid automatized naming (letters or digits and objects or colors), writing or writing one's name, phonological memory, concepts about print, print knowledge, reading readiness, oral language, visual processing, performance IQ, and arithmetic skills.

However, in contrast to our study, the NELP review did not use meta-analytic SEM to investigate these relations. In contrast, the results reported in the NELP (2008) review were based on univariate meta-analyses, which are associated with severe methodological limitations. Additionally, the NELP (2008) review reported reading comprehension outcomes from kindergarten and preschool levels, while our review examines reading comprehension during formal schooling. Furthermore, several years have passed since the NELP (2008) review was undertaken. However, to the best of our knowledge, no similar reviews have been conducted since then. The most recent reference included in the NELP (2008) review was published in 2004. However, the authors of the NELP (2008) review did not specify which studies are included in which analysis; thus, it is unclear whether this reference was part of the review's prediction of reading comprehension. Consequently, an updated review of previous research on early predictors of reading comprehension is needed.

The review by **García and Cain (2014)** assessed the relation between decoding and reading comprehension, and they restricted their review to these two measures. In other words, García and Cain's (2014) review was considerably less comprehensive than both the NELP (2008) review and the present study. Additionally, in contrast to our review, the García and Cain (2014) review studied concurrent relations between the included variables; that is, the measures used to calculate the correlations were administered at the same time point. Our review assesses the longitudinal correlational relations between the predictor variables in preschool and reading comprehension at school age after reading instruction has begun.

The systematic reviews conducted by NELP (2008) and García and Cain (2014) included published studies retrieved from searches conducted in two databases: PsycINFO and ERIC (Educational Resources Information Center). Additionally, supplementary studies located through, for instance, manual searches of relevant journals and reference checks of past literature reviews were utilized in the NELP (2008) review. The same databases and sources are used in this study; however, five additional databases are used in the electronic search. Consistent with the guidelines of a Campbell review, our review also includes a systematic search for unpublished reports (to avoid publication bias). This search is one of the strengths of this present study, as such a search was not conducted in the other two reviews, which included only studies published in refereed journals.

Although the NELP (2008) review does not state that it restricted the included samples to typical monolingual children, the García and Cain (2014) review excluded bilingual children and those who were learning English as a second language. The present review also uses this as a criterion. García and Cain (2014) stated that studies conducted with special populations were discarded if they did not include a typically developing control sample. The only exception for this criterion was if the study included participants with reading disabilities. In the NELP (2008) review, the sample criterion was children who represented the normal range of abilities and disabilities that would be common to regular classrooms. In this regard, these reviews differ from our review, which includes only typical children; we do not include children with a special group affiliation, such as children with reading disabilities.

In a recently published PhD dissertation, **Quinn (2016)** sought to examine the components of the simple view of reading via a meta-analytic structural equation modeling approach. Quinn (2016) included 155 studies conducted with English-speaking students so that difference in orthography did not influence the relations between the reading-related predictors and outcomes, i.e., reading comprehension. In addition, special population (e.g., intellectual disabilities and hearing impairment) samples with behavioral problems and Second Language Learners were excluded. Studies were grouped in two groups, a younger cohort (age <11 years) and an older cohort (age >=11 years), in the moderator analyses. Only correlations between concurrent measures were included. Neither of the additional predictors (working memory, background knowledge, and reasoning and inference making) accounted for additional variance beyond that of linguistic comprehension and decoding. One element in particular separates this review from the current review: Quinn (2016) included predictors assessed after the onset of formal reading instruction, whereas our review is limited to the predictors of abilities assessed prior to the onset of formal reading instruction.

Objectives

The primary objective for this systematic review is to summarize the available research on the correlation between reading-related preschool predictors and later reading comprehension skills.

In this review, we aim to answer the following research questions:
1) To what extent do phonological awareness, rapid naming, and letter knowledge correlate with later decoding and reading comprehension skills?
2) To what extent do linguistic comprehension skills in preschool correlate with later reading comprehension skills?
3) To what extent do domain-general skills in preschool correlate with later reading comprehension skills, and do these skills uniquely contribute to reading comprehension skills beyond decoding and linguistic comprehension?
4) To what extent do preschool predictors of reading comprehension correlate with later reading comprehension skills after concurrent decoding ability has been considered?
5) To what extent do other possible influential moderator variables (e.g., age, test types, SES, language, country) explain any observed differences between the studies included?

To answer our research questions, we have summarized available research on the topic by conducting a meta-analysis.

Methods

Criteria for considering studies for this review

Types of studies

The studies included in this review are longitudinal observational non-experimental studies that follow a group of children from preschool age into school age. In addition, business-as-usual controls in experimental studies were included (i.e., only the untreated controls, not the intervention samples). To be included, studies had to report data from at least two assessment time points: one at preschool age, before formal reading instruction has begun (predictors), and one at school age, after formal reading instruction has been implemented (outcome: reading comprehension).

However, because of the different traditions concerning the start of formal reading instruction (ranging from the ages of 3 to 6 years), we were somewhat lenient in respect to this criterion. That is, studies were included as long as the predictor assessment was conducted within 6 months of the onset of reading instruction. The minimum length of duration between the first and second assessments was set to one year, although we accepted predictor assessments conducted early in the fall semester paired with outcome assessments late in the spring semester.

Types of participants

The study population consists of samples of mainly monolingual typically developing children who were not selected for study participation because of a special group affiliation (e.g., special diagnosis or bilingualism). This inclusion criterion was chosen to avoid an overrepresentation of children with a risk of reading difficulties, which could yield biased estimates of the predictors of reading comprehension.

Types of outcome measures

The included studies reported analyses of data on (1) at least one of the predictors (vocabulary, grammar, phonological awareness, letter knowledge, RAN, memory, and nonverbal intelligence) and (2) reading comprehension as measured by standardized or researcher-designed tests.

Types of effect sizes

The primary focus of this meta-analysis is the predictive relations between different language and cognitive abilities and later reading comprehension. The studies therefore had to report a Pearson's *r* correlation between predictive measures and reading comprehension. From the studies that met this criterion, we also extracted correlations between the reported predictors and word recognition abilities. However, we did not include studies that only reported correlations between the predictors and word recognition. In addition to correlations between predictors and outcomes, we also extracted correlations between the predictors when provided.

Types of settings

Studies reported in a broad base of research literature, including journal articles, book chapters, unpublished reports, conference papers, and dissertations, were eligible for the meta-analysis. However, we did limit our search by publication year. Only studies published in the past thirty years (since 1985) were considered for inclusion. Moreover, although studies conducted in any country and language were relevant for inclusion, the studies had to be reported in English.

Search methods for the identification of studies

Electronic searches

To identify all relevant empirical studies, we established a comprehensive search strategy that was developed in collaboration with information retrieval specialist librarians at the Humanities and Social Sciences Library at the University of Oslo. Because the search settings differed somewhat between the databases, we had to use slightly different combinations of search terms for the seven databases listed below. The complete search strategy is located in online supplement 1.

The electronic search in each of the seven databases listed below was conducted on 17 January 2015. The search was also updated in February 2016.

- Google Scholar
- PsycINFO via OVID
- ERIC (Ovid)
- Web of Science
- ProQuest Dissertations and Theses
- OpenGrey.eu
- Linguistics and Language Behavior Abstracts

Searching other resources

We identified studies included in previous reviews: the NELP (2008) review and the García and Cain (2014) review. Second, a manual review of the tables of contents of key journals was conducted. The selected journals were those that had the largest number of articles cited in the studies included in this review based on an electronic search:

- Journal of Educational Psychology
- Scientific Studies of Reading (The Official Journal of the Society for the Scientific Study of Reading)
- Developmental Psychology

Data collection and analysis

Selection of studies

Studies were selected in three phases: In the first phase, the candidate studies located in their respective channels were imported to Endnote (The EndNote Team, 2013) and organized in separate folders. From there, the references were imported into the internet-based software DistillerSR (Evidence Partners, Ottawa, Canada).

In the second phase, title and abstract screening, the first (HNH) and second author (EIB) independently screened the candidate studies for relevance (see page 26 for the inter-rater reliability). A form with five questions was created in DistillerSR to determine the relevance of each reference.

1) Does the reference appear to be a longitudinal non-experimental study (or have a non-treatment control group)? Response options: Yes/No/Can't tell

2) Does the reference appear to include a study of mainly monolingual typical children (i.e., not simply included because of a special group affiliation)? Response options: Yes/No/Can't tell

3) Does the reference appear to have data from both preschool and school? Response options: Yes/No/Can't tell

4) Does the reference appear to include data on at least one of the predictors and on later reading comprehension? Response options: Yes/No/Can't tell

5) Should this reference be included at this stage? Response options: Yes/No

If any of the answers to the first four questions was "No", the reference was excluded at this stage. If the abstracts did not provide sufficient information to determine inclusion or exclusion (i.e., “can’t tell” on these questions), the reference was included in the next stage (full text screening) in order to consider information given in the full text.

In the third phase, two of the authors (HNH and EIB) retrieved the full texts and independently screened the method and results sections in each of the candidate studies to determine whether they met the inclusion criteria. In addition to the five questions above, we had sufficient information to evaluate whether the candidate studies reported bivariate correlation between the predictor(s) and outcome.

Inter-rater reliability

The first (HNH) and second author (EIB) independently double screened 25% of the references to establish coder reliability both in the second and third phases of study selection. We used the last question “Should this reference be included at this stage?” to calculate inter-rater reliability. Cohen’s κ, the inter-rater reliability for inclusion or exclusion, was satisfactory at both stages, with coefficients of .92 and .95, respectively. Any disagreements were resolved by discussing and consulting the original paper. After establishing inter-rater reliability, the two raters divided the remaining 75% of the references evenly amongst themselves for further screening.

Data extraction and management

With the exception of a few general characteristics, determining how to code information from single studies is not always clear. For this reason, we developed a coding scheme describing the data extraction procedure. The purpose of this procedure was twofold: (a) to ensure that the coding was reliable and comparable across studies and (b) to preserve the statistical independence of the data in our analysis (i.e., to not allow each study to contribute more than a single effect size for each included predictor-outcome relation).

The coding scheme was as follows. First, if data from one sample were reported in several publications, all of the publications were treated as a single study, and data were extracted across the reports. Second, if a study included more than two points of measurement, we coded the correlation between the first and last time points. Third, if a study reported several measures of a single construct (e.g., vocabulary), measurement features were considered and coded as either a receptive or an expressive measure, or a composite measure was computed. This fine-grained coding procedure was employed to enable the option of using them as separate indicators in the analysis if the amount of information allowed it. Later, composites were created by calculating the mean correlation from the receptive and expressive measures to impute a broader measure of the ability. We used Microsoft Excel to extract data on study characteristics, study quality and correlations. Two of the review’s authors (HNH and EIB) independently extracted data on 37.5% of the studies (24/64) to check for accuracy and reliability of coding. The inter-rater reliability as calculated by Pearson r correlation was r = .95. After this reliability was established, the first author (HNH) extracted data from the remaining studies.

Unit of analysis issues

In some cases, multiple observations existed for the same outcome. In such cases, we calculated a mean correlation based on these measures. This calculation was performed to gain a broad measure of the abilities that we wanted to study. Additionally, in some cases, the children were measured at more than one time point in school or in kindergarten. In those cases, we chose the first assessment time point in preschool and the last time point in school.

Details of study coding categories

Between-study variation in the relation between the predictors and reading comprehension may be related to systematic differences in study design, sample characteristics or other methodological factors. We therefore coded several moderator variables in an attempt to account for any differential effects between the studies in our meta-analysis. These moderator variables can be described within three broad categories: participant characteristics, measurement characteristics, and methodological quality.

Participant characteristics and educational setting:

Our inclusion criteria and coding scheme allowed a potentially broad age range of participants to be included in the study. Because developmental factors may influence the strength of predictive relations, age at each time point was coded as a moderator variable. In addition, we coded the number of months between the predictor and outcome assessments. To examine the impact of educational factors, we coded the amount of time (in months) that the participants had been exposed to formal reading instruction when the outcome variables were measured. We also coded the language that the children spoke and learned to read as a potential moderator of the variation in correlation size. Thus, we could determine whether the studies varied based on whether the orthography of a language was transparent or opaque.

Measurement characteristics:

We coded measurement characteristics to examine whether the predictive relations varied in strength depending on how the constructs were assessed. More specifically, we coded whether the measures were researcher created or standardized (all variables) and whether the measures were timed or untimed (outcome variables). We also coded whether reading comprehension was assessed through open – or closed-item formats (e.g., open-ended comprehension questions and free recall or multiple-choice and closed tasks, respectively). A description of the measures can be found in online supplement 2.

Methodological quality:

To further examine methodological characteristics as plausible explanations for heterogeneity in effect sizes, we coded several indicators of study quality. Tools for assessing quality in clinical trials are well developed, but much less attention has been given to similar tools for observational studies (Sanderson, Tatt, & Higgins, 2007). Thus, it was not possible to find one single quality-rating scheme that fitted our studies. We therefore developed a set of quality indicators based on some of the few rating scales that exist in medicine for observational studies. These scales include the following: the STROBE statement for improving reporting in observational studies (Vandenbroucke et al., 2014), The Newcastle-Ottawa Scale for assessing the quality of non-randomized studies (NOS, Wells et al., 2015) and a checklist for the assessment of the methodological quality of both randomized and non-randomized studies of health care interventions (Downs & Black, 1998). Based on these scales, we adapted the following quality indicators for observational studies:

- Sampling: Sampling procedure is coded when reported in the studies. The two categories are random and convenience sampling. In addition, we coded whether the sample in each study was selected or unselected. A study was coded as a selected sample if there was a set of criteria that guided the selection process (e.g., received special need education, developmental disorders or second language learners).
- Instrument quality: All included measures are coded as either a standardized or a researcher-made instrument. The studies are coded as including only standardized measures, a combination of standardized or researcher-made measures or only researcher-made measures.
- Test reliability: We have coded whether or not the test reliability of the measures used is reported in the studies.
- Floor or ceiling effect: When it was possible with the information provided in the studies, we coded whether any of the measures showed any floor or ceiling effect.
- Attrition: We calculated the percentage attrition from first assessment and last assessment. In some instances, this could not be obtained because only sample size at one time point was reported.
- Missing data: We coded what action was taken to deal with missing data. Listwise deletion represented a higher risk of bias than, for instance, other approaches commonly used in SEM – analyses (e.g., full information minimum likelihood estimation).
- Latent variables: We coded whether the studies used latent variables.
- Statistical power/sample size: Statistical power in multivariate studies depends upon many factors. However, as a general rule, sample sizes smaller than 70 will yield unstable estimates and in general have low power to detect relationships of the size that are of interest here (Little, 2013). We therefore coded sample size in three categories, below 70, 70-150 and above 150. Notably, the preferred option would be to use sample size as a continuous variable. However, this distribution deviated from normality.

Each study was given a value on the abovementioned quality indicators. The value 0 indicated a low risk of bias on that indicator, whereas a higher value reflected a higher risk of bias. Failure to report also represented a higher risk. The complete coding procedures are provided in online supplement 3. The values were combined into a total score for each study and were used in moderator analyses when applicable. Notably, some of these quality indicators (missing data, latent variables and statistical power) were not listed in the protocol of this review and, thus, this choice represents both an addition to and a deviation from the protocol.

Dealing with missing data in the review

We identified several types of missing data: 1) missing correlation matrices, 2) missing paths in correlation matrices, 3) missing sample characteristics (e.g., age in months or years), and 4) missing information pertaining to methodological quality (e.g., information regarding sampling procedure and measurement characteristics).

When a study that met our inclusion criteria failed to report an uncorrected bivariate correlation matrix (1), we contacted the corresponding authors and requested the necessary data. However, because a large number of studies did not report correlations, we contacted only the authors of studies published since 2010. We sent 11 emails requesting missing data and received 2 responses with additional statistics.

Most studies did not report data for all of the paths that we have specified in the model in Figure 1. For missing paths in the correlation matrices (2) in the meta-analytic SEM, we used the full information maximum likelihood (ML) procedure to handle missing data under the assumption that they were missing at random (Enders, 2010).

When data were missing on variables concerning sample characteristics (3) or methodological quality (4), the study with missing data was excluded from the moderator analysis.

Assessment of reporting biases

Publication bias refers to the notion that a mean effect size can be upwardly biased because only studies with large or significant effect estimates are published (i.e., the file-drawer problem with entire studies) or because authors exclude non-significant effect estimates from their study (often referred to as p-hacking, or the file-drawer problem for parts of studies; see Simmons, Nelson, & Simonsohn, 2011; Simonsohn, Nelson, & Simmons, 2014). Intervention studies are particularly vulnerable to publication bias because they often test one specific hypothesis, and a positive result is usually regarded as more interesting than null results, which are often difficult to interpret (Rothstein, Sutton, & Borenstein, 2005).

Although publication bias related to intervention studies has been repeatedly demonstrated, less is known about how publication bias affects multivariate studies that are purely observational in nature, such as those included in the present meta-analysis (see, however, Egger, Schneider, & Smith, 1998). As opposed to intervention studies, multivariate correlational studies are usually focused on patterns of relations between different variables, and they often test more than one hypothesis. Hence, a correlational study does not necessarily have a specific desired outcome; different patterns of results may be interesting for various reasons. One could thus argue that multivariate correlational studies are less prone to publication bias than intervention studies are. However, as with intervention studies, large observational studies showing clear and easily interpretable findings could be expected to be published more easily. This expectation could motivate researchers to publish only some of their data and to exclude variables that affect their analyses in particular ways or those that do not add anything to the analyses. However, the question of how observational studies are affected by publication bias is difficult to answer, since it has not been examined empirically.

Nevertheless, in line with recommendations for meta-analyses, we made special efforts to retrieve studies from the grey literature by conducting searches in databases with grey literature, and we used publication status as a moderator when possible (Higgins & Green, 2011). Additionally, to statistically estimate the impact from publication bias, researchers have commonly used funnel plots in combination with a trim-and-fill analysis. We also use this procedure in this case. However, notably, the validity of the funnel plot/trim-and-fill method is associated with several problems (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006), especially when it is used in the presence of large between-study variation (Terrin, Schmid, Lau, & Olkin, 2003). Therefore, the results from the funnel plot/trim and fill analysis must be interpreted with caution.

Statistical procedures and data synthesis

The analyses were conducted in two steps. First, the bivariate predictive relationships between the preschool predictors and later reading skills were analyzed using the Comprehensive Meta-Analysis software (CMA) Version 3 (Borenstein, Hedges, Higgins, & Rothstein, 2014). These relations were further explored by analyzing the correlation matrices from the included studies using a meta-analytic SEM approach.

Statistical approach applied for the analysis of bivariate correlations and moderators of these correlations.

The present meta-analysis includes only studies reporting correlational data. As previously noted, we therefore used Pearson's *r* as an effect size index for all outcomes. These first analyses were conducted using the software Comprehensive meta-analysis. As is typical for correlational meta-analysis, the analysis was performed using Fisher's *z* (Borenstein, Hedges, Higgins, & Rothstein, 2009), but this was calculated to be presented as Pearson's *r* in the results. To determine the strength of a relation between two variables, benchmarks are needed to compare the effect sizes. Thus, it is common to adopt Cohen's (1988) suggested standards: correlations of .50 indicate a strong relation, correlations between .30 and .49 indicate a moderate relation, and correlations below .30 indicate a weak relation.

Importantly, Cohen's general benchmarks do not necessarily apply to a given field without comparing it with prior effect sizes reported in a field. Determining to what extent an effect size is of practical significance and substantial in the current field is crucial. To examine the generalizability of the benchmarks set by Cohen (1988), Bosco, Aguinis, Singh, Field, and Pierce (2015) proposed empirical effect size benchmarks within 20 common research domains in applied psychology by extracting correlations reported in two psychology journals from 1980 to 2010. The median effect size was found to be *r* = .16, and upper and lower boundaries for medium effect size were *r* = .09 and .26. Comparing the distribution exhibited with the benchmarks set by Cohen (1988) for small, medium, and large ESs (i.e., *r* =.10, .30, .50.) indicated a correspondence with approximately the 33rd, 73rd and 90th percentiles. Although this correspondence is not directly transferable to the present study, it signifies the importance of considering the benchmarks used and of using more-context-specific benchmarks where applicable (Bosco et al., 2015; Cohen, 1988). For instance, because there are as far as we know no existing benchmarks for interpreting the size of the correlation between preschool abilities and later reading ability, we will in addition refer to a comparable field and the correlation between socioeconomic status (SES) and academic achievement. A meta-analysis by Sirin (2005) reports an average ES for this relationship of *r* =.29. The author suggests that this figure represents a strong effect in comparison with a review of more than 300 meta-analyses by Lipsey and Wilson (1993). Thus, an effect size in the present study (i.e., correlation) that would be regarded as moderate (medium) by Cohen's standards might be interpreted as strong in this context (alternatively, moderate to strong).

Moreover, in our analysis, effect sizes were averaged across studies using a random-effects model, in which correlations from independent samples were weighted by sample size. We preferred a random-effects model to a fixed-effect model because it does not assume that all studies in the meta-analysis share a common true effect size (Borenstein et al., 2009). In other words, a random-effects model takes into account that variation in effect sizes between studies may be due to both random error and systematic differences in study characteristics.

A formal test of the heterogeneity in effect sizes was conducted using the Q-statistic. The Q-statistic and its *p*-value in a random effects model is only a test of significance and reflect whether the variance is significantly different from zero. The null hypothesis is that the studies share a common effect size. Thus, a *p*-value set at 0.05 leading to a rejection of the null hypothesis suggests that the studies do not share a common effect size (Borenstein et al., 2009). A statistically significant Q *could* indicate a substantial amount of observed dispersion; however, it could also indicate a minor amount of observed dispersion with precise studies (Borenstein et al., 2009). In addition, the Q statistic is highly dependent on sample size. In addition to calculating Q, we therefore used Tau^2 to examine the magnitude of variation in effect sizes between studies (Hedges & Olkin, 1985). Notably, Tau is used to assign weights under the random effects model; thus, the total variance for a study is the sum of the within-study variance and the between-studies variance. This method for estimating the variance between studies is known as the method of moments (Borenstein et al., 2009). Given a Tau^2 of 0.01, the estimated Tau (i.e., estimated standard deviation) is 0.1. In other words, with a mean correlation of .40, 95% of the studies fall within the range .20 to .60 (.40±.2 (2 SD)), which signifies a large variation in effect size. This threshold was based on typical population SDs in applied psychology being approximately .1 to .2 (see Bosco et al., 2015), thus Tau^2=.01 and above can be considered large.

We also used the I^2-statistic to quantify the amount of true variability in the effect sizes. More specifically, the I^2-statistic indicates the proportion of variance in effects that can be attributed to true heterogeneity versus random error (0% = no systematic differences between studies: variation is primarily due to chance; 100% = no chance variation: variation is primarily due to true heterogeneity). Note that I^2 does not assess the size of the variation between studies. That is, the proportion of true heterogeneity between studies may be large even if the total between-study variation is small. Nevertheless, the presence of true heterogeneity is a prerequisite for conducting moderator analyses. We considered moderator analyses to be appropriate if the Q-statistic was significant and if I^2 was greater than 25%.

We used meta-regression to analyze continuous moderator variables. The meta-regression based on the method of moments for random-effects models was used to predict variations in effect size across studies from the moderator variables. The percentage of between-study variance explained (R^2) was used as a measure of the effect size of the moderator. The meta-regression was not conducted when there were fewer than six studies. The rule of thumb concerning the number of covariates in the meta-regression analyses is ten studies for each covariate. Although there is no clear boundary, the CMA-software notifies when the number of covariates is exceeded. We considered age and months of formal reading instruction the most relevant moderators to examine in relation to the variance shown in the studies. In addition, these variables were the most complete data from the included studies. Therefore, of the variables that were coded, these were prioritized and considered the most crucial for determining the strength of the relations between the predictors and outcome. To test the adequacy of the model, we first examined the Q_M (model) indices to determine whether one of the regression coefficients (not including the intercept) was different from zero as indicated by a significant *p*-value. A second indication of model adequacy is Q_R (residual), which shows whether there is more residual variance than would be expected if the model "fits" the data. A significant Q_R indicates that there is additional residual variation to explain that is not accounted for in the model. K represents the number of studies/effect sizes.

Statistical approach applied for the meta-analytic structural equation modelling

The correlations extracted from the studies were merged as correlation matrices and summarized in an Excel document. These matrices were then imported into R to obtain a pooled correlation matrix (R Core Team, 2013). In R, we used the statistical package metaSEM (version 0.9.14) to pool the correlation matrices and perform further SEM (Cheung, 2015). More specifically, we used correlation-based, meta-analytical structural equation modeling (c-MASEM) through a two-stage structural equation modeling approach (TSSEM, see Cheung & Chan, 2005). In the first stage, we combined the correlation matrices based on a random-effects model (Cheung, 2014). The resultant pooled correlation matrix formed the basis for the second stage, in which the hypothesized structural equation model was specified.

Importantly, the c-MASEM approach has gained considerable attention because it overcomes several f limitations associated with, for instance, univariate or generalized least squares meta-analysis (for further details, please refer to Cheung, 2015 or Jak, 2015). By allowing for variation in the correlations (i.e., random effects), this approach enables the precision of the pooled correlation matrix to be explicitly considered in the second stage of analysis—it thus improves the estimates of the relations among constructs and helps to avoid otherwise conflicting research results (Cheung, 2015). Compared with multivariate meta-analysis or meta-analytic SEM approaches, in which correlation matrices are aggregated by simply aggregating all correlations across studies individually, the c-MASEM approach results in accurate parameter estimates and standard errors. This accuracy is achieved by accounting for the (correct) sample sizes when aggregating the correlation matrices. The alternative approaches are oftentimes based on arithmetic or harmonic means of sample sizes across all studies, thus under- or overestimating parameters and standard errors.

Results

This section consists of three parts: first, we present the flow chart for the process of selecting studies for inclusion and exclusion, as well as a description of the final sample of studies that were included in the present meta-analysis. Second, we present a series of analyses of the bivariate relationships shown in the theoretical model (Figure 1). The results of a set of corresponding moderator analyses will also be presented. Finally, we present the empirical models based on meta-analytic SEM.

Note that we will present both the bivariate analyses and the structural equation models because a number of studies could not be included in the structural equation models because of missing data on different paths. However, these studies could add information to the bivariate analyses.

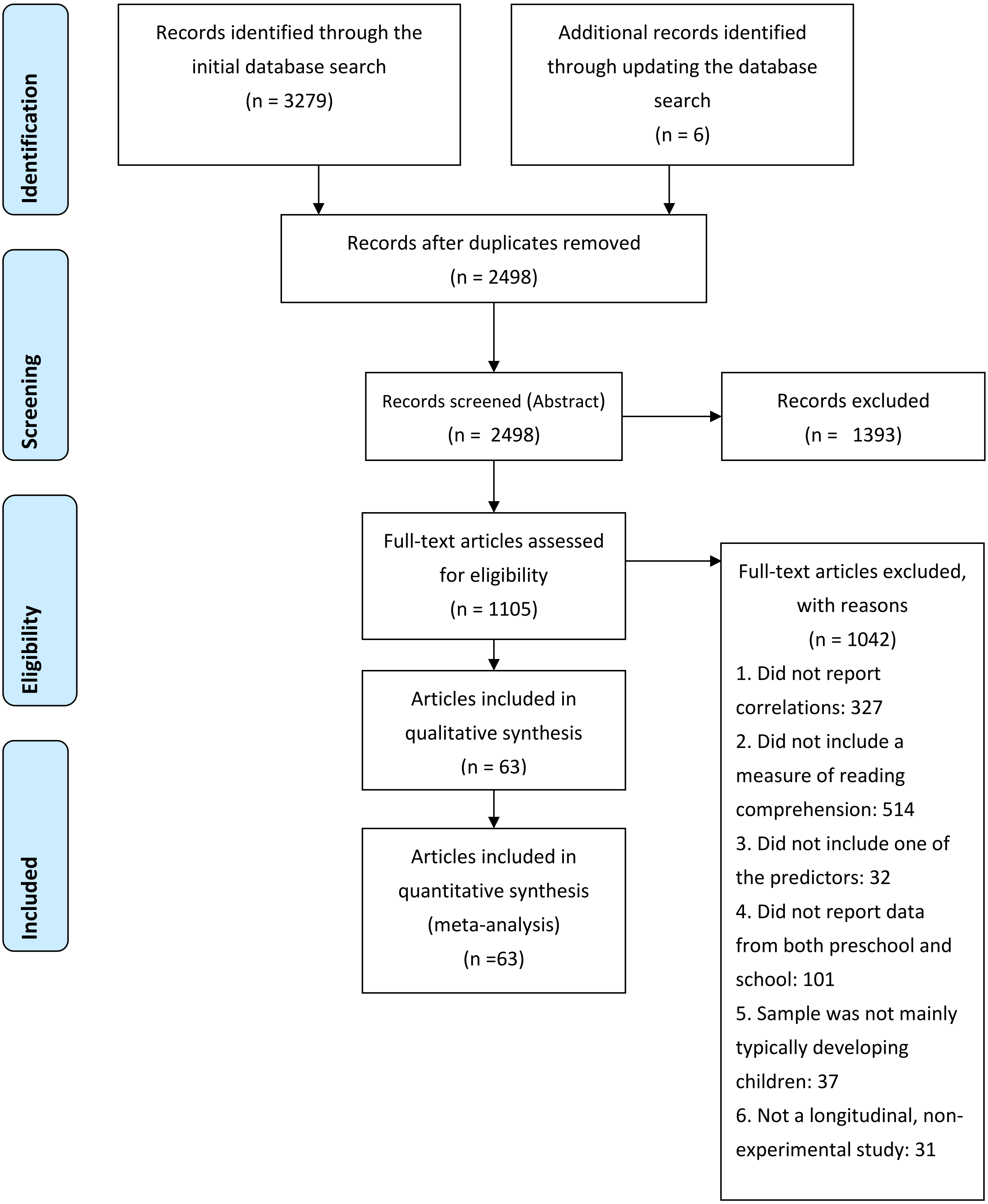
Description of studies

Results of the search

The electronic search conducted on 17 January 2015 yielded 3,279 references from seven different databases (the databases are listed on p. 24). In addition, the search was updated in February 2016, resulting in six additional studies. Duplicate studies (i.e., the same reference located in different databases) were placed under quarantine with the duplication detector application embedded in the DistillerSR software. After we removed duplicates, the number of references decreased to 2,498. By screening the abstracts of the 2,498 references, we further excluded 1,393, leaving 1105 full articles to be read and evaluated for inclusion. In the end, 64 studies (with 63 articles with 64 independent samples) met the eligibility criteria and were included in the analyses. For ease of reading, we further refer to this set as the 64 included studies. Figure 2 is a flow chart illustrating the selection of studies for inclusion.

In addition to the electronic search, we also conducted a manual search by crosschecking references from previously published reviews and meta-analyses (García & Cain, 2014; NELP, 2008) and reviewing the tables of contents of key journals (the Journal of Educational Psychology, Developmental Psychology, Scientific Studies of Reading). This procedure did not reveal any additional eligible studies that we had not already located through the database search.

Figure 2: Flow chart of the inclusion and exclusion of studies



Included studies

A table with all of the included studies and the main study characteristics is provided in online supplement 4. A summary overview with some key factors follows below.

Location

The United States was the country of origin for 24 studies. Eight studies were conducted in Canada. Six studies were conducted in Israel. In five studies, the country of origin was Finland. Four studies were conducted in England. Three studies were conducted in Australia. Two studies were conducted in France and two in Germany. One study each was conducted in the Netherlands, Croatia, Spain (Canary Islands), Brazil, New Zealand, Austria and Norway. In addition, one study was conducted in both the USA and Australia. Furthermore, in one study, the location was both the USA and Canada. Finally, one study was conducted in both Norway and Sweden.

Sample sizes

The number of participants in the sample ranged from 16-9165. Of the 64 included studies, 16 studies had more than 150 participants, 28 studies had from 70-150 participants, and 20 studies were conducted with fewer than 70 participants.

Measures

Forty-five studies included a measure of vocabulary. Thirty-six studies included a measure of phoneme awareness. Twenty-six studies included a measure of letter knowledge. Twenty-one studies included non-verbal intelligence. Seventeen studies included RAN as a measure. Sixteen studies included a measure of grammar. Fifteen studies reported on rhyme awareness. Nine studies included a measure of sentence memory, whereas seven studies reported on non-word repetition.

Excluded studies

Most of the studies that came close to inclusion were eliminated because relevant statistics were not reported. Other important reasons for exclusion were the lack of any predictor measure or reading comprehension measure, the lack of longitudinal design, and sample characteristics that did not fit the eligibility criteria. For an overview of a number of references excluded for different reasons, please see the Flow chart (Figure 2). Note that some studies may be excluded for several reasons; for instance, they only reported a broad reading (e.g., reading accuracy and comprehension) score and did not report bivariate correlations.

