

Journal Pre-proof

Preschool pathways to reading comprehension: A systematic meta-analytic review

Hanne Næss Hjetland, Ellen Iren Brinchmann, Ronny Scherer, Charles Hulme,
Monica Melby-Lervåg



PII: S1747-938X(18)30635-3

DOI: <https://doi.org/10.1016/j.edurev.2020.100323>

Reference: EDUREV 100323

To appear in: *Educational Research Review*

Received Date: 20 December 2018

Revised Date: 15 February 2020

Accepted Date: 17 February 2020

Please cite this article as: Hjetland, Hanne.Næ., Brinchmann, E.I., Scherer, R., Hulme, C., Melby-Lervåg, M., Preschool pathways to reading comprehension: A systematic meta-analytic review, *Educational Research Review* (2020), doi: <https://doi.org/10.1016/j.edurev.2020.100323>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier Ltd.

Author statement:

Hanne Næss Hjetland: Conceptualization, Methodology, Validation, Formal analysis, Resources, Software, Writing – original draft, Writing – review & editing, Visualization, Project administration.

Ellen Iren Brinchmann: Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing.

Ronny Scherer: Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing, Visualization.

Charles Hume: Conceptualization, Methodology, Writing – original draft, Writing – review & editing.

Monica Melby-Lervåg: Conceptualization, Methodology, Formal analysis, Supervision, Writing – original draft, Writing – review & editing, Project administration.

Title Page

Preschool Pathways to Reading Comprehension: A Systematic Meta-analytic Review

Hanne Næss Hjetland^{1,2}

Ellen Iren Brinchmann¹

Ronny Scherer³

Charles Hulme^{1,4}

Monica Melby-Lervåg¹

¹Department of Special Needs Education, University of Oslo

²Nordic Institute for Studies in Innovation, Research and Education

³Department of Teacher Education and School Research, University of Oslo

⁴Department of Education, University of Oxford

Preschool Pathways to Reading Comprehension: A Systematic Meta-Analytic Review

Journal Pre-proof

Abstract

The ability to construct meaning from texts is the core of reading. We report a meta-analysis and a systematic review of 64 longitudinal studies tracing the development of reading comprehension from preschool. Previous research showed that linguistic comprehension and code-related abilities in preschool correlate moderately with reading comprehension, but the results across studies are inconsistent. Meta-analytic structural equation modelling showed two distinct but related pathways from preschool linguistic comprehension abilities to reading comprehension in school. One pathway consists of code-related skills (letter knowledge and phonological awareness), and it affects reading comprehension through word recognition. A second pathway consists of linguistic comprehension skills (vocabulary and grammar), and it has a direct influence on reading comprehension. Early interventions to facilitate language development appear to provide a promising approach to facilitate the later development of reading comprehension skills.

Keywords: Reading comprehension, language development, longitudinal study designs, meta-analysis, systematic review

1. Introduction

The ability to simultaneously extract and construct meaning through interaction and involvement with texts is the core of reading (RAND Reading Study Group, 2002). Reading comprehension is critical for all aspects of education and for participation in society. It has long been accepted that the foundations of reading comprehension are laid long before children start learning to read (Teale & Sulzby, 1986). To understand this process, it is crucial to conduct longitudinal studies that trace the precursors of reading comprehension from preschool into the school years. Such studies enhance our theoretical understanding of reading comprehension and also provide the basis for developing methods of teaching to improve reading comprehension.

In the last 15 years, the number of longitudinal studies of reading comprehension has increased rapidly. Here, we present a systematic review that summarises these studies. Given how important reading comprehension is for learning outcomes in school, understanding the factors that promote or impede this ability is critical for educational practice. An important contribution of this review is to evaluate the consistency of findings across studies. Summarizing these studies will not only have important theoretical value but could also give direction for educational practice in terms of assessments and interventions. In addition, this review provides convergent robust evidence across different samples, languages, and contexts on the association between linguistic comprehension ability in preschool and reading comprehension in school.

1.1. Theoretical Frameworks that Inform the Development of Reading Comprehension

Several theoretical frameworks describe the complexity of reading comprehension (Cromley & Azevedo, 2007; McNamara & Kintsch, 1996; Perfetti & Stafura, 2014). For children in the early stages of reading development, the most influential theoretical framework is the simple view of reading. According to the simple view, reading

comprehension is the product of word recognition (decoding) and linguistic (language) comprehension (Gough & Tunmer, 1986). Word recognition refers to the ability to translate printed words into speech, independent of their meaning. Linguistic comprehension refers to the ability to understand the meaning of spoken language. Notably, Gough and Tunmer used listening comprehension as a synonym for linguistic comprehension, which is in line with studies showing that these two constructs are highly related (Lervåg, Hulme, & Melby-Lervåg, 2017; Protopapas, Mouzaki, Sideridis, Kotsolakou, & Simos, 2012). It has also been suggested that listening comprehension and linguistic comprehension are best understood as one construct (Justice et al., 2017). In their original 1986 article, Gough and Tunmer sought to clarify the role of word recognition in reading and reading disability, resulting in several claims that may inform our understanding of reading comprehension development. For instance, Gough and Tunmer (1986) argued that although both word recognition and linguistic comprehension are necessary conditions for reading to occur, their contributions to reading comprehension are independent. Consequently, their contributions are not necessarily equal – their relative importance may change across time, and there may even be a discrepancy between a reader's word recognition skills and linguistic comprehension ability. All of these predictions can be tested empirically at different developmental stages and at various levels of reading proficiency. However, none of the predictions concern the development of reading comprehension before the onset of word recognition. Thus, in the present study we extend the scope of the simple view of reading by investigating the early foundations of reading comprehension. More specifically, we examine the extent to which preschool measures of linguistic comprehension and code-related skills (i.e., precursors of word recognition) predict later reading comprehension. We perform correlation-based meta-structural equation modelling (MASEM) — a method that allows structural equation models to be fitted to meta-analytic datasets (Cheung, 2015).

Previous research has provided strong support for the simple view of reading. Recent studies show that the components of the simple view explain as much as 94-98 % of the variance in reading comprehension in early primary school (Foorman, Koon, Petscher, Mitchell, & Truckenmiller, 2015; Lervåg et al., 2017). However, in some longitudinal studies the two components have explained a relatively small percentage of the variance in reading comprehension (e.g., Torppa et al., 2016). The reason for this difference in results may to some extent be explained by different orthographies and the included measures (Florit & Cain, 2011). When it comes to the two components—word recognition and linguistic comprehension—the relative strength changes across development, thus the length of the studies may also be one factor in explaining the differences in studies which may lead to differences in results. In the early stages of learning to read (e.g., Grades 1-3), word recognition is a major constraint on reading comprehension (Lervåg et al., 2017), but later, when children have mastered word recognition, linguistic comprehension becomes a more important influence on reading comprehension (Verhoeven & van Leeuwe, 2012; see also Geva & Farnia, 2012; Storch & Whitehurst, 2002).

Storch and Whitehurst (2002) reported a seminal study seeking to predict reading comprehension in the 2nd to 4th grade (4 to 9 years of age) from oral language and code-related measures in preschool. This study provided strong support for the simple view of reading with two distinct pathways from children's early language abilities to later reading comprehension: a direct linguistic comprehension pathway and a code-related pathway that drives reading comprehension via word recognition skills. In the earliest grades, word recognition had the greatest influence on reading comprehension, but from the third grade onwards, language comprehension made a significant contribution. There was also a strong association between children's preschool linguistic comprehension and code-related skills, indicating a close relation between these component skills at an early developmental stage.

However, the strength of this association decreased with age and the children's linguistic comprehension abilities showed higher longitudinal stability than did their code-related skills.

1.2. Unresolved Issues in Longitudinal Studies Predicting Reading Comprehension from Preschool

In the wake of Storch and Whitehurst's (2002) seminal study, the number of longitudinal studies of reading comprehension has increased rapidly. However, there are large variations between studies in terms of the preschool predictors they include. The most common predictors are, in line with Storch and Whitehurst's take on the simple view of reading, vocabulary, grammar, phoneme awareness, letter knowledge and rapid automatized naming (RAN; Fricke, Szczerbinski, Fox-Boyer, & Stackhouse, 2016; Hulme, Nash, Gooch, Lervåg, & Snowling, 2015). However, although results from a broad range of predictors are reported in the literature, studies that include a combination of measures that adequately assess both code-related and linguistic comprehension skills are rare (Hjetland, Brinchmann, Scherer, & Melby-Lervåg, 2017). Moreover, despite Gough and Tunmer's (1986) suggestion to use listening comprehension as a measure of linguistic comprehension, few studies involving young children have included this type of assessment (Hjetland et al., 2017). On the other hand, a number of studies include additional predictor variables that are not based on the original account of the simple view of reading, such as working memory, socio-economic background, and nonverbal intelligence (Roth, Speece, & Cooper, 2002; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004).

The results from previous longitudinal studies of reading comprehension are inconsistent in several respects. For instance, some studies have found a strong predictive relationship between preschool vocabulary and later reading comprehension (Roth et al., 2002), whereas others have only found a weak relationship (Fricke et al., 2016). The variation in the size of the bivariate correlations between measures, coupled with differences in the type

of predictors that have been assessed in different studies, has led to variations in the conclusions drawn on the development of reading comprehension.

Several of the inconsistencies in the results of prior studies, may stem from differences in their methodological approach. For instance, one issue that may explain the between-study variation is sample characteristics, including the age of the participants and how long they have been receiving reading instruction at the point at which reading comprehension is assessed (Hjetland et al., 2017). In studies where reading comprehension is measured early, we would expect to find a relatively weak relation between preschool vocabulary and reading comprehension (and a relatively stronger relationship between reading comprehension and preschool precursors of word recognition, such as phoneme awareness and letter-sound knowledge). In contrast, we would expect to find a stronger relation between preschool vocabulary and reading comprehension in older children who have mastered word recognition.

Previous studies have also shown that measurement issues, such as the type of reading comprehension test used influences the strength of the correlations between reading comprehension, word recognition and linguistic comprehension. Keenan and Betjemann (2006) showed that tests using multiple-choice questions are typically not good measures of reading comprehension, as children may be able to answer many questions on such tests using background knowledge without reading the passage. Keenan, Betjemann, and Olson (2008) showed that tests with open-ended questions are more dependent on linguistic comprehension skills than tests with multiple-choice questions or a cloze procedure. Thus, the type of reading comprehension test used may also have led to inconsistencies in the results of previous studies.