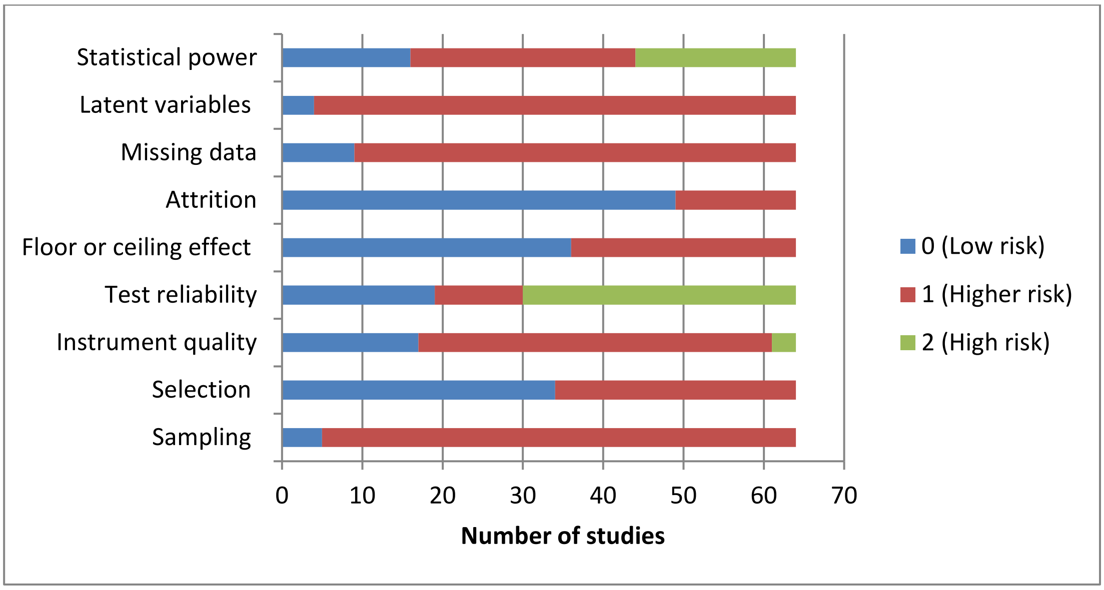
Risk of bias in the included studies

Risk of bias issues

- Sampling: Of the 64 included studies, 5 used random sampling, whereas 59 used convenience sampling. Moreover, 34 of the studies included an unselected sample, whereas 30 studies had a selected sample. Importantly, the coding of a sampling procedure (convenience/random or selected/unselected) contains great uncertainty due to a lack of sufficient reporting. In particular, information about the sample (e.g., if there was a set of selection criteria in place) was often not reported. Thus, in those instances, we were unsure whether this omission meant that the sampling approach was in fact unselected or that the authors failed to report the approach.
- Instrument quality: All included measures were coded as either a standardized or a researcher-made instrument. Ordinarily, a mixture of both types of measures are used in the studies (N = 44). Of the 64 studies, 17 of them only used standardized measures. In three studies, only researcher-made instruments were used.
- Test reliability: Most often, test reliability was not reported, the authors only reported reliability from the test manual (N = 34), or it was reported on some of the measures (N =11) or on all measures (N = 19).
- Floor or ceiling effect: 28 studies included one or more measures for which either floor or ceiling effect, or the necessary statistics (M and SD) or number of items (maximum score) were not reported. In the remaining 36, we could detect neither floor nor ceiling effect.
- Attrition: We calculated the percentage attrition from first assessment and last assessment. In some instances (N=15), this measure was not possible to obtain because sample size at only one time point was reported. Often, the longer the study, the more attrition that can be expected due to normal mobility (for example, moving to other school districts or changing teachers). The highest percentage of attrition in the included studies was 59%, and that study spanned ten years.
- Missing data analysis: In a number of studies, only samples that had completed all of the measurement time points were included in the analyses. Although this point was not usually mentioned in the article, we assumed listwise deletion was used. Nine of the studies used a technique to handle missing data (e.g., full information maximum likelihood estimation). The remaining studies used listwise deletion or did not report on this aspect.
- Latent variables: Four of the included studies used latent variables in the analyses.
- Statistical power: Sixteen studies had more than 150 participants, 28 studies had from 70-150 participants, and 20 studies were conducted with fewer than 70 participants.

The overall study quality is summarized in figure 3. As the figure shows, there is a risk of bias in several of the included studies on the risk of bias issues outlined above. Notably, some of the indicators have two values (0 and 1), whereas others have three available values (0, 1 and 2) (please see table 3 in online supplement). The individual scores for each of the included studies are provided in online supplement 5.

Figure 3: study quality in the included studies



Analyses on study quality

As previously mentioned, we calculated a total score based on the indicators. These were used as moderators in separate meta-regressions for the relations that were examined with reading comprehension as the outcome. Notably, this process was not conducted with the two analyses on verbal short-term memory because of a low number of studies (sentence memory) and non-word repetition (low degree of variance between the studies). The mean total score was 6.9 (SD = 1.8). The range was 1-10, with 1 representing the lowest risk of bias and 10 the highest risk. The highest obtainable score was 14. Analyses using study quality as moderator showed that the quality score was not significantly related to the effect size. The result from the meta-regressions is provided in online supplement 6.

Synthesis of results for bivariate relations and moderators

The results of the meta-analysis are organized in sections aligned with the research questions that were presented in chapter 2 (Objectives, p. 22):

1) We present the summarized correlations between the preschool predictors and reading comprehension and the moderators of these relations.
2) We report the summarized correlations of the code-related predictors and later word recognition abilities and the moderators of these relations. Word recognition ability is coded concurrently with the assessment of reading comprehension.
3) We present the results of the meta-analytic SEM.

A summary of correlations between the predictors and outcomes is presented in table 1.

We also present results for core moderators that we would expect to have the largest impact in the predicting of reading comprehension (age, onset of formal reading instruction and time between the two assessments). A table showing the meta-regression results is provided in online supplement 7.

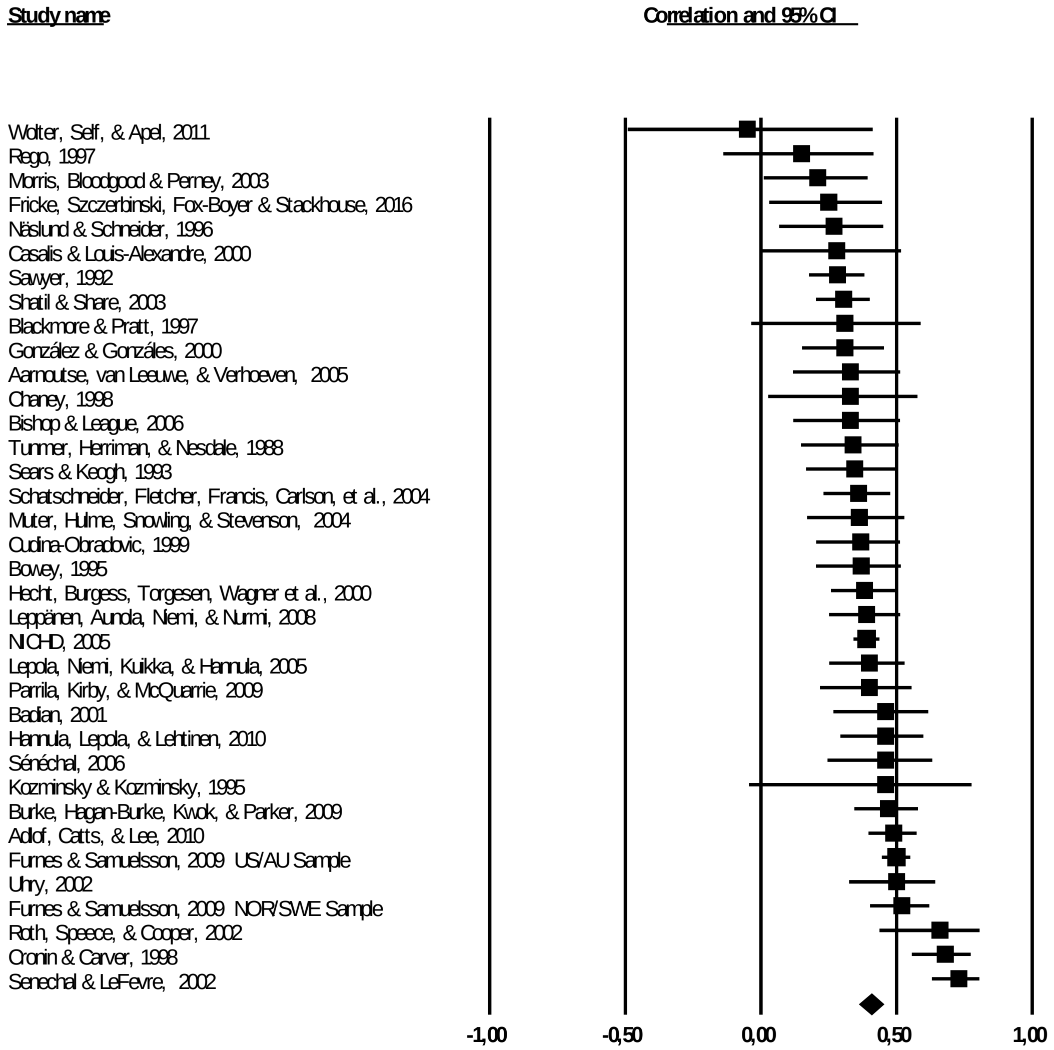
Table 1: Summary table of correlations between predictors and outcomes

Outcome	Predictor	Average correlation	Number of studies k
Reading comprehension	Phoneme awareness	*r* =.40	36
	Rhyme awareness	*r* =.39	15
	Letter knowledge	*r* =.42	26
	Rapid naming	*r* = -.34	17
	Vocabulary	*r* =.42	45
	Grammar	*r* =.41	16
	Sentence memory	*r* =.36	9
	Non-word repetition	*r* =.17	7
	Non-verbal intelligence	*r* =.35	21
Word recognition	Phoneme awareness	*r* =.37	28
	Rhyme awareness	*r* =.32	13
	Letter knowledge	*r* =.38	16
	Rapid naming	*r* =-.37	14

The longitudinal correlation between phoneme awareness and reading comprehension

Thirty-six studies reported a bivariate correlation between measures of phoneme awareness and reading comprehension. The total number of participants across these studies was 6,626. The participants' mean age was 5.5 years (*SD* = 0.7) at the time of the initial assessment and 8.4 years (*SD* = 1.7) when reading comprehension was measured. Analysis 1 shows the overall mean correlation between phoneme awareness and reading comprehension. Analysis 1 also shows the correlation coded from each study, with a 95% confidence interval (CI). As is apparent from Analysis 1, the mean correlation between preschool phoneme awareness and later reading comprehension is moderate to strong (r = .40; CI [.36, .44]) and statistically significant ($z[35] = 17.05$; $p < .001$). The correlation coefficients among the studies varied from r = -.05 to .73. This variation was significant ($Q[35] = 99.07$; $p < .001$) and represented a substantial proportion of true heterogeneity (I^2 = 64.7%). The total amount of variation between studies, as indicated by Tau^2 = 0.01, is large. The estimated Tau (i.e., estimated standard deviation) is 0.1. In other words, with a mean correlation of .40, 95% of the studies fall within the range .20 to .60 (.40±.2 (2 SD)), which signifies a large variation in effect size.

Analysis 1: Forest plot of the correlation between phoneme awareness and reading comprehension

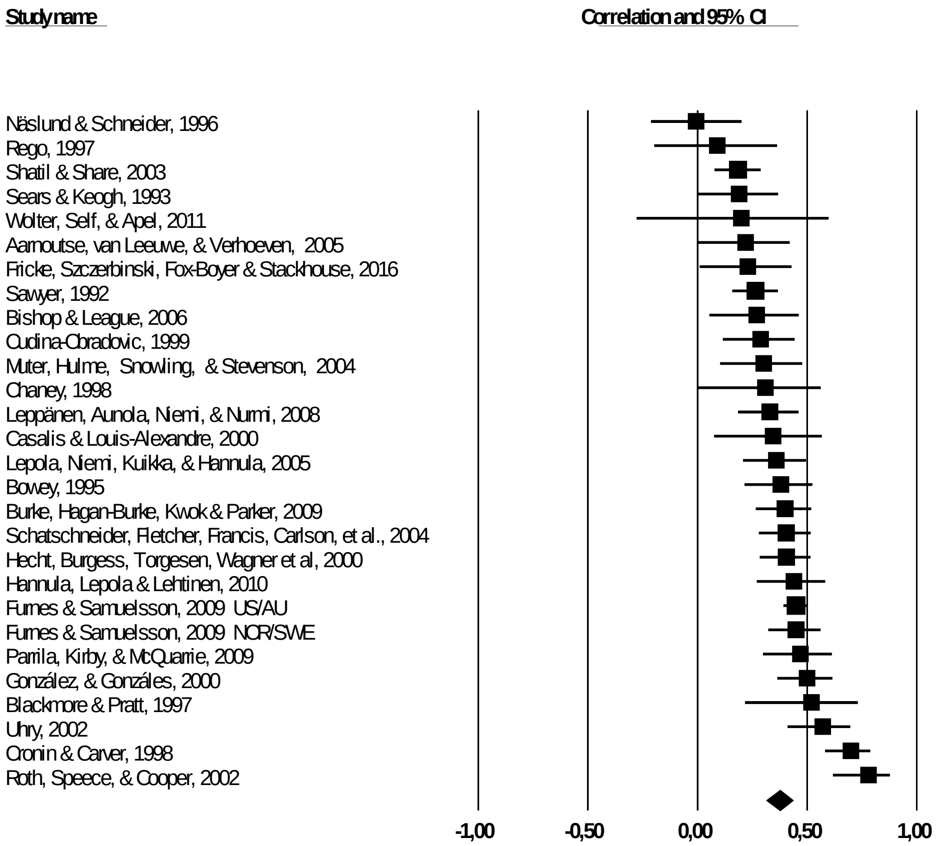


We conducted a meta-regression to explain the between-study variation. This method allowed us to determine whether age at initial assessment, age at reading comprehension assessment and months of reading instruction at the time of the reading comprehension assessment could predict the variation in correlation size between studies. However, the meta-regression was not significant ($Q[3] = 3.44$; $p = .329$), and neither age nor months of reading instruction could explain the variance in effect sizes between the studies ($R^2 = .00$).

The longitudinal correlation between phoneme awareness and word recognition

Among the 36 studies that reported a correlation between phoneme awareness and reading comprehension, 28 studies also included a measure of word recognition. The total number of participants across these studies was 4,772. The participants' mean age was 5.5 years (SD = 0.6) at the time of the initial assessment, and 8.0 years (SD = 1.2) when word recognition was last measured. The overall mean correlation between phoneme awareness and word recognition is presented in Analysis 2 with a 95% CI. As shown in the figure, the mean correlation between preschool phoneme awareness and later word recognition is moderate to strong (r = .37; CI [.31, .43]) and significant (z[27] = 11.43; p < .001). The correlation coefficients among the studies varied from r = -.01 to .78. This variation was significant (Q[27] = 103.27, p < .001) and represented a substantial proportion of true heterogeneity (I^2 = 73.9%), and the total amount of variation between studies is large (Tau^2 = 0.02).

Analysis 2: Forest plot of the correlation between phoneme awareness and word recognition

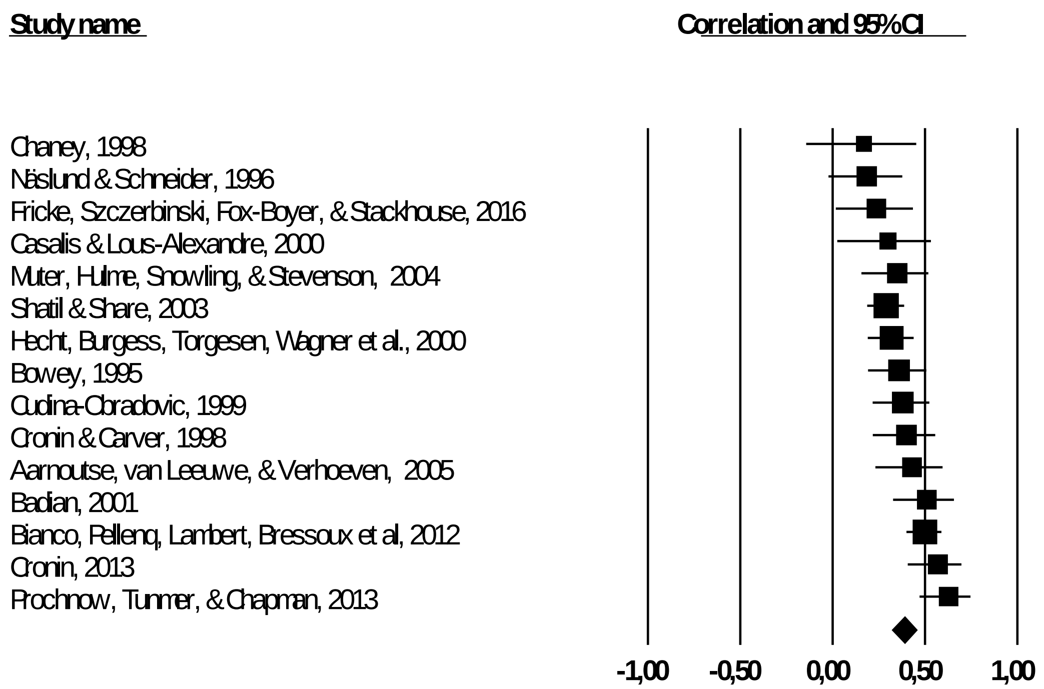


We conducted a meta-regression analysis to explain the between-study variation. An analysis of a model with age at initial assessment, age at reading comprehension assessment and months of reading instruction at the time of the reading comprehension assessment as predictors was not significant (Q[3] = 6.30, p = .098). Age at the last assessment and months of reading instruction at the time of this assessment were significantly associated with word recognition ability. The model explained 6.29% of the total variance in effect sizes between the studies.

The longitudinal correlation between rhyme awareness and reading comprehension

A total of 15 studies reported a bivariate correlation between rhyme awareness and reading comprehension. The total number of participants across these studies was 1,741. The participants' mean age was 5.3 years (*SD* = 0.8) at the time of the first rhyme awareness assessment and 8.3 years (*SD* = 1.8) when reading comprehension was measured. Analysis 3 shows the overall mean correlation between rhyme awareness and reading comprehension. Analysis 3 also shows the correlation coded from each study with a 95% CI. As is apparent in Analysis 3, the mean correlation between preschool rhyme awareness and later reading comprehension is moderate to strong (r = .39; CI [.32, .45]) and statistically significant ($z[14] = 10.40$; $p < .001$). The correlation coefficients among the studies varied from r = .17 to .63. This variation was significant ($Q(14]$ = 33,22 p = .003) and represented a substantial proportion of true heterogeneity (I^2 = 57.9%). The total amount of variation between studies was large (Tau^2 = 0.01).

Analysis 3: Forest plot of the correlation between rhyme awareness and reading comprehension

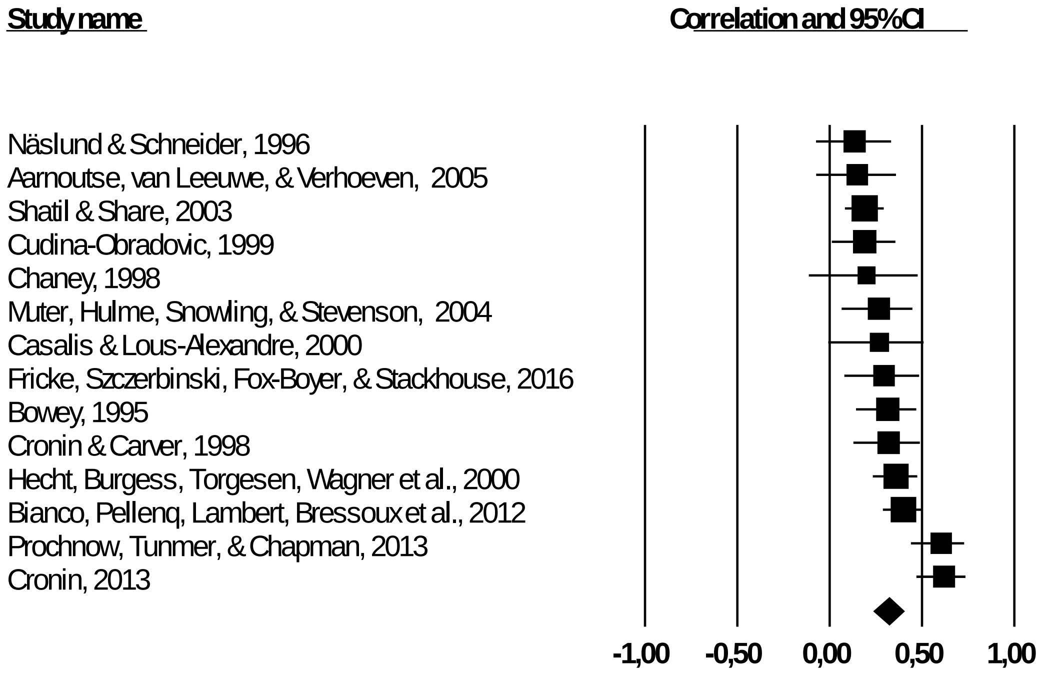


We conducted a meta-regression analysis to explain the between-study variation. By doing so, we could observe whether age at initial assessment and age at reading comprehension assessment could predict the variation in correlation size between studies. The meta-regression analysis did yield a significant result ($Q[2]$ = 7.53, p = .023). A further examination of the unique contribution of each covariate showed that age at reading assessment could predict the heterogeneity in effect sizes (p = .020), whereas age at rhyme assessment could not (p = .310). In other words, when the other covariates were controlled, a higher effect size is associated with higher age at the last assessment.

The longitudinal correlation between rhyme awareness and word recognition

From the 14 studies that reported a correlation between rhyme awareness and reading comprehension, 13 studies also included a measure of word recognition. The total number of participants across these studies was 1,662. The participants' mean age was 5.4 years (*SD* = 0.8) at the time of the initial assessment and 8.0 years (*SD* = 1.3) when word recognition was last measured. Analysis 4 shows the overall mean correlation between rhyme awareness and word recognition. Analysis 4 also shows the correlation coded from each study with a 95% CI. As shown in the figure, the mean correlation between preschool rhyme awareness and later word recognition is moderate to strong (r = .32; CI [.24, .40]) and significant (z[13] = 7.24; p < .001). The correlation coefficients among the studies varied from r = .14 to .62. This variation was significant (Q[13] = 39.75, p < .001) and represented a substantial proportion of true heterogeneity (I^2 = 67.3%). The total amount of variation between studies was large (Tau^2 = 0.02).

Analysis 4: Forest plot of the correlation between rhyme awareness and word recognition

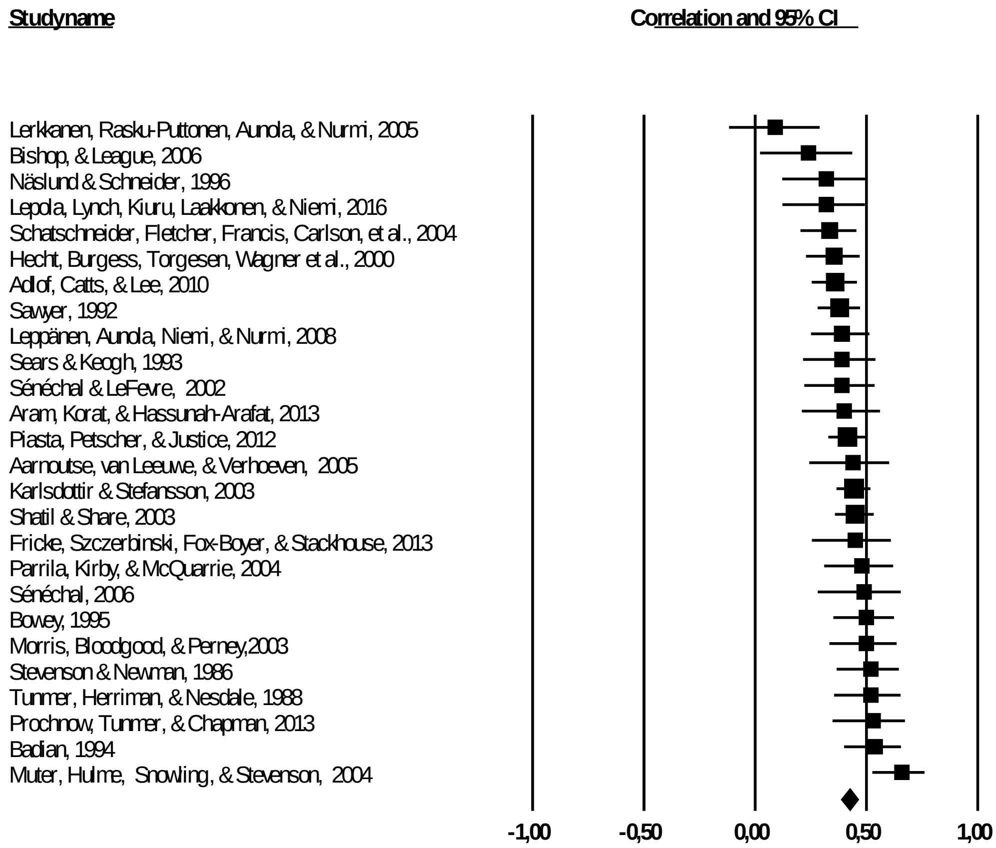


In the next step, we sought to test whether age at the two assessments could explain the between-study variation. The meta-regression analysis including age at rhyme assessment and age at word recognition assessment was significant (Q[2] = 18.53; p< .001), indicating that the effect size is related to at least one of the covariates. A further examination of the unique contribution of each covariate showed that age at reading assessment could predict the heterogeneity in effect sizes (p < .001), whereas age at rhyme assessment could not (p = .062). In other words, when the other covariates were controlled, a greater effect size was associated with older age at the last assessment. The model with the included covariates together explained 83.7% of the total variance in effect sizes between the studies.

The longitudinal correlation between letter knowledge and reading comprehension

A total of 26 studies reported a bivariate correlation between measures of letter knowledge and reading comprehension. The total number of participants across these studies was 3,869. The participants' mean age was 5.6 years (*SD* = 0.7) at the time of the first letter knowledge assessment and 9.0 years (*SD* = 2.2) when reading comprehension was last measured. Analysis 5 shows the overall mean correlation between letter knowledge and reading comprehension. Analysis 5 also shows the correlation coded from each study with a 95% CI. As shown in Analysis 5, the mean correlation between preschool letter knowledge and later reading comprehension is moderate to strong (r = .42; CI [.38, .46]) and statistically significant (z[25] = 19.87; p < .001). The correlation coefficients among the studies varied from r = -.13 to .67. This variation was significant (Q[25] = 44.00, p = .011) and represented a substantial proportion of true heterogeneity (I^2 = 43.2%), although the total amount of variation between studies was large (Tau² = 0.01).

Analysis 5: Forest plot of the correlation between letter knowledge and reading comprehension

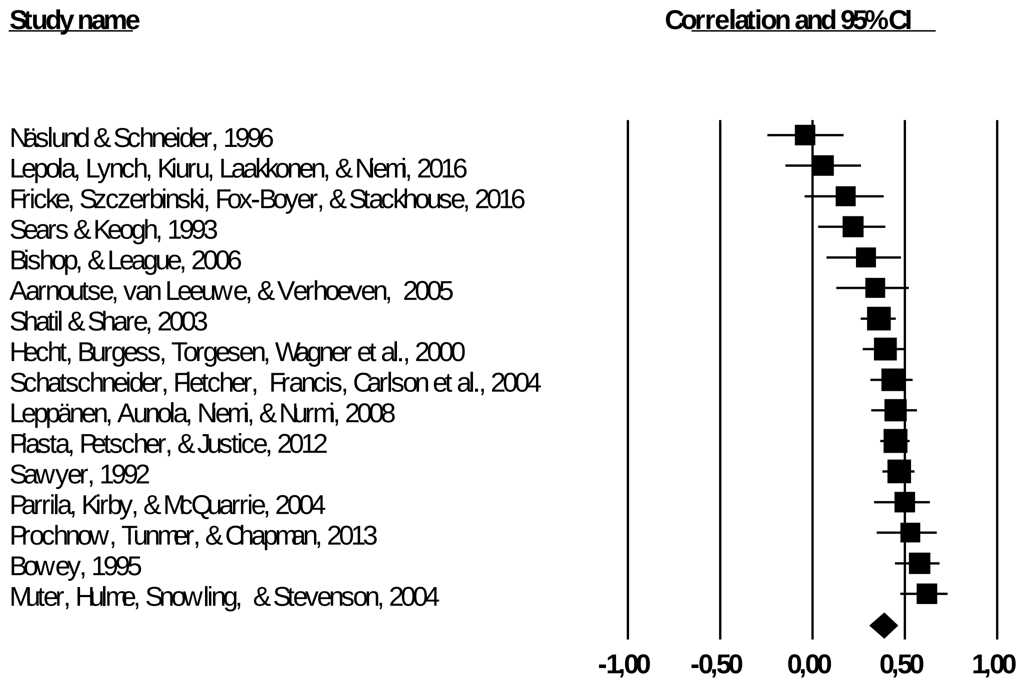


We conducted a meta-regression analysis to explain the between-study variation. By doing so, we could determine whether age at initial assessment, age at reading comprehension assessment and months of reading instruction at reading comprehension assessment could predict the variation in correlation size between studies. However, the meta-regression analysis was not significant ($Q[3] = 3.31$; $p = .346$), and neither age nor months of reading instruction could explain the variance in effect sizes between the studies ($R^2 = .00$).

The longitudinal correlation between letter knowledge and word recognition

From the 26 included studies that reported a correlation between letter knowledge and reading comprehension, 16 studies also included a measure of word recognition. The total number of participants across these studies was 2,423. The participants' mean age was 5.4 years ($SD = 0.6$) at the time of the initial assessment and 8.5 years ($SD = 1.6$) when word recognition was last measured. Analysis 6 shows the overall mean correlation between letter knowledge and word recognition. In Analysis 6, the correlation coded from each study is shown with a 95% CI. As is apparent in Analysis 6 the mean correlation between preschool letter knowledge and later word recognition is moderate to strong ($r = .38$; CI [.31, .45]) and significant ($z[15] = 9.22$; $p < .001$). The correlation coefficients among the studies varied from $r = -.04$ to .62. This variation was significant ($Q[15] = 62.97$, $p < .001$) and represented a substantial proportion of true heterogeneity ($I^2 = 76.2\%$), although the total amount of variation between studies was large ($Tau^2 = 0.02$).

Analysis 6: Forest plot of the correlation between letter knowledge and word recognition

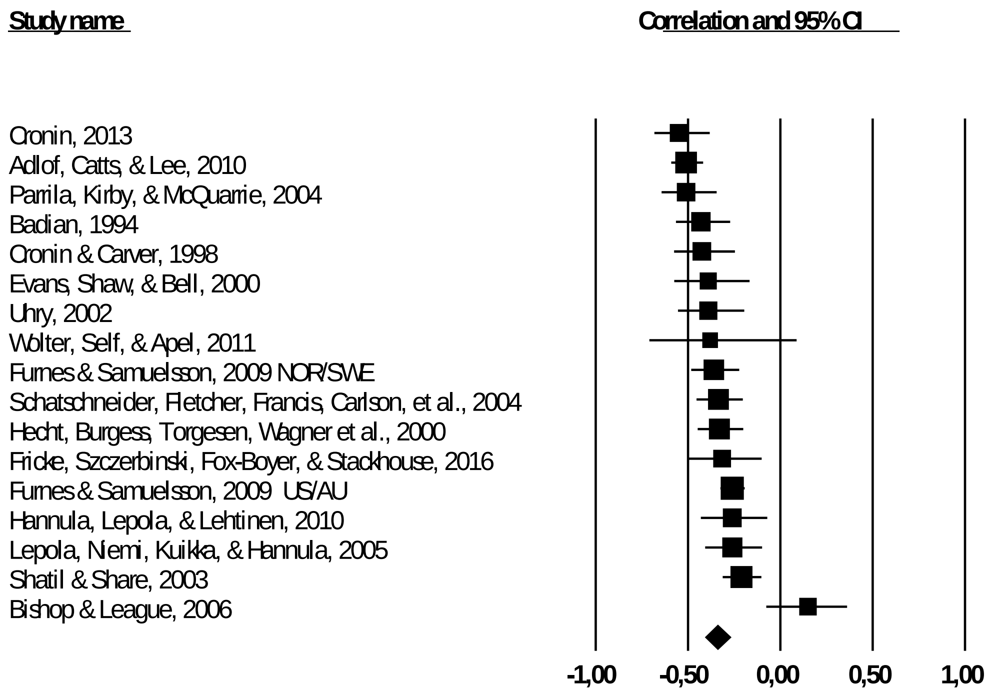


In the next step, we conducted a meta-regression analysis to explain the between-study variation. However, a meta-regression analysis including age at letter knowledge assessment and age at word recognition assessment was not significant ($Q[2] = 1.12$, $p = .560$), an indication that neither age at the initial assessment nor age at last assessment could explain the variance in effect sizes between the studies. The model explained 0% of the variance (R^2 = .00).

The longitudinal correlation between RAN and reading comprehension

Seventeen studies reported a bivariate correlation between measures of RAN and reading comprehension. The total number of participants across these studies was 3,746. The participants' mean age was 5.6 years (*SD* = 0.4) at the time of the first assessment and 8.4 years (*SD* = 1.8) when reading comprehension was measured. Analysis 7 shows the overall mean correlation between rapid naming and reading comprehension. Additionally, Analysis 7 shows the correlation coded from each study with a 95% CI. As shown in the figure, the mean correlation between preschool RAN and later reading comprehension is moderate to strong (r = -.34; CI [-.41, -.28]) and statistically significant (z[16] = -9.28; p< .000). Moreover, the predictive relation is negative because the faster one is (the less time one uses) at naming objects, colors, letters or digits, the better one is at reading (the higher score one receives). The correlation coefficients among the studies varied from r = -.55 to .15. This variation was significant (Q[16] = 56.18, p< .001) and represented a substantial proportion of true heterogeneity (I^2 = 71.5%), which was also indicated by the fact that the total amount of variation between studies was large (Tau^2 = 0.02).

Analysis 7: Forest plot of the correlation between RAN and reading comprehension

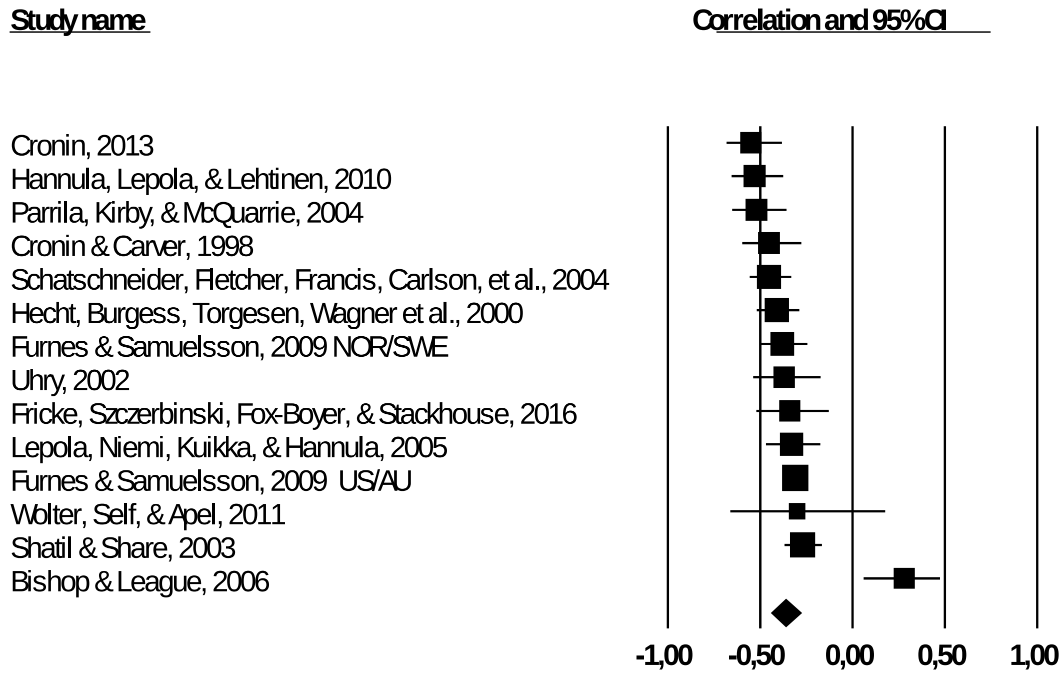


We conducted a meta-regression analysis to explain the between-study variance. However, a model with age at the two assessment time points did not yield a significant result (Q[2] = 4.74; p = .094). The model with the two covariates explained 14.75% of the variance.

The longitudinal correlation between RAN and word recognition

From the 17 studies that reported a correlation between RAN and reading comprehension, 14 studies also included a measure of word recognition. The total number of participants across these studies was 3,285. The participants' mean age was 5.6 years (*SD* = 0.4) at the time of the initial assessment and 8.2 years (*SD* = 1.1) when word recognition was last measured. Analysis 8 shows the overall mean correlation between RAN and word recognition. Analysis 8 also shows the correlation coded from each study with a 95% CI. As shown in Analysis 8, the mean correlation between preschool RAN and later word recognition is moderate to strong (r = -.37; CI [-.44, -.28]) and significant (z[13] = -8.22; p < .001). The correlation coefficients among the studies varied from r = -.55 to .28. This variation was significant (Q[13] = 54.77, p < .001) and represented a substantial proportion of true heterogeneity (I^2 = 76.3%). The total amount of variation between studies was also large (Tau^2 = 0.02).

Analysis 8: Forest plot of the correlation between RAN and word recognition

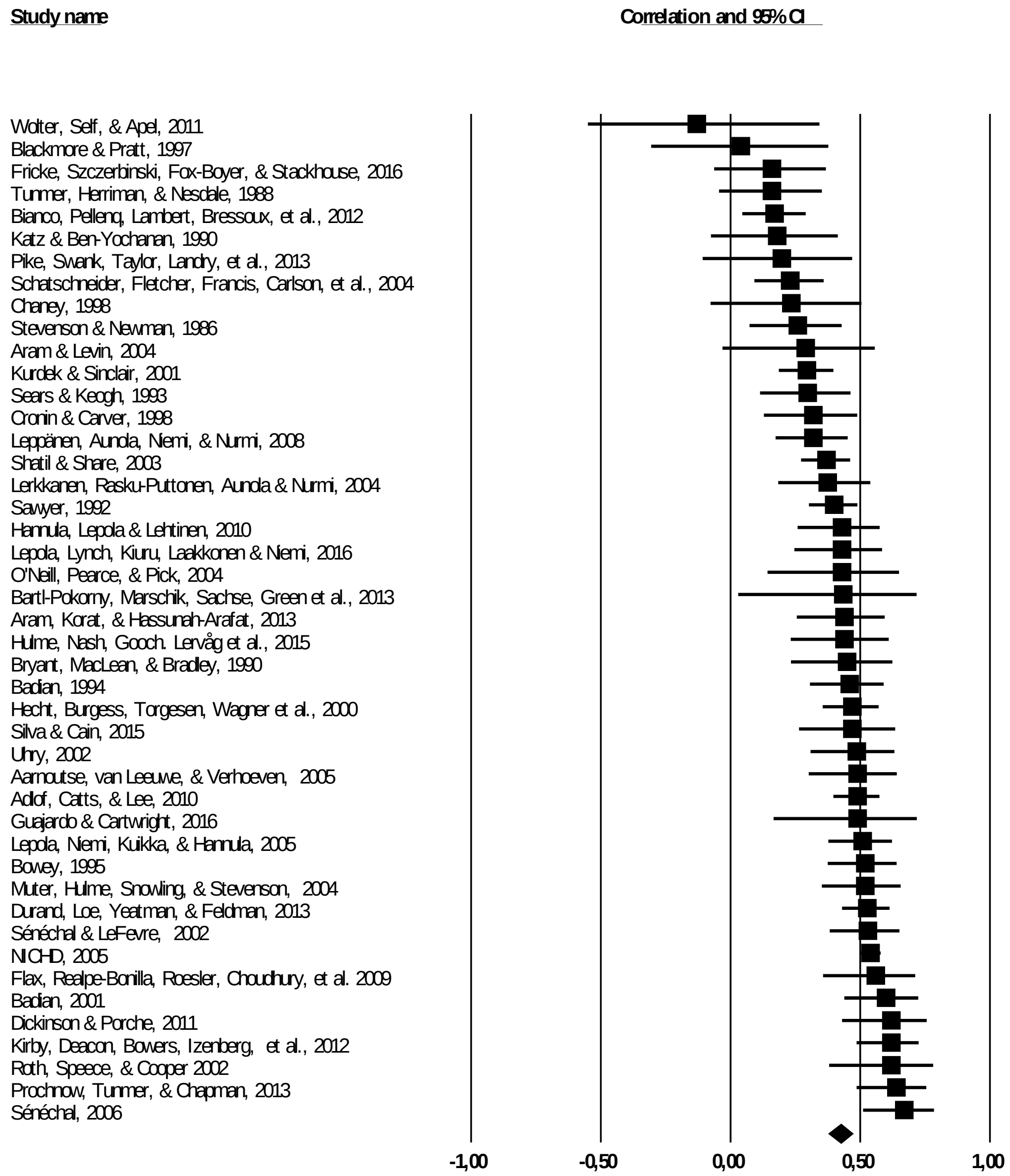


In the next step, we conducted a meta-regression analysis to explain the between-study variation. However, an analysis including age at RAN assessment and age at word recognition assessment as predictors was not significant (Q[2] = 2.09, p = .351). Moreover, the model with the two covariates explained 0% of the variance (R^2 = .00). To test whether other covariates could predict the variation in correlation size between studies, we conducted an analysis with a second model. However, a meta-regression with the number of months between the two assessments and the number of months with formal reading instruction at the reading assessment could not predict the variation in correlation size between studies (Q[2] = 2.20, p = .333).

The longitudinal correlation between vocabulary and reading comprehension

In the 45 studies reporting a bivariate correlation between measures of vocabulary and reading comprehension, the total number of participants was 5,907. The participants' mean age was 5.2 years (*SD* = 1.1) at the time of the vocabulary assessment and 9.0 years (*SD* = 2.3) when reading comprehension was measured. Analysis 9 shows the overall mean correlation between vocabulary and reading comprehension. Additionally, Analysis 9 shows the correlation coded from each study with a 95% CI. As is apparent in Analysis 9, the mean correlation between preschool vocabulary and later reading comprehension is strong (moderate) ($r = .42$; CI [.38, .46]) and statistically significant ($z[44] = 16.76$; $p < .001$). The correlation coefficients among the studies varied from $r = -.13$ to .67. This variation was significant ($Q[44] = 153.13$, $p < .001$) and represented a substantial proportion of true heterogeneity ($I^2 = 71.3\%$). In addition, the total amount of variation between studies was large ($Tau^2 = 0.02$).

Analysis 9: Forest plot of the correlation between vocabulary and reading comprehension

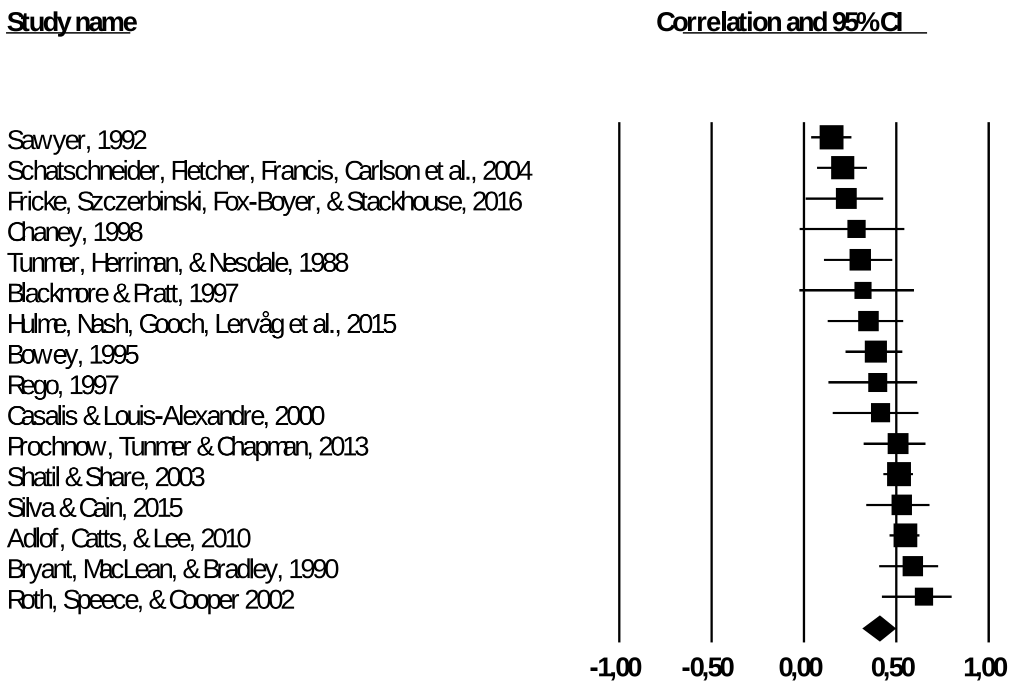


We conducted a meta-regression analysis to explain the between-study variation. Because this was the analysis with the greatest number of included studies, we could test whether age at vocabulary assessment, age at reading comprehension assessment, months of reading instruction at reading comprehension assessment and the type of reading comprehension assessment (open-ended/retelling vs. multiple-choice or cloze tasks) could predict the variation in correlation size between the studies. However, the meta-regression was not significant ($Q[4] = 4.53$, $p = .339$). The model with the abovementioned covariates together explained 11% of the between study variance ($R^2 = .11$).

The longitudinal correlation between grammar and reading comprehension

The 16 studies reporting a bivariate correlation between measures of grammar and reading comprehension totaled 1,857 participants. The mean age was 5.2 years (*SD* = 0.9) at initial grammar assessment and 8.1 years (*SD* =2.0) when reading comprehension was last assessed. Analysis 10 shows the overall mean correlation between early grammar and later reading comprehension. Analysis 10 also shows the correlation coded from each study with a 95% CI. As is apparent in Analysis 10, the mean correlation between preschool grammar and later reading comprehension is strong (moderate) (r = .41; CI [.32, 49]) and significant (z[15]= 8,26, p < .001). The correlation coefficients among the studies varied from r = .15 to .65. The differences between the studies in the magnitude of correlation was significant (Q[15] = 63.87, p < .001) and represented a substantial proportion of true heterogeneity (I^2 = 76.5%). In addition, the total amount of variation between studies was very large (Tau^2 = 0.03).

Analysis 10: Forest plot of the correlation between grammar and reading comprehension



We conducted a meta-regression to attempt to explain the between-study variation. Because the number of included studies was 16, we included age at initial assessment and age at last assessment as the two covariates. However, the meta-regression was not significant (Q[2] = 0,36; p =.837), indicating that neither age at grammar assessment nor age at reading comprehension assessment had a significant effect on predicting the variation in correlation size between studies. The model with the included covariates together explained 0% of the variance (R^2 = .00).

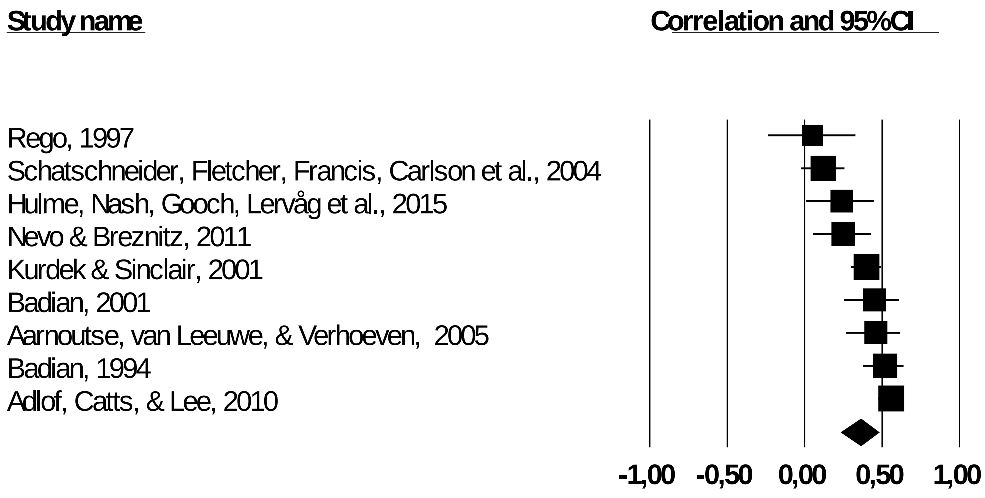
The longitudinal correlation between verbal short-term memory and reading comprehension

The most frequent measures of verbal short-term memory were sentence memory and non-word repetition. These measures were explored in separate analyses. Notably, an insufficient number of studies reported a correlation between preschool working memory and later reading comprehension.

Sentence repetition

Nine studies reported a bivariate correlation between measures of sentence repetition and reading comprehension. These studies had 1,237 participants. The mean age at initial assessment was 5.3 years (SD = 0.7), and the mean age at reading comprehension assessment was 9.1 years (SD = 2.9). Analysis 11 shows the overall mean correlation between sentence repetition and reading comprehension. In Analysis 11, we can observe the correlation coded from each study with a 95% CI. As shown in the forest plot, the overall mean correlation between preschool sentence repetition and later reading comprehension is moderate to strong (r = .36; CI [.23, .47]) and significant (z[8] = 5.35, p < .001). The correlation coefficients among the studies varied from r = .05 to .56. This variation was significant (Q[8] = 43.39, p < .001) and represented a substantial proportion of true heterogeneity (I^2 = 81.5%), although the total amount of variation between studies was very large (Tau^2 = 0.04).

Analysis 11: Forest plot of the correlation between sentence memory and reading comprehension

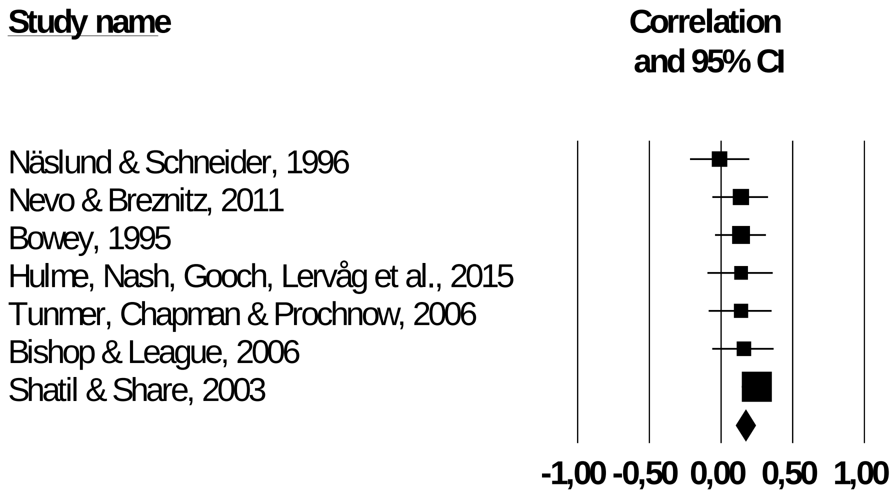


Because of the low number of included studies, we were able to include only one covariate in the analysis. A meta-regression with age at reading comprehension assessment generated a significant result (Q[1] = 4.14, p = .042). That is, the older the children were at the last follow-up, the higher correlations those studies tended to report (R^2 = .46).

Non-word repetition

Seven studies reported a bivariate correlation between measures of non-word repetition and reading comprehension. However, one of these studies (Bishop & League, 2006) included a composite measure of digit span and non-word repetition. The total number of participants across these studies was 841. The mean age at initial assessment was 5.2 years (*SD* = 0.9), and the age at reading comprehension assessment was 8.3 years (*SD* = 1.7). The overall mean correlation is presented in Analysis 12. As shown in the forest plot, the overall summarized correlation is weak to moderate (r = .17; CI [.10, .23]) and significant (z[6] = 4.87, p < .001). The correlation coefficients among the studies varied from r = -.01 to .25. This variation was not significant (Q[8] = 5.35, p = .499), there were no systematic differences between studies (I^2 = 0%), and the total amount of variation between studies was minimal (Tau^2 = 0.00). Thus, there was no need to conduct a moderator analysis to account for the nearly non-existent variance.

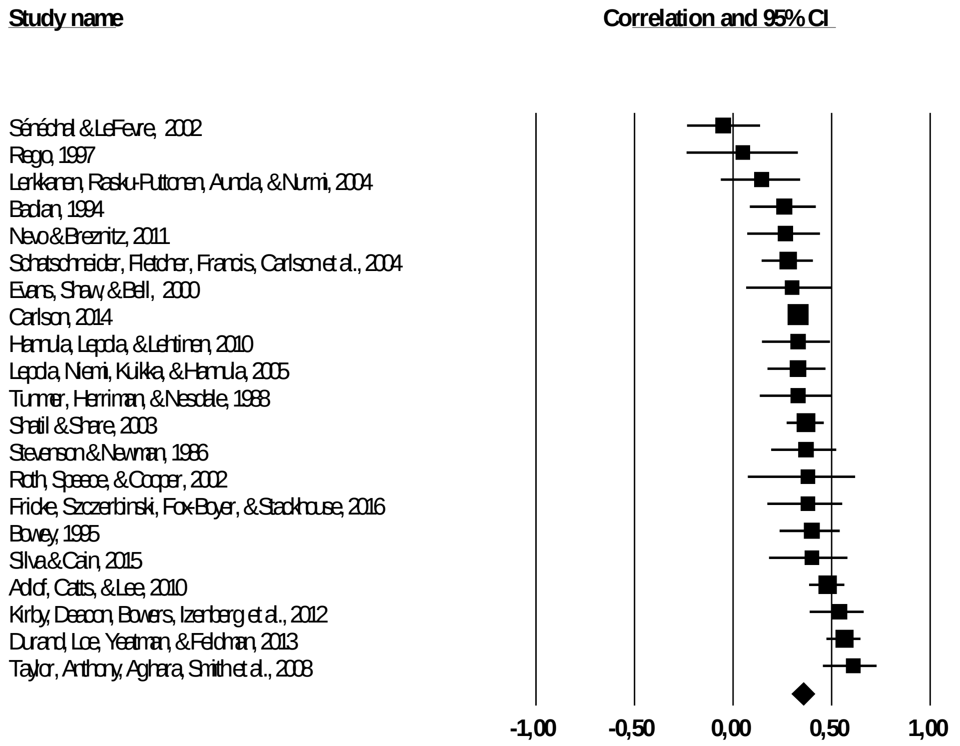
Analysis 12: Forest plot of the correlation between non-word repetition and reading comprehension



The longitudinal correlation between non-verbal intelligence and reading comprehension

In the 21 studies reporting a bivariate correlation between measures of non-verbal intelligence and reading comprehension, the total number of participants was 11,632. The mean age at initial assessment was 5.5 years (*SD* = 0.9), and the mean age at reading comprehension assessment was 8.7 years (*SD* = 2.5). The overall mean correlation is presented in Analysis 13. As shown in the forest plot, the overall summarized correlation is moderate to strong (r = .35; CI [.30, .41]) and statistically significant (z[20] = 11.48, p < .001). The correlation coefficients among the studies varied from r = -.05 to .61. This variation was significant (Q[20] = 73.75, p < .001) and represents a substantial proportion of true heterogeneity (I^2= 72.8%), although the total amount of variation between studies was large (Tau^2 =0.01).

Analysis 13: Forest plot of the correlation between non-verbal intelligence and reading comprehension



We conducted a meta-regression analysis to explain the between-study variation. An analysis with age at non-verbal intelligence assessment and age at reading comprehension assessment was significant ($Q[2] = 14.91$, $p < .001$). A further examination of each covariates contribution revealed that the size of the correlation is related to age at initial assessment. The negative regression coefficient implies that a higher correlation is associated with lower age. The model explained 22% of the variance between the studies ($R^2 = .22$).

Synthesis of results: meta-analytic structural equation modeling

Two-stage SEM stage 1: Combining the correlation matrices

In the first step, we checked the correlation matrices for their positive definiteness —a matrix is considered positive definite if all its eigenvalues are positive (Wothke, 1993). Only matrices that are positive definite contributed to this step of combining them into a pooled matrix. A matrix is positive definite if all of its eigenvalues are positive (Wothke, 1993). In the context of c-MASEM, which is based on maximum likelihood (ML) estimation procedures, a non-positive definite correlation matrix may cause serious problems in the estimation of the model-implied covariance or correlation matrices. This probably arises mainly because ML estimation inverts this matrix and maximizes its similarity with the input matrix.

Consequently, 21 of the 64 studies are excluded, and the data set was therefore reduced. The likely reason that these studies do not provide positive definite correlation matrices is the missing data in many of the correlations among the constructs or the low variation across studies in some of the correlations (Cheung, 2015). Table 2 shows the corresponding numbers of studies for which correlations were available.

Table 2: Coverage of correlations within the 42 selected studies

	PHONEME	LK	VOC	GRA	WDEC
LK	13				
VOC	18	13			
GRA	7	5	8		
WDEC	18	10	19	6	
RC	26	17	30	8	28

Using the resultant 43 studies, we combined the correlation matrices. The 43 included studies are marked with "MASEM" in the table provided in the online supplement 4. For the homogeneity of correlations between the studies, although most correlations do not show significant variation across the 43 studies (probably because of the small number of studies for some correlation matrices), at least three correlations show significant variation: the correlations of the outcome variable, reading comprehension, with phoneme awareness ($r = .43$, CI [.38, .47], $z[40] = 18.12$, $p < .001$; $I^2 = 62.5\%$, $Tau^2 = 0.01$), vocabulary ($r = .42$, CI [.36, .47], $z[40] = 15.06$, $p < .001$; $I^2 = 75.2\%$, $Tau^2 = 0.01$), and concurrent word decoding ($r = .73$, CI [.67, .79], $z[40] = 23.90$, $p < .001$; $I^2 = 95.6\%$, $Tau^2 = 0.02$). Moreover, the overall test of the 43 correlation matrices indicates heterogeneity in the data, $Q[207] = 919.70$, $p < .001$. These findings indicate the need to consider variation in correlations within the matrices across studies and support our decision to specify a random-effects model in this stage. Table 3 shows the estimated correlation matrix from these 43 studies that resulted from a random-effects model.

Table 3: Pooled correlation matrix estimated in the two-stage SEM, stage1 (random-effects model)

	PHONEME	LK	VOC	GRA	WDEC	RC
PHONEME	1.00					
LK	.45	1.00				
VOC	.33	.33	1.00			
GRA	.39	.34	.42	1.00		
WDEC	.43	.49	.34	.34	1.00	
RC	.43	.42	.42	.36	.73	1.00

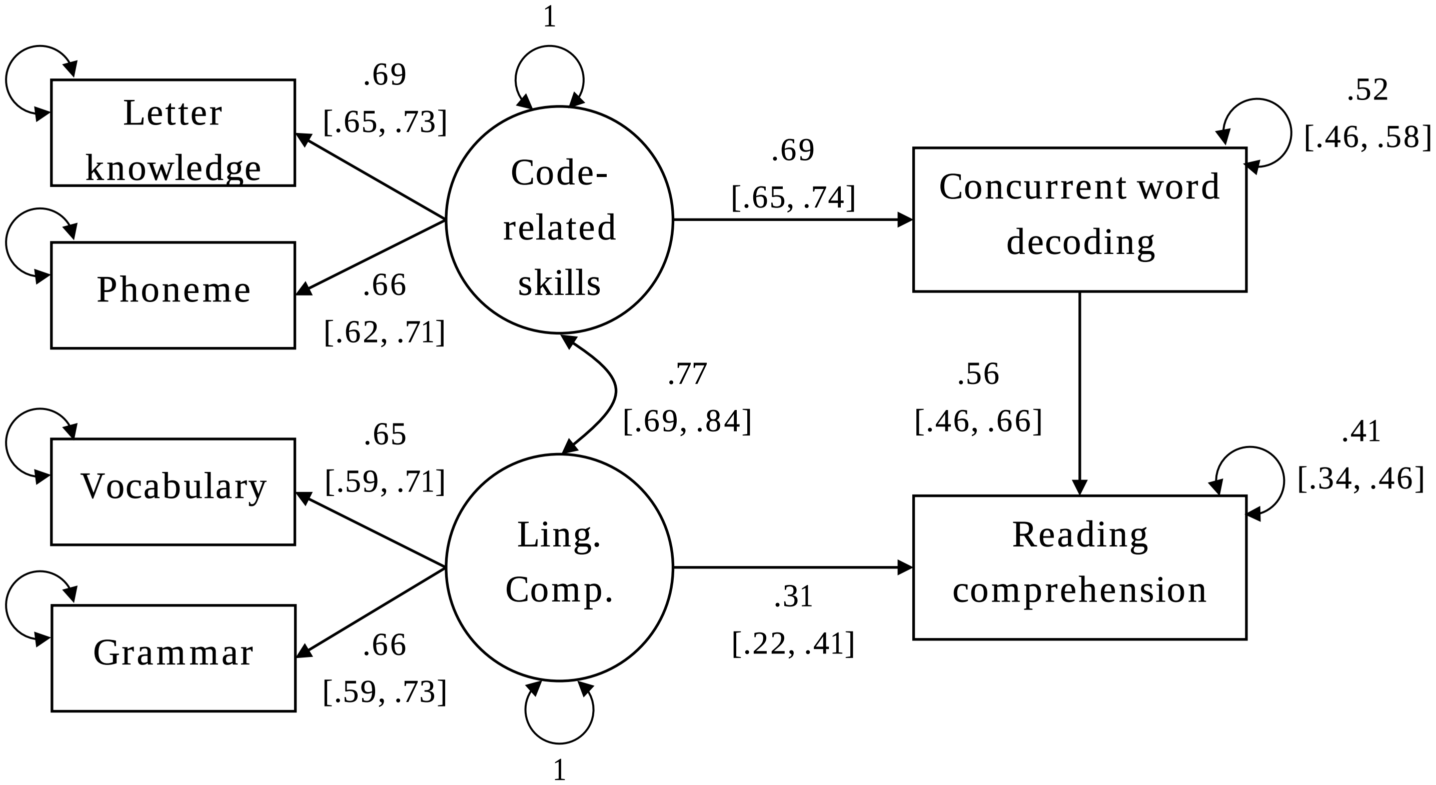
Note: Phoneme = phoneme awareness, LK = letter knowledge, VOC = vocabulary and listening comprehension (verbal ability), GRA = grammar, WDEC = concurrent word decoding, RC = reading comprehension

Two-stage SEM, Stage 2: Structural equation modelling

Based on the pooled correlation matrix, a structural equation model is fitted. The result is shown in Figure 4 below. The model fitted the data well; $\chi 2$ [7] = 7.62, p = .37, RMSEA = .004, CFI = 1.000, TLI = .999, SRMR = .021, AIC = -6.38, and BIC = -54.04. Moreover, a significant indirect effect of code-related skills on reading comprehension via consecutive word decoding existed, b = .39 [.31, .46]. The overall variance explanation in reading comprehension was 59.5%; that of consecutive word decoding was 47.6%. Note that, given the relatively small sample of studies (i.e., 43 studies corresponding to 6,696 participants in total), the 95% likelihood-based confidence intervals are shown (for further details, please refer to Cheung, 2015).

Notably, the two-stage approach we use here requires positive definite correlation matrices for all studies, thus limiting the number of studies that can be considered. As outlined above, because of the non-positive definiteness of some correlation matrices, 21 of the 64 studies had to be excluded from the TSSEM. Using all studies in the current sample of studies to perform SEM through alternative approaches (i.e., methods based on harmonic means of sample sizes across all studies) might not provide accurate parameter estimates and standard errors. However, to test for potential bias, we ran SEM models for the entire sample of 64 studies (i.e., independently of the definiteness of correlation matrices). The results from these analyses showed results comparable to those of the TSSEM models and did not alter the main conclusion. See online supplement 8 for analyses with methods that are alternatives to TSSEM.

Figure 4: Meta-analytic structural equation model in the two-stage SEM stage 2



Two- stage SEM sub-group analysis

Using the hypothesized model, we perform further sub-group analyses in which the grouping variable was the number of years of reading instruction to which the children had been exposed at the last assessment time point. The group of studies named "Early reading" included the studies that assessed reading comprehension after the children had received 1-2 years of formal reading instruction. The "Later reading" group included the studies in which the children had received more than two years of formal reading instruction. This grouping was used to determine whether the predictive relations changed when the children became more-experienced readers.

The hypothesis was that the studies that measured reading comprehension after the children had been exposed to more than two years of reading instruction would exhibit a higher correlation between linguistic comprehension abilities (vocabulary and grammar) than the other studies would. To examine this hypothesis, the meta-analytic structural equation model can be extended to a multi-group model. This extension, however, must be performed during the pooling of correlation matrices in the stage 1 analysis. In the case of a random-effects model, the correlation matrices for each of the sub-groups (i.e., group 1: early reading; group 2: later reading) are combined separately (Jak, 2015). This step divides the sample of correlation matrices into two groups, reduces the number of studies per group, and therefore decreases the variation between studies within groups. In the current meta-analysis, we specified two random-effects models for studies focusing on early and later reading separately to pool the correlation matrices.

In the first stage, the correlation matrices are combined for each of the study design groups. Table 4 details the pooled matrices.