Finally, some of the discrepancies between studies may stem from the failure to deal adequately with measurement error. Measurement error attenuates the correlation between

variables, and in multivariate studies, differences in reliability can have unpredictable consequences for the estimation of regression scores because a predictor with good reliability will explain more variance than a competing predictor with poor reliability (Cole & Preacher, 2014). Most prior studies have used single measures of language constructs, such as vocabulary, grammar and phonological awareness. Using multiple measures of each construct is far preferable, as it allows the use of latent variables free of measurement error (e.g., Little, 2013). The use of single measures may cause further inconsistencies between studies due to the large variation in the choice of instruments that are used to measure different language constructs. For instance, vocabulary has previously been assessed by different measures, such as word definition tasks, picture naming, and picture pointing (e.g., the Peabody Picture Vocabulary Test [PPVT]). The same is the case for code-related skills (such as word recognition and phonological awareness) and working memory. The assumption that various tasks are equally representative of the higher-order constructs that they are designed to measure has rarely been tested in prior studies.

To sum up, although developmental studies of reading are increasing in number, variation in the results of these studies limits our understanding of the early pathways to reading comprehension. In the present study we use a MASEM-approach to overcome the limitations of individual studies. Conducting a meta-analysis on previous studies will increase statistical power to detect meaningful associations among constructs, and enable us to generalize findings across different samples, settings and assessment types. The use of latent variables in MASEM allows us to deal effectively with measurement error across individual studies.

Although reproducibility in experimental psychology has attracted much attention in recent years (Open Science Network, 2015), much less emphasis has been put on issues of robustness in multivariate observational studies. Our systematic literature review, coupled

with a meta-analysis of key structural relations, will pinpoint which findings are robust and how we can refine future studies.

1.3. Previous Systematic Reviews

Three prior reviews are of particular interest: Quinn and Wagner (2018) used MASEM to examine the components of the simple view of reading in concurrent correlational studies including both younger students (from 3.5 years) and adults. Their meta-analytic structural equation model included three latent variables: word recognition, linguistic comprehension, and a construct they referred to as cognitive abilities, consisting of working memory and inferencing skills. The model explained 56.8 % of the variance in the students' reading comprehension, but only word recognition and linguistic comprehension had statistically significant independent contributions to reading comprehension. Garcia and Cain (2014) found a sizeable concurrent correlation between word recognition skills and reading comprehension ($r = .74$), with age and listening comprehension moderating this relationship. These reviews were limited to studies conducted with English speaking samples. Finally, the National Early Literacy Panel (NELP, 2008) summarised longitudinal studies of reading comprehension, including studies up to 2004. The average correlations between language-related variables and reading comprehension ranged from $r = .20$ (concept knowledge) to $r = .59$ (reading readiness). However, only few primary studies were included, and this review did not use a model-based meta-analysis to analyse multivariate relations.

1.4. The Current Study

In the present study we investigate the early foundations of reading by summarizing preschool predictors of reading comprehension. The selection of preschool predictors in this review was guided by the simple view of reading and included linguistic comprehension and code-related measures such as vocabulary, grammar, phonological awareness, letter knowledge, and RAN. However, we also included predictors such as working memory,

nonverbal IQ, and socio-economic background, which by some accounts may represent useful additions to the simple view of reading (Quinn & Wagner, 2018). We will summarise the size of the bivariate relations between language skills in preschool and later reading comprehension and examine factors that possibly moderate them (e.g., age, test type). As type of reading comprehension assessment has been shown to be associated with the strength of correlation between the reading comprehension and the other two main components in the simple view of reading, we hypothesize that tests with open-ended questions are more dependent on language comprehension skills than tests using a multiple-choice or a cloze procedure. Importantly, tests that use similar procedure to tap the student's comprehension of text may differ in other aspect. The current study summarizes studies of the correlation between preschool skills and later reading comprehension. However, most studies in this review only include a single reading comprehension test involving open-ended questions and it remains for future studies to clarify whether the type of reading comprehension test has important effects on the relative importance of word recognition and linguistic comprehension as predictors of reading comprehension. Tests with open-ended questions may be more dependent on linguistic comprehension skills than tests using a multiple-choice or a cloze procedure. Importantly, tests that use similar procedures to assess the students' reading comprehension may differ in other aspects (i.e., vocabulary, testing procedure). Our meta-analysis summarizes the correlations between abilities in preschool and later reading comprehension across different types of reading comprehension assessment, examining the possible moderating effects of the type of assessment. Explaining variation across different assessment types is one of the key strengths of a meta-analysis that synthesizes several primary studies.

However, to gain knowledge about how reading comprehension develops concurrent bivariate correlations are not sufficient. We therefore model the relationship between

preschool skills and reading comprehension and examine which separate measures go together to define common constructs. We control for measurement error and use latent variables when examining how preschool factors influence later reading comprehension. To accomplish this, we utilize correlation-based MASEM (Cheung, 2015). This approach first synthesizes entire correlation matrices and then performs structural equation modelling on the resultant, aggregated correlation matrix. In contrast to performing separate meta-analyses for each individual correlation, this two-step approach accounts for the dependencies between correlations within studies and thus provides more accurate parameter and variance estimates (Cheung & Cheung, 2016). MASEM also handles missing data efficiently since incomplete correlation matrices need not be excluded—the maximum-likelihood estimation procedures account for missing data efficiently (for details, please refer to Cheung & Chan, 2009). Overall, MASEM provides a powerful approach to testing a model that describes the pathways to reading comprehension. At the same time, it requires correlation matrices to be positive definite and may thus limit the selection of primary studies eligible for structural equation modelling.

This meta-analytic review has the following aims:

(1) We seek to establish accurate estimates of the relationships between preschool code-related skills, and later word recognition and reading comprehension skills, as well as between linguistic comprehension skills in preschool and later reading comprehension.

(2) We seek to establish accurate estimates of the relationships between domain general skills, such as nonverbal intelligence and working memory and later reading comprehension. In addition, we seek to examine the relationship between background factors related to socio-economic background (SES) and later reading comprehension.

(3) We formulate and evaluate a structural equation model with two distinct but related pathways comprising code-related skills and linguistic comprehension as predictors of later reading comprehension skills.

(3a) In this model, we expect code-related skills in preschool to have a significant indirect effect on reading comprehension in school through word recognition skills.

(3b) The hypothesised model describing the pathways from code-related skills and linguistic comprehension to reading comprehension should apply to both younger and more experienced readers and to studies using different types of assessment.

2. Method

This study is based on a preregistered review conducted within the Campbell collaboration framework (see Hjetland et al., 2017). A Campbell review comprises three elements: (1) the title registration, (2) the protocol, and finally (3) the review. All undergo an extensive peer review process before being published online as open access. The current paper is an extension of the Campbell systematic review, as we present additional moderator analyses (e.g., the possible moderators are examined individually instead of a regression analyses with multiple covariates) and subgroup analyses (e.g., reading comprehension measure: multiple-choice vs. open-ended assessments of reading comprehension). We conducted both the bivariate analyses and the MASEM technique, and further introduced children's socio-economic background in the primary samples as a predictor of reading comprehension.

2.1. Inclusion Criteria

This review included longitudinal non-experimental studies that follow children from preschool into school. To be included, studies had to report data from at least two assessment time points: one in preschool, before formal reading instruction had begun, and one in school,

after formal reading instruction had started. Because formal reading instruction begins at different ages in different countries, we use the term preschool in this study to refer to measurements taken before the start of formal reading instruction. Children in preschool in studies reviewed here span a range of ages from 3 to 6 years. In addition to these conditions, we applied the following inclusion criteria:

- *Sample.* The study population consisted of samples of mainly monolingual typically developing children who were not subject to restrictive selection criteria (e.g., a diagnosed learning difficulty or bilingualism).
- *Measures of constructs.* At least one of the predictors (vocabulary, grammar, phonological awareness, letter knowledge, RAN, verbal working memory, and nonverbal intelligence) and reading comprehension, as measured by standardised or researcher-designed tests, were included.
- *Measures of association.* The studies had to report a Pearson's r correlation between predictors and reading comprehension, along with the sample size N .
- *Publication year.* Only studies published after 1985 were considered for inclusion.
- *Language of reporting.* The studies had to be reported in English.

2.2. Constructs and Measures

Before including studies in the review, we set up different criteria that had to be met for the indicators to be considered valid measures of the constructs in which we were interested.

2.2.1. Outcome variables. A test was considered to measure *reading comprehension* if the child had to read sentences and/or passages and either answer questions (either open ended or closed) about the content of the text or identify a missing keyword. For *word recognition*, indicators had to be measures in which the child read, timed or untimed, either strings of words or passages of words.

2.2.2. Predictors of word recognition. A task was considered a measure of *phonological awareness* if it involved detecting, manipulating or analysing the phonological structure of spoken language (including the ability to distinguish or segment words, syllables or phonemes). Two indicators of phonological awareness are included here: phoneme awareness and rhyme awareness. Measures of *letter knowledge* had to tap knowledge of the names and sounds associated with printed letters. *RAN tasks* were tasks that assessed the ability to name a sequence of randomly repeating pictures (e.g., 'car', 'tree', 'house', 'man'), colours, letters or digits as rapidly as possible.

2.2.3. Predictors of linguistic comprehension. A task was considered a measure of *vocabulary* if the child was asked to define a word, locate a picture corresponding to a spoken word, name a picture, fill in a missing word in a sentence or answer questions after listening to a story. Composite measures of language ability were also included. In measures of *grammatical skills*, children had to demonstrate knowledge about how words and their component parts are combined to form coherent sentences (i.e., morphology and syntax).

In tests measuring *verbal working memory*, children had to repeat spoken sentences or non-words verbatim. Non-word repetition and sentence memory were coded in separate categories, as their relations with memory and language are debated, and it is not entirely clear into which category they fit. As for *socio-economic background*, indicators that measure parental educational level, occupation, free/reduced school lunch and more broad indices of socio-economic status were included. *Nonverbal IQ* measures the ability to analyse information and solve problems without using language-based reasoning.

In addition, we coded methodological quality in the studies. The criteria for this coding were as follows: *Sampling procedure* was coded when reported in the studies. Two categories were used, namely random and convenience sampling. *Instrument type* was coded based on whether the studies included only standardised instruments, a combination of

standardised and researcher-made instruments, or only researcher-made instruments. **Test reliability** was coded based on whether or not the reliability of the measures used was reported in the studies, that is, whether reliability was reported for all measures, some measures or whether reliability was not reported/reported from test manual only. The occurrence of **floor or ceiling effects** was coded based on whether any of the measures showed floor or ceiling effects as indicated by the reported means and standard deviations. **Attrition** was coded based on whether or not attrition was reported in the article. We also calculated the percentage of attrition from the first and last assessment. **Missing data** was separated into two categories based on how studies dealt with missing data (listwise deletion or other procedure better than listwise). **Latent variables** were coded based on whether the studies used this (yes/no). **Statistical power/sample size** depends on many factors. However, as a general rule, samples below 70 will yield unstable estimates and, in general, have low power to detect relationships of the size that is 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, so this was not possible.