Table 4: Pooled correlation matrices across study designs (fixed-effects model)

	PHONEME	LK	VOC	GRA	WDEC	RC
			Early reading (n = 16 studies, N = 2,426)			
PHONEME	1.00					
LK	.40	1.00				
VOC	.27	.34	1.00			
GRA	.42	.38	.36	1.00		
WDEC	.44	.51	.32	.37	1.00	
RC	.41	.44	.34	.39	.74	1.00
			Later reading (n = 26 studies, N = 4,270)			
PHONEME	1.00					
LK	.47	1.00				
VOC	.34	.31	1.00			
GRA	.35	.31	.43	1.00		
WDEC	.41	.47	.34	.23	1.00	
RC	.43	.41	.46	.34	.72	1.00

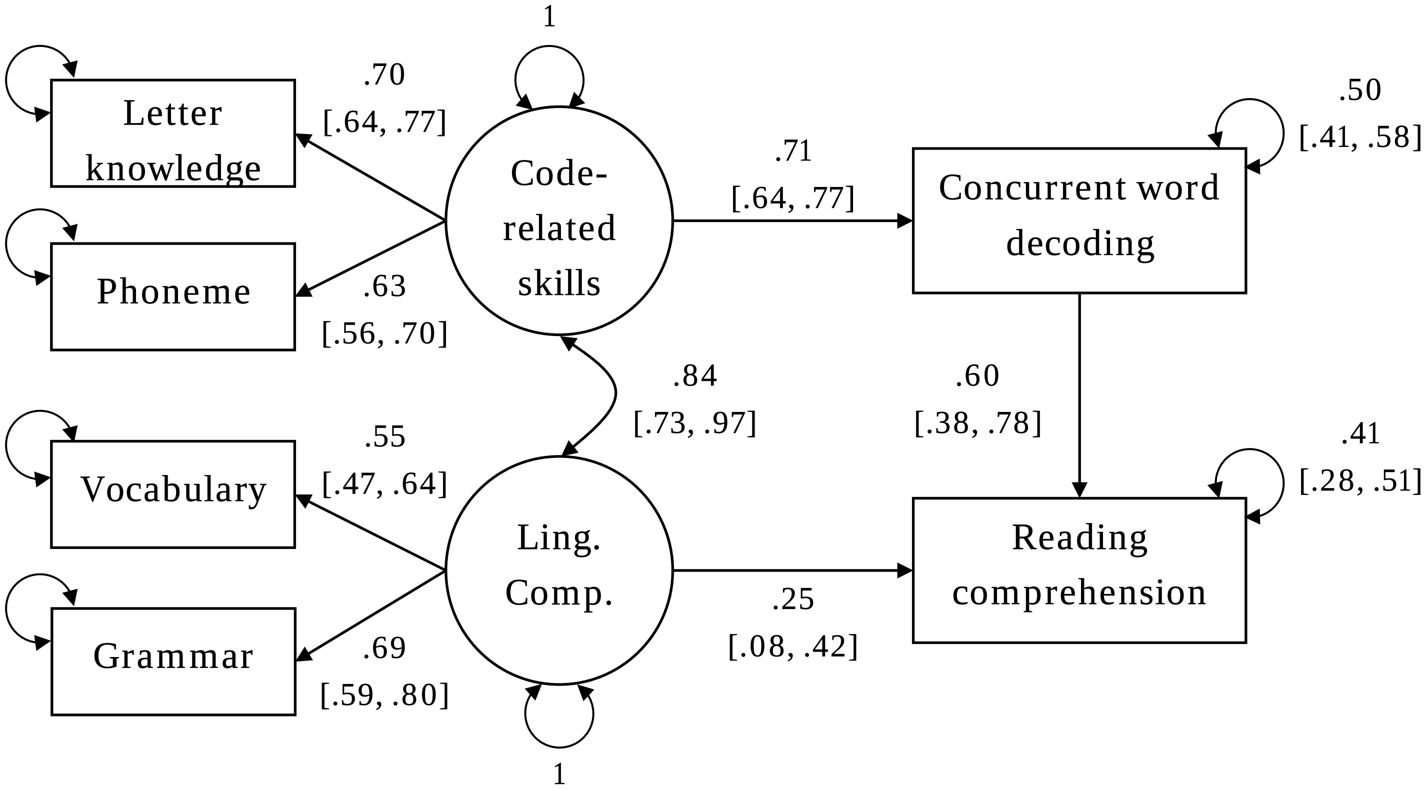
In the second stage, a multi-group SEM is separately specified based on the pooled correlations and the hypothesized model structure for each group.

The resultant model fitted the data well for studies in the early reading group (χ^2 [7] = 6.28, *p* = .51, RMSEA = .000, CFI = 1.000, TLI = 1.002, SRMR = .026, AIC = -7.7, and BIC = -48.3) and studies in the later reading group (χ^2 [7] = 5.38, *p* = .61, RMSEA = .000, CFI = 1.000, TLI = 1.003, SRMR = .033, AIC = -8.6, and BIC = -53.1). Figure 5 details the model parameters, accompanied by their 95% Likelihood-based confidence intervals, for each group. Because the confidence intervals of the model parameters overlap, we cannot be certain that the subgroup differences are statistically significant.

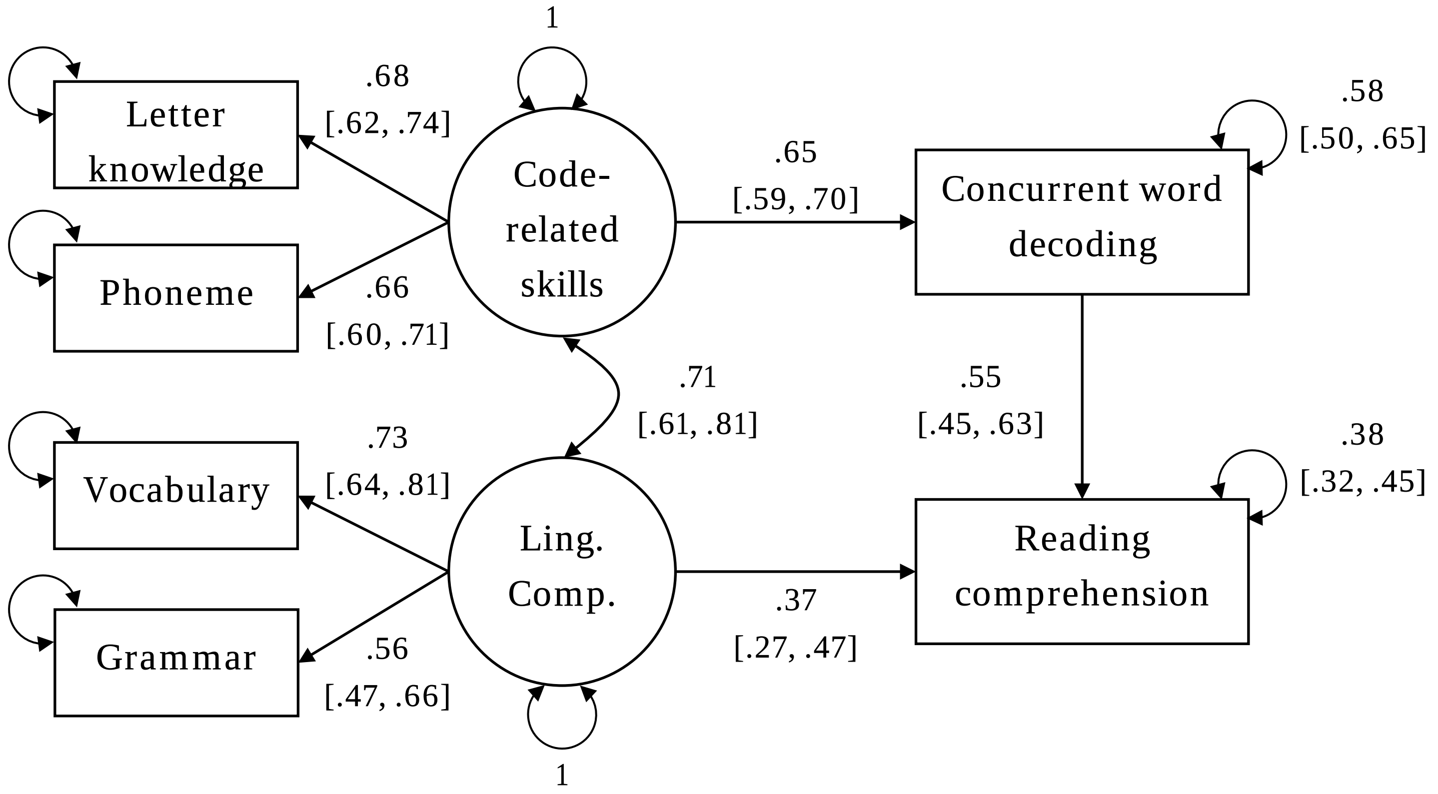
For the total sample of 42 studies, the indirect effect of code-related skills on reading comprehension via word decoding was observed in studies in which children had 1-2 years of reading instruction (*b* = .42 [.27, .57]) and studies in which children had more than two years of reading instruction (*b* = .35 [.29, .42]). These results suggest that the hypothesized model holds even across the two groups.

Figure 5: Multi-group, meta-analytic structural equation model in the two-stage SEM with study design as the grouping variable

Early reading:



Later reading:



Reflections on meta-analytic structural equation modeling

As noted earlier, the c-MASEM approach has several advantages relative to univariate or generalized least squares analyses. Meta-analyses of correlational studies often analyze only bivariate correlations or use methods of merging data that do not consider that the different paths in a correlation matrix are covered by an unequal number of studies. MASEM is a novel and important way to address the shortcomings that have been present in most previous meta-analyses of correlational studies.

Nevertheless, c-MASEM has some limitations: first, using this approach requires a reasonable number of studies, such that the coverage of correlations in the pooled correlation matrix is sufficient. This limitation may become particularly problematic for models with a large number of variables and constructs. Second, although the full information ML procedure can generally handle missing data fairly well (Enders, 2010), high numbers of missing correlations may cause serious convergence problems, particularly in the first stage of analysis. In the current review, we were not able to specify a more complex SEM that could have included further constructs or even measurement occasions, because many correlations were completely missing in all studies. Based on the coverage and on prior research in the field, we selected the most important variables. Third, the estimation of likelihood-based confidence intervals may not necessarily work equally well for different types of structural equation models, although they are generally preferred over Wald's *z*-based confidence intervals (Cheung, 2009). In some models, the lower and upper bounds may be out of the possible range (Cheung, 2015). Fourth, c-MASEM might not be suitable to explain variation of model parameters (e.g., path coefficients) across studies—parameter-based MASEM (p-MASEM) seems to be a reasonable alternative (Cheung and Cheung, 2016).Despite these limitations, meta-analytic SEM still represents a powerful and promising approach to test complex hypotheses on the structural relations among constructs.

Discussion

Summary of main results

First, all the included predictors, except for non-word repetition, had a moderate correlation with later reading comprehension, as shown in the bivariate analyses. Non-word repetition had only a weak contribution to later reading comprehension ability.

Second, the results showed a significant indirect effect of code-related skills on reading comprehension via consecutive word recognition. Moreover, the overall individual variance in reading comprehension explained by the model was 59.5%; that of consecutive word recognition was 47.6%.

Third, as hypothesized, linguistic comprehension had a larger contribution in predicting reading comprehension ability when children became more-experienced readers.

Fourth, the results revealed a high correlation between code-related skills and linguistic comprehension skills in preschool. The correlation is even higher (r = .84) between the two latent variables in the studies in which the children's reading comprehension was assessed within the two first years of formal reading instruction.

With respect to the generalizability of the findings, this review included studies of typically developing monolingual children. In other words, the findings are generalizable to this group but not necessarily to children with learning difficulties or second language learners. However, longitudinal studies show that the main predictive pattern is similar between these groups and typically developing children (Lervåg & Aukrust, 2010). In addition to apparent differences in levels between the groups, there could be group differences in the strength of the predictors across development. Another caveat that can affect generalization is that previous studies use convenience sampling rather than random sampling.

Overall completeness and applicability of the evidence

With this review, we sought to gather all the available empirical research on the longitudinal relation between language skills in preschool and later reading comprehension ability. After a comprehensive search and a thorough screening process, we obtained 64 included studies, all of which followed a sample of typically developing, mainly monolingual children from preschool and over time into school. Notably, the term "typically developing" might be to some degree misleading because some of the included studies have unselected samples, with all the distribution such samples entail, whereas the samples in the other included studies represent selected samples (i.e., a comparison sample with, for example, criteria of not being impaired or twins). Because longitudinal studies take time to conduct and publish, there will certainly be ongoing studies that were not included in this review but that will be eligible for subsequent updates to this review. In addition, through this work, we have identified a number of reporting weaknesses that should be addressed in future studies. The failure to report important study characteristics is unfortunate and complicates the interpretation of the results because we do not have sufficient information about the included studies.

Quality of the evidence

As previously noted, research in the field of language and reading development has proliferated. Thus, a wealth of information is available on the predictive relations that are the focus in this review. Although some of the included studies are large and provide much information, others are smaller, with typically developing children serving as a comparison group. The strength of the evidence in this review is in the updated overall summarized correlations and the application of the meta-analytic SEM approach.

Evaluating the quality of the evidence is challenging. Although we, as authors of a systemic review, would want and expect all the included studies to report the information needed for our analyses and coding schemes, the authors of primary studies must follow the guidelines of the journals in which that they seek to publish. Although the analyses using study quality as a moderator proved insignificant, this result does not necessarily mean that study quality is not related to the size of the correlations shown in the included studies. It is plausible to believe that study quality can be a factor that introduces bias, but it is difficult to determine how and to predict the direction of its effect. Moreover, the coding of study quality might not have been sufficiently sensitive to the variation in study quality, and we might have differentiated (e.g., used more range in values within each indicator) to a greater extent. The concern was that this approach might cause some quality indicators to have a greater effect on the total score than the other dichotomous indicators. The coding of study quality showed that there are concerns related to the study quality in the included studies.

First, most of the studies use convenience samples rather than random samples. In other words, we cannot be sure that the results are actually generalizable to the population. Notably, the aim of this review is to examine relations between preschool predictors and later reading comprehension but not differences in levels. If the sample is biased, for instance, with respect to socio-economic background, this bias is perhaps likely to have a stronger effect on levels rather than the strength of the relations. However, such a line of argument is merely speculation; as long as the samples are not randomly selected, the lack of random sampling can cause bias with respect to generalizability (Vandenbroucke et al., 2014).

A second issue that has not been sufficiently addressed in most previous studies is measurement error. Only 19 of the 64 studies actually report reliability for all of the included measures. In combination with only 4 of the 64 studies dealing with measurement issues by using latent variables, such measurement issues can bias the results (Cole & Preacher, 2014). Notably, because we use latent variables in the meta-SEM, this issue is more pertinent at the primary study level than in the review.

Another important source of bias is attrition of participants from the study. In longitudinal studies, some amount of attrition is expected because children move or are absent on the day of assessment. Therefore, the longer a study is ongoing, the higher its odds of attrition. However, addressing the reason for attrition is important because it may not be completely random. Fifteen studies do not report sample sizes at the two time points and only include information on the number of participants that completed the entire data collection.

In addition, few studies reported employing methods more reliable than listwise deletion to address missing data. Notably, a common approach to handle attrition is to compare the remaining sample with the group that did not complete the study to ascertain to what extent the remaining sample differs significantly on the included variables. High levels of attrition combined with listwise deletion can cause bias in the analyses.

Notably, although we did search the grey literature, only one study that would be considered grey literature, a PhD dissertation, is included in the analyses.
The quality of evidence will be further addressed in the next section.

Limitations and potential biases in the review process

Issues of measurement

Some limitations must be considered when drawing conclusions from the present study. One of these limitations concerns the reliability of the measures that were coded. From the studies that we included in our meta-analysis, we extracted simple raw correlations between measures of predictor and outcome variables, which imply that none of the effect sizes were corrected for measurement error in the bivariate analyses. Because measurement error can lead to the attenuation of effect sizes, the strength of our summarized correlation coefficients could be somewhat underestimated (Schmidt & Hunter, 2004). However, this possibility was addressed with the application of the meta-analytic SEM, in which the predictor's phoneme awareness, letter knowledge, vocabulary and grammar were corrected for measurement error by the inclusion of two latent variables.

Importantly, replicability issues in experimental psychology have received much attention in recent years (Open Science Collaboration, 2015). However, much less attention has been directed toward replicability in multivariate observational studies such as the ones we review here. Although we have used latent variables to deal with measurement error in our synthesis of studies, measurement error is an important source of bias in the primary studies. As mentioned above, for bivariate relations, measurement error attenuates the correlations. However, measurement error has unpredictable consequences for multivariate relations (Cole & Preacher, 2014). For example, if predictors with observed variables differ in their reliability, the most reliable predictor is likely to surpass the others in explaining unique variation (Cole & Preacher, 2014). A striking feature of the primary studies reviewed here is that they reach highly different conclusions concerning which variables are important and explain unique variation in reading comprehension. For instance, working memory, syntax, nonverbal IQ, exposure to books and socioeconomic background are all examples of variables that explain unique variation in one or more of the primary studies (e.g., Bowey, 1995; Hecht, Burgess, Torgesen, Wagner, & Rashotte, 2000; Roth, Speece, & Cooper, 2002; Sénéchal & LeFevre, 2002) but that are not replicated in other studies that include these variables. Although replicability issues can have several explanations, dealing with measurement error is clearly important because this factor is likely to strongly affect the replicability level of findings.

Furthermore, many of the included studies use the same measures. For instance, more than half of the effect sizes included in the present meta-analysis on vocabulary represented a correlation between the two outcome variables and highly similar tests: the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007) and the British Picture Vocabulary Scale (BPVS; Dunn, Dunn, Whetton, & Burley, 1997). Arguably, the vocabulary component of the average correlations may therefore represent a single test type rather than a broad theoretical construct. The narrow range of test types thus reflects a tendency in the field to prefer measures such as PPVT and BPVS above other measures of vocabulary.

The same tendency is noticeable with regard to measures of reading ability. Most of the studies in our analysis measured reading comprehension and word recognition by using different editions of the Woodcock-Johnson test battery (e.g., Woodcock-Johnson III; Woodcock, McGrew, & Mather, 2001). Although these measures are known for their good psychometric properties, it is unfortunate that theoretical constructs have become equated with certain types of tests. For example, there is reason to believe that measures of reading comprehension using cloze procedures, such as Woodcock-Johnson's measure of passage comprehension, rely heavily upon word recognition processes (Francis et al., 2006; Keenan, Betjemann, & Olson, 2008). Consequently, word recognition abilities may be overrepresented in this particular operationalization of reading comprehension. Within the context of a single study, this issue of construct validity is often an acceptable limitation. However, when similar measures are systematically favored by reading researchers, we might find a skewed perception of reading comprehension ability in the field. We hope that researchers will consider these theoretical limitations when choosing measures in future studies. A latent variable approach with multiple indicators may be a good alternative to the use of single measures, and this approach is becoming more common in large longitudinal studies.

Statistical power

Inclusion criteria

First, we could have increased the number of studies in our meta-analysis by adjusting our eligibility criteria. For instance, we could have included studies that reported measurements of the predictor variables after the onset of formal schooling and/or studies with concurrent measurements of the predictor and outcome variables. Such a method would probably have increased the total number of eligible studies substantially.

However, by ensuring that the predictor variables were present before the acquisition of conventional literacy skills, we were able to establish an important criterion for causal inference, namely, temporal precedence. Although great caution must be exercised when making causal inferences based on predictive correlations, temporal precedence represents a minimum requirement for indicating causal order. Concurrent correlations, however, provide virtually no evidence for the purpose of causal interpretation. Thus, by including concurrent data in our study, we may have indeed gained statistical power for the moderator analyses, but the interpretation of our main analyses would also be obscured.

Moreover, by including only samples with mainly typically developing children, we excluded multiple studies that could have increased the number of studies and the sample size. In addition, we excluded studies in which most children had attended Head Start because that would have represented an additional intervention for these children.

Missing data

A second condition relevant to the issue of statistical power concerns the information provided by the research reports in our study. More specifically, not all reports included information relevant to the moderator analyses that we conducted. It is therefore difficult to exclude the possibility that the amount of missing data may have undermined some of the analyses, thus creating a skewed image of the importance of the individual moderator variables. Furthermore, some of the moderator analyses that were originally planned could not be conducted because too few studies reported relevant information. In particular, this issue concerned various characteristics of reading comprehension measures, such as the genre of the reading material (e.g., narrative vs. expository), whether the text was read silently or aloud, and the availability of the text while receiving comprehension questions. In addition, many authors did not report the age of children at all follow-up time points; therefore, we resorted to estimating the age of participants at the missing time points in many instances. Moreover, it is important to provide information concerning when children in the study began their formal reading instruction. We would therefore like to conclude by encouraging researchers to always provide generous descriptions of study characteristics, especially with regard to the nature of the measurements being used. Generous descriptions will not only increase the replicability of individual studies but also facilitate systematic analyses of research in the field.

Agreements and disagreements with other studies or reviews

The review conducted by NELP (2008) is the most similar to ours and is thus the one that we will mainly compare with ours in terms of agreement and disagreement.

To what extent do phonological awareness, rapid naming, and letter knowledge correlate with later decoding and reading comprehension skills?

Phonological awareness is one of the key predictors of early reading ability (Melby-Lervåg et al., 2012b). As children become more experienced readers, their other abilities, such as vocabulary and grammar, are expected to explain more of the variance in their reading comprehension. Thus, we hypothesize that the correlation with reading comprehension will decrease over time. Because our review aimed to focus on the longitudinal predictive relation and thus coded the last follow-up assessment in the included studies, we would expect the longitudinal contribution of phonological awareness to reading comprehension to be lower than the one reported in the NELP (2008) review because reading abilities were assessed earlier. Because phoneme awareness and rhyme awareness have demonstrated unique contributions (Melby-Lervåg et al., 2012b) in predicting later reading development, we chose to separate these two in the analysis. The authors of the NELP (2008) review also presented the different subcategories of phonological awareness. First, the average correlation between phoneme awareness and reading comprehension was reported to be r = .44 in the 2008 review compared to r = .40 in the present review. This small difference in size might be related to some degree to the time of reading comprehension assessment and the greater number of studies included in the current review. Second, the predictive relation between rhyme and reading comprehension, r = .39, is identical in the NELP (2008) and the present review.

As previously noted, phoneme awareness and rhyme awareness have a special contribution to the technical side of reading, decoding. However, depending on how much time passed between the two assessments, we would expect a higher correlation in the NELP (2008) review than in ours. Although assessments that are performed closer in time are likely to be more highly correlated than assessments with more time elapsed between them, developmental changes also affect the strength of this correlation.

First, the average correlation between phoneme awareness and word recognition in the NELP (2008) review was reported to be r = .42, while that in the current study was r = .37. Second, the average correlation between rhyme and word recognition was r = .29 in the NELP (2008) review and r = .32 in the current review. Our assumption was confirmed with phoneme awareness, but the contribution of rhyme was almost identical. Rhyme was the weakest predictor in both the NELP (2008) study and the current review.

Another influential predictor of early reading ability is letter knowledge. In the NELP (2008) review, the average correlation between letter knowledge and reading comprehension was *r* =.48, while that in the current study was *r* = .42. The difference between the two reviews is greater with respect to letter knowledge and word recognition. The NELP (2008) review reported a strong correlation of *r* = .50, while the present study reported a moderate correlation of *r* = .38. We hypothesize that this correlation can be attributed to a weaker contribution because of the longer time between assessments and, hence, more experience with reading.

RAN is the third predictor that is particularly related to early reading. In the NELP (2008) review, RAN is divided in two subcategories: alphanumeric (naming of letters and digits) and non-alphanumeric (naming of objects and colors). However, in the present review, we instead chose to create one composite measure of RAN. The longitudinal contributions of RAN to reading comprehension in the NELP (2008) review are respectively *r* =.43 and *r* = .42 for the two above-mentioned subcategories, while it is *r* = -.34 in the current study. For RAN and word recognition in the NELP (2008) review, the average correlation between the two subcategories is somewhat different, with *r* = .40 for the naming of letters and digits and *r* = .32 for the naming of objects and colors. In the present review, the correlation is *r* = -.37, which can be interpreted as an average of the two subcategories in the previous review. In the NELP (2008) review, the authors chose to present a positive correlation between RAN and the two outcomes, whereas we chose to present it as a negative correlation. In our review, the faster one is at naming (smaller number of seconds), the better reading comprehension (the higher the score) one has. In the NELP (2008) review, the RAN score refers to the number of items per second (the higher the score, the better one is).

To what extent do linguistic comprehension skills in preschool correlate with later reading comprehension ability?

The NELP (2008) review reported an average correlation of *r* = .33 between oral language in kindergarten or earlier and reading comprehension. Oral language here includes both measures of vocabulary and grammar. In the present review, the bivariate analysis showed a correlation of *r* = .42 for vocabulary and *r* = .41 for grammar. This difference in results may be attributed to a number of factors.

One important factor is the kind of measures included. In the breakdown of results into different subcategories of oral language measures, the NELP (2008) review reported an average correlation of *r* = .25 between measures of receptive vocabulary in preschool and reading comprehension in kindergarten. This predictive correlation is fairly weak and is actually the weakest of the included subcategories, and NELP's (2008) finding was therefore somewhat unexpected considering the central role of word knowledge in theories of comprehension (Perfetti & Stafura, 2014). Although we chose to create composites when the primary studies included both receptive and expressive measures, 26 of the 45 included studies had only a receptive vocabulary measure (e.g., PPVT). In the NELP (2008) review, all the other oral language measures (e.g., listening comprehension, verbal IQ, expressive vocabulary) had a higher average correlation with reading comprehension, with a range from r = .31 to .70. Consequently, we chose to create composites that reflected broader vocabulary ability in the analysis. The establishment of vocabulary as a robust predictor of reading comprehension in the present study is thus consistent with theoretical expectations.

In contrast, the present review reports a lower correlation between grammar and reading comprehension than the NELP (2008) review, which reports a strong average correlation of r = .64. Drawing inferences about the reasons for this difference is difficult, but one possible explanation is the measures included. As the different oral language measures in the NELP (2008) review suggest, the decision whether to include receptive or expressive measures or make composites may have influenced the strength of the predictive relations that the measures reported.

Because we wanted to focus on the longitudinal contribution of vocabulary and grammar to reading comprehension, we selected the last reported follow-up in the primary studies. We hypothesized that predictors related to linguistic comprehension would have a larger contribution when children had become more-experienced readers. Thus, the moderator analyses and sub-group analysis in the meta-analytical SEM addressed this issue. To test the impact of age, the authors of the NELP (2008) review sought to group the studies examining oral language in two groups: one group with the studies that assessed reading comprehension in kindergarten and the other group with the studies assessing reading comprehension in first or second grade. However, since fewer than three studies assessed reading comprehension in first or second grade, there were only a sufficient number of studies measuring reading comprehension in kindergarten. Furthermore, it is stated, “This comparison indicated that oral language was a significantly stronger predictor when reading comprehension was measured in first and second grade” (p.72). From our perspective, however, it is unclear how they reached this conclusion, since this is not included in the analysis.

Furthermore, more years had passed between the measurements of vocabulary and reading comprehension in the present study than in the NELP (2008) review. Typically, correlations are expected to diminish over time; thus, the difference in results between the present study and the NELP (2008) review opposes the general empirical pattern. However, as we suggested in the introduction, these results must be interpreted in light of developmental theories of reading. For instance, according to the simple view of reading, reading comprehension is the product of word recognition and linguistic comprehension (Gough & Tunmer, 1986). Although both components are equally important, their independent contributions to reading comprehension change over the course of development (Gough, Hoover, & Peterson, 1996). Thus, the different magnitude in the effect sizes found in the present study versus that in the NELP (2008) review may represent a developmental trend rather than conflicting results. Following this line of argument, it might seem surprising that age did not emerge as an important moderator in our analysis. However, the fact that age was not a significant moderator in either the present study or the NELP (2008) review could also reflect the limited variation in the age of the participants included in each of the meta-analyses. Overall, the combined results of the two meta-analyses support the simple view of reading. Although this conclusion is not particularly newsworthy, it represents an important empirical validation of central theoretical assumptions regarding the development of reading comprehension.

To what extent does verbal short-term memory in preschool correlate with later reading comprehension ability?

In the NELP (2008) review, the average correlation between phonological short-term memory was reported to be r = .39. Thirteen studies were included in that analysis, and these included a measure that assessed the ability to remember spoken information for a short period of time (e.g., digit span, sentence repetition or non-word repetition). In the present review, we chose to separate two of these assessments, sentence memory and non-word repetition. Our review showed different results, with average correlations of r = .36 from the nine studies including sentence memory and r = .17 from the seven studies on non-word repetition. Sentence memory has a stronger predictive relation to later reading comprehension than non-word repetition does. Previous research has shown that remembering sentences places high demands on abilities related to linguistic comprehension (i.e., vocabulary and grammar) (Klem et al., 2015). Moreover, sentence repetition shares attributes that are typical of assessing reading comprehension, including asking questions about the text, which requires children to remembering related information. The weak correlation between non-word repetition and reading comprehension indicates that the longitudinal contribution from repeating non-words is not highly related to later reading comprehension. The difference in the results in the two reviews may also be attributed to the measures included; the phonological short-term memory predictor variable in the NELP (2008) review may involve a greater number of studies with sentence repetition than those with non-word repetition.

To what extent does non-verbal intelligence in preschool correlate with later reading comprehension ability?

The longitudinal contribution between non-verbal intelligence and reading comprehension was shown to be moderate in both the previous NELP (2008) review and the present review. The average correlation from the five studies included in the NELP (2008) review was r = .34, and that from the twenty studies included in our review was r = .35.

To what extent do preschool predictors of reading comprehension correlate with later reading comprehension skills after concurrent decoding ability has been considered?

As previously stated, a latent variable approach is rare, but its use is increasing. From the 64 included studies, only two used a SEM approach with latent variables to analyze their data. However, a number of studies were excluded because they did not report bivariate correlations and instead performed SEM.

A study by Hulme, Nash, Gooch, Lervåg, and Snowling (2015) is one of the two included studies that have used a SEM approach. In this study, the model with speech and language at age 3 ½ and RAN, letter knowledge and phoneme awareness at age 4 ½ accounted for 47% of the variance in word-level skills at age 5 ½ and 12% of the variance in reading comprehension ability at age 8. Reading comprehension at age 8 was predicted by language at 3 ½ years and word-level literacy at 5 ½ years. Here, the regression coefficient from language (with the observed variables of sentence repetition, vocabulary, sentence structure and basic constructs) is $\beta = .26$. The direct effect from language to later reading comprehension showed that reading comprehension is also strongly linked to variation in linguistic comprehension at an early age, even after decoding has been considered.

Considering the longitudinal aspect of reading is also important. Different factors and abilities make significant contributions to the development process at different times. Phonological awareness, letter knowledge and RAN have been shown to be important in the beginning, when a child is learning to match sounds to letters. Later, when the decoding has become automatized, capacities are freed for the linguistic comprehension components. The present review includes studies that have measured reading comprehension ability at different ages. Some studies have assessed reading comprehension in second grade, while others have assessed it in tenth grade. Thus, decoding ability may be a factor to varying degrees, depending on the children's exposure to and amount of experience with reading. Moreover, this relation changes with age. Cain and Oakhill (2007) referred to longitudinal studies showing that correlations between reading and linguistic comprehension are generally low in beginning readers, but these correlations gradually increase when decoding differences are small.

To what degree do other possible influential moderator variables (e.g., age, test types, SES, language, country) contribute to explaining any observed differences between the studies included?

All 64 included studies have their own study characteristics. A number of factors may explain the between-study variation in the reported effect sizes. First, the studies are conducted in different educational systems that may have implications for the approach to formal reading instruction. Thus, children may be exposed to varying degrees of school readiness activities in preschool that are difficult to account for in the review because the studies include little information about the extent of this activity. Second, children also start school at different ages. Although we can attempt to control for this in the analyses, most studies did not report this data; therefore, the estimate was less precise than we would have wanted in the moderator analyses. Third, age was used as a moderator because we expected this to be a variable that could predict some of the heterogeneity shown in the studies. As expected, age proved to be a significant moderator in a number of analyses.

Authors' conclusions

Implications for practice and policy

The present review provides compelling evidence of the predictive relation between children's early oral language skills and the development of reading comprehension. Although the correlational nature of this evidence provides a limited basis for causal inference, we argue that the results of the study have important practical implications. First, by gaining insight into the developmental variation in children's oral language skills, preschool educators can more confidently monitor children's progress toward literacy. We must provide educators with well-developed assessment tools targeting the precursors of reading comprehension and the knowledge of how to understand and use the results of such measures. Importantly, children identified as at risk for later reading difficulties should receive appropriate intervention to promote their literacy development.