The abovementioned broad set of quality indicators was coded to be able to give precise recommendations concerning future studies. Each study was given a value on the 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 (see Table S1 in online supplemental material for coding procedures). In addition, type of reading comprehension test, sample size, attrition and reliability were included as categorical moderators. It should be noted that the use and interpretation of the study quality indicators should be handled with some caution. For instance, a standardized test is not automatically of better quality than a researcher-made test. There are also differences in the availability and

suitability of standardized instruments when it comes to specific language domains and countries that should be taken into account.

2.3. Search

The electronic search consisted of seven databases (Google Scholar, PsycINFO via OVID, ERIC via Ovid, Web of Science, ProQuest Dissertations and Theses, OpenGrey.eu Linguistics and Language Behavior Abstracts), and it was conducted in February 2016. For a full list of search strategies and search terms, see Table S2 in the online supplemental material.

The electronic search yielded a total of 3285 references from the seven databases (2498 after excluding duplicates). After screening abstracts, a further 1393 were further excluded, leaving 1105 full articles to be read and evaluated for inclusion. Just 64 studies met the eligibility criteria and were included in the meta-analysis. The search and flow of studies through the review is shown in Figure 1.

2.4. Data Extraction and Coding

When screening abstracts, as well as in reviewing the full text of articles, 25 % of the references was double-screened. Cohen's κ , the inter-rater reliability for inclusion or exclusion, was satisfactory at both stages, with coefficients of Cohen's $\kappa = .92$ and $.95$, respectively. Any disagreements between raters were resolved by discussing and consulting the original paper. After establishing inter-rater reliability, the two raters (i.e. the first and second author of this review) divided the remaining 75 % of the references evenly amongst themselves for further screening.

To standardise the extraction of data from the primary studies, we developed a coding scheme describing the data extraction procedure. The first and the second author independently extracted data from 37.5 % of the studies (24/64) to check the accuracy and reliability of coding. The interrater reliability was good, Pearson's $r = .95$, with level of

agreement of 96 % on the extracted effect sizes. After reliability had been established, the first author extracted data from the remaining studies.

In some cases, multiple observations existed for the same outcome. In such cases, we calculated an average 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, children were measured at more than one time point in school or in preschool. In those cases, we chose the first assessment in preschool and the last assessment in school.

2.5. Missing Data

We identified several types of missing data: incomplete correlation matrices, missing sample characteristics and missing information pertaining to methodological quality. When a study met our inclusion criteria but failed to report an uncorrected bivariate correlation matrix, we contacted the corresponding authors and requested the necessary data. In fact, most studies did not report all correlations among the variables under consideration—hence, correlation matrices were often incomplete. Conducting MASEM, we used the maximum likelihood procedures to handle missing data under the assumption that missing values occurred randomly (Cheung, 2015; Enders, 2010). When data were missing from variables concerning sample characteristics or methodological quality, the study with missing data was excluded from the moderator analysis but was retained for the overall effect size estimation.

2.6. Statistical Analyses

To test our hypotheses and examine possible moderation effects, we took two analytic steps. First, we synthesised the bivariate correlations among the preschool predictors and later reading skills using the Comprehensive Meta-Analysis software version 3 (BioStat; Borenstein, Hedges, Higgins, & Rothstein, 2014). This step resulted in a single, aggregated matrix of correlation among all constructs, which was further subjected to the analysis of

publication bias and possible moderation effects. Second, we synthesised the correlation matrices retrieved from the primary studies to test whether the pathways of the reading comprehension model were supported. This step involved correlation-based MASEM (e.g., Hong & Cheung, 2015).

2.6.1. Meta-analysis of bivariate correlations. We used Pearson's r as the measure of effect size; the analyses were conducted using Fisher's z but results were transformed back to Pearson's r for the purposes of reporting. To calculate an average effect size, we used random-effects models that account for the fact that variation in effect sizes between studies may be due to both random error and systematic differences in study characteristics. To determine the degree of heterogeneity between studies, we tested whether the heterogeneity was different from zero using the chi-squared-based Q -test. Because the Q -test is highly dependent on sample size, we also used τ^2 to examine the magnitude of variation in effect sizes between studies (Hedges & Olkin, 1985).

Notably, τ^2 is used to assign weights under the random-effects model; thus, the total variance in 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, Hedges, Higgins, & Rothstein, 2009, p. 72). As a rule, we considered the variation between studies large if τ^2 exceeded 0.1—this criterion was based on the fact that the typical population standard deviations (SDs) in applied psychology are around 0.1 to 0.2 (see Bosco, Aguinis, Singh, Field, & Pierce, 2015), corresponding to a range of τ^2 between 0.01 and 0.04. Finally, we used the I^2 -statistic, a statistic indicating the proportion of variance in effects that can be attributed to true heterogeneity versus random error, to further indicate the heterogeneity of correlations within the range of 0 % (no systematic differences between studies: variation is primarily due to chance) to 100 %.

As for the size of an average correlation, determining the effect size to be deemed of practical significance in the current field is crucial. Thus, an effect size that would be considered moderate by the standards set by Cohen (1988) could be interpreted as strong when compared to the average effect sizes shown in education.

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.

To estimate publication bias, we applied Duval and Tweedie's trim-and-fill method (Duval & Tweedie, 2000) and inspected funnel plots. However, the validity of the funnel plot/trim-and-fill method has been questioned (Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006), especially when it is used in the presence of a large between-study variation (Terrin, Schmid, Lau, & Olkin, 2003). Therefore, the results from these analyses should be interpreted with caution.

2.6.2. Meta-analytic structural equation modelling (MASEM). We further tested a structural equation model that represented the simple view of reading and contained multiple correlations at the same time. Given that the primary studies did not only contribute one correlation between certain variables but multiple, dependencies among the correlations (i.e., their nesting within studies) occurred. These dependencies would violate the assumptions of traditional, univariate meta-analysis (Borenstein et al., 2009), and hence a multivariate approach was needed to account for them in a meta-analytic model. Correlation-based MASEM represents such an approach (Cheung, 2015). Specifically, correlation-based MASEM can accommodate multiple correlations that were extracted from the primary studies (i.e., correlation matrices) by considering the covariance between the correlation coefficients and their variances (Cheung, 2013). Correlation-based MASEM can be implemented in

multiple ways; although a two-stage structural equation modelling (TSSEM) is the dominant approach (Cheung & Cheung, 2016). This procedure comprises two stages (Cheung & Chan, 2005; Cheung, 2015): In the first stage, the correlation matrices are combined under a random-effects model to a pooled (overall) correlation matrix (TSSEM stage 1). Similar to univariate meta-analysis in which single correlations are pooled under fixed- or random-effects models, the pooling of correlation matrices can also be based on fixed- or random-effects models (Jak, 2015). Oftentimes, the assumption of random effects is more realistic than that of fixed effects, because it allows the inclusion of between-study variances and covariances (Cheung, 2016).

In the second stage, the pooled correlation matrix is submitted to structural equation modelling (TSSEM stage 2). In this stage, the structural equation model is specified, estimated, and evaluated according to its model fit. Moreover, competing models can be compared using chi-square difference testing or information criteria. In this study, we evaluated the fit of the structural equation models on the basis of the common guidelines for an acceptable model fit (i.e., $CFI \geq .95$, $RMSEA \leq .08$, and $SRMR \leq .10$; Hu & Bentler, 1999; Marsh, Hau, & Grayson, 2005). We performed correlation-based MASEM using the R package 'metaSEM' (version 0.9.8; Cheung, 2015).

3. Results

Table S3 in the online supplemental material shows the characteristics of the included studies (with correlations coded from each study, sample size and average age of participants). Notably, of the 64 included studies, 40 were conducted with English-speaking children. We first present average bivariate correlations between reading comprehension and all the predictors that were included in the study. Second, the resultant correlations and their variances are explored through moderator analyses and analyses of publication bias. Third, we synthesise entire correlation matrices using MASEM to evaluate the fit and applicability of a

'pathways to reading comprehension' model. Extending the MASEM approach, we further test the fit and applicability of the model for selected subgroups of studies (i.e., early [two years of formal reading instruction or less] vs late reading [more than two years of formal reading instruction], multiple-choice vs open-ended assessments).

3.1. Longitudinal Correlations between Preschool Predictors, Word recognition and Reading Comprehension

Table 1 shows the mean longitudinal correlations among preschool predictors, word recognition and reading comprehension. That the correlation between RAN and reading comprehension is negative indicates that students with better reading comprehension have shorter RAN times. There is a moderate correlation among all predictors and the outcomes of word recognition and reading comprehension. An exception is non-word repetition, which shows a smaller correlation. Importantly, there is a large variation among individual studies in the size of the correlations that are reported. Table 1 shows that for most measures, the true heterogeneity between studies (I^2) is significant and above 70 %. Thus, despite the fact that all of the studies examine preschool predictors of word recognition and reading comprehension, it is evident that they reach different conclusions about the size of these relations. It is also clear that moderators can be operating and that mean effect sizes should be interpreted with some caution.

A closer look at the correlations among measures of linguistic comprehension, vocabulary, grammar and reading comprehension illustrates this heterogeneity. Figure 2 shows a forest plot of the correlations between vocabulary and reading comprehension. The mean correlation is moderate, but there is large variation between studies. The vast majority of studies use picture vocabulary tasks, such as the PPVT (Dunn & Dunn, 2007), whereas most reading comprehension measures use multiple choice questions or close tasks, such as the Passage comprehension subtest from the Woodcock-Johnson test battery (Woodcock,

McGrew, & Mather, 2001). Figure 3 shows a forest plot of the correlation between grammar and reading comprehension, which again shows a moderate correlation but with a large variation between studies. Finally, the two best-established predictors of word recognition (letter knowledge [Figure 4] and phoneme awareness [Figure 5]) both show moderate correlations with reading comprehension, but large variation between individual studies. There were also moderate correlations between letter knowledge and phoneme awareness and word recognition skills (see Table 1).

3.2. Study Quality

Figure 6 shows an overview of the risk of methodological bias in the studies. As presented in Figure 6, there is moderate-to-high risk of bias in several different aspects of the studies. As for sampling, of the 64 included studies, five used random sampling, whereas 59 used convenience sampling. Importantly, many studies lacked vital information about how samples were recruited. As for measurement instruments, a mixture of standardised (norm-referenced) and researcher-made instruments was typically used in the studies ($n = 44$). Of the 64 studies, 17 used only standardised measures. In three studies, only researcher-made instruments were used.