Furthermore, knowledge of young children's oral language abilities and early literacy skills can provide preschool teachers guidance for adapting instructional activities to children's developmental levels. Finally, we would like to emphasize that the results of the bivariate analyses revealed that a wide range of oral language predictors served as stable indicators of children's reading comprehension development. The meta-analytic SEM analyses further demonstrated that the shared contribution from children's semantic, grammatical and code-related language skills could explain the better part of the variance in their later reading comprehension ability. These results strongly indicate the need for a broad and comprehensive focus on oral language in early childhood education. In summary, we argue that the results of the present review may strengthen preschool practices and increase our ability to provide children rich opportunities for literacy learning.

Implications for research

Based on the risk of bias analyses we conducted, it is clear that previous longitudinal studies had risks of bias that is important to address in future studies. The most pertinent are the following:

- In general, many of the studies lacked transparency with respect to important information and did not report matrices with uncorrected bivariate correlations, means and standard deviations so that the results could be used in a meta-analysis or reanalyzed. A number of journals now allow for online supplement material, where authors can include large correlational matrices for all measures at all time points, means and standard deviations so that the covariance matrix can be reproduced and information can be easily coded in future reviews.
- Many studies had small samples (below 70), were clearly underpowered, and did not report attrition. Furthermore, most of the studies handled missing data by using listwise deletion. These are important aspects to improve in future studies.
- Few studies reported reliability, and even fewer dealt with measurement error by using latent variables. This approach can cause bias and is important to address in future studies.
- Most of the studies included cognitive measures, but only a minority of the studies included measures of potentially important variables such as socio-economic background, home literacy environment and background knowledge. These are potentially important variables to consider in future studies.
- Most of the studies were based on convenience sampling and not on randomized samples. This choice could affect the generalizability of the findings.

References

References to included studies

Aarnoutse, C., van Leeuwe, J., & Verhoeven, L. (2005). Early literacy from a longitudinal perspective. *Educational Research and Evaluation*, *11*, 253-275. doi:10.1080/08993400500101054.

Adlof, S. M., Catts, H. W., & Lee, J. (2010). Kindergarten predictors of second versus eighth grade reading comprehension impairments. *Journal of Learning Disabilities*, *43*, 332-345. doi:10.1177/0022219410369067.

Aram, D., Korat, O., & Hassunah-Arafat, S. (2013). The contribution of early home literacy activities to first grade reading and writing achievements in Arabic. *Reading and Writing*, *26*, 1517-1536. doi:10.1007/s11145-013-9430-y.

Aram, D., & Levin, I. (2004). The role of maternal mediation of writing to kindergartners in promoting literacy in school: A longitudinal perspective. *Reading and Writing*, *17*, 387-409. doi:10.1023/B:READ.0000032665.14437.e0.

Badian, N. A. (1994). Preschool prediction: Orthographic and phonological skills, and reading. *Annals of Dyslexia*, *44*, 1–25. doi:10.1007/BF02648153.

Badian, N. A. (2001). Phonological and orthographic processing: Their roles in reading prediction. *Annals of Dyslexia*, *51*, 177–202. doi:10.1007/s11881-001-0010-5.

Bartl-Pokorny, K. D., Marschik, P. B., Sachse, S., Green, V. A., Zhang, D., Van Der Meer, L., ... Einspieler, C. (2013). Tracking development from early speech-language acquisition to reading skills at age 13. *Developmental Neurorehabilitation*, 16, 188-195. doi:10.3109/17518423.2013.773101.

Bianco, M., Pellenq, C., Lambert, E., Bressoux, P., Lima, L., & Doyen, A. (2012). Impact of early code-skill and oral-comprehension training on reading achievement in first grade. *Journal of Research in Reading*, *35*, 427-455. doi:10.1111/j.1467-9817.2010.01479.x.

Bishop, A. G., & League, M. B. (2006). Identifying a multivariate screening model to predict reading difficulties at the onset of kindergarten: A longitudinal analysis. *Learning Disability Quarterly*, *29*, 235-252. doi:10.2307/30035552.

Blackmore, A. M., & Pratt, C. (1997). Grammatical awareness and reading in grade 1 children. *Merrill-Palmer Quarterly*, *43*, 567-590.

Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology*, *87*, 476-487. doi:10.1037/0022-0663.87.3.476.

Bryant, P., MacLean, M., & Bradley, L. (1990). Rhyme, language, and children's reading. *Applied Psycholinguistics*, *11*, 237-252. doi:10.1017/S0142716400008870.

Burke, M. D., Hagan-Burke, S., Kwok, O., & Parker, R. (2009). Predictive validity of early literacy indicators from the middle of kindergarten to second grade. *Journal of Special Education*, *42*, 209-226. doi:10.1177/0022466907313347.

Carlson, A. G. (2014). *Kindergarten fine motor skills and executive function: Two non-academic predictors of academic achievement*. Dissertation Abstracts International: Section B: The Sciences and Engineering, 74(10-B(E)). Washington DC: George Mason University.

Casalis, S., & Louis-Alexandre, M.-F. (2000). Morphological analysis, phonological analysis and learning to read French: A longitudinal study. *Reading and Writing*, *12*, 303-335. doi:10.1023/A:1008177205648.

Chaney, C. (1998). Preschool language and metalinguistic skills are links to reading success. *Applied Psycholinguistics*, *19*, 433-446. doi:10.1017/S0142716400010250.

Cronin, V. S. (2013). RAN and double-deficit theory. *Journal of Learning Disabilities*, *46*, 182-190. doi:10.1177/0022219411413544.

Cronin, V., & Carver, P. (1998). Phonological sensitivity, rapid naming, and beginning reading. *Applied Psycholinguistics*, *19*, 447-461. doi:10.1017/S0142716400010262.

Cudina-Obradovic, M. (1999). Determinants of reading achievement in the 1st grade: One year follow-up study. *Hrvatska Revija za Rehabilitacijska Istrazivanja*, *35*, 79-94.

Dickinson, D. K., & Porche, M. V. (2011). Relation between language experiences in preschool classrooms and children's kindergarten and fourth-grade language and reading abilities. *Child Development*, *82*, 870-886. doi:10.1111/j.1467-8624.2011.01576.x.

Durand, V. N., Loe, I. M., Yeatman, J. D., & Feldman, H. M. (2013). Effects of early language, speech, and cognition on later reading: A mediation analysis. *Frontiers in Psychology*, *4*, 586. doi:10.3389/fpsyg.2013.00586.

Evans, M. A., Shaw, D., & Bell, M. (2000). Home literacy activities and their influence on early literacy skills. *Canadian Journal of Experimental Psychology*, *54*, 65-75. doi:10.1037/h0087330.

Flax, J. F., Realpe-Bonilla, T., Roesler, C., Choudhury, N., & Benasich, A. (2009). Using early Standardized language measures to predict later language and early reading outcomes in children at high risk for language-learning impairments. *Journal of Learning Disabilities*, *42*, 61-75. doi:10.1177/0022219408326215.

Fricke, S., Szczerbinski, M., Fox-Boyer, A., & Stackhouse, J. (2016). Preschool predictors of early literacy acquisition in German-speaking children. *Reading Research Quarterly*, *51*, 29-53. doi:10.1002/rrq.116.

Furnes, B., & Samuelsson, S. (2009). Preschool cognitive and language skills predicting kindergarten and Grade 1 reading and spelling: A cross-linguistic comparison. *Journal of Research in Reading*, *32*, 275-292. doi:10.1111/j.1467-9817.2009.01393.x.

González, J. E., & González, M. R. (2000). Metalinguistic awareness and reading acquisition in the Spanish language. *Spanish Journal of Psychology*, *3*, 37-46. doi:10.1017/S1138741600005527.

Guajardo, N. R., & Cartwright, K. B. (2016). The contribution of theory of mind, counterfactual reasoning, and executive function to pre-readers' language comprehension and later reading awareness and comprehension in elementary school. *Journal of Experimental Child Psychology*, *144*, 27-45. doi:10.1016/j.jecp.2015.11.004.

Hannula, M. M., Lepola, J., & Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *Journal of Experimental Child Psychology*, *107*, 394-406. doi:10.1016/j.jecp.2010.06.004.

Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). Explaining Social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, Rate of access, and print knowledge. *Reading and Writing*, *12*, 99–128. doi:10.1023/A:1008033824385.

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science*, *26*, 1877-1886. doi:10.1177/0956797615603702.

Karlsdottir, R., & Stefansson, T. (2003). Predicting performance in primary school subjects. *Perceptual and Motor Skills*, *97*, 1058–1060. doi:10.2466/pms.2003.97.3f.1058.

Katz, Y. J., & Ben-Yochanan, A. (1990). Reading and cognitive abilities: Research note on a longitudinal study in Israel. *Perceptual and Motor Skills*, *70*, 47-50.

Kirby, J. R., Deacon, S. H., Bowers, P. N., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2012). Children's morphological awareness and reading ability. *Reading and Writing, 25,* 389-410. doi:10.1007/s11145-010-9276-5.

Kozminsky, L., & Kozminsky, E. (1995). The effects of early phonological awareness training on reading success. *Learning and Instruction, 5*, 187-201. doi:10.1016/0959-4752(95)00004-M doi:10.1016/0959-4752(95)00004-M.

Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology, 93,* 451–455. doi:10.1037/0022-0663.93.3.451.

Lepola, J., Lynch, J., Kiuru, N., Laakkonen, E., & Niemi, P. (2016). Early oral language comprehension, task orientation, and foundational reading skills as predictors of Grade 3 reading comprehension. *Reading Research Quarterly*, 51, 373–390. doi:10.1002/rrq.145.

Lepola, J., Niemi, P., Kuikka, M., & Hannula, M. M. (2005). Cognitive-linguistic skills and motivation as longitudinal predictors of reading and arithmetic achievement: A follow-up study from kindergarten to Grade 2. *International Journal of Educational Research, 43,* 250-271. doi:10.1016/j.ijer.2006.06.005.

Leppänen, U., Aunola, K., Niemi, P., & Nurmi, J. (2008). Letter knowledge predicts Grade 4 reading fluency and reading comprehension. *Learning and Instruction, 18*, 548-564. doi:10.1016/j.learninstruc.2007.11.004.

Lerkkanen, M., Rasku-Puttonen, H., Aunola, K., & Nurmi, J. (2004). Predicting reading performance during the first and the second year of primary school. *British Educational Research Journal, 30*, 67-92. doi:10.1080/01411920310001629974.

Morris, D., Bloodgood, J., & Perney, J. (2003). Kindergarten predictors of first- and second-grade reading achievement. *The Elementary School Journal, 104*, 93-109. doi:10.1086/499744.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical 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.

Näslund, J. C., & Schneider, W. (1996). Kindergarten letter knowledge, phonological skills, and memory processes: Relative effects on early literacy. *Journal of Experimental Child Psychology, 62,* 30-59. doi:10.1006/jecp.1996.0021.

Nevo, E., & Breznitz, Z. (2011). Assessment of working memory components at 6 years of age as predictors of reading achievements a year later. *Journal of Experimental Child Psychology*, *109*, 73-90. doi:10.1016/j.jecp.2010.09.010.

NICHD Early Child Care Research Network Rockville MD US. (2005). Pathways to reading: The role of oral language in the transition to reading. *Developmental Psychology, 41*, 428-442. doi:10.1037/0012-1649.41.2.428.

O'Neill, D. K., Pearce, M. J., & Pick, J. L. (2004). Preschool children's narratives and performance on the Peabody individualized achievement test--Revised: Evidence of a relation between early narrative and later mathematical ability. *First Language*, *24*, 149-183. doi:10.1177/0142723704043529.

Parrila, R., Kirby, J. R., & McQuarrie, L. (2004). Articulation rate, naming speed, verbal short-term memory, and phonological awareness: Longitudinal predictors of early reading development? *Scientific Studies of Reading*, *8*, 3-26. doi:10.1207/s1532799xssr0801_2.

Piasta, S. B., Petscher, Y., & Justice, L. M. (2012). How many letters should preschoolers in public programs know? The diagnostic efficiency of various preschool letter-naming benchmarks for predicting first-grade literacy achievement. *Journal of Educational Psychology*, *104*, 954–958. doi:10.1037/a0027757.

Pike, M., Swank, P., Taylor, H., Landry, S., & Barnes, M. A. (2013). Effect of preschool working memory, language, and narrative abilities on inferential comprehension at school-age in children with spina bifida myelomeningocele and typically developing children. *Journal of the International Neuropsychological Society*, *19,* 390-399. doi:10.1017/S1355617712001579.

Prochnow, J. E., Tunmer, W. E., & Chapman, J. W. (2013). A longitudinal investigation of the influence of literacy-related skills, reading self-perceptions, and inattentive behaviours on the development of literacy learning difficulties. *International Journal of Disability, Development and Education*, *60*, 185-207. doi:10.1080/1034912X.2013.812188.

Rego, L. L. B. (1997). The connection between syntactic awareness and reading: Evidence from Portuguese-speaking children taught by a phonic method. *International Journal of Behavioral Development*, *20*, 349-365. doi:10.1080/016502597385379.

Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and Early reading. *Journal of Educational Research*, *95*, 259-272. doi:10.1080/00220670209596600.

Sawyer, D. J. (1992). Language abilities, reading acquisition, and developmental dyslexia: A discussion of hypothetical and observed relationships. *Journal of Learning Disabilities*, *25*, 82-95. doi:10.1177/002221949202500201.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, *96*, 265-282. doi:10.1037/0022-0663.96.2.265.

Sears, S., & Keogh, B. (1993). Predicting reading performance using the Slingerland procedures. *Annals of Dyslexia*, *43*, 78-89. doi:10.1007/BF02928175.

Sénéchal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to Grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific Studies of Reading*, *10*, 59-87. doi:10.1207/s1532799xssr1001_4.

Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development*, *73*, 445-460. doi:10.1111/1467-8624.00417.

Shatil, E., & Share, D. L. (2003). Cognitive antecedents of Early reading ability: A test of the modularity hypothesis. *Journal of Experimental Child Psychology*, *86*, 1-31. doi:10.1016/S0022-0965(03)00106-1.

Silva, M., & Cain, K. (2015). The relations between lower and higher level comprehension skills and their role in prediction of early reading comprehension. *Journal of Educational Psychology*, *107*, 321-331. doi:10.1037/a0037769.

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development*, *57*, 646-659. doi:10.2307/1130343.

Taylor, H. B., Anthony, J. L., Aghara, R., Smith, K. E., & Landry, S. H. (2008). The interaction of early maternal responsiveness and children's cognitive abilities on later decoding and reading comprehension skills. *Early Education and Development*, *19*, 188-207. doi:10.1080/10409280701839304.

Tunmer, W. E., Chapman, J. W., & Prochnow, J. E. (2006). Literate cultural capital at school entry predicts later reading achievement: A seven year longitudinal study. *New Zealand Journal of Educational Studies*, *41*, 183.

Tunmer, W. E., Herriman, M. L., & Nesdale, A. R. (1988). Metalinguistic abilities and beginning reading. *Reading Research Quarterly*, *23*, 134-158. doi:10.2307/747799.

Uhry, J. K. (2002). Kindergarten phonological awareness and rapid serial naming as predictors of grade 2 reading and spelling. In E. Witruk, A. D. Friederici, & T. Lachmann (Eds.), *Basic functions of language, reading and reading disability* (pp. 299-313). Dordrecht, Netherlands: Kluwer Academic Publishers. doi:10.1007/978-1-4615-1011-6_18.

Wolter, J. A., Self, T., & Apel, K. (2011). Initial mental graphemic representation acquisition and later literacy achievement in children with language impairment: A longitudinal study. *Journal of Learning Disabilities*, *44*, 543-555. doi:10.1177/0022219410392042.

Additional references

Bollen, K. A. (1989). A new incremental fit index for general structural equation models. *Sociological Methods and Research*, *17*, 303-316. doi:10.1177/0049124189017003004.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex, UK: John Wiley & Sons.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2014). Comprehensive Meta-analysis Version 3. Software. Engelwood, NJ: Biostat.

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*, *85*, 1346-1356. doi:10.1111/cdev.12192.

Bosco, F. A., Aguinis, H., Singh, K., Field, J. G., & Pierce, C. A. (2015). Correlational effect size benchmarks. *Journal of Applied Psychology*, *100*, 431-449. doi:10.1037/a0038047.

Brimo, D., Apel, K., & Fountain, T. (2017). Examining the contributions of syntactic awareness and syntactic knowledge to reading comprehension. *Journal of Research in Reading*, *40*, 57–74. doi:10.1111/1467-9817.12050.

Cain, K., & Oakhill, J. (2007). Reading comprehension difficulties: Correlates, causes, and consequences. In K. Cain, & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language: A cognitive perspective* (pp. 41-75). New York, NY: Guilford Press.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*, 31–42. doi:10.1037/0022-0663.96.1.31.

Carroll, J. M. (2011). *Developing language and literacy: Effective intervention in the early years*. Chichester, West Sussex: Wiley-Blackwell.

Chen, R. S., & Vellutino, F. R. (1997). Prediction of reading ability: A cross-validation study of the simple view of reading. *Journal of Literacy Research*, *29*, 1-24. doi:10.1080/10862969709547947.

Cheung, M. W. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods, 41,* 425-438. doi:10.3758/BRM.41.2.425.

Cheung, M. W. (2014). Fixed- and random-effects meta-analytic structural equation modeling: Examples and analyses in R. *Behavior Research Methods, 46,* 29-40. doi:10.3758/s13428-013-0361-y.

Cheung, M. W. (2015). *Meta-analysis: A structural equation modeling approach.* Chichester, West Sussex: John Wiley & Sons, Ltd.

Cheung, M. W. L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10,* 40-64. doi:10.1037/1082-989X.10.1.40.

Cheung, M. W.-L., & Cheung, S. F. (2016). Random-Effects Models for Meta-Analytic Structural Equation Modeling: Review, Issues, and Illustrations. *Research Synthesis Methods, 7*(2), 140-155. doi:10.1002/jrsm.1166

Clarke, P. J., Truelove, E., Hulme, C., & Snowling, M. J. (2014). *Developing reading comprehension*. Oxford: John Wiley & Sons.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Cole, D. A., & Preacher, K. J. (2014). Manifest variable path analysis: Potentially serious and misleading consequences due to uncorrected measurement error. *Psychological Methods, 19,* 300–315. doi:10.1037/a0033805.

Coltheart, M. (2006). Dual route and connectionist models of reading: An overview. *London Review of Education, 4,* 5-17. doi:10.1080/13603110600574322.

Conners, F. A. (2009). Attentional control and the simple view of reading. *Reading and Writing, 22,* 591-613. doi:10.1007/s11145-008-9126-x.

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99,* 311-325. doi:10.1037/0022-0663.99.2.311.

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10,* 277-299. doi:10.1207/s1532799xssr1003_5.

DistillerSR, Evidence Partners, Ottawa, Canada

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health, 52*, 377-384. doi:10.1136/jech.52.6.377.

Dunn, L., & Dunn, L. (2007). *Peabody picture vocabulary test* (4th ed.). Minneapolis, MN: Pearson.

Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). *The British picture vocabulary scale II.* Windsdor: GL assessment.

Egger, M., Schneider, M., & Smith, G. D. (1998). Spurious precision? meta-analysis of observational studies. *BMJ, 316*, 140–144. doi:10.1136/bmj.316.7125.140.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Florit, E., Roch, M., Altoè, G., & Levorato, M. C. (2009). Listening comprehension in preschoolers: The role of memory. *British Journal of Developmental Psychology*, *27*, 935-951. doi:10.1348/026151008X397189.

Foorman, B. R., Koon, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). Examining general and specific factors in the dimensionality of oral language and reading in 4th–10th grades. *Journal of Educational Psychology*, *107*, 884–899. doi:10.1037/edu0000026.

Francis, D. J., Snow, C. E., August, D., Carlson, C. D., Miller, J., & Iglesias, A. (2006). Measures of reading comprehension: A latent variable analysis of the diagnostic assessment of reading comprehension. *Scientific Studies of Reading*, *10*, 301-322. doi:10.1207/s1532799xssr1003_6.

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*, 74-111. doi:10.3102/0034654313499616.

Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL. *Reading and Writing*, *25*, 1819-1845. doi:10.1007/s11145-011-9333-8.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi, & J. V. Oakhill (Eds.), *Reading comprehension difficulties* (pp. 1-13). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Hatcher, P. J., Hulme, C., & Snowling, M. J. (2004). Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *45*, 338–358. doi:10.1111/j.1469-7610.2004.00225.x.

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

Higgins, J. P. T., & Green, S. (Eds.). (2011). Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Retrieved from www.cochrane-handbook.org.

Hjetland, H. N., Lervåg, A., Lyster, S., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (Under review). Pathways to reading comprehension: A Longitudinal Study from 4 to 9 years of age.

Høien, T., & Lundberg, I. (2000). *From theory to intervention*. Norwell, MA: Kluwer Academic Publishers.

Høien-Tengesdal, I., & Høien, T. (2012). The reading efficiency model: An extension of the componential model of reading. *Journal of Learning Disabilities*, *45*, 467-479. doi:10.1177/0022219411432688.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*, 127-160. doi:10.1007/BF00401799.

Jak, S. (2015). *Meta-analytic structural equation modelling*. Utrecht, The Netherlands: Springer Verlag.

Johnston, T. C., & Kirby, J. R. (2006). The contribution of naming speed to the simple view of reading. *Reading and Writing*, *19*, 339-361. doi:10.1007/s11145-005-4644-2.

Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, *21*, 85-97. doi: 10.1080/02702710050084428.

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*, 281-300. doi:10.1080/10888430802132279.

Kim, A. E., Oines, L. D., & Sikos, L. (2016). Prediction during sentence comprehension is more than a sum of lexical associations: The role of event knowledge. *Language, Cognition and Neuroscience*, *31*, 597–601. doi:10.1080/23273798.2015.1102950.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*, 163-182. doi:10.1037/0033-295X.95.2.163.

Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S. A., Gustafsson, J. E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, *18*, 146-154. doi:10.1111/desc.12202.

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

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. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *51*, 612-620. doi:10.1111/j.1469-7610.2009.02185.x.

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology*, *45*, 764-781. doi:10.1037/a0014132.

Lervåg, A., & Hulme, C. (2009). Rapid automatized naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science*, *20*, 1040-1048. doi:10.1111/j.1467-9280.2009.02405.x.

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2017). Unpicking the developmental delationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*. doi:10.1111/cdev.12861.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *American Psychologist*, *48*, 1181–1209. doi:10.1037/0003-066X.48.12.1181.

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.

Lundberg, I., Frost, J., & Petersen, O. (1988). Effects of an extensive program for stimulating phonological awareness in pre-school children. *Reading Research Quarterly*, *23*, 263–284. doi:10.1598/RRQ.23.3.1.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247-288. doi:10.1080/01638539609544975.

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of Research in Reading*, *34*, 114-135. doi:10.1111/j.1467-9817.2010.01477.x.

Melby-Lervåg, M., Lervåg, A., Lyster, S. H., Klem, M., Hagtvet, B., & Hulme, C. (2012a). Nonword-repetition ability does not appear to be a causal influence on children's vocabulary development. *Psychological Science, 23*, 1092-1098. doi:10.1177/0956797612443833.

Melby-Lervåg, M., Lyster, S. A., & Hulme, C. (2012b). Phonological skills and their role in learning to read: A meta-analytic review. *Psychological Bulletin, 138*, 322-352. doi:10.1037/a0026744.

National Early Literacy Panel (NELP). (2008). *Developing early literacy: Report of the National Early Literacy Panel.* Washington, DC: National Center for Family Literacy.

Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16,* 91-121. doi:10.1080/10888438.2010.529219.

Open Science Collaboration. (2015). PSYCHOLOGY. Estimating the reproducibility of psychological science. *Science, 349,* aac4716. doi:10.1126/science.aac4716.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading, 18*, 22-37. doi:10.1080/10888438.2013.827687.

Quinn, J. M. (2016). *Predictors of reading comprehension: A model-based meta-analytic review*. (10161140 Ph.D. Dissertation). Tallahassee, FL: The Florida State University. Retrieved from https://search.proquest.com/docview/1842240600?accountid=14699.

R Core Team. (2013). R: A language and environment for statistical comuting. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/.

Rothstein, H., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta analysis: Prevention, assessment, and adjustments*. Chichester, UK: Wiley.

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*, 666-676. doi:10.1093/ije/dym018.

Schmidt, F., & Hunter, J. (2004). *Methods of Meta-analysis: Correcting error and Bias in Research findings* (2nd ed.). Thousand Oaks, CA: Sage Publishing House.

Seidenberg, M. S. (2005). Connectionist models of word reading. *Current Directions in Psychological Science, 14*, 238-242. doi:10.1111/j.0963-7214.2005.00372.x.

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. doi:10.1177/0956797611417632.

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). p-Curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science, 9,* 666-681. doi:10.1177/1745691614553988.

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75,* 417-453. doi:10.3102/00346543075003417.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine, 22,* 2113-2126. doi:10.1002/sim.1461.

The EndNote Team. (2013). *EndNote.* Philadelphia, PA: Thomson Reuters.

Tunmer, W. E., Chapman, J. W., & Prochnow, J. E. (2006). Literate cultural capital at school entry predicts later reading achievement: A seven year longitudinal study. *New Zealand Journal of Educational Studies, 41,* 183.

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., . . . Egger, M. (2014). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *International Journal of Surgery, 12,* 1500-1524. doi: 10.1016/j.ijsu.2014.07.014.

Wells, G., Shea, B., O'Connell, D., Peterson, J., Welch, V., Losos, M., & Tugwell, P. (2015). The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Ottawa Hospital Research Institute.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities, 33,* 387-407. doi:10.1177/002221940003300409.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside.

Wothke, W. (1993). Nonpositive definite matrices in structural modeling. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 256-293). Newbury Park, CA: Sage Publishing House.

Information about this review

Review authors

Lead review author

Name:	Hanne Næss Hjetland
Title:	PhD student
Affiliation:	University of Oslo, Department of Special Needs Education
Address:	P.O. Box 1140 Blindern
City, State, Province or County:	Oslo
Post code:	N-0318
Country:	Norway
Phone:	+47-22858044
Mobile:	+47-97545122
Email:	h.n.hjetland@isp.uio.no

Co-authors

Name:	Ellen Iren Brinchmann
Title:	Postdoctoral Fellow
Affiliation:	University of Oslo, Department of Special Needs Education
Address:	P.O. Box 1140 Blindern
City, State, Province or County:	Oslo
Post code:	0318
Country:	Norway

Phone:	+47-22854877
Email:	e.i.brinchmann@isp.uio.no

Name:	Ronny Scherer
Title:	Postdoctoral Fellow
Affiliation:	University of Oslo, Centre for Educational Measurement
Address:	P.O. Box 1161 Blindern
City, State, Province or County:	Oslo
Post code:	0318
Country:	Norway
Phone:	+47-22844402
Email:	ronny.scherer@cemo.uio.no

Name:	Ronny Scherer
Title:	Postdoctoral Fellow
Affiliation:	University of Oslo, Centre for Educational Measurement
Address:	P.O. Box 1161 Blindern
City, State, Province or County:	Oslo
Post code:	0318
Country:	Norway
Phone:	+47-22844402
Email:	ronny.scherer@cemo.uio.no

Name:	Monica Melby-Lervåg
Title:	Professor
Affiliation:	University of Oslo, Department of Special Needs Education
Address:	P.O. Box 1140 Blindern
City, State, Province or County:	Oslo

Post code:	0318
Country:	Norway
Phone:	+47-22858138
Email:	monica.melby-lervag@isp.uio.no

Roles and responsibilities

The members of the review team possess substantial expertise in terms of both content and methodology. All of the contributors to this review are working in the field of language and reading comprehension. Professor Monica Melby-Lervåg has extensive experience with conducting meta-analyses and has the required statistical analysis competence. The first and last authors have also completed a two-day course on meta-analysis with Michael Borenstein (October 2013), using Comprehensive Meta-Analysis version 3. All authors have knowledge of and experience with structural equation modeling, and they attended a two-day workshop on meta-analytic structural equation modeling with Associate Professor Mike Cheung (National University of Singapore) (Oslo, 6-7 October 2015). In addition, the review team has experience with electronic database retrieval and coding.

Responsibilities:

- Content: H. N. Hjetland, E. Brinchmann, & M. Melby-Lervåg
- Systematic review methods: H. N. Hjetland, E. Brinchmann, & M. Melby-Lervåg
- Statistical analysis: H. N. Hjetland, E. Brinchmann, R. Scherer (MASEM), & M. Melby-Lervåg
- Information retrieval: H. N. Hjetland, E. I. Brinchmann & M. Melby-Lervåg

Sources of support

The review team has not received extra funding to conduct this review.

Declarations of interest

The review team has no conflicts of interest.

Plans for updating the review

A new search will be conducted every other year. The first (Hjetland) and last author (Melby-Lervåg) will be responsible for updating the review.

Online supplements

List of online supplements

1. Search strategy
2. Description of measures
3. Coding procedure – quality indicators
4. Study characteristics
5. Study quality scores (coding)
6. Results of analysis of study quality
7. Results of meta-regression analyses
8. Alternative SEM approach
9. Funnel plots and trim and fill analyses

Online supplement 1: Search strategy

Database	Filters	Search strategy
Google Scholar	Filters: Limit to yr="1986 -Current" Language: English	(vocabulary OR «word knowledge» OR «language abilit*» OR «oral language» OR «linguistic comprehension») AND (reading OR «text comprehension») AND (kindergarten* OR preschool*) AND (longitudinal* OR «prospective stud*» OR prediction)
PsychINFO via Ovid	Filters: Limit to yr="1986 - Current" Language: English	1 exp Vocabulary/ or vocabulary.tw. or "word knowledge".tw. 2 exp Oral Communication/ or "oral adj2 language".tw. or "oral communication".tw or "speech communication".tw. 3 (linguistic adj2 comprehension).tw. 4 exp Verbal Comprehension/ or "verbal comprehension".tw. 5 exp Word Recognition/ or "word recognition".tw. 6 decod*.tw. 7 exp Listening Comprehension/ or "listening comprehension".tw. 8 exp Language Development/ or "language development".tw. 9 "language processing".tw. 10 exp Language Proficiency/ or "language proficiency".tw. 11 exp Phonics/ or phonics.tw. 12 (phonem* adj2 aware*).tw. 13 exp Phonological Awareness/ or (phonolog* adj2 aware*).tw. 14 "phoneme grapheme correspondence".tw. 15 exp Semantics/ or semantic*.tw. 16 (letter adj2 knowledge).tw. 17 "lexical access".tw. 18 "speech skills".tw. 19 exp Speech Perception/ or "speech perception".tw. 20 exp Naming/ or naming.tw. 21 naming task.id. 22 naming response.id. 23 exp Grammar/ or grammar.tw. 24 exp Syntax/ or syntax.tw. or syntactic*.tw. 25 exp "Morphology (Language)"/ or morpholog*.tw. or morphem*.tw.

	26 exp Nonverbal Ability/ or "non verbal intelligence".tw. or "nonverbal intelligence".tw. or "non verbal ability".tw. or "nonverbal ability".tw. or "non verbal iq".tw. or "nonverbal iq".tw. 27 exp Short Term Memory/ or "short term memory".tw. or "working memory".tw. or "verbal memory".tw. or "visual memory".tw. or "nonverbal memory".tw. 28 blending.tw. 29 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 (217825) 30 exp Reading/ or reading.tw. 31 exp Reading Comprehension/ 32 "text comprehension".tw. 33 exp Sentence Comprehension/ or "sentence comprehension".tw. 34 "passage comprehension".tw. 35 exp Reading Ability/ 36 exp Reading Skills/ 37 exp Reading Achievement/ 38 "literacy skills".tw. 39 30 or 31 or 32 or 33 or 34 or 35 or 36 or 37 or 38 40 exp Kindergartens/ or kindergarten*.tw. 41 exp Preschool Students/ or preschool*.tw. 42 "early childhood education".tw. 43 exp Primary School Students/ or "primary education".tw. or "primary school students".tw. 44 "160".ag. 45 40 or 41 or 42 or 43 or 44 46 exp Cohort Analysis/ or "cohort stud*".tw. or "cohort analysis".tw. 47 exp Longitudinal Studies/ or "longitudinal*".tw. or longitudinal study.md. 48 exp Followup Studies/ or "followup stud*".tw. or "follow up stud*".tw. or followup study.md. 49 exp Prospective Studies/ or "prospective stud*".tw. or prospective study.md. 50 exp Academic Achievement Prediction/ or exp Prediction/ or prediction.tw. 51 46 or 47 or 48 or 49 or 50

		52 29 and 39 and 45 and 51 53 limit 52 to (english language and yr="1986 -Current")
ERIC (OVID)	Filters: Limit to yr="1986 - Current" Language: English	1 exp Vocabulary/ or vocabulary.tw. or "word knowledge".tw. 2 exp Speech Communication/ or "oral adj2 language".tw. or "oral communication".tw. 3 (linguistic adj2 comprehension).tw. 4 "verbal comprehension".tw. 5 exp Word Recognition/ or "word recognition".tw. 6 exp "Decoding (Reading)"/ or decod*.tw. 7 exp Listening Comprehension/ or "listening comprehension".tw. 8 exp Language Development/ or "language development".tw. 9 exp Language Processing/ or "language processing".tw. 10 exp Language Proficiency/ or "language proficiency".tw. 11 exp Vocabulary Development/ 12 exp Vocabulary Skills/ 13 exp Phonics/ or phonics.tw. 14 exp Phonemic Awareness/ or (phonem* adj2 aware*).tw. 15 exp Phonological Awareness/ or (phonolog* adj2 aware*).tw. 16 exp Phoneme Grapheme Correspondence/ or "phoneme grapheme correspondence".tw. 17 exp Semantics/ or semantic*.tw. 18 (letter adj2 knowledge).tw. 19 "lexical access".tw.

	20 exp Speech Skills/ or "speech skills".tw. 21 "speech perception".tw. 22 exp Naming/ or naming.tw. 23 naming task.id. 24 naming response.id. 25 exp Grammar/ or grammar.tw. 26 exp Syntax/ or syntax.tw. or syntactic*.tw. 27 exp "Morphology (Language)"/ or morpholog*.tw. or morphem*.tw. 28 exp Nonverbal Ability/ or "non verbal intelligence".tw. or "nonverbal intelligence".tw. or "non verbal ability".tw. or "nonverbal ability".tw. or "non verbal iq".tw. or "nonverbal iq".tw. 29 27 exp Short Term Memory/ or "short term memory".tw. or "working memory".tw. or "verbal memory".tw. or "visual memory".tw. or "nonverbal memory".tw. 30 blending.tw. 31 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21 or 22 or 23 or 24 or 25 or 26 or 27 or 28 or 29 or 30 32 exp Reading/ or reading.tw. 33 exp Reading Comprehension/ 34 "text comprehension".tw. 35 "sentence comprehension".tw. 36 "passage comprehension".tw. 37 exp Reading Fluency/ 38 exp Reading Ability/ 39 exp Reading Skills/ 40 exp Reading Achievement/ 41 "literacy skills".tw. 42 32 or 33 or 34 or 35 or 36 or 37 or 38 or 39 or 40 or 41 43 exp Kindergarten/ or kindergarten*.tw. 44 exp Preschool Children/ or exp Preschool Education/ or preschool*.tw. 45 exp Early Childhood Education/ or "early childhood education".tw. 46 exp Primary Education/ or "primary education".tw. or "primary school students".tw. 47 kindergarten.el. 48 preschool education.el.

		49 early childhood education.el. 50 43 or 44 or 45 or 46 or 47 or 48 or 49 (91624) 51 exp Cohort Analysis/ or "cohort stud*".tw. or "cohort analysis".tw. 52 exp Longitudinal Studies/ or "longitudinal*".tw. 53 exp Followup Studies/ or "followup stud*".tw. or "follow up stud*".tw. 54 "prospective stud*".tw. 55 exp Prediction/ or prediction.tw. 56 51 or 52 or 53 or 54 or 55 57 31 and 42 and 50 and 56 58 limit 57 to (english language and yr="1986 -Current")
Web of Science	Filters: Limit to yr= 1986 - 2015 Languages: English	TS=(vocabulary OR "word knowledge" OR "oral communication" OR oral NEAR/2 language OR "speech communication" OR linguistic NEAR/2 comprehension OR "verbal comprehension" OR "word recognition" OR decod* OR "listening comprehension" OR "language development" OR "language processing" OR language proficiency" OR phonics OR phonem* NEAR/2 aware* OR phonolog* NEAR/2 aware* OR "phoneme grapheme correspondence" OR semantic* OR letter NEAR/2 knowledge OR "lexical access" OR "speech skills" OR "speech perception" OR naming OR grammar OR syntax OR syntactic* OR morpholog* OR morphem* OR "nonverbal ability" OR "non verbal ability" OR "nonverbal intelligence" OR "non verbal intelligence" OR "nonverbal iq" OR "non verbal iq" OR "short term memory" OR "working memory" OR "verbal memory" OR nonverbal memory" OR "visual memory" OR blending) AND TS=(reading OR "text comprehension" OR "sentence comprehension" OR "passage comprehension" OR "literacy skills") AND TS=(kindergarten* OR preschool* OR "early childhood education" OR "primary school students" OR "primary education") AND TS=("cohort analysis" OR "cohort stud*" OR longitudinal* OR "followup stud*" OR "follow up stud*" OR "prospective stud*" OR prediction

ProQuest Dissertations and Theses	Filters: Limit to yr="1986 - Current" Language: English	ALL((vocabulary OR "word knowledge" OR "oral communication" OR oral NEAR/2 language OR "speech communication" OR linguistic NEAR/2 comprehension OR "verbal comprehension" OR "word recognition" OR decod* OR "listening comprehension" OR "language development" OR "language processing" OR language proficiency" OR phonics OR phonem* NEAR/2 aware* OR phonolog* NEAR/2 aware* OR "phoneme grapheme correspondence" OR semantic* OR letter NEAR/2 knowledge OR "lexical access" OR "speech skills" OR "speech perception" OR naming OR grammar OR syntax OR syntactic* OR morpholog* OR morphem* OR "nonverbal ability" OR "non verbal ability" OR "nonverbal intelligence" OR "non verbal intelligence" OR "nonverbal iq" OR "non verbal iq" OR "short term memory" OR "working memory" OR "verbal memory" OR nonverbal memory" OR "visual memory" OR blending) AND (reading OR "text comprehension" OR "sentence comprehension" OR "passage comprehension" OR "literacy skills") AND (kindergarten* OR preschool* OR "early childhood education" OR "primary school students" OR "primary education") AND ("cohort analysis" OR "cohort stud*" OR longitudinal* OR "followup stud*" OR "follow up stud*" OR "prospective stud*" OR prediction))
OpenGrey.eu	Filters: Limit to yr="1986 - Current" Language: English	(vocabulary OR "word knowledge" OR "oral communication" OR oral NEAR/2 language OR "speech communication" OR linguistic NEAR/2 comprehension OR "verbal comprehension" OR "word recognition" OR decod* OR "listening comprehension" OR "language development" OR "language processing" OR language proficiency" OR phonics OR phonem* NEAR/2 aware* OR phonolog* NEAR/2 aware* OR "phoneme grapheme correspondence" OR semantic* OR letter NEAR/2 knowledge OR "lexical access" OR "speech skills" OR "speech perception" OR naming OR grammar OR syntax OR syntactic* OR morpholog* OR morphem* OR "nonverbal ability" OR "non verbal ability" OR "nonverbal intelligence" OR "non verbal intelligence" OR "nonverbal iq" OR "non verbal iq" OR "short term memory" OR "working memory" OR "verbal memory" OR nonverbal memory" OR "visual memory" OR blending) AND (reading OR "text comprehension" OR "sentence comprehension" OR "passage comprehension" OR "literacy skills") AND (kindergarten* OR preschool* OR "early childhood education" OR "primary school students" OR "primary education") AND ("cohort analysis" OR "cohort stud*" OR longitudinal* OR "followup stud*" OR "follow up stud*" OR "prospective stud*" OR prediction)

Linguistics and Language Behavior Abstracts	Filters: Limit to yr="1986 - Current" Language: English	ALL((vocabulary OR "word knowledge" OR "oral communication" OR oral NEAR/2 language OR "speech communication" OR linguistic NEAR/2 comprehension OR "verbal comprehension" OR "word recognition" OR decod* OR "listening comprehension" OR "language development" OR "language processing" OR language proficiency" OR phonics OR phonem* NEAR/2 aware* OR phonolog* NEAR/2 aware* OR "phoneme grapheme correspondence" OR semantic* OR letter NEAR/2 knowledge OR "lexical access" OR "speech skills" OR "speech perception" OR naming OR grammar OR syntax OR syntactic* OR morpholog* OR morphem* OR "nonverbal ability" OR "non verbal ability" OR "nonverbal intelligence" OR "non verbal intelligence" OR "nonverbal iq" OR "non verbal iq" OR "short term memory" OR "working memory" OR "verbal memory" OR nonverbal memory" OR "visual memory" OR blending) AND (reading OR "text comprehension" OR "sentence comprehension" OR "passage comprehension" OR "literacy skills") AND (kindergarten* OR preschool* OR "early childhood education" OR "primary school students" OR "primary education") AND ("cohort analysis" OR "cohort stud*" OR longitudinal* OR "followup stud*" OR "follow up stud*" OR "prospective stud*" OR prediction))

Online supplement 2: Description of measures

Measure	Description
Reading comprehension	"Measures of comprehension of meaning of written language passages. Typically measured with standardized test, such as the Passage Comprehension subtest of the Woodcock Reading Mastery Test" (NELP, 2008, p. 43). Both tests designed for passage comprehension and sentence comprehension will be coded. The type of test will be reported to control for the sensitivity of the measures: • Whether comprehension is measured by asking open ended/retell or multiple choice test/ cloze questions If the primary study includes several follow-ups, the last assessment will be coded.
Decoding	"Decoding words: Use of symbol-sound relations to verbalize real words or use of orthographic knowledge to verbalize sight words (e.g., 'have,' 'give,' 'knight')" (NELP, 2008, p. 42). Typically assessed with a standardized measure, such as word Identification subtest of the Woodcock Reading Mastery Test and subtest Form A –- Sight Word Efficiency (SWE) of the Test of Word Reading Efficiency (TOWRE). "Decoding non-words: Use of symbol-sound relations to verbalize pronounceable non-words (e.g., 'gleap,' 'taip'). Typically measured with a standardized measure, such as the Word attack subtest of the Woodcock Reading Mastery test" (NELP, 2008, p. 42). Decoding ability will be coded the first time it is assessed in the primary study (which can be after the predictors are assessed) and concurrently with the outcome measure. If the studies include decoding of both single word and non-word reading, both will be coded. In addition, if the primary study reports a composite score of decoding (i.e., a mix of real words and non-words), this score will be coded in its own category.
Vocabulary	Preschool vocabulary can include standardized or research-designed measures of vocabulary. Tests that tap receptive and/or expressive vocabulary and vocabulary composites will be coded. If the included studies have several assessment time points, the first time point in preschool will be coded. Vocabulary is typically assessed with a standardized test, such as the Peabody Picture Vocabulary scale (receptive).

Grammar – syntax	Grammar tests, which assess the child's knowledge about how words or other elements of sentence structure are combined to form grammatical sentences, will be coded. Tests that tap receptive and/or expressive grammar and composites will be coded. If the included studies have several assessment time points, the first time point in preschool will be coded. Grammar is typically measured with a standardized test, such as the Test for Reception of Grammar (TROG) (receptive).
Phonological awareness	"Ability to detect, manipulate or analyze components of spoken words independent of meaning. Examples include detection of common onsets between words (alliteration detection) or common rime units (rhyme detection); combining syllables, onset rimes, or phonemes to form words; deleting sounds from words; counting syllables or phonemes in words; or reversing phonemes in words. Often assessed with a measure developed by the investigator, but sometimes assessed with a standardized test, such as the Comprehensive Test of Phonological Processing" (NELP, 2008, p. 42). In the present study, tests that tap rhyme, phoneme awareness and composites will be coded. If the included studies have several assessment time points, the first time point in preschool will be coded.
Letter knowledge	"Knowledge of letter names or letter sounds, measured with recognition or naming test. Typically assessed with measure developed by investigator" (NELP, 2008, p. 42). If the included studies have several assessment time points, the first time point in preschool will be coded.
Rapid automatized naming	Rapid naming of sequentially repeating random sets of pictures of objects, objects, letters or digits. Typically measured with researcher-created measure (NELP, 2008). If the primary study includes several measures, a composite score will be calculated – one for alphanumeric RAN (letters and digits) and one for non-alphanumeric RAN (symbols and colors. Cases in which RAN ability is reported in the correlation matrix as one composite will be coded in a separate category.
Memory	Short-term memory: "Ability to remember spoken information for a short period of time. Typical tasks include digit span, sentence repetition, and non-word repetition from both investigator-created measures and standardized tests" (NELP, 2008, p. 43). Working memory: "the capacity to store information while engaging in other cognitively demanding activities" (Florit et al., 2009, p.936)". Examples of tests include sentence span tests. These tests measure the ability to store and process sentences/ numbers and non-word repetition and to recall them. Both STM and WM will be

	coded. A composite will not be computed; instead, single test scores will be used because they often are not highly correlated.
Non-verbal intelligence	"Scores from nonverbal subtests or subscales from intelligence measures, such as the Wechsler Preschool and Primary Scales of Intelligence or Stanford-Binet Intelligence Scale" (NELP, 2008, p. 43). As long as there is a non-verbal component included in the measure, it will be included (e.g., full-scale IQ)

Online supplement 3: Coding procedure – quality indicators

Risk of bias indicators	Categories	Value
Sampling	Random	0
	Convenience	1
Instrument quality	Only standardized	0
	Combination	1
	Only researcher made	2
Test reliability	Reports on all measures	0
	Reports on some measures	1
	Reports from test manual or does not report reliability	2
Floor or ceiling effect	No floor or ceiling effect	0
	Floor or ceiling effect on one or more measures or does not report the necessary statistics	1
Attrition	Reports attrition	0
	Does not report attrition (sample size at both time points)	1
Missing data	Other (better than listwise)	0
	Listwise deletion	1
Latent variables	Yes	0
	No	1
Statistical power/sample size	Above 150	0
	70-150	1
	Below 70	2

Online supplement 4: Study characteristics (in alphabetical order)

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Aarnoutse, van Leeuwe, & Verhoeven (2005)	78	Phoneme: Initial phoneme test Rhyme: Rhyming test Letter knowledge: Letter test Vocabulary: Vocabulary test Sentence memory: Sentence recall test Concurrent word recognition: One minute test Reading comprehension: Reading comprehension test	[1] *r* = .33 [2] *r* = .22 [3] *r* =.43 [4] *r* =.15 [5] *r* = .44 [6] *r* =.34 [9] *r* =.49 [11] *r* =.46 [MASEM]	Age t1: Spring semester second year of kindergarten (Estimated 70 months) Age t2: Fall semester second grade (Estimated 85 months) Reading instruction: 15 months (Estimated) Country: the Netherlands Language: Dutch SES: N/A Attrition: 67.9% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Adlof, Catts, & Lee (2010)	276	Phoneme: Syllable/Phoneme deletion Letter knowledge: Letter Identification (Woodcock Reading Mastery Test-Revised) RAN: Naming of animals (Woodcock Reading Mastery Test-Revised) Vocabulary: Picture Vocabulary + Oral Vocabulary (Subtests from Test of Language Development-2: Primary – TOLD-2) Grammar: Grammatical Understanding + Grammatical completion (Subtests from TOLD-2:P) Sentence repetition: Sentence Imitation (Subtest from TOLD-2:P) Non-verbal Intelligence: Composite of Block Design and Picture Completion (WPPSI-R) Reading comprehension: Passage Comprehension (WRMT-R), Comprehension subtest from Gray Oral Reading Test-3 (GORT-3) and passage comprehension subtest from Qualitative Reading Inventory-2 (QRI-2). (Composite made by authors of the original paper)	[1] *r* =.49 [5] *r* =.36 [7] *r* = -.51 [9] *r* =.49 [10] *r* =.55 [11] *r* =.56 [13] *r* =.48 [MASEM]	Age t1: Kindergarten (estimated 70 months) Age t2: Eight grade (estimated 166 months) Reading instruction: (estimated 108 months) Country: USA Language: English SES: Years of maternal education Attrition: 54.3% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Aram, Korat, & Hassunah-Arafat (2013)	88	Letter knowledge: Letter naming Vocabulary: PPVT (Dunn & Dunn, 1981) Reading comprehension: A translated version of Shatil and Nevos' (2007) test of reading comprehension	[5] *r* =.40 [9] *r* =.44 [MASEM]	Age t1: 68.36 months Age t2: End of first grade, one year after initial assessment (estimated 80 months) Reading instruction: (estimated 12 months) Country: Israel Language: Palestinian Arabic SES: Mother's education, Father's education, Parental profession and occupation Attrition: 1.12% Reading comprehension assessment format: multiple choice/cloze
Aram & Levin (2004)	38	Vocabulary: Definitions task Reading comprehension: Sentence comprehension + Story comprehension (Composite made by authors of the original paper.)	[9] *r* =.29 [MASEM]	Age t1: 69.59 months Age t2: Last month of second grade (estimated 93.59 months) Reading instruction: (estimated 24 months) Country: Israel Language: Hebrew SES: Parents professional qualification and current occupation Attrition: 7% Reading comprehension assessment format: open ended/retell

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Badian (1994)	118	Letter knowledge: Letters (13 upper case letters) RAN: RAN objects (changed to a negative correlation because a higher score indicated a better performance) Vocabulary: Short form Verbal IQ (WPPSI Information and Arithmetic) Sentence comprehension: WPPSI sentences Non-verbal intelligence: Draw-a-person Reading comprehension: Stanford Achievement Test (SAT), Primary 1, Form J	[5] *r* =.54 [7] *r* =-.43 [9] *r* =.37 [11] *r* =.52 [13] *r* =.26	Age t1: 60.20 months Age t2: 84.20 months Reading instruction: (estimated 18 months) Country: USA Language: English SES: Parental occupation Attrition: 22.88% Reading comprehension assessment format: multiple choice/cloze
Badian (2001)	79	Phoneme: Syllable Segmentation Rhyme: Rhyme Detection Vocabulary: Verbal IQ WPPSI – subtests: Information, Arithmetic, and Similarities. Sentence repetition: WPPSI sentences Reading comprehension: Stanford Achievement Test – Passage comprehension	[1] *r* =.46 [3] *r* =.51 [9] *r* =.60 [11] *r* =.45 [MASEM]	Age t1: 60 months Age t2: 157.2 months Reading instruction: (Estimated 96 months) Country: USA Language: English SES: Parental occupation Attrition: 17.71% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Bartl-Pokorny et al. (2013)	23	Vocabulary: PPVT + Productive Vocabulary Test Reading comprehension: Reading Comprehension Test (LGVT)	[9] *r* =.44 [MASEM]	Age t1: 55 months Age t2: 162 months Reading instruction: Estimated 108 months Country: Austria Language: Austrian – German SES: N/A Attrition: 62.9% Reading comprehension assessment format: N/A
Bianco et al. (2012)	236	Rhyme: Includes syllable parsing, rhyming and phonological discrimination Vocabulary: Test de Vocabulaire Actif et Passif (TVAP) Concurrent word recognition: Lexical score Reading comprehension: composite of sentence and text reading	[3] *r* =.50 [4] *r* =.40 [9] *r* =.17 [MASEM]	Age t1: 54 months Age t2: First grade (Estimated 108.94 months) Reading instruction: Estimated 12 months Country: France Language: French SES: Parental occupation Attrition: 33.33% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Bishop & League (2006)	79	Phoneme: Phoneme elision, blending, and sound matching from CTOPP Letter knowledge: Letter identification (lowercase) RAN: Naming of objects and colors (CTOPP) Non-word repetition: Composite of memory of digits and non-word repetition (CTOPP) Concurrent word recognition: TOWRE: sight word efficiency Reading comprehension: The Qualitative Reading Inventory-II	[1] *r* =.33 [2] *r* =.27 [5] *r* = .24 [6] *r* =.29 [7] *r* =.15 [8] *r* =.28 [12] *r* =.16 [MASEM]	Age t1: Fall kindergarten (estimated 55 months) Age t2: End of fourth grade (Estimated 108 months) Reading instruction: Estimated 60 months Country: USA Language: English SES: Federal school lunch Attrition: 23.3% Reading comprehension assessment format: open ended/retell
Blackmore & Pratt (1997)	33	Phoneme: Phoneme deletion test Vocabulary: Form M of the PPVT Grammar: Grammatical awareness: Grammatical correction task + Oral cloze Concurrent word recognition: Concept about print test, followed by the eight-word lists from the IRAS Reading comprehension: Passage A from the IRAS	[1] *r* =.31 [2] *r* =.52 [9] *r* =.04 [10] *r* =.32 [MASEM]	Age t1: 66 months Age t2: 12 months after initial testing (Estimated 78 months) Reading instruction: 12 months Country: Australia Language: English SES: N/A Attrition: 17.5% Reading comprehension assessment format: open ended/retell

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Bowey (1995)	116	Phoneme: Phoneme oddity Rhyme: Rhyme oddity Letter knowledge: Letter knowledge (uppercase and lower case) Vocabulary: PPVT Grammar: Grammatical understanding subtest of the revised Test of Oral Language Development – Primary Non-word repetition: Non-word repetition test Non-verbal intelligence: Block design subtest of the revised Wechsler Preschool and Primary Scale of Intelligence Concurrent word recognition: Word identification from Form H of the Woodcock Reading Mastery Tests + St. Lucia Word Reading comprehension: Passage comprehension from Form H of the Woodcock Reading Mastery Tests	[1] *r* =.37 [2] *r* =.38 [3] *r* =.36 [4] *r* =.32 [5] *r* = .50 [6] *r* =.58 [9] *r* =.52 [10] *r* = .39 [12] *r* =.14 [13] *r* =. 40 [MASEM]	Age t1: 5 years (Estimated 60 months) Age t2: End of first grade (Estimated 82 months) Reading instruction: 12 months Country: Australia Language: English SES: Australian Standard Classification of Occupation Scales (ASCO) Attrition: 52.85% Reading comprehension assessment format: multiple choice/cloze
Bryant, MacLean, & Bradley (1990)	66	Vocabulary: BPVS Grammar: Expressive language Reynell Developmental Language Scale Reading comprehension: France Primary Reading Test (understanding of words and simple sentences)	[9] *r* =.45 [10] *r* =.59 [MASEM]	Age t1: 40.8 months Age t2: 80.4 months Reading instruction: 18 months Country: England Language: English SES: N/A Attrition: 1.52% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Burke, Hagan-Burke, Kwok, & Parker (2009)	167	Phoneme: Initial sound fluency + Phoneme segmentation fluency from DIBELS Concurrent word recognition: oral reading fluency Reading comprehension: WRMT-R: Passage Comprehension	[1] *r* =.47 [2] *r* =.40 [MASEM]	Age t1: Midpoint of the kindergarten school year (Estimated 70 months) Age t2: Second grade (Estimated 94 months) Reading instruction: 30 months Country: USA Language: English SES: Free/reduced-priced lunches Attrition: 23.39% Reading comprehension assessment format: multiple choice/cloze
Carlson (2014) *ECLS-K dataset*	9165	Non-verbal intelligence: Fine motor skills – seven items from the Early Screening Inventory – Revised (ESI-R) Reading comprehension: Respond to multiple passages of text	[13] *r* =.30 [MASEM]	Age t1: Fall Kindergarten (Estimated 65 months) Age t2: Spring Grade 8 (Estimated 161 months) Reading instruction: 108 months Country: USA Language: English SES: N/A Attrition: 7.39% Reading comprehension assessment format: N/A

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Casalis & Louis-Alexandre (2000)	50	Phoneme: Made a composite of Phoneme deletion test and syllable deletion test Rhyme: Rhyme choice test Grammar: Composite: Sentence completion with an affixed word, Segmentation, synthesis, Feminine/word, Verb tense/word, Feminine/pseudowords, Verb/ pseudoword. Concurrent word recognition: Alouette Reading comprehension: Ecosse (Sentence reading): [1]Changed to positive correlations because the score on the reading comprehension was number of errors rather than number of correct answers.	[1] *r* =.28[1] [2] *r* =.35 [3] *r* =.30[1] [4] *r* = .27 [10] *r* =.42[1]	Age t1: 68 months Age t2: Second grade, 92 months Reading instruction: 24 months Country: France Language: French SES: N/A Attrition: 0% Reading comprehension assessment format: multiple choice/cloze
Chaney (1998)	41	Phoneme: Initial sound Rhyme: Rhyme task Vocabulary: Preschool Language Scale Revised (PLS) + PPVT-R Grammar: Mean correlation of two tests: Sentence structure + Structural awareness test Concurrent word recognition: Word Identification from Woodcock Reading Mastery Test (WRMT) Reading comprehension: Passage Comprehension from WRMT	[1] *r* =.33 [2] *r* =.31 [3] *r* =.17 [4] *r* =.20 [9] *r* = .24 [10] *r* =.29 [MASEM]	Age t1: 44 months Age t2: 87 months Reading instruction: 24 months (After completing first grade) Country: USA Language: English SES: N/A Attrition: 4.65% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	**Sample size at T2**	**Measures**	**Analysis [ ], and correlation to outcome**	**Study characteristics (N/A= Not available)**
Cronin (2013)	84	Rhyme: Rhyming and end-sound discrimination. Composite made by the authors of the original paper. RAN: Object naming. (Authors of the original paper scored items named per second. We changed it from a positive to a negative correlation.) Concurrent word recognition: WMRT-R Word identification Reading comprehension: WRMT-R Passage comprehension	[3] *r* =.57 [4] *r* =.62 [7] *r* =-.55 [8] r =-.55 [MASEM]	Age t1: 60.96 months Age t2: Spring fourth grade (Estimated 108,94 months) Reading instruction: 48 months Country: Canada Language: English SES: Income level Attrition: 35.58 % Reading comprehension assessment format: multiple choice/cloze
Cronin & Carver (1998) *Primary cohort*	95	Phoneme: Initial consonant discrimination task Rhyme: Rhyme Discrimination task RAN: Picture naming + Letter and number naming Vocabulary: PPVT Concurrent word recognition: Woodcock Word Identification Reading comprehension: Woodcock Passage Comprehension	[1] *r* =.68 [2] *r* =.70 [3] *r* =.40 [4] *r* =.32 [7] *r* =.43 [8] *r* =.45 [9] *r* =.32 [MASEM]	Age t1: 67.56 months Age t2: First grade, spring (Estimated 79.56 months) Reading instruction: 12 months Country: Canada Language: English SES: N/A Attrition: 16.66% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Cudina-Obradovic (1999)	119	Phoneme: First phoneme recognition (phoneme identity task), Word blending, Word segmentation, pseudoword blending, phoneme elision Rhyme: Onset-rhyme task Concurrent word recognition: Reading aloud a short story – The cat is fat – Accuracy – corrected and uncorrected together Reading comprehension: Reading aloud a short story – The cat is fat	[1] *r* =.37 [2] *r* =.29 [3] *r* =.38 [4] *r* =.19 [MASEM]	Age t1: 79 months Age t2: End of first grade (Estimated 91 months) Reading instruction: 12 months Country: Croatia Language: Croatian SES: N/A Attrition: 4.8% Reading comprehension assessment format: open ended/retell
Dickinson & Porche (2011)	57	Vocabulary: PPVT Reading comprehension: subtest from the California Achievement test	[9] *r* =.62 [MASEM]	Age t1: 67.3 months Age t2: 116.4 months Reading instruction: 48 months Country: USA Language: English SES: Years of maternal education Attrition: 22.97% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Durand, Loe, Yeatman, & Feldman (2013) *Association cohort*	233	Vocabulary: PPVT Non-verbal intelligence: The McCarthy Scales of Children's Abilities (MSCA) Reading comprehension: Woodcock Reading Mastery Tests Revised Norms Updated (WRMTR/NU) Passage Comprehension	[9] *r* =.53 [13] *r* =.57	Age t1: Age 3 (Estimated 36 months) Age t2: Age 9-11 (Estimated 120 months) Reading instruction: 8th grade (Estimated 108 months) Country: USA Language: English SES: N/A Attrition: 3.32% Reading comprehension assessment format: multiple choice/cloze
Evans, Shaw & Bell (2000)	67	RAN: RAN colors Non-verbal intelligence: Block Design subtest of the Wechsler Preschool and Primary Scales of Intelligence – Revised (WPPSI-R) Reading comprehension: Woodcock Reading Mastery Tests – Revised – Passage Comprehension	[7] *r* =-.39 [13] *r* =.30	Age t1: 71 months Age t2: December Grade 2, 90 months Reading instruction: 18 months Country: Canada Language: English SES: Parent Education Attrition: 14.1% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Flax, Realpe-Bonilla, Roesler, Choudhury, & Benasich (2009) *Control group*	59	Vocabulary: Auditory Comprehension – Preschool Language Scale-3 (PLS-3) Reading comprehension: Woodcock Reading Mastery – Revised: Passage Comprehension	[9] *r* =.56	Age t1: 3 years (Estimated 36 months) Age t2: 7 years (Estimated 84 months) Reading instruction: 96 months Country: USA Language: English SES: Hollingshead SES Attrition: 0% Reading comprehension assessment format: multiple choice/cloze

Fricke, et al. (2016)	78	Phoneme: Test for Phonological Awareness skills. Subtests: Syllable Segmentation Output, Sound Identification Beginning Output, Sound Identification Beginning Input, Sound Blending Output, Sound Blending Input, Sound Deletion, and Sound Deletion Input Rhyme: Test for Phonological Awareness skills. Subtests: Rhyme Production Output, Rhyme Identification Input, Onset-Rhyme-Blending Output, and Onset-Rhyme-Blending Input Letter knowledge: Letter Knowledge: uppercase and lowercase RAN: Naming objects + naming colors. (Authors of the original paper scored items named per second. We changed it from a positive to a negative correlation.) Vocabulary: Test for naming and understanding nouns and verbs Grammar: Test for Reception of Grammar – German version Non-verbal intelligence: The booklet version of Raven's Colored Progressive Matrices Concurrent word recognition: Composite made by the authors of original paper. 30 frequent words, a short text of 30 words, 24 legal pseudowords dissimilar to real words, and 30 legal pseudowords similar to real words. Reading comprehension: The paper version of the Leseverständnistest für Erstbis Sechstklässler (ELFE 1-6) (reading comprehension test for first to sixth graders)	[1] r =.25 [2] r =.23 [3] r =.22 [4] r =.30 [5] r =.45 [6] r =.18 [7] r =-.32 [8] r =-.34 [9] r = .16 [10] r =.23 [13] r =.38	Age t1: 71 months (5 y., 11 m.) Age t2: 94 months (7 y., 10 m.) Reading instruction: Grade 2, 24 months Country: Germany Language: German SES: Neighborhood characteristics, educational and employment levels Attrition: 11% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Furnes & Samuelsson (2009)	US/AU: 737 NOR/SWE: 169	Phoneme: Composite of syllable and phoneme blending, word elision, syllable and phoneme elision, sound matching, rhyme and final sounds and phoneme identity training RAN: Naming of objects and colors (CTOPP) Concurrent word recognition: TOWRE: sight word efficiency Reading comprehension: WRMT-R: Passage Comprehension	[1] US/AU: *r* =.50 NOR/SWE: *r* =.52 [2] US/AU: *r* =.45 NOR/SWE *r* =. 45 [7] US/AU: *r* =-.26 NOR/SWE: *r* = -.36 [8] US/AU: *r* =.-.31 NOR/SWE: *r* = -.38 [MASEM]	Age t1: US/AU: 58 months. NOR/SWE: 61 months Age t2: US/AU: 88 months. NOR/SWE: 92 months Reading instruction: US/AU: 24 months NOR/SWE: 12 months Country: USA, Australia, Norway, & Sweden Language: English, English, Norwegian, & Swedish SES: US/AU: Parents' mean years of education NOR/SWE: Parents' mean years of education Attrition: 0% Reading comprehension assessment format: multiple choice/cloze
González & González (2000)	136	Phoneme: Syllabic awareness: isolating syllables, syllable synthesis, syllabic segmentation, syllable deletion (Prueba de Conocimientos sobre el Lenguaje Escrito, CLE) Concurrent word recognition: Word reading – Prueba de Lectura. (Authors of the original paper scored number of errors. We changed it from a negative to a positive correlation.) Reading comprehension: The "Subtest de Comprehensión Lectora, Nivel II" from "Test de Análisis de Lectura y Escritura”	[1] *r* = .31 [2] *r* =.50 [MASEM]	Age t1: 67.2 months Age t2: Two years later. End of first grade (Estimated 91,20 months) Reading instruction: 12 months Country: Canary Islands, Spain Language: Spanish SES: N/A Attrition: 0% Reading comprehension assessment format: open ended/retell