Test reliability was not commonly reported. Typically, only reliabilities from the test manual were reported ($n = 34$), although some reported reliability coefficients based on their own data for some of their measures ($n = 11$). Nineteen of the 64 studies reported reliability estimates for all measures. Only four studies used latent variables to control for measurement error. There were also noteworthy issues with missing data. Many studies had problems with attrition; the highest percentage of attrition in the included studies was 59 %, although this particular study spanned 10 years. In addition, for 15 of the studies, there was no available information about attrition because only the sample size at the first time point was reported. Moreover, most of the studies did not report information about how they dealt with missing

data, and only nine of the studies reported using a specific technique to handle this issue (e.g. full information maximum likelihood estimation). Finally, several of the studies had issues concerning statistical power, and as many as 32 % ($n = 20$) of the studies had fewer than 70 participants.

3.3. Moderators of the Longitudinal Relationships between Preschool Predictors, Word Recognition and Reading Comprehension

Due to the significant true heterogeneity between the studies, we examined whether moderator variables could explain the variation in effect sizes shown in the studies. Table 2 shows the results from the continuous moderator analyses for the correlations between the main preschool predictors and reading comprehension in school. Notably, we did not perform moderator analyses with non-word repetition and SES because the meta-analyses did not show statistically significant variation between the studies. The results of the analyses revealed that age of assessment and months of reading instruction when reading comprehension was measured could explain the variation between studies in the predictive value of sentence repetition, RAN and rhyme awareness. However, after correcting for multiple significance tests, these moderator effects were no longer significant.

Tables S4-S7 in the online supplementary material shows the results from the tested categorical moderator analyses for the correlations between phoneme awareness, letter knowledge, vocabulary and grammar and reading comprehension, respectively. These moderator analyses have only been performed on the relations that are further explored in the MASEM analyses. The tested categorical moderators are type of reading comprehension test, sample size, attrition and reliability. However, the results show that despite inconsistent findings and the large heterogeneity between studies, none of these moderators explained significant variance between studies.

3.4. Publication Bias

We examined publication bias by analysing funnel plots using the trim-and-fill method. Across all bivariate relationships, nine studies were filled in for the correlation between phoneme awareness and reading comprehension ($r = .40$ before adjustment, $.43$ after adjustment), three studies for letter knowledge/reading comprehension ($r = .42$ before adjustment, $.40$ after adjustment), six studies for rapid naming/reading comprehension ($r = .34$ before adjustment, $.27$ after adjustment), four studies for non-word repetition/reading comprehension ($r = .17$ before adjustment, $.20$ after adjustment) and five studies for nonverbal intelligence/reading comprehension ($r = .35$ before adjustment, $.40$ after adjustment). Thus, there were indications of publication bias for some bivariate correlations, but the pattern was not consistent and the changes in the size of the correlations were generally small.

3.5 Results from Meta-analytic Structural Equation Modelling

3.5.1 Aggregating correlation matrices. One of the key prerequisites for synthesising correlation matrices for structural equation modelling is that matrices are positive definite (Cheung, 2015). Wothke (1993) considered matrices to be positive definite if all their eigenvalues are positive. Of the 64 available correlation matrices, 42 met this criterion, and the remaining matrices were excluded from MASEM. Given this reduction, we could only include four preschool predictors in the model: phoneme awareness, letter knowledge, vocabulary and grammar. This, of course, reduced the sample of available primary studies that formed the basis for the structural equation modelling step. However, as noted earlier, the alternative approach, aggregating correlations individually by performing separate meta-analyses of correlations, does not take into account the multivariate nature of the data (i.e., correlation matrices instead of single correlations) and would have introduced severe bias into the estimation of model parameters (Cheung, 2015; Jak, 2015; Sheng, Kong, Cortina, & Hou,

2016). Ultimately, we did not consider this approach to circumvent erroneous conclusions from biased model parameters (Cheung & Hong, 2017).

Table 3 summarises the resultant, overall correlation matrix along with the relevant statistics. This correlation matrix exhibited the presence of heterogeneity between studies (i.e., random effects), $Q(207) = 919.70, p < .001$. Specifically, three correlations within the pooled matrix varied significantly across studies: the correlations between reading comprehension and (1) phoneme awareness, (2) vocabulary and (3) concurrent word recognition. All other between-study variance estimates did not differ significantly from zero. Notably, the variation between studies was smaller in the subsample of 42 studies than in the entire sample of 64 studies.

3.5.2 Performing structural equation modelling. The pooled correlation matrix was submitted to structural equation modelling. The ‘pathways to reading comprehension’ model is shown in Figure 7. It considers reading comprehension and word recognition to be key outcomes, which are predicted by code-related skills and linguistic comprehension. Furthermore, as outlined previously, word recognition serves as a predictor of reading comprehension. In this model, code-related skills and linguistic comprehension are represented as latent variables with two manifest indicators each. Letter knowledge and phoneme awareness measure the former, vocabulary and grammar the latter. The proposed model resulted in a very good fit, $\chi^2(7) = 7.62, p = .37, RMSEA = .004, CFI = 1.000, TLI = .999, SRMR = .021, AIC = -6.38, BIC = -54.04$. Figure 7 shows the resultant parameter estimates, along with their likelihood-based confidence intervals (LBCIs). About 59.5 % of the variance in reading comprehension could be explained by linguistic comprehension and code-related skills in preschool and about 47.6 % of the variance in word recognition. This model also revealed a significant indirect effect of code-related skills on reading comprehension through word recognition, $b = 0.39, 95\% \text{ LBCI} = [0.31, 0.46]$.

Notably, this model also exhibited a substantial correlation between the two predictor variables linguistic comprehension and code-related skills, $\rho = 0.77$. This correlation may bias the resultant path coefficients to reading comprehension and word decoding due to possible multicollinearity (e.g., Marsh, Dowson, Pietsch, & Walker, 2004). We therefore estimated the variance inflation factor (VIF) of code-related skills to test whether multicollinearity issues existed. The regression model underlying this estimation showed a very good fit to the data, $\chi^2(1) = 13.37, p < .01, RMSEA = .043, CFI = .997, TLI = .984, SRMR = .008$. Overall, 62.4 % of the variance in linguistic comprehension was explained, resulting in a VIF = 2.66. Hence, there was no evidence for a multicollinearity issue provided by the VIF (VIF < 5 as a possible criterion; O'Brien, 2007). To further substantiate this finding, we re-specified the structural equation model as a Cholesky-decomposed hierarchical model using phantom variables. This model controlled for the high correlation between the two predictor variables through so-called "phantom variables" (e.g., Dang et al., 2015). The model did not result in substantially different path coefficients (see Supplementary Material S10). In support of our expectations, the positive and significant relations between the predictors and the outcome variables increased slightly and showed a larger indirect effect, $b = 0.59, 95\% \text{ LBCI} = [0.55, 0.63]$. Please find more details in the Supplementary Material S10. Overall, the high correlation between the two predictors did not compromise our conclusions drawn from the meta-analytic structural equation model.

3.5.3 MASEM subgroup analyses. To examine further the extent to which the model applies to certain subgroups of studies, we conducted MASEM subgroup analyses, differentiating between early and late reading studies and studies administering multiple-choice and open-ended assessments.

3.5.3.1 Early vs. late reading. First, we grouped studies with respect to the years of reading instruction to which children had been exposed at the last assessment time point. This

grouping resulted in 16 ‘early reading’ studies ($n = 2426$) that assessed reading comprehension after the children had received 1-2 years of formal reading instruction and 26 ‘later reading’ studies ($n = 4270$) when the children had received more than two years of formal reading instruction. The correlation matrices of both subgroups were combined under random-effects models and used to test the structural equation model. Table S8 in the online supplemental materials shows the pooled correlation matrices for each subgroup. Overall, significant variation in the correlation matrices across studies within subgroups existed (early reading studies: $Q[83] = 353.1, p < .001$; late reading studies: $Q[109] = 435.0, p < .001$).

The structural equation model fit the data very well for early reading studies ($\chi^2 [7] = 6.28, p = .51, RMSEA = .000, CFI = 1.000, TLI = 1.002, SRMR = .026, AIC = -7.7, BIC = -48.3$) and late reading studies ($\chi^2 [7] = 5.38, p = .61, RMSEA = .000, CFI = 1.000, TLI = 1.003, SRMR = .033, AIC = -8.6, BIC = -53.1$). Figures 8a and b show the corresponding model parameters and their confidence intervals. The indirect effects of code-related skills on reading comprehension through word recognition were significant for early ($b = 0.42, 95\% \text{ LBCI} = [0.27, 0.57]$) and late readers ($b = 0.35, 95\% \text{ LBCI} = [0.29, 0.42]$). Considering model fit and the comparability of effects within the model, these findings provide evidence that the proposed model fits the data for both reading groups.

3.5.3.2 Multiple-choice vs. open-ended assessments of reading comprehension. Next, we differentiated between two types of assessments: multiple-choice (closed) vs open-ended assessments of reading comprehension. Given that only eight correlation matrices were available for the latter, we could only specify fixed- instead of random-effects models to achieve maximal comparability between the two subgroups. Two of the 42 eligible studies (i.e., those with positive definite correlation matrices) did not provide sufficient information on the type of assessment tasks used and had to be excluded.

Table S9 in the online supplemental materials shows the pooled correlation matrices for studies using multiple-choice and open-ended assessments of reading comprehension. These matrices exhibited some differences in the associations of variables. For instance, while the correlation between phoneme awareness and vocabulary was moderate for studies with multiple-choice assessments ($r = .38$), studies administering open-ended assessments exhibited a weaker correlation ($r = .14$). Despite these differences, the proposed model of reading comprehension fit the data well for both subgroups of studies (multiple-choice assessments: $\chi^2 [7] = 38.0$, $p < .01$, RMSEA = .027, CFI = .998, TLI = .996, SRMR = .023, AIC = 24.0, BIC = -22.9; open-ended assessments: $\chi^2 [7] = 12.1$, $p = .10$, RMSEA = .033, CFI = .990, TLI = .978, SRMR = .063, AIC = -1.9, BIC = -33.5), thus supporting the applicability of the model to both groups of studies. Model parameters are shown in Figures 9a and b. These parameters indicated some differences, for instance, in the relation between word recognition and reading comprehension (i.e., larger effects for multiple-choice assessments) and the relation between linguistic and reading comprehension (i.e., larger effects for open-ended assessments). Considering the variance explained in reading comprehension, about 68.2 % could be explained in studies with multiple-choice assessments, whereas 43.3 % could be explained in studies with open-ended assessments. The indirect effect of code-related skills on reading comprehension via word recognition was $b = 0.42$ ($p < .05$) and $b = 0.21$ ($p < .05$) for multiple-choice and open-ended assessments, respectively. However, given the small sample size available from open-ended assessment studies, the interpretation of these differences must be treated with caution.

4. Discussion

The current study gives robust evidence of which factors in preschool are related to the later development of reading comprehension. To address some of the limitations of issues in previous primary studies and reviews we first used MASEM to account to measurement error.

Second, since the demand of linguistic comprehension and word recognition have been shown to change throughout reading development, we examined whether the strength of association changed when the studies were grouped according to years of reading instruction. Third, to examine whether the type of reading comprehension test affected the strength of association between the components in the simple view of reading we grouped the studies according to open ended/retell versus multiple choice/cloze procedure. Our findings have important theoretical and applied implications and also highlight some directions for future research.

A number of key findings surfaced:

(1) The correlations between reading comprehension and preschool measures were moderate in longitudinal studies and varied across primary studies.

(2) In a meta-analytic structural equation model with latent variables, we found that code-related skills and linguistic comprehension were separable predictors of later reading comprehension skills, with code-related predictors in preschool having a significant indirect effect on reading comprehension in school through word recognition skills.

(3) The hypothesised model describing the pathways from code-related skills and linguistic comprehension to reading comprehension fit the data well from both younger and more experienced readers and from studies using different types for reading comprehension assessment. However, the relationship between reading comprehension and linguistic comprehension in preschool was stronger in older readers. In addition, reading comprehension measures with open-ended questions showed a stronger relationship with linguistic comprehension in preschool than did multiple-choice tests.

4.1 The Simple View of Reading and Pathways to Reading Comprehension

According to the simple view of reading, reading comprehension is the product of word recognition and linguistic comprehension (Gough & Tunmer, 1986). In this study,

MASEM provides strong support for the simple view. A consistent finding across the 42 studies included in the MASEM was that word recognition and linguistic comprehension explained the majority of the variation in reading comprehension (around 60 %). This is quite remarkable when we consider the limited number of indicators and latent constructs that we were able to include. It is also important to keep in mind that results from a meta-analysis may differ from a single study because it includes data from over 40 studies, thus there will be variation amongst the studies, in terms of the age groups studied, educational practice, and the different assessments used. Our results align well with the Quinn and Wagner's (2018) review of concurrent studies.

In addition to giving strong support to the simple view, our study extends it by showing two distinct pathways leading from preschool language and code-related skills to reading comprehension. These findings align well with two previous large-scale studies: Storch and Whitehurst (2002), who studied a low socio-economic sample, and Hulme et al. (2015), who presented findings from a mixed sample of children at family risk of dyslexia or with preschool language difficulties. Taken together, these studies highlight the importance of a broad focus on language skills from an early age, both for children at risk and for typically developing children.

4.2 Relationships between Linguistic Comprehension Skills and Reading

Comprehension

The current study confirms a key role of linguistic comprehension skills in the development of reading comprehension. A consistent finding across the 42 studies in the MASEM models was that there was a moderate relationship between linguistic comprehension and reading comprehension. However, the average relationship across these studies was weaker than that found in many primary studies using latent variables. There may be several explanations for this. One is that most of the children in these studies were young

readers, in Grade 3 or below, and English speaking. Thus, at that age and in an inconsistent orthography such as English, many children will still be struggling to master fluent word recognition skills (Caravolas, Lervåg, Defior, Málková, & Hulme, 2013) and have not yet reached the stage at which language comprehension comes into play as a major determinant for reading comprehension (Hjetland, Lervåg, Lyster, Hagtvet, Hulme, & Melby-Lervåg, 2018; Lervåg et al., 2017). This interpretation is supported by the moderator analyses, which indicated that the relationship between reading comprehension and linguistic comprehension is stronger in older readers. Another possible explanation is that due to the limited set of measures in primary studies, the latent linguistic comprehension variable here only included measures of two aspects of language ability (vocabulary and grammar). In the individual studies that reported the strongest relationships between reading comprehension and linguistic comprehension, the latent linguistic comprehension variable typically included loadings from a wider range of measures, including inference skills, different aspects of vocabulary, such as receptive and expressive vocabulary, and listening comprehension (Storch & Whitehurst, 2002; Foorman et al., 2015; Lervåg et al., 2017). Thus, the latent variable in the present study was narrower than in some previous studies, and this may have reduced the strength of its correlation with reading comprehension.

4.3 Relationship among Background Variables, Domain-General Skills and Reading Comprehension

We found a moderate correlation between socio-economic background variables and reading comprehension. SES was determined with a large variety of measures (e.g., parental educational level, occupation, free/reduced school lunch and other broader measures, such as number and use of books at home etc.), but this result was consistent across studies. One explanation for this relationship is that educational background affects the home literacy environment provided to the children (Hoff, 2006). However, it is also likely that at least part

of the relationship between parental background measures and children's reading comprehension can be accounted for by heritable influences and is not purely a direct influence from parental education and the home literacy environment on children's reading comprehension (van Bergen, van Zuijen, Bishop, & de Jong, 2017; Puglisi, Hamilton, Hulme, & Snowling, 2017). Domain-general skills, such as nonverbal IQ, also showed a moderate correlation with reading comprehension, although the sizes of such correlations varied widely between studies (r 's ranged between $-.05$ and $.61$). For measures of working memory, non-word repetition had a smaller correlation with reading comprehension than the other predictors we examined, while sentence repetition showed a moderate correlation. The magnitude of the correlation was in line with a recent meta-analysis, which shows that working memory correlated with reading comprehension (Peng et al., 2018). However, Peng et al. found that the relationship between working memory and reading comprehension is fully accounted for by word recognition skills and vocabulary. Because working memory could not be included in the MASEM analyses, we could not examine this here, but similar findings have also been demonstrated in other studies (Lervåg et al., 2017).

4.4 Relationships between Linguistic Comprehension Skills and Word Recognition Skills

Gough and Tunmer (1986) argued that the two components of the simple view of reading make independent contributions to reading comprehension, which is line with the distinct linguistic comprehension and code-related pathways we identified in the present study. Yet, a different set of findings may nuance our understanding of these component skills. The individual studies reviewed here vary in the strength of the relationship found between code-related predictors and linguistic comprehension skills (Dickinson, McCabe, Anastasopoulos, Peisner-Feinberg, & Poe, 2003; Metsala, 1999). However, when synthesising the 42 studies in our MASEM model, we found a strong relationship between code-related skills and linguistic comprehension. This result aligns with observations from recent studies

demonstrating that early in development, there is a strong indirect effect of linguistic comprehension skills on word recognition via letter/sound skills (Hjetland et al., 2018; Hulme et al., 2015). Thus, linguistic comprehension creates the foundation for both code-related skills and word recognition.

Our MASEM results show an indirect path to reading comprehension from code-related skills through word recognition. This supports the idea that the foundations for word recognition are created in early childhood and in turn affect reading comprehension (Hjetland et al., 2018; Storch & Whitehurst, 2002). This result also supports findings from previous studies showing that poor word recognition skills and poor code-related skills prior to reading instruction can act as a bottleneck that hampers the development of reading comprehension (Hjetland et al., 2018; Storch & Whitehurst, 2002).

4.5 The Influence of Assessment Type

In our models there were clear differences in the strength of the path estimates associated with the use of open-ended and multiple-choice question formats. Reading comprehension measures with open-ended questions showed a stronger relationship with linguistic comprehension in preschool compared with multiple-choice tests. This finding is in line with studies showing that tests using a multiple-choice format allowed children to answer test questions based on background knowledge without reading the passage (Keenan & Betjemann, 2006) and that tests with open-ended questions are more dependent on linguistic comprehension skills than tests with multiple-choice or a cloze procedure (Keenan, Betjemann, & Olson, 2008). Importantly, most of the studies in our analysis measured reading comprehension by using different editions of the Woodcock-Johnson test battery. This may limit the operationalisation of the theoretical reading comprehension construct and effect the explained variance of the two components both in the primary studies and reviews. Thus, the

type of reading comprehension measure can potentially have a large impact on the conclusions reached in a study.

4.6 Methodological Reflections

Besides synthesizing correlations, quantifying the between-study variation, and explaining this variation by study features in the first part of our study, we tested a specific model that represented the simple view of reading in the second part. As noted earlier, when multiple correlations between multiple variables are synthesized (i.e., correlation matrices), their multivariate nature must be considered to circumvent severe bias in the parameters of the tested models (Cheung, 2015). This step, however, required positive definite correlation matrices and ultimately reduced the pool of available primary studies. On the one hand, this reduction of the overall sample size can be considered a key limitation of the MASEM approach we have taken (i.e., the TSSEM approach), as it limits not only the accuracy of the pooled correlation matrices but also the possibilities to examine moderation effects through, for instance, subgroup analyses (Jak & Cheung, 2018). Researchers should therefore be aware that aggregating multiple correlations in the first stage of the TSSEM approach can reduce the number of studies included in their meta-analyses. Moreover, results obtained from this approach and those obtained from aggregating multiple correlations through separate meta-analyses (i.e., the so-called univariate approach; Cheung, 2015) may not be comparable and are likely to show substantial differences (Cheung & Hong, 2017). On the other hand, excluding correlation matrices from the pool of primary studies is necessary for both the pooling of correlation matrices (stage-1 TSSEM) and the specification of any structural equation model (stage-2 TSSEM; see also Jak, 2015). In fact, including correlation matrices that do not fulfil this criterion can not only result in estimation or convergence issues in both TSSEM stages, but also compromise the validity of parameter and model interpretations (Cheung, 2015).

Overall, despite its current limitations, correlation-based MASEM through TSSEM represents a valuable approach to testing hypotheses based on structural equation models (Cheung & Cheung, 2016)—“valuable” because (a) it correctly aggregates multiple correlations between multiple variables extracted from the primary studies, (b) it allows researchers to test and compare structural equation models, (c) it is based on the correct sample sizes (as opposed to the univariate approach that is typically based on the harmonic mean of the sample sizes reported in the primary studies). TSSEM, as a correlation-based MASEM approach, brings together meta-analysis and structural equation modelling and addresses the shortcomings of existing approaches. We believe that future methodological research will extend the repertoire of types of structural equation models (e.g., allowing for the inclusion of latent interaction terms) and advance the field of quantitative meta-analyses.

4.7 Implications for Instruction and Recommendations for Future Studies

Our study has several important messages and implications for educational practice. First, we have established that reading comprehension grows out of skills established at an early age. Thus, in Grade 1, there are already large differences between children in their abilities that are prerequisites for learning to read. Randomised controlled trials have also demonstrated that language skills can be improved before starting school through small group language teaching (Hagen, Melby-Lervåg, & Lervåg, 2017; Rogde, Melby-Lervåg, & Lervåg, 2015; Fricke, Bowyer-Crane, Haley, Hulme, & Snowling, 2013). Moreover, after starting school, it is vital to ensure that children develop well-functioning word recognition skills, because poor word recognition creates a bottleneck for the development of reading comprehension. In short, our study shows clearly that programmes aimed at providing language enrichment for children in the preschool years are likely to have beneficial effects on the development of both word recognition and reading comprehension.