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Guajardo & Cartwright (2016)	31	Vocabulary: Vocabulary Subscale of the Test of Auditory Comprehension of Language – III (TACL-3) Reading comprehension: The WRMT-R Passage Comprehension subtest, Form G.	[9] *r* =.49 [MASEM]	Age t1: 52.16 months Age t2: 97 months Reading instruction: 6-9 years at time 2 Country: USA Language: English SES: N/A Attrition: 0% Reading comprehension assessment format: multiple choice/cloze
Hannula, Lepola, & Lehtinen (2010)	102	Phoneme: Initial phoneme and phoneme blending RAN: Object and color naming Vocabulary: Listening comprehension Non-verbal intelligence: Raven´s colored matrices Concurrent word recognition: Decoding fluency (YTTE). (The authors of the original paper scored time per word. We changed it from a negative to a positive correlation.) Reading comprehension: Two subtests of the Standardized Reading Test for Primary School	[1] *r* =.46 [2] *r* =.44 [7] *r* =-.26 [8] *r* = -.53 [9] *r* =.43 [13] *r* =.33	Age t1: 68 months Age t2: 102 months Reading instruction: 20 months Country: Finland Language: Finnish SES: N/A Attrition: 24.46% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Hecht et al. (2000) *Subset of Wagner et al. 1994, 1997*	197	Phoneme: composite of phoneme elision, sound categorization, first sound comparison, blending onset & blending phonemes into words, and blending phonemes into non-words. Rhyme: Rime Letter knowledge: knowledge of letter names and knowledge of letter sounds RAN: naming digits, naming letters, and naming digits & letters. (Authors scored number of items per second. We changed it to a negative correlation.) Vocabulary: Stanford-Binet Vocabulary (word definition) Concurrent word recognition: Word Identification from Woodcock Reading Mastery Test (WRMT) Reading comprehension: Passage Comprehension from WRMT	[1] r =.38 [2] r =.41 [3] r =.32 [4] r =.36 [5] r = .36 [6] r =.40 [7] r =-.41 [8] r =-.33 [9] r =.47 [MASEM]	Age t1: 68.3 months (SD: 4.3 months) Age t2: 122 months (SD: 4.2 months) Reading instruction: 60 months Country: USA Language: English SES: Hollingshead and Redlich (1958) index of social class Attrition: 0% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Hulme et al. (2015)	71	Vocabulary: the Clinical Evaluation of Language Fundamentals – Preschool – Expressive vocabulary Grammar: The Clinical Evaluation of Language Fundamentals – Preschool – Sentence Structure Sentence repetition: The Preschool Repetition subtest from the Early Repetition Battery – Sentence repetition Non-word repetition: The Preschool Repetition subtest from the Early Repetition Battery – Non-word repetition Reading comprehension: Passage Reading subtest from the YARC	[9] *r* =.44 [10] *r* =.35 [11] *r* =.24 [12] *r* =.14	Age t1: 44.69 months Age t2: 104.40 months Reading instruction: 36 months Country: England Language: English SES: N/A Attrition: 0% Reading comprehension assessment format: multiple choice/cloze
Karlsdottir & Stefansson (2003)	407	Letter knowledge: Letter naming Reading comprehension: Silent Reading Comprehension test of Gjessing	[5] *r* =.45 [MASEM]	Age t1: School start in Grade 1 (Estimated 72 months) Age t2: Fifth Grade (Estimated 132 months) Reading instruction: 60 months Country: Norway Language: Norwegian SES: N/A Attrition: 0% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Katz & Ben-Yochanan (1990)	60	Vocabulary: Verbal subtests of the Wechsler Preschool and Primary Scale of Intelligence Reading comprehension: Israel Reading Comprehension Test (upper Class version)	[9] *r* =.18 [MASEM]	Age t1: Age 5, final kindergarten year (Estimated 60 months) Age t2: Age 13, End of grade 8 (Estimated 156 months) Reading instruction: 96 months Country: Israel Language: Hebrew SES: N/A Attrition: 17.81% Reading comprehension assessment format: N/A
Kirby et al. (2012)	103	Vocabulary: PPVT Non-verbal intelligence: Raven Colored Progressive Matrices Reading comprehension: Passage Comprehension subtest from Woodcock Reading Mastery Test	[9] *r* = .62 [13] *r* =.54 [MASEM]	Age t1: 67 months Age t2: 97 months Reading instruction: 36 months Country: Canada Language: English SES: N/A Attrition: 51.87% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Kozminsky & Kozminsky (1995) *Control group*	16	Phoneme: Lindamood Auditory Conceptualization Test Reading comprehension: Reading Comprehension Test	[1] *r* = .46 [MASEM]	Age t1: 62.6 months (Beginning of kindergarten) Age t2: End of third grade (Estimated 110.6 months) Reading instruction: 36 months Country: Israel Language: Hebrew SES: N/A Attrition: 54.29% Reading comprehension assessment format: multiple choice/cloze
Kurdek & Sinclair (2001)	281	Vocabulary: Kindergarten Diagnostic instrument – subtests: General Information + Verbal association + Verbal opposites + Vocabulary (word definitions) Sentence memory: Kindergarten Diagnostic instrument – subtest Auditory memory Reading comprehension: Ohio proficiency-based assessments (CTB)	[9] *r* = .30 [11] *r* =.40 [MASEM]	Age t1: Kindergarten entry (Estimated 65 months) Age t2: 134.65 months Reading instruction: 36 months Country: USA Language: English SES: N/A Attrition: 0% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Lepola, Lynch, Kiuru, Laakkonen, & Niemi (2016)	90	Letter knowledge: Letter knowledge (uppercase and lower case) Vocabulary: An adaption of the vocabulary test in the third edition of the Finnish Wechsler Intelligent Scale for Children Concurrent word recognition: 78-word narrative text adapted from a reading test battery Reading comprehension: Two narrative texts from a reading test battery	[5] *r* =.32 [6] *r* =.06 [9] *r* =.43	Age t1: 54 months Age t2: Age 9 February–March Grade 3 (Estimated 104.51 months) Reading instruction: 32 months Country: Finland Language: Finnish SES: N/A Attrition: 33% Reading comprehension assessment format: multiple choice/cloze
Lepola, Niemi, Kuikka, & Hannula (2005)	139	Phoneme: Initial Phoneme Recognition test + Writing of the alphabet test RAN: Finnish adaptation of the Rapid Automatized Naming Vocabulary: Comprehension of Instructions from the Developmental Neuropsychological Assessments Non-verbal intelligence: Raven Concurrent word recognition: A 120-word reading-aloud test (Accuracy) Reading comprehension: Two sub-tests of the Standardized Reading Test for Primary School	[1] *r* = .40 [2] *r* =.36 [7] *r* = -.26 [8] *r* = -.33 [9] *r* =.51 [13] *r* =.33	Age t1: 68 months Age t2: Spring second grade (Estimated 104 months) Reading instruction: 24 months Country: Finland Language: Finnish SES: N/A Attrition: 6.71% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Leppänen, Aunola, Niemi & Nurmi (2008)	158	Phoneme: Two subtests of the Diagnostic Test 1: Reading and Writing. Composite of Recognizing the Initial Sound of a Word subtest and Naming the Initial Sound of a Word subtest Letter knowledge: Composite of Naming Letters Test (developed by a school) and Writing Letters test Vocabulary: Sentence Test Concurrent word recognition: Oral Reading Fluency Test Reading comprehension: The Reading Comprehension Test – subtest of the Primary School Reading Test	[1] *r* =.39 [2] *r* =.33 [5] *r* =.39 [6] *r* =.45 [9] *r* =.32 [MASEM]	Age t1: 75 months Age t2: Spring grade 4 (Estimated 123) Reading instruction: 48 months Country: Finland Language: Finnish SES: Mother's education Attrition: 23.67% Reading comprehension assessment format: multiple choice/cloze
Lerkkanen, Rasku-Puttonen, Aunola & Nurmi (2004)	90	Letter knowledge: Letter Knowledge test: Diagnostic tests 1: reading and spelling Vocabulary: Listening comprehension from the Finnish School Beginners' Test battery Non-verbal intelligence: General concept ability Reading comprehension: Literal Text Comprehension + Inferential Text Comprehension: Finnish Reading Test for Primary School	[5] *r* = .09 [9] *r* =.38 [13] *r* =.15 [MASEM]	Age t1: 87 months Age t2: March year 2 (Estimated 111 months) Reading instruction: 20 months Country: Finland Language: Finnish SES: Educational level of the parents Attrition: 21.05% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Morris, Bloodgood & Perney (2003)	95	Phoneme: Beginning consonant awareness: Oral segmentation task and consonant sorting + Phoneme segmentation Letter knowledge: Alphabet recognition (upper and lower case) Reading comprehension: Passage reading task	[1] *r* =.21 [5] *r* =.50 [MASEM]	Age t1: Beginning of kindergarten (Estimated 65 months) Age t2: Second grade (Estimated 89 months) Reading instruction: 36 months Country: USA Language: English SES: Free/reduced -price lunch Attrition: 6.86% Reading comprehension assessment format: multiple choice/cloze
Muter et al., (2004)	90	Phoneme: Subtests from the Phonological Abilities Test – Phoneme completion, Beginning Phoneme Deletion & Ending Phoneme Deletion Rhyme: Subtests from the Phonological Abilities Test – Rhyme detection, Rhyme production & Rhyme Oddity Letter knowledge: Letter Knowledge subtests from the Phonological Abilities Test Vocabulary: BPVS II Concurrent word recognition: Mean correlation of Hatcher Early Word Recognition Test + Word Reading Test from British Abilities Scales II + Neale reading accuracy Reading comprehension: Neale Analysis of Reading Ability II	[1] *r* =.36 [2] *r* =.30 [3] *r* =.35 [4] *r* =.27 [5] *r* =.66 [6] *r* =.62 [9] *r* =.52 [MASEM]	Age t1: 57 months Age t2: Beginning of third grade Reading instruction: 24 months Country: England Language: English SES: Standard Occupational Classification Attrition: 8.91% Reading comprehension assessment format: open ended/retell

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Näslund & Schneider (1996)	89	Phoneme: Syllable count + Sound-in-word detect + syllable segment + phoneme oddity (Bradley and Bryant, 1985: middle sound oddity+ end sound oddity + onset-sound oddity) Rhyme: Rhyme detection + onset/rime blend Letter knowledge: Letter knowledge Non-word repetition: Pseudoword repeat Concurrent word recognition: Word decoding speed Reading comprehension: Reading comprehension test developed by first author	[1] *r* =.27 [2] *r* =-.01 [3] *r* =.19 [4] *r* =.14 [5] *r* =-.04 [6] *r* =.32 [12] *r* = -.01	Age t1: 73.2 months Age t2: Age 8 (Estimated 96 months) Reading instruction: 24 months Country: Germany Language: German SES: N/A Attrition: 33.58% Reading comprehension assessment format: multiple choice/cloze
Nevo & Breznitz (2011)	97	Sentence repetition: Sentence recall from Automated Working Assessment (AWMA) test suite Non-word repetition: Non-word recall task Non-verbal intelligence: Wechsler Intelligence Scale for Children – Block Design Reading comprehension: Silent reading of sentences + Oral paragraph reading + Silent paragraph reading + sentences reading-Elul	[11] *r* = .25 [12] *r* = .14 [13] *r* = .27	Age t1: 73 months Age t2: one year later (Estimated 85 months) Reading instruction: 12 months Country: Israel Language: Hebrew SES: N/A Attrition: 9.35% Reading comprehension assessment format: open ended/retell

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
NICHD (2005)	1137	Phoneme: Incomplete Words Subtest from the WJ-R Vocabulary: Preschool Language Scale (PLS-3) + Picture Vocabulary Subtest from WJ-R Reading comprehension: WJ-R: Passage Comprehension	[1] *r* = .39 [9] *r* = .54 [MASEM]	Age t1: 54 months Age t2: Third grade (Estimated 96 months) Reading instruction: 48 months Country: USA Language: English SES: Years of maternal education Attrition: 16.64% Reading comprehension assessment format: multiple choice/cloze
O'Neill, Pearce, & Pick (2004)	41	Vocabulary: TELD-2 Reading comprehension: Peabody Individualized Achievement Test – Revised (PIAT-R) subtest: Reading Comprehension	[9] *r* = .43	Age t1: 37.32 months Age t2: 74.4 months Reading instruction: 1-3 years variations in grade level Country: Canada Language: English SES: N/A Attrition: 24.07% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Parrila, Kirby, & McQuarrie (2004)	95	Phoneme: Sound Isolation and Blending Phonemes Letter knowledge: Letter Identification test RAN: Color naming Concurrent word recognition: Woodcock Reading Mastery Test (WMRT-R): Word Identification. Form H. Reading comprehension: WMRT-R: Passage Comprehension	[1] *r* =.40 [2] *r* =.47 [5] *r* =.48 [6] *r* =.50 [7] *r* = -.51 [8] *r* =-.52	Age t1: 66.7 months (senior kindergarten) Age t2: Third grade (Estimated 102.7 months) Reading instruction: 36 months Country: Canada Language: English SES: N/A Attrition: 40.99% Reading comprehension assessment format: multiple choice/cloze
Piasta, Petscher, & Justice (2012)	371	Letter knowledge: Letter naming Uppercase and lowercase Concurrent word recognition: Letter word identification Reading comprehension: WJ-III: passage comprehension	[5] *r* = .42 [6] *r* =.45 [MASEM]	Age t1: Spring preschool, 52 months Age t2: Spring first grade (Estimated 84 months) Reading instruction: 24 months Country: USA Language: English SES: Average yearly income/Level of maternal education Attrition: 32.67% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Pike, Swank, Taylor, Landry, & Barnes (2013)	43	Vocabulary: Auditory comprehension (Preschool language scale) Reading comprehension: Passage Comprehension Woodcock Johnson	[9] *r* =.20 [MASEM] *Correlations were sent on request by e-mail*.	Age t1: 36 months Age t2: 114 months Reading instruction: 60 months Country: USA and Canada Language: English SES: Not specified how this was measured Attrition: 0% Reading comprehension assessment format: multiple choice/cloze
Prochnow, Tunmer, & Chapman (2013)	76	Rhyme: Onset-rime segmentation + Sound matching Letter knowledge: Letter Identification subtest of the Diagnostic Survey (uppercase and lowercase letters) Vocabulary: the Peabody Picture Vocabulary Test – Form M Grammar: Oral Cloze + Word-order correction Concurrent word recognition: Reading subtest of the Wide Range Achievement Test Reading comprehension: the Comprehension subtest of the Neale Analysis of Reading Ability, Revised	[3] *r* = .63 [4] *r* =.60 [5] *r* =.53 [6] *r* =.53 [9] *r* =.64 [10] *r* =.51 *Correlations were sent on request by e-mail.*	Age t1: 61 months Age t2: 141 months Reading instruction: 84 months Country: New Zealand Language: English SES: Elley-Irving Socio-Economic Index: 2001 Census Revision Attrition: 50% Reading comprehension assessment format: open ended/retell

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Rego (1997)	48	Phoneme: Alliteration task Grammar: The Syntactic Awareness Task Sentence repetition: The Verbal Memory Task Non-verbal intelligence: The Raven's Progressive Matrices Concurrent word recognition: The Word Reading Task Reading comprehension: The Reading Comprehension Task	[1] *r* =.15 [2] *r* = .09 [10] *r* = .40 [11] *r* =.05 [13] *r* =.05	Age t1: 68 months Age t2: 80 months Reading instruction: 12 months Country: Brazil Language: Portuguese SES: N/A Attrition: 20% Reading comprehension assessment format: open ended/retell
Roth et al. (2002)	39	Phoneme: Blending and elision Vocabulary: PPVT + Oral Vocabulary subtest – TOLD-2, Boston naming test Grammar: Test of Auditory Comprehension o1f Language-Revised (TACL-R) + Formulated Sentences subtest of the Clinical Evaluation of Language Fundamentals – Revised (CELF-R) Non-verbal intelligence: Raven Colored Progressive Matrices Concurrent word recognition: WJ-R: Letter-Word Identification Reading comprehension: WJ-R: Passage Comprehension	[1] *r* =.66 [2] *r* =.78 [9] *r* =.62 [10] *r* =.65 [13] *r* =.38	Age t1: 66 months Age t2: Grade 2 (Estimated 90 months) Reading instruction: 36 months Country: USA Language: English SES: Free/reduced-priced lunches Attrition: 40.91% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Sawyer (1992)	300	Phoneme: Auditory Segmenting Ability – Test of Awareness of Language Segments – Words in sentences or sounds in words + Gates-MacGinitie Reading Tests: Readiness Skills – Subtests Auditory Discrimination and Auditory Blending Letter knowledge: Letter Name Knowledge Vocabulary: PPVT-R Grammar: Test of Auditory Comprehension Concurrent word recognition: Slosson Oral Reading Test & Iowa Tests of Basic Skills Reading comprehension: Iowa Tests of Basic Skills – Subtest Reading	[1] *r* =.28 [2] *r* = .27 [5] *r* = .38 [6] *r* =.47 [9] *r* =.40 [10] *r* =.15 [MASEM]	Age t1: July prior to kindergarten (Estimated 64 months) Age t2: May in Third grade (Estimated 98 months) Reading instruction: 48 months Country: USA Language: English SES: N/A Attrition: 0% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Schatschneider, Fletcher, Francis, Carlson, & Foorman (2004)	189	Phoneme: Blending onset and rime, Blending phonemes into words, Blending phonemes into non-words, First sound comparison, Phoneme elision, Phoneme segmentation, Sound categorization Letter knowledge: Letter names and sounds RAN: naming object + naming letter. (Authors scored items per second. We changed it to negative correlation.) Vocabulary: PPVT Grammar: Sentence Structure subtest from CELF-R Sentence repetition: The Recalling Sentences subtest of the CELF-R Non-verbal intelligence: The Recognition-Discrimination test Concurrent word recognition: TOWRE (SWE) & Letter word Identification (WJ-R) Reading comprehension: WJ-R passage comprehension	[1] *r* =.36 [2] *r* = .41 [5] *r* = .34 [6] *r* =.44 [7] *r* =-.34 [8] *r* =-.45 [9] *r* =.23 [10] *r* =.21 [11] *r* =.12 [13] *r* =.28 [MASEM]	Age t1: October kindergarten (Estimated 66 months) Age t2: End of second grade (Estimated 84 months) Reading instruction: 36 months Country: USA Language: English SES: Hollingshead scale Attrition: 50.78% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Sears & Keogh (1993)	104	Phoneme: The Revised Slingerland Pre-Reading Screening procedure – test 12: phonological awareness Letter knowledge: The Revised Slingerland Pre-Reading Screening procedure – test 6 – letter name knowledge Vocabulary: The Revised Slingerland Pre-Reading Screening procedure – tests 5 and 8 – listening Comprehension Concurrent word recognition: The Stanford Reading achievement Test – word study Reading comprehension: The Stanford Reading achievement Test – interpret pictures and recall both explicit and implicit meaning in passages	[1] r =.35 [2] r =.19 [5] r = .39 [6] r = .22 [9] r = .30	Age t1: Kindergarten (Estimated 70 months) Age t2: Fifth grade (Estimated 130 months) Reading instruction: 66 months Country: USA Language: English SES: School attended at Kindergarten Attrition: 75.98% Reading comprehension assessment format: multiple choice/cloze
Sénéchal (2006)	65	Phoneme: Phoneme deletion Letter knowledge: Composite of Letter-name knowledge and letter-sound knowledge Vocabulary: French-Canadian version of PPVT-R Reading comprehension: Reading Comprehension subtest from the Test de Rendement pour Francophones	[1] r =.46 [5] r =.49 [9] r =.67 [MASEM]	Age t1: 72 months (SD = 6 months) Age t2: 120 months (SD = 3 months) Reading instruction: 48 months Country: Canada Language: French SES: Years of parental education Attrition: 26% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Sénéchal & LeFevre (2002)	66	Phoneme: Sound categorization task of the Stanford Early School Achievement Test (SESAT; Psychological Corporation, 1989) Letter knowledge: Alphabet knowledge – name 15 letters Vocabulary: PPVT-R (Dunn & Dunn, 1981) Non-verbal intelligence: Analytic intelligence – animal house subtest of the Wechsler Preschool and Primary Scale of Intelligence – Revised (Wechsler, 1989) Reading comprehension: Gates-MacGinitie Reading Test (Level C, Form 3; MacGinitie & MacGinitie, 1991) – Vocabulary and comprehension subtest	[1] *r* =.73 [5] *r* = .39 [9] *r* =.53 [13] *r* = -.05 [MASEM]	Age t1: 4-5 years (Estimated 78 months) Age t2: grade 3 (Estimated 102 months) Reading instruction: 36 months Country: Canada Language: English SES: N/A Attrition: 40% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Shatil & Share (2003)	313	Phoneme: Initial consonant isolation, initial consonant match, phonemic blending, phonological word production Rhyme: Rhyme detection and production Letter knowledge: Letter naming RAN: Serial naming picture and colors Vocabulary: PPVT Grammar: Syntactic awareness: sentence correction + sentence completion Non-word repetition: Pseudoword repetition Non-verbal intelligence: Raven's colored matrices – sets A and B Concurrent word recognition: Oral word recognition Reading comprehension: Silent reading comprehension. Composite of Reading vocabulary + Paragraph comprehension, expository + Paragraph comprehension, and Narrative + Comprehension monitoring. (Composite made by authors of the original paper)	[1] *r* =.31 [2] *r* =.19 [3] *r* =.29 [4] *r* = .19 [5] *r* =.45 [6] *r* =.36 [7] *r* =-.21 [8] *r* = -.27 [9] *r* =.37 [10] *r* =.52 [12] *r* = .25 [13] *r* =.37	Age t1: 72 months (Kindergarten) Age t2: end of grade 1 (Estimated 84 months) Reading instruction: 12 months Country: Israel Language: Hebrew SES: Home literacy: Hebrew versions of the Author Recognition Test and the Magazine Recognition Test (Stanovich & West, 1989) + mothers rated the frequency of story reading and literacy activities at home Attrition: 10.3% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Silva & Cain (2015)	69	Vocabulary: British Picture Vocabulary Scale – II Grammar: The Test for Reception of Grammar (2nd ed.) Non-verbal intelligence: The Matrix Reasoning subtest from the Wechsler Preschool and Primary Scale of Intelligence (3rd ed.) Reading comprehension: The Neale Analysis of Reading Ability – II	[9] *r* =.47 [10] *r* =.53 [13] *r* =.40	Age t1: One-half is 62 months. The other half is 74 months; *M* = 68 months Age t2: One year after initial assessment (Estimated 80 months) Reading instruction: 12 months Country: England Language: English SES: Parental education Attrition: 15.85% Reading comprehension assessment format: open ended/retell
Stevenson & Newman (1986)	105	Letter knowledge: Naming letters (WRAT) Vocabulary: PPVT Non-verbal intelligence: Draw a person Test Reading comprehension: Portions of the Gates-MacGinitie Reading Comprehension Test	[5] *r* =.52 [9] *r* = .26 [13] *r* =.37	Age t1: 64.8 months (summer before kindergarten entry) Age t2: Several months after they entered the tenth grade (Estimated 184.8 months) Reading instruction: 120 months Country: USA Language: English SES: Parental education Attrition: 58.82% Reading comprehension assessment format: multiple choice/cloze

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Taylor, Anthony, Aghara, Smith, & Landry (2008)	83	Non-verbal intelligence: Stanford-Binet Intelligence Scale (4th ed.): Full test administered Reading Comprehension: Woodcock-Johnson Revised Test of Cognitive Ability: Passage Comprehension	[13] *r* = .61 [MASEM]	Age: 48 months Age t2: 96 months Reading instruction: 48 months Country: USA Language: English SES: Maternal education + The Hollingshead (1975) Four Factor Index of Social Status Attrition: 25% Reading comprehension assessment format: multiple choice/cloze
Tunmer, Chapman, & Prochnow (2006) *Same sample as Prochnow, Tunmer, & Chapman (2013)*	76	Non-word repetition: Non-word repetition task Reading comprehension: Comprehension subtest of the Neale Analysis of Reading Ability, Revised	[12] *r* =.14	Age t1: 61 months Age t2: 141 months Reading instruction: 84 months Country: New Zealand Language: English SES: Elley-Irving Socio-Economic Index: 2001 Census Revisions (Elley & Irving, 2003) Attrition: 50% Reading comprehension assessment format: open ended/retell

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Tunmer, Herriman, & Nesdale (1988)	92	Phoneme: Phonological awareness Test Letter knowledge: Letter identification test Vocabulary: PPVT Grammar: Pragmatic awareness test (a modified version of one devised for an earlier study) + Oral correction task Non-verbal intelligence: Concrete operativity test Reading comprehension: Reading Comprehension subtest from IRAS	[1] *r* =.34 [5] *r* =.52 [9] *r* =.16 [10] *r* =.31 [13] *r* =.33 [MASEM]	Age t1: 68 months Age t2: End of second grade (Estimated 92 months) Reading instruction: 24 months Country: Australia Language: English SES: N/A Attrition: 22.03% Reading comprehension assessment format: open ended/retell
Uhry (2002)	86	Phoneme: The test of Auditory Analysis Skills (TAAS) RAN: The Rapid Automized Naming Test – colors, numbers, pictured objects, and letters (Authors scored items per second. Changed to negative correlation.) Vocabulary: PPVT Concurrent word recognition: The Word Identification subtest of Reading Mastery Test (WRMT) + Word-Reading Accuracy in text Reading comprehension: Oral comprehension (oral reading of passages) + Silent comprehension	[1] *r* = .57 [2] *r* = .50 [7] *r* = -.39 [8] *r* =-.37 [9] *r* =.49 [MASEM]	Age t1: 70.4 months Age t2: 92.36 months Reading instruction: 36 months Country: USA Language: English SES: N/A Attrition: 21.1% Reading comprehension assessment format: open ended/retell

Study (in alphabetical order)	Sample size at T2	Measures	Analysis [ ], and correlation to outcome	Study characteristics (N/A= Not available)
Wolter, Self, & Apel (2011)	19	Phoneme: Phonological awareness test and the Rosner´s auditory analysis test RAN: Naming animals Vocabulary: Vocabulary subtest from TACL-3 Concurrent word recognition: WMRT-R: Word Identification Reading comprehension: WRMT-R: Passage Comprehension	[1] *r* = -.05 [2] *r* =.20 [7] *r* = -.38 [8] *r* = -.30 [9] *r* = -.13 [MASEM]	Age t1: Second semester kindergarten (Estimated 72 months) Age t2: 123 months Reading instruction: 54 months Country: USA Language: English SES: N/A Attrition: 0% Reading comprehension assessment format: multiple choice/cloze

Note: additional correlations between the different predictors and between word recognition and reading comprehension that are included in the correlation matrices used in the MASEM are not included in this table.

Online supplement 5: Study quality scores (coding)

Study	Sampling	Selection	Instrument quality	Test reliability	Floor or ceiling effect	Attrition	Missing data	Latent variables	Statistical power/ sample size	Total score
Aarnoutse et al., 2005	0	0	1	0	1	0	1	1	1	5
Adlof et al., 2010	0	0	1	2	1	0	1	1	0	6
Aram et al., 2013	1	1	1	0	1	0	1	1	1	7
Aram & Levin, 2004	1	1	2	1	0	0	1	1	2	9
Badian, 1994	1	1	1	2	0	0	1	1	1	8
Badian, 2001	1	1	1	2	0	0	1	1	1	8
Bartl-Pokorny et al., 2013	1	1	0	2	1	0	1	1	2	9
Bianco et al., 2013	1	1	1	0	0	0	1	1	0	5
Bishop & League, 2006	1	0	1	0	0	0	1	1	1	5
Blackmore & Pratt, 1997	1	1	1	1	1	0	1	1	2	9
Bowey, 1995	1	1	1	2	1	0	1	1	1	9
Bryant el al., 1990	1	1	1	2	0	0	1	1	2	9
Burke et al., 2009	1	0	0	2	0	0	0	1	0	4
Carlson, 2014	1	0	1	1	0	0	0	1	0	4
Casalis & Louis Alexandre, 2000	1	1	1	2	0	1	1	1	2	10
Chaney, 1998	1	1	1	2	0	0	1	1	2	9

Cronin, 2013	1	0	1	1	0	0	1	1	1	6
Cronin & Carver, 1998	1	0	1	2	0	0	1	1	1	7
Cudina-Obradovic, 1999	1	0	2	2	1	0	1	1	1	9
Dickinson & Porche, 2011	1	1	0	2	0	0	1	1	2	8
Durand et al., 2013	1	0	0	2	0	0	1	1	0	5
Evans et al., 2000	1	1	1	2	0	0	1	1	2	8
Flax et al., 2009	1	1	0	0	0	1	1	1	2	7
Fricke et al., 2016	1	1	1	0	1	1	0	1	1	7
Furnes & Samuelsson, 2009 (US/AU)	1	1	1	0	0	1	1	1	0	6
Furnes & Samuelsson, 2009 (NOR/SWE)	1	1	1	0	0	1	1	1	0	6
González & González, 2000	1	0	1	2	0	1	1	1	1	8
Guarjardo & Cartwright, 2016	1	0	1	2	1	0	1	1	2	9
Hannula et al., 2010	1	1	1	1	1	0	1	1	1	8
Hecht et al., 2000	1	0	1	0	1	1	1	0	0	5
Hulme et al., 2015	1	1	0	1	0	0	0	0	1	4
Karlsdottir & Stefansson, 2003	1	0	1	2	1	1	1	1	0	8
Katz & Ben-Yochanan, 1990	1	0	1	2	1	0	1	1	2	9
Kirby et al., 2012	1	0	0	2	0	1	1	1	1	7

Kozminsky, & Kozminsky, 1995	1	0	1	2	0	0	0	1	2	7
Kurdek & Sinclair, 2001	1	0	0	0	0	1	1	1	0	4
Lepola et al., 2016	1	0	1	0	1	0	0	0	1	4
Lepola et al., 2005	1	1	1	1	0	0	1	1	1	7
Leppänen et al., 2008	1	0	1	0	1	0	0	1	0	4
Lerkkanen et al., 2004	1	0	1	2	1	0	0	1	1	7
Morris et al., 2003	1	0	1	2	1	0	1	1	1	8
Muter et al., 2004	1	1	0	0	1	0	1	1	1	6
Näslund & Schneider, 1996	1	0	2	1	1	0	1	1	1	8
Nevo & Breznitz, 2011	1	1	1	1	1	0	1	1	1	8
NICHD, 2005	0	0	0	1	0	0	0	1	0	2
O'Neill et al., 2004	1	0	0	2	0	0	1	1	2	7
Parrila et al., 2004	1	0	1	2	1	0	1	1	1	8
Piasta et al., 2012	0	1	0	0	1	0	1	1	0	4
Pike et al.,	1	1	0	2	0	1	1	1	2	9
Prochnow et al., 2013 (Tunmer et al., 2006 *Same sample)*	1	0	1	0	0	0	1	1	1	5
Rego, 1997	1	0	1	2	1	0	1	1	2	9
Roth et al., 2002	1	1	1	2	0	1	1	1	2	10
Sawyer, 1992	1	0	1	2	1	1	1	0	0	7

Schatschneider et al., 2004	0	0	1	2	0	0	1	1	0	5
Sears & Keogh, 1993	1	0	0	2	1	0	1	1	1	7
Sénéchal, 2006	1	0	1	0	1	0	1	1	2	7
Sénéchal & LeFevre, 2002	1	1	0	0	1	0	1	1	2	7
Shatil & Share, 2003	1	0	1	1	0	0	1	1	0	5
Silva & Cain, 2015	1	1	0	0	0	0	1	1	2	6
Stevenson & Newman, 1986	1	0	1	2	1	0	1	1	1	8
Taylor et al., 2008	1	1	0	2	0	1	1	1	1	8
Tunmer et al., 1988	1	1	1	2	0	0	1	1	1	8
Uhry, 2002	1	0	1	2	0	0	1	1	1	7
Wolter et al., 2011	1	1	1	0	0	1	1	1	2	8

Online supplement 6: Results of analysis of study quality

Analysis	Number of studies *K*	Q_M	Covariate	Individual coefficients	Q_R	R-sq
1. Phoneme awareness – reading comprehension	36	Q[1] = 0.20; *p* = .657	Study quality	*β* = -.0064; *p* = .657	Q[34] = 99.05; *p* =.000	0%
3. Rhyme awareness – reading comprehension	15	Q[1] = 1.83, *p* = .176	Study quality	*β* = -.0316; *p* = .176	Q[13] = 31.34; *p* =.003	0%
5. Letter knowledge – reading comprehension	26	Q[1] = 2.32; *p* = .127	Study quality	*β* = .0225; *p* = .127	Q[24] = 40.63; *p* =.018	6.65%
7. RAN – reading comprehension	17	Q[1] = 2.78.; *p* = .095	Study quality	*β* =-.0554; *p* =.095	Q[15] = 48.39; *p* = .000	11.59%
9. Vocabulary – reading comprehension	45	Q[1] = 0.39, *p* = .532	Study quality	*β* =-.0094; *p* = .532	Q[43] = 143.54; *p* =.000	0 %
10. Grammar – reading comprehension	16	Q[1] = 0.08; *p* =.777	Study quality	*β*= .0081; *p* =.777	*Q*[14] = 63.49; *p* =.000	0%
13. Non-verbal intelligence – reading comprehension	21	Q[1] = 0.68, *p* =.409	Study quality	*β*= -.0210; *p* =.409	Q[19] = 73.38; *p* =.000	0%