In addition, our review detected risks of bias in several primary studies due to low statistical power, attrition, reliability issues and the lack of quality of assessment instruments. Although replication problems have received most attention in experimental studies (Open science Collaboration, 2015), replication is also critical to longitudinal studies. Despite the limitations associated with their design, longitudinal observation studies still create a vital foundation for generating hypotheses and refining theories that can be tested in randomised controlled trials. Thus, it is important that the conclusions of longitudinal studies be reliable. A recent recommendation pointed out that, for observational studies of clinical groups and in educational research, large samples and the use of latent variables are critical to avoid bias (Tackett et al., 2017). Based on the risks of bias demonstrated in this review, this is an important recommendation for future studies to ensure our knowledge base is as robust as possible.

5. References

References marked with an asterisk () indicate primary studies that were included in the meta-analysis.*

- *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.
- 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*. Englewood, NJ: BioStat.
- 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
- *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
- Caravolas, M., Lervåg, A., Defior, S., Málková, G. S., & Hulme, C. (2013). Different patterns, but equivalent predictors, of growth in reading in consistent and inconsistent orthographies. *Psychological science, 24*, 1398–1407. doi: 0956797612473122
- *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
- Cheung, M. W.-L. (2013). Multivariate meta-analysis as structural equation models. *Structural Equation Modeling, 20*(3), 429–454. doi:10.1080/10705511.2013.797827
- Cheung, M. W.-L. (2015). *Meta-analysis: A structural equation modeling approach*. Chichester, West Sussex: John Wiley & Sons, Ltd.
- Cheung, M. W.-L. (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

- 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*, 140–155. doi:10.1002/jrsm.1166
- Cheung, M. W., & Chan, S. F. (2009). A two-stage approach to synthesizing covariance matrices in meta-analytic structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 16*, 28–53. doi:10.1080/10705510802561295
- Cheung, M. W.-L., & Hong, R. Y. (2017). Applications of meta-analytic structural equation modeling in health psychology: Examples, issues, and recommendations. *Health Psychology Review, 11*(3), 265-279. doi:10.1080/17437199.2017.1343678
- 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
- *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
- 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

- *Cudina-Obradovic, M. (1999). Determinants of reading achievement in the 1st grade: One year follow-up study. *Hrvatska Revija za Rehabilitacijska Istraživanja*, 35, 79–94.
- Dang, C.-P., Braeken, J., Colom, R., Ferrer, E., & Liu, C. (2015). Do processing speed and short-term storage exhaust the relation between working memory capacity and intelligence? *Personality and Individual Differences*, 74, 241-247.
doi:10.1016/j.paid.2014.10.012
- Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner-Feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology*, 95(3), 465–481. doi: 10.1037/0022-0663.95.3.465
- *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
- Dunn, L., & Dunn, L. (2007). *Peabody picture vocabulary test* (4th ed.). Minneapolis, MN: Pearson.
- *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
- Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel Plot–Based Method of Testing and Adjusting for Publication Bias in Meta–Analysis. *Biometrics*, 56(2), 455-463. doi:10.1111/j.0006-341X.2000.00455.x
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

- *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
- Florit, E., & Cain, K. (2011). The simple view of reading: Is it valid for different types of alphabetic orthographies? *Educational Psychology Review*, *23*(4), 553–576. doi: 10.1007/s10648-011-9175-6.
- 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
- *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
- Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, *54*(3), 280–290. doi: 10.1111/jcpp.12010
- *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

- García, J. R., & Cain, K. (2014). Word recognition 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
- *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
- Gough, P. B., & Tunmer, W. E. (1986). Word recognition, reading, and reading disability. *Remedial and Special Education, 7*, 6–10. doi: 10.1177/074193258600700104
- *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
- Hagen, Å.M. Melby-Lervåg, M., & Lervåg, A. (2017). Improving language comprehension in preschool children with language difficulties: A cluster randomized trial. *Journal of Child Psychiatry and Psychology, 58*. doi: 10.1111/jcpp.12762
- *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

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

Hjetland, H. N., Brinchmann, E. I., Scherer, R., & Melby-Lervåg, M. (2017). Preschool predictors of later reading comprehension ability: a systematic review. *Campbell systematic reviews, 14*. doi: 10.4073/csr.2017.14.

Hjetland, H. N., Lervåg, A., Lyster, S.-A. H., Hagtvet, B. E., Hulme, C., & Melby-Lervåg, M. (2019). Pathways to Reading Comprehension: A Longitudinal Study From 4 to 9 Years of Age. *Journal of Educational Psychology, 111*, 751- 763. doi: 10.1037/edu0000321

Hong, R. Y., & Cheung, M. W.-L. (2015). The structure of cognitive vulnerabilities to depression and anxiety: evidence for a common core etiologic process based on a meta-analytic review. *Clinical Psychological Science, 3*, 892–912. doi:10.1177/2167702614553789

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review, 26*, 55–88.

Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

- *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
- Jak, S. (2015). *Meta-analytic structural equation modelling*. Utrecht, The Netherlands: Springer Verlag.
- Jak, S., & Cheung, M. W.-L. (2018). Testing moderator hypotheses in meta-analytic structural equation modelling using subgroup analysis. *Behavior Research Methods*, 50, 1359–1373. doi: 10.3758/s13428-018-1046-3
- Justice, L. M., Lomax, R., O’Connell, A., Pentimonti, J., Petrill, S. A., Piasta, S. B., ... Language and Reading Research Consortium (LARRC) (2017). Oral language and listening comprehension: Same or different constructs? *Journal of Speech, Language, and Hearing Research*, 60, 1273–1284. doi: 10.1044/2017_JSLHR-L-16-0039
- *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.
- Keenan, J. M. & Betjemann, R. S. (2006). Comprehending the Gray Oral Reading Test without reading it: Why comprehension tests should not include passage-independent items. *Scientific Studies of Reading*, 10(4), 363–380. doi: 10.1207/s1532799xssr1004_2
- Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on word recognition and oral comprehension. *Scientific Studies of Reading*, 12, 281–300. doi:10.1080/10888430802132279

- *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
- *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
- 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
- *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.
- Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2017). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, 1–18. doi:10.1111/cdev.12861
- Little, T. D. (2013). *Longitudinal structural equation modeling*. New York: Guilford Press.
- Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. (2004). Why Multicollinearity Matters: A Reexamination of Relations Between Self-Efficacy, Self-Concept, and Achievement. *Journal of Educational Psychology*, *96*(3), 518-522. doi:10.1037/0022-0663.96.3.518
- Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary Psychometrics* (pp. 275-340). Mahwah, NJ: Lawrence Erlbaum.
- Metsala J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, *91*, 3–19. doi:10.1037/0022-0663.91.1.3
- 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
- *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

*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. doi:10.1037/0012-1649.40.5.665

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

*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'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, *41*(5), 673–690. doi:10.1007/s11135-006-9018-6

*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

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

*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

Peng, P., Barnes, M., Wang, C., Wang, W., Li, S., Swanson, H. L., . . . Tao, S. (2018). A meta-analysis on the relation between reading and working memory. *Psychological Bulletin*, 144(1), 48–76. doi:10.1037/bul0000124

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

*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

Protopapas, A., Simos, P. G., Sideridis, G. D., & Mouzaki, A. (2012). The components of the simple view of reading: A confirmatory factor analysis. *Reading Psychology*, 33(3), 217–240. doi:10.1080/02702711.2010.507626

- Puglisi, M. Hamilton, L., Hulme, C. & Snowling, M. (2017). The home literacy environment is a correlate, but perhaps not a cause, of variations in children's language and literacy development. *Scientific Studies of Reading*. doi: 10.1080/10888438.2017.1346660
- Quinn, J. & Wagner, R. K. (2018). Using meta-analytic structural equation modeling to study developmental change in relations between language and literacy. *Child Development*. Advance online publication. doi: 10.1111/cdev.13049
- RAND Reading Study Group. (2002). *Reading for understanding: Towards an R&D program in reading comprehension*. Washington, DC: RAND Education
- *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
- Rogde, K., Melby-Lervåg, M., & Lervåg, A. (2016). Improving the general language skills of second-language learners in kindergarten: A randomized controlled trial. *Journal of Research on Educational Effectiveness*, 9, 150–170. doi:10.1080/19345747.2016.1171935
- *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
- Sheng, Z., Kong, W., Cortina, J. M., & Hou, S. (2016). Analyzing matrices of meta-analytic correlations: Current practices and recommendations. *Research Synthesis Methods, 7*, 187-208. doi:10.1002/jrsm.1206
- *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

- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology*, 38(6), 934–947. doi:10.1037/0012-1649.38.6.934
- Tackett, J. L., Lilienfeld, S. O., Patrick, C. J., Johnson, S., L. Krueger, R. F. Miller, D. J., . . . Shrout, P. E. (2017). It's time to broaden the replicability conversation: Thoughts for and from clinical psychological science. *Perspectives on Psychological Science*, 12, 742–756. doi:10.1177/1745691617690042
- 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
- *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 word recognition and reading comprehension skills. *Early Education and Development*, 19, 188–207. doi:10.1080/10409280701839304
- Teale, W. H., & Sulzby, E. (1986). *Emergent Literacy as a Perspective for Examining How Young Children Become Writers and Readers*. Norwood, NJ: Ablex.
- Torppa, M., Georgiou, G. K., Lerkkanen, M.-K., Niemi, P., Poikkeus, A.-M., & Nurmi, J.-E. (2016). Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *Merrill-Palmer Quarterly*, 62(2), 179-206. doi: 10.13110/merrpalmquar1982.62.2.0179
- *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
- van Bergen, E., van Zuijen, T., Bishop, D., de Jong, P. F. (2017). Why are home literacy environment and children's reading skills associated? What parental skills reveal. *Reading Research Quarterly*, *52*, 147–160. doi:10.1002/rrq.160
- Verhoeven, L., & van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and Writing*, *25*, 1805–1818. doi:10.1007/s11145-011-9346-3
- *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
- 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.

Tables

Table 1

Average correlations between the outcomes reading comprehension and word decoding and the predictors

<i>Outcome variable</i>	<i>Predictor</i>	<i>Number of studies (k) and study participants (n_k)</i>	<i>Average correlation with 95 % confidence intervals</i>	<i>Average age range between measurement occasions of the predictor and the outcome variable [years]</i>	<i>Variation in correlations reported in the studies (minimum and maximum r)</i>	<i>Heterogeneity (I²) with significance test of heterogeneity</i>
<i>Reading comprehension</i>	Phoneme awareness	36 (6,626)	.40 [.36, .44]	5.5–8.4	-.05 to .73	64.7 %**
	Rhyme awareness	15 (1,741)	.39 [.32, .45]	5.3–8.3	.17 to .63	57.9 %**
	Letter knowledge	26 (3,869)	.42 [.38, .46]	5.6–9.0	-.13 to .67	43.2 %**
	Rapid naming	17 (3,746)	-.34 [-.41, -.28]	5.6–8.4	-.55 to .15	71.5 %**
	Vocabulary	45 (5,907)	.42 [.38, .46]	5.2–9.0	-.13 to .67	71.3 %**
	Grammar	16 (1,857)	.41 [.32, .49]	5.2–8.1	.15 to .65	76.5 %**
	Sentence memory	9 (1,237)	.36 [.23, .47]	5.3–9.1	.05 to .56	81.5 %**
	Non-word repetition	7 (841)	.17 [.10, .23]	5.2–8.3	-.01 to .25	0.0 %
	Non-verbal intelligence	21 (11,632)	.35 [.30, .41]	5.5–8.7	-.05 to .61	72.8 %
	Socio-economic	12 (1,595)	.36 [.30, .41]	5.5–8.9	.10 to .46	31.0 %

Running head: PRESCHOOL PATHWAYS TO READING COMPREHENSION

<i>Word recognition</i>	status Phoneme awareness	28 (4,772)	.37 [.31, .43]	5.4–8.0	.01 to .78	73.9 %**
	Rhyme awareness	13 (1,662)	.32 [.24, .40]	5.4–8.5	.14 to .62	67.3 %**
	Letter knowledge	16 (2,432)	.38 [.31 -.45]	5.6–8.2	-.04 to .62	76.2 %**
	Rapid naming	14 (3,285)	-.37 [-.44, -.45]	5.4–8.0	-.55 to .28	76.3 %**

Note. ** $p < .01$.

Running head: PRESCHOOL PATWAYS TO READING COMPREHENSION

Table 2

Continuous Moderator analysis

Correlations with reading comprehension	Moderator											
	Age initial assessment				Age at reading assessment				Months of reading instruction at reading assessment			
	β	R ²	<i>k</i>	p-value	β	R ²	<i>k</i>	p-value	β	R ²	<i>k</i>	p-value
Phoneme awareness	.0014	.00	36	.634	.0006,	.00	36	.618	.0012	.00	36	.254
Rhyme awareness	-.0051	.19	15	.241	.0039*	.43	15	.012	.0027	.09	15	.063
Letter knowledge	-.0051	.09	26	.063	-.0005	.00,	26	.544	-.0002	.00	26	.784
Rapid naming	.0087	.00	17	.352	-.0028	.30	17	.075	-.0032*	.44	17	.012
Vocabulary	-.0018	.07	45	.394	-.0001	.00	45	.945	.0010	.05	43	.229
Grammar	-.0007	.00	16	.893	.0012,	.00	16	.560	.0011	.00	16	.562
Sentence repetition	.0018	.00	9	.852	.0034*	.46	9	.042	.0028	.27	9	.115
Nonverbal intelligence	-.0104**	.52	21	< .001	.0010	.00	21	.388	.0017	.00	21	.077

Note. * $p < .05$, ** $p < .01$.

Table 3

Pooled correlation matrix under a random-effects model (k = 42 correlation matrices, m = 222 correlations, N = 6,696 children)

	PHONEME	LK	VOC	GRA	WDEC
LK					
<i>r</i>	.45*				
95 % CI	[.40, .50]				
τ^2	0.0035				
<i>SE</i> (τ^2)	0.0026				
<i>I</i> ²	38.7 %				
VOC					
<i>r</i>	.33*	.33*			
95 % CI	[.27, .38]	[.28, .38]			
τ^2	0.0058	0.0029			
<i>SE</i> (τ^2)	0.0036	0.0038			
<i>I</i> ²	50.6 %	32.3 %			
GRA					
<i>r</i>	.39*	.34*	.42*		
95 % CI	[.31, .47]	[.26, .41]	[.36, .48]		
τ^2	0.0051	0.0027	0.0007		
<i>SE</i> (τ^2)	0.0048	0.0038	0.0026		
<i>I</i> ²	45.3 %	29.8 %	10.9 %		
WDEC					
<i>r</i>	.43*	.49*	.34*	.34*	
95 % CI	[.39, .48]	[.45, .52]	[.29, .38]	[.26, .41]	
τ^2	0.0051	0.0002	0.0023	0.0029	
<i>SE</i> (τ^2)	0.0029	0.0008	0.0023	0.0044	
<i>I</i> ²	48.9 %	3.4 %	28.1 %	31.7 %	
RC					
<i>r</i>	.43*	.42*	.42*	.36*	.73*
95 % CI	[.38, .47]	[.39, .46]	[.36, .47]	[.27, .44]	[.67, .79]
τ^2	0.0079	0.0001	0.0143	0.0084	0.0218
<i>SE</i> (τ^2)	0.0035	0.0017	0.0054	0.0059	0.0071
<i>I</i> ²	62.5 %	0.0 %	75.2 %	58.1 %	95.6 %

Note. Phoneme = phoneme awareness, LK = letter knowledge, VOC = vocabulary and listening comprehension (verbal ability), GRA = grammar, WDEC = concurrent word decoding, RC = reading comprehension; *r* = aggregated correlation, τ^2 = variance between correlation matrices (i.e., study samples), *I*² = heterogeneity coefficient based on the *Q*-statistic (Higgins & Green, 2011). Wald's confidence intervals (CIs) are shown. * *p* < .05

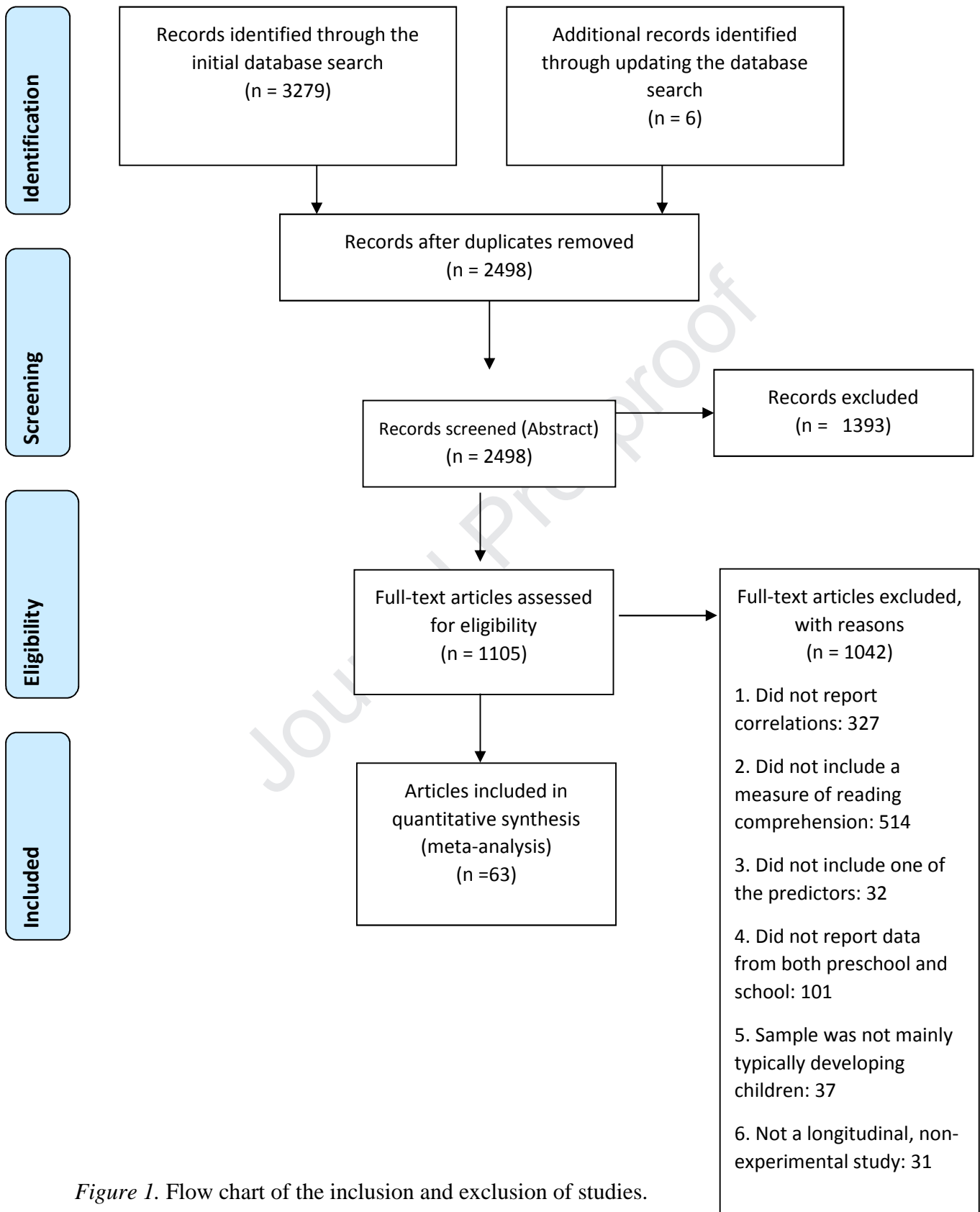


Figure 1. Flow chart of the inclusion and exclusion of studies.