Online supplement 7: Results of meta-regression analyses

Analysis	Number of studies K	Q_M	Covariates	Individual coefficients	Q_R
1. Phoneme awareness – reading comprehension	36	Q[3] = 3.44; $p = .329$	Age at initial assessment Age at reading comprehension assessment Months of reading instruction	$\beta = .0035$; $p =.283$ $\beta =-.0043$; $p =.160$ $\beta =.0047$; $p = .080$	Q[32] = 94.05; $p <.001$
2. Phoneme awareness – word recognition	28	Q[3] = 6.30; $p = .098$	Age at initial assessment Age at reading comprehension assessment Months of reading instruction	$\beta = -.0002$; $p = .969$ $\beta = -.0086$; $p = .021$ $\beta =.0065$; $p = .038$	Q[24] = 83.40; $p <.001$
3. Rhyme awareness – reading comprehension	15	Q[2] = 7.53; $p = .023$	Age at initial assessment Age at reading comprehension assessment	$\beta = -.0040$; $p = .310$ $\beta = .0036$; $p = .020$	Q[12] = 19.74; $p =.072$
4. Rhyme awareness – word recognition	14	Q[2] = 18.53; $p <.001$	Age at initial assessment Age at reading comprehension assessment	$\beta = -.0065$; $p = .062$ $\beta = .0064$; $p < .001$	Q[11] = 14.48; $p =.201$
5. Letter knowledge – reading comprehension	26	Q[3] = 3.31; $p = .346$	Age at initial assessment Age at reading comprehension assessment Months of reading instruction	$\beta = -.0049$; $p = .163$ $\beta = -.0005$; $p = .890$ $\beta = .0003$; $p = .919$	Q[22] = 40.12; $p =.011$
6. Letter knowledge – word recognition	16	Q[2] = 1.16; $p =.560$	Age at initial assessment Age at reading comprehension assessment	$\beta = -.0061$; $p = .316$ $\beta = -.0004$; $p = .882$	Q[13] = 58.35, $p <.001$

Analysis	Number of studies *K*	Q_M	Covariates	Individual coefficients	Q_R
7. RAN – reading comprehension	17	Q[2] = 4.74; *p* = .094	Age at initial assessment Age at reading comprehension assessment	*β* =.0119; *p* =.154 *β* = -.0033; *p* = .052	Q[14] = 38.50; *p* <.001
8. RAN – word recognition	14	Q[2] = 2.09, *p* =.351	Age at initial assessment Age at reading comprehension assessment	*β* =.0157; *p* =.198 *β* =-.0033; *p* =.443	Q[11] = 50.12; *p* <.001
		Q[2] = 2.20, *p* = .333	Number of months between the two assessments Number of months with formal reading instruction	*β* =-.0025; *p* = .567 *β* = -.0034; *p* = .371	Q[11] = 46.25; *p* <.001
9. Vocabulary – reading comprehension	40	Q[4] = 4.53, *p* = .339	Age at initial assessment Age at reading comprehension assessment Months of reading instruction Type of reading comprehension assessment	*β* =-.0006; *p* = .833 *β* =-.0024; *p* = .192 *β* = .0030; *p* = .060 *β* =-.0349; *p* = .676	Q[35] = 112.49; *p* <.001
10. Grammar – reading comprehension	16	Q[2] = 0.36; *p* =.837	Age at initial assessment Age at reading comprehension assessment	*β*= -.0013; *p* =.827 *β* =.0013; *p* =.561	*Q*[13] = 60.01; *p* <.001
11. Verbal short-term memory – reading comprehension	9	Q[1] = 4.14, *p* = .042	Age at reading comprehension assessment	*β* = .0034; *p* =.042	*Q*[7] = 22.70; *p* =.002
13. Non-verbal intelligence	21	Q[2] = 14.91, p < .001	Age at initial assessment Age at reading comprehension assessment	*β*= -.0105; *p* =.000 *β* =.0008; *p* =.374	Q[18] = 43.45; *p* =.001

MASEM approach in M*plus*

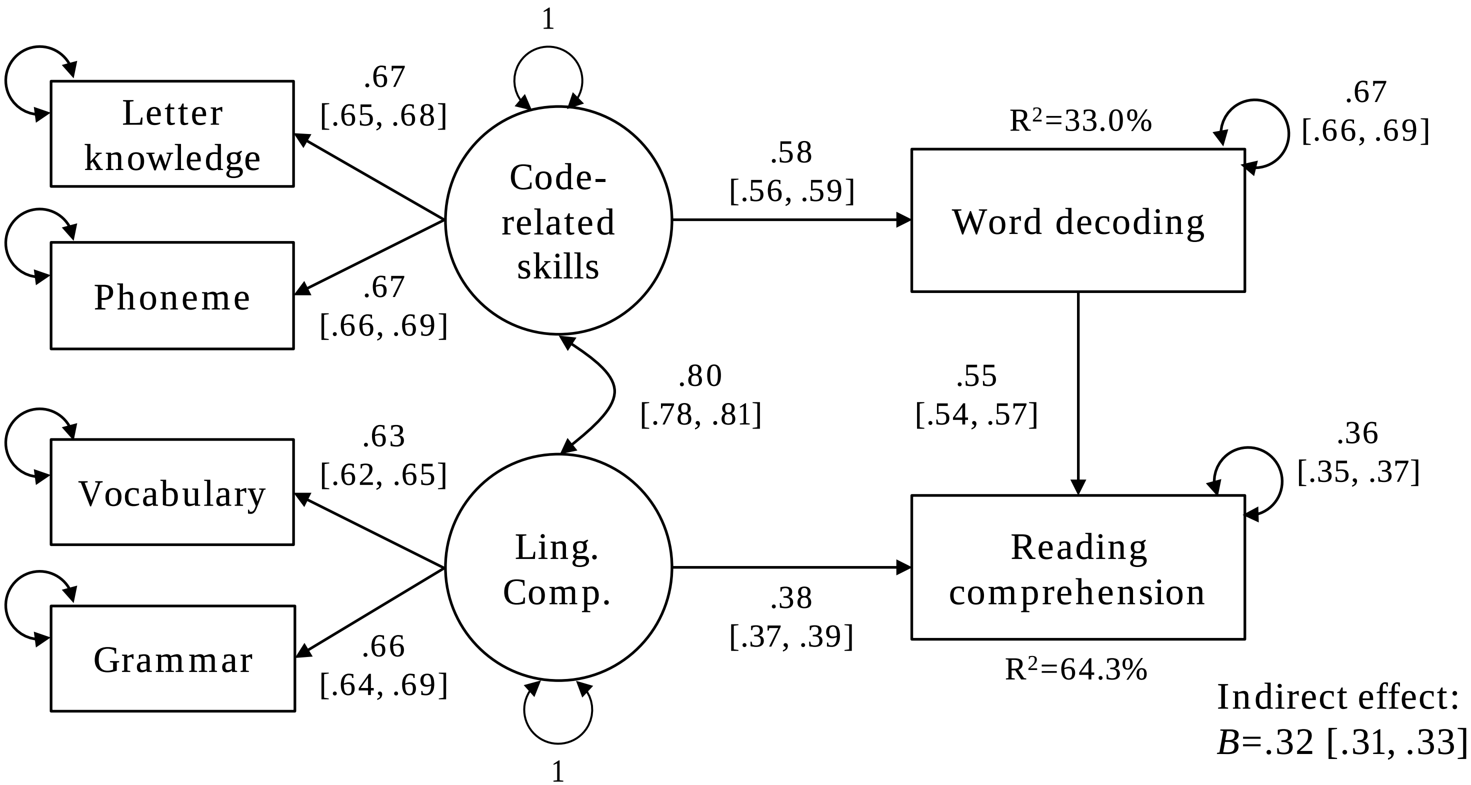
Combined correlation matrix

	PHONEME	LK	VOC	GRA	WDEC	RC
PHONEME	1.0000					
LK	0.4550	1.0000				
VOC	0.3188	0.3200	1.0000			
GRA	0.3896	0.3321	0.4151	1.0000		
WDEC	0.3726	0.3873	0.3008	0.3116	1.0000	
RC	0.4006	0.4230	0.4216	0.4058	0.7291	1.0000

Number of available correlations

	PHONEME	LK	VOC	GRA	WDEC	RC
PHONEME						
LK	15					
VOC	21	14				
GRA	8	6	10			
WDEC	28	17	30	12		
RC	36	26	45	16	32	

The simple MASEM approach in M*plus*



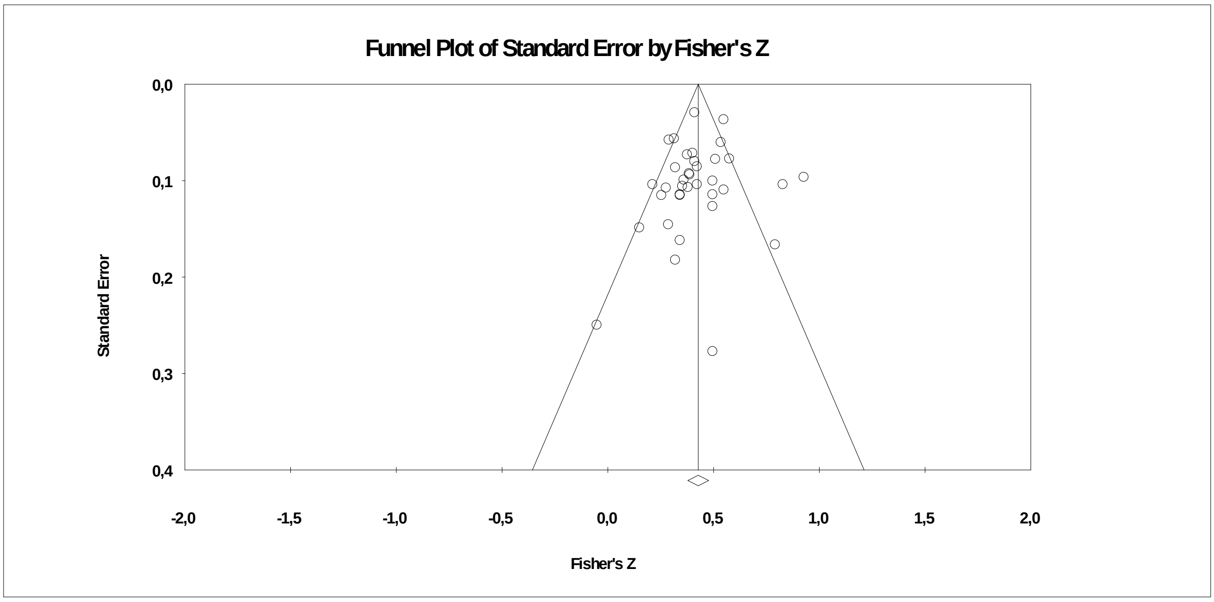
Model fit (N=17981, 64 studies):

Chi.sq[7]=183.7, p<.001, CFI=0.995, RMSEA=0.037, 90%-CI RMSEA=[0.033, 0.042], SRMR=0.013

Online supplement 9: Funnel plots and trim and fill analyses

Phoneme – reading comprehension

Funnel plot:



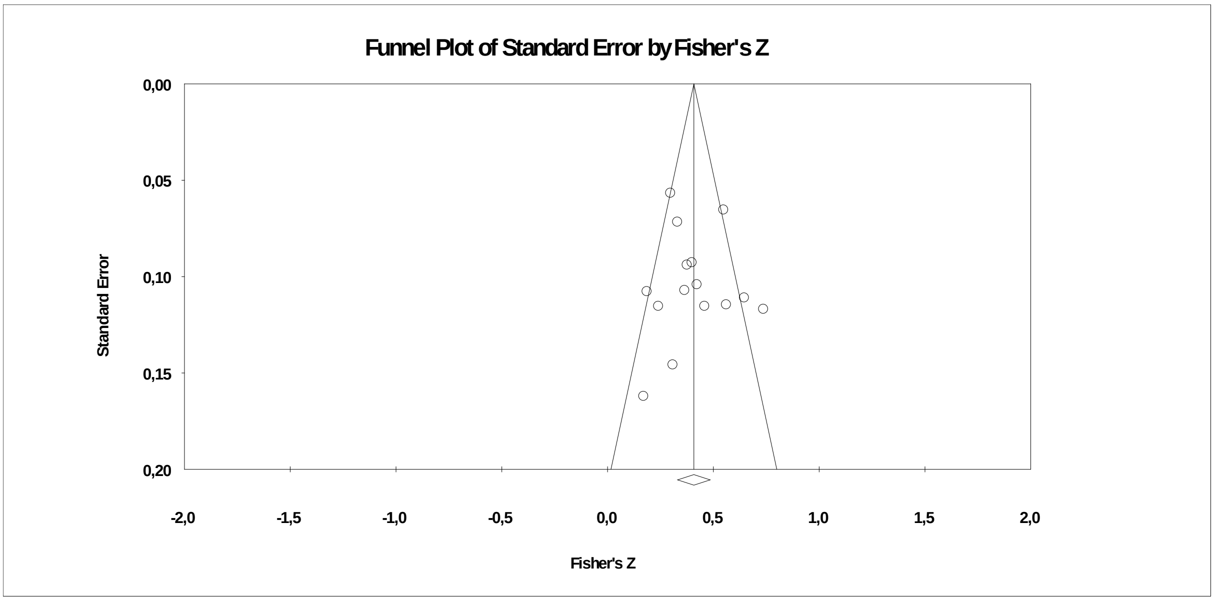
Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0,40996	0,38800	0,43146	0,40326	0,36132	0,44357	99,06462
Adjusted values	9	0,43930	0,41966	0,45852	0,44235	0,40184	0,48113	150,33122

Note: Adjusted values to the right of the mean (zero-adjusted values to the left of the mean).

Rhyme – reading comprehension

Funnel plot:



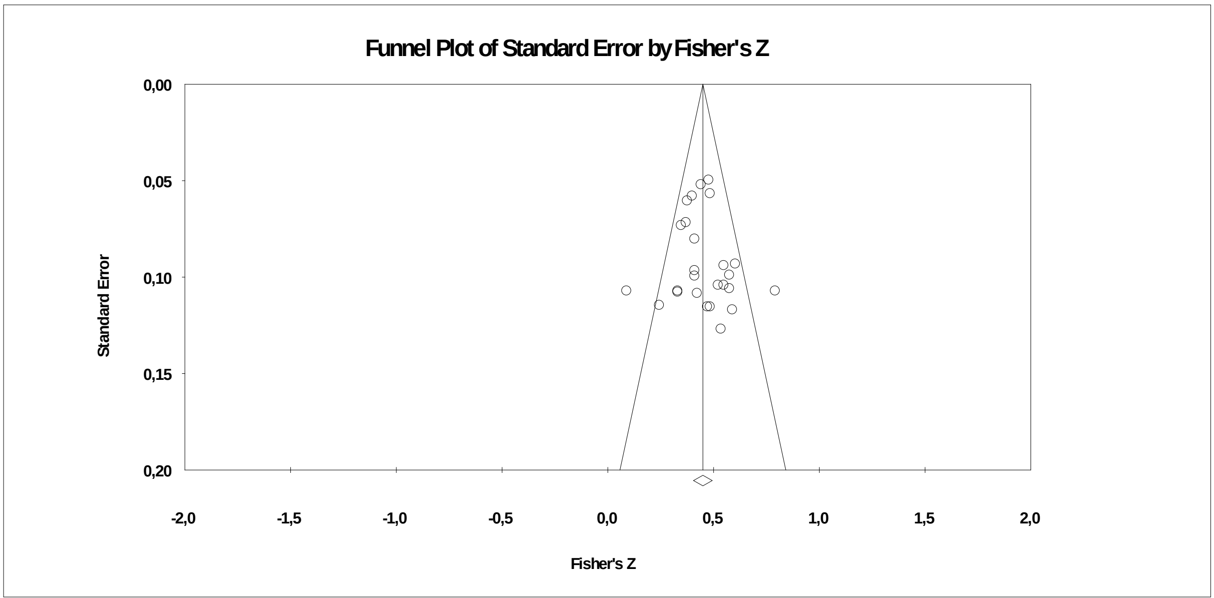
Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0,38253	0,34118	0,42240	0,38661	0,31934	0,45000	33,22361
Adjusted values	1	0,38739	0,34667	0,42665	0,39496	0,32853	0,45750	35,74119

Note: Adjusted value to the right of the mean (zero-adjusted values to the left of the mean).

Letter knowledge – reading comprehension

Funnel plot:



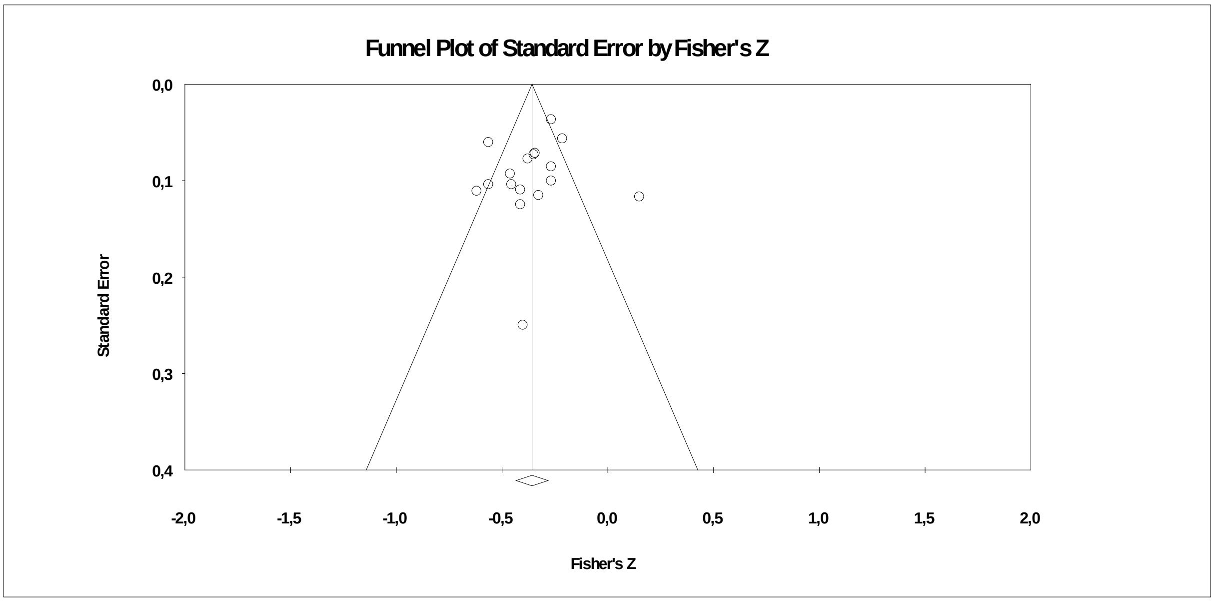
Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0,41871	0,39211	0,44461	0,42139	0,38425	0,45716	44,00412
Adjusted values	3	0,40466	0,37864	0,43004	0,40313	0,36214	0,44256	62,29775

Note: Adjusted values to the left of the mean (zero-adjusted values to the right of the mean).

RAN-reading comprehension

Funnel plot:



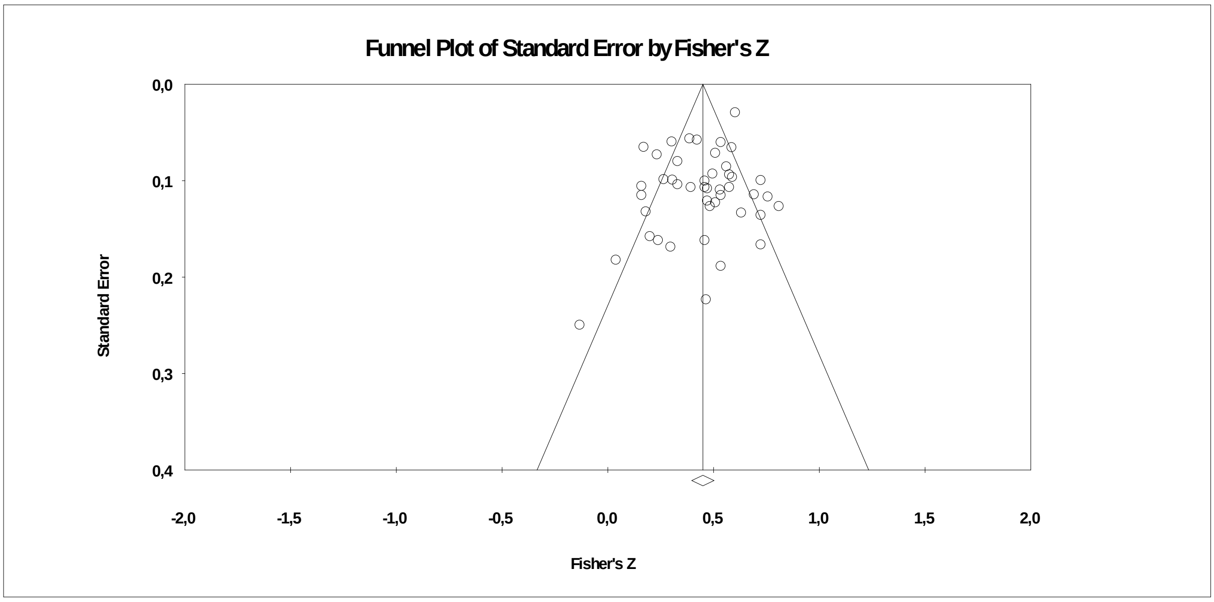
Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		-0,32749	-0,36021	-0,29397	-0,34380	-0,40871	-0,27542	56,18539
Adjusted values	6	-0,26988	-0,30029	-0,23892	-0,27006	-0,34243	-0,19451	113,79947

Note: Adjusted values to the right of the mean (zero-adjusted values to the left of the mean).

Vocabulary – reading comprehension

Funnel plot:



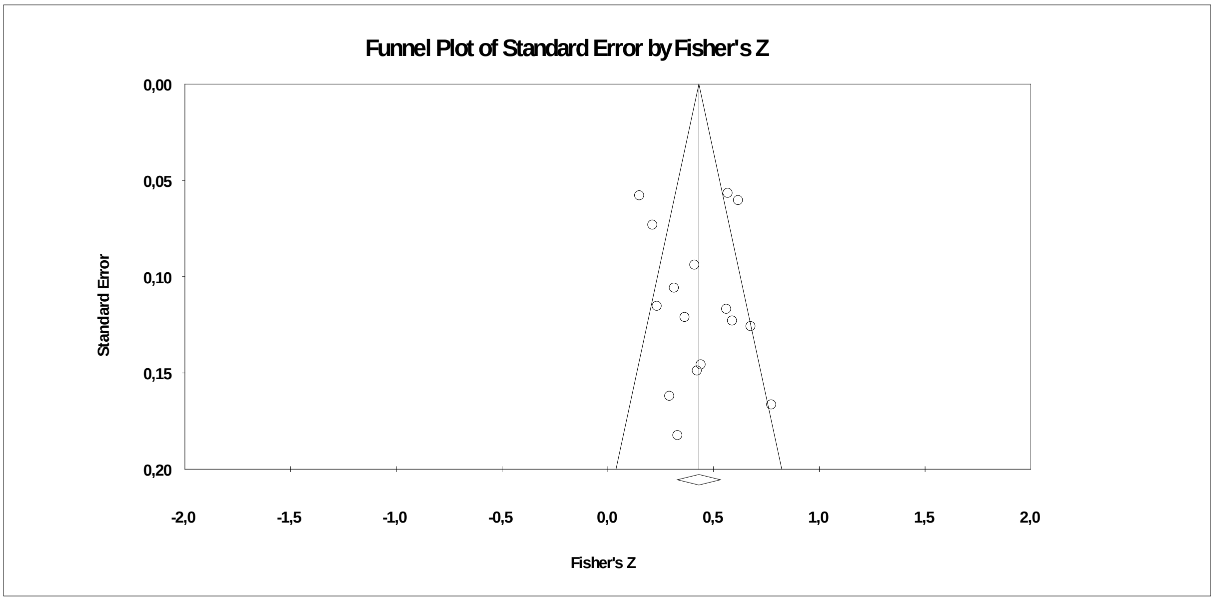
Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0,43727	0,41619	0,45789	0,42142	0,37725	0,46368	153,13103
Adjusted values	0	0,43727	0,41619	0,45789	0,42142	0,37725	0,46368	153,13103

Note: No adjusted values to either side of the mean.

Grammar – reading comprehension

Funnel plot:



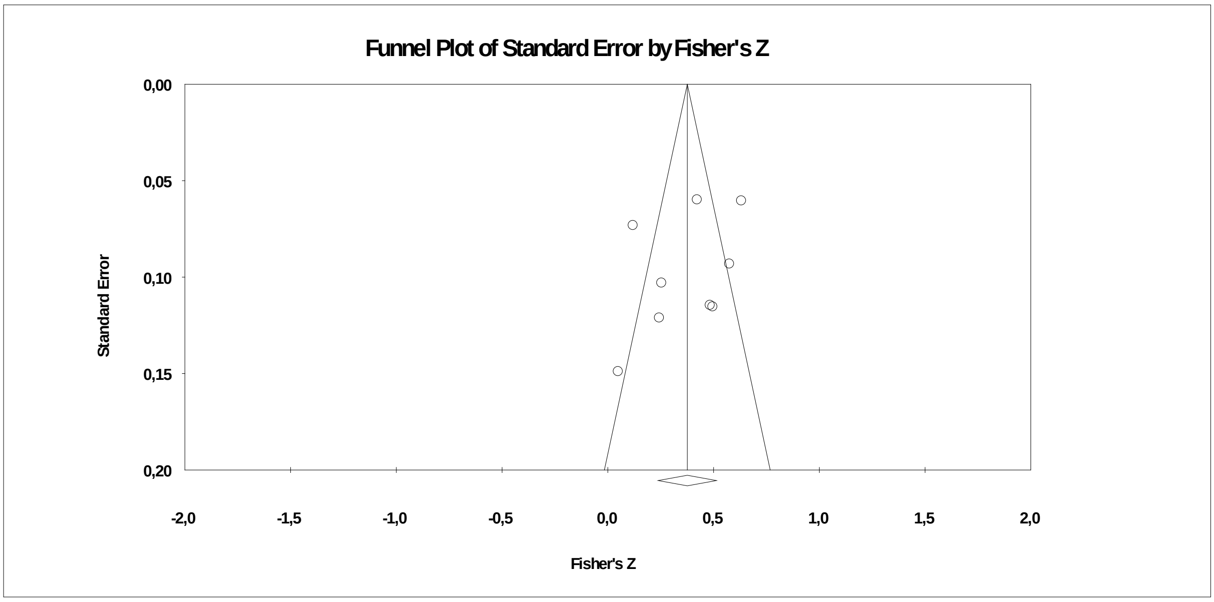
Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0,39630	0,35676	0,43442	0,40573	0,31700	0,48741	63,87184
Adjusted values	0	0,39630	0,35676	0,43442	0,40573	0,31700	0,48741	63,87184

Note: No adjusted values to either side of the mean.

Sentence memory – reading comprehension

Funnel plot:



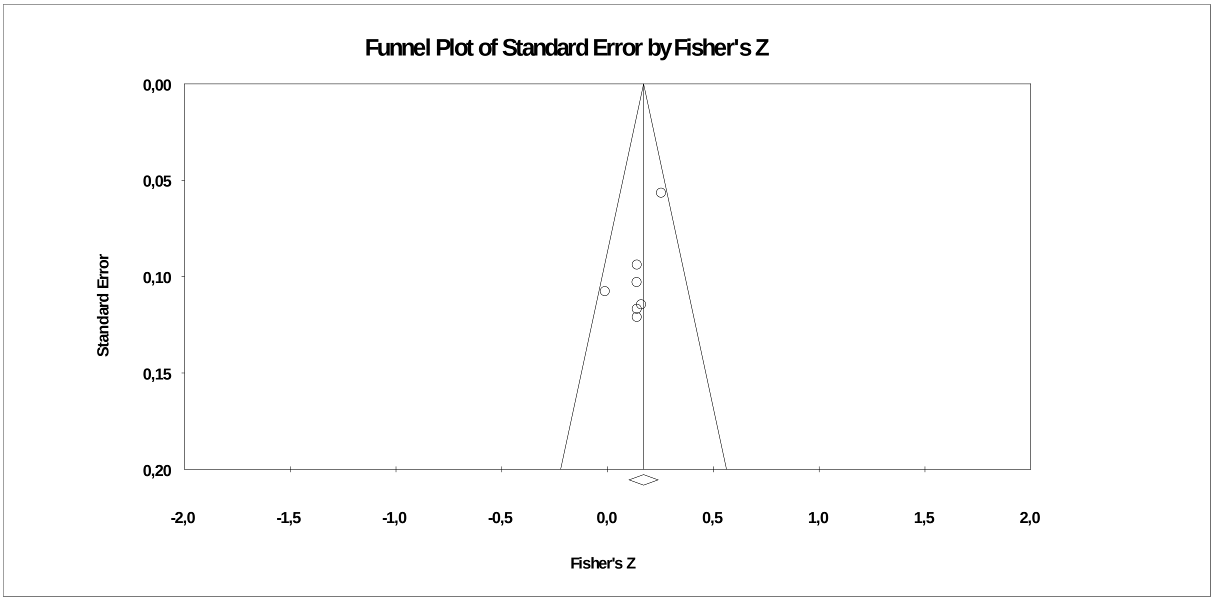
Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0,38860	0,33975	0,43537	0,35899	0,23364	0,47261	43,29857
Adjusted values	0	0,38860	0,33975	0,43537	0,35899	0,23364	0,47261	43,29857

Note: No adjusted values to either side of the mean.

Non-word repetition – reading comprehension

Funnel plot:



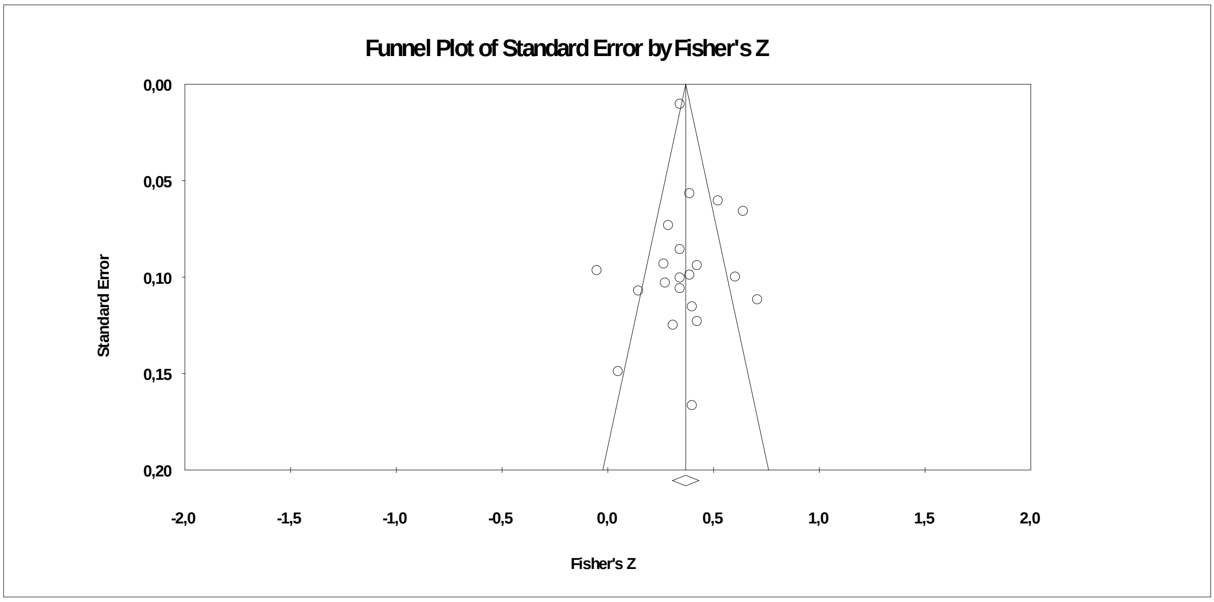
Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0,16853	0,10136	0,23418	0,16853	0,10136	0,23418	5,35343
Adjusted values	4	0,21191	0,15586	0,26659	0,20848	0,14316	0,27199	12,68431

Note: Adjusted values to the right of the mean (zero-adjusted values to the left of the mean).

Non-verbal intelligence – reading comprehension

Funnel plot:



Duval and Tweedie's trim and fill

		Fixed Effects			Random Effects			Q Value
	Studies Trimmed	Point Estimate	Lower Limit	Upper Limit	Point Estimate	Lower Limit	Upper Limit	
Observed values		0,33866	0,32243	0,35469	0,35253	0,29630	0,40633	73,75480
Adjusted values	5	0,35039	0,33461	0,36597	0,40454	0,34527	0,46059	136,40746

Note: Adjusted values to the right of the mean (zero-adjusted values to the left of the mean).

About this review

Determining how to provide the best instruction to support children's reading comprehension requires an understanding of how reading comprehension actually develops. To promote our understanding of this process, this review summarizes evidence from observations of the development of language and reading comprehension from the preschool years into school. The main outcome in this review is reading comprehension skills.

Understanding the development of reading comprehension and its precursors can help us develop hypotheses about what effective instruction must comprise to facilitate well-functioning reading comprehension skills. These hypotheses can be tested in randomized controlled trials.

The Campbell Collaboration
info@campbellcollaboration.org
Phone: (+47) 23 25 50 00

Mailing address:
P.O. Box 4404, Nydalen
N-0403 Oslo, Norway

Website:
www.campbellcollaboration.org