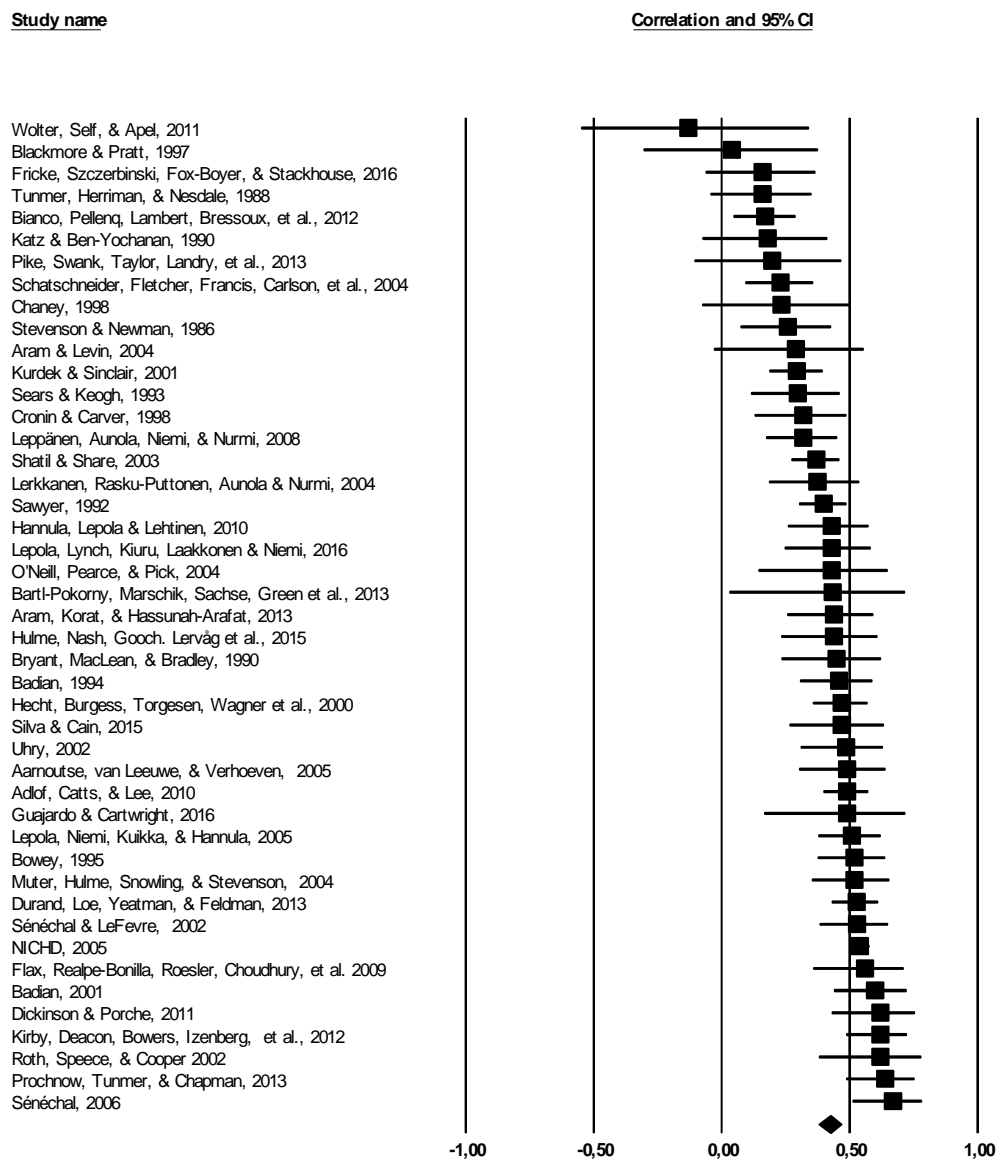


Figure 2. The correlation between vocabulary and reading comprehension.

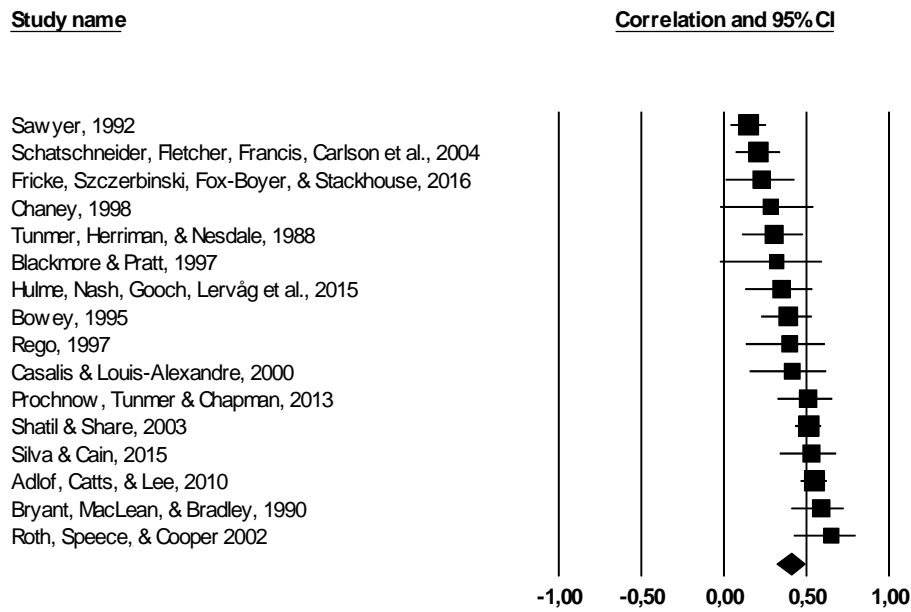


Figure 3. The correlation between grammar and reading comprehension.

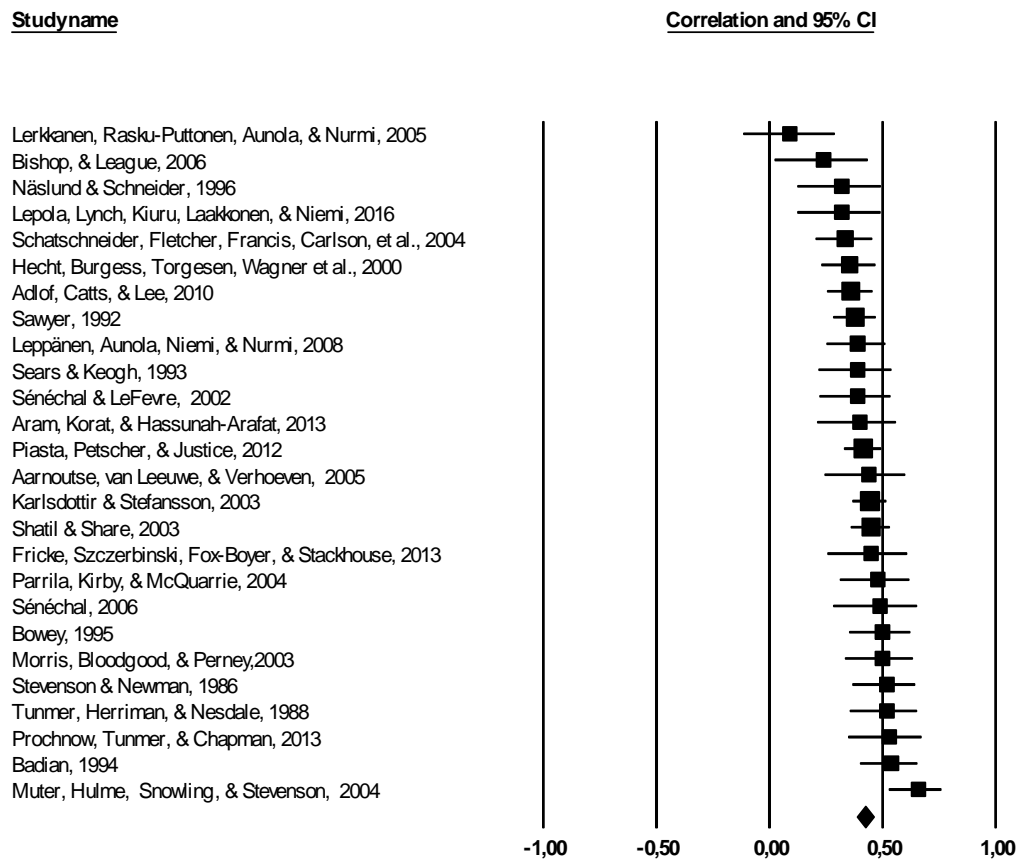


Figure 4. The correlation between letter knowledge and reading comprehension.

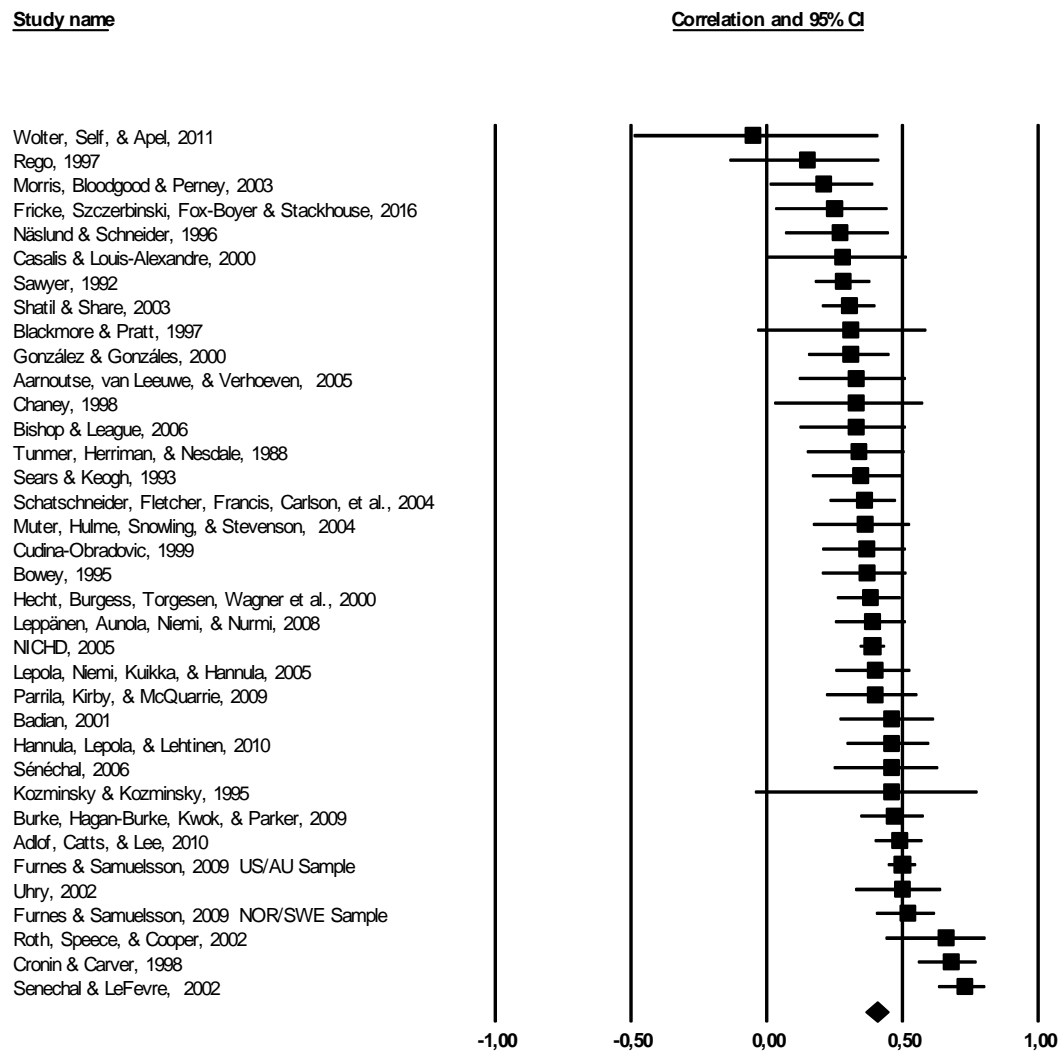


Figure 5. The correlation between phoneme awareness and reading comprehension.

PRESCHOOL PREDICTORS OF READING COMPREHENSION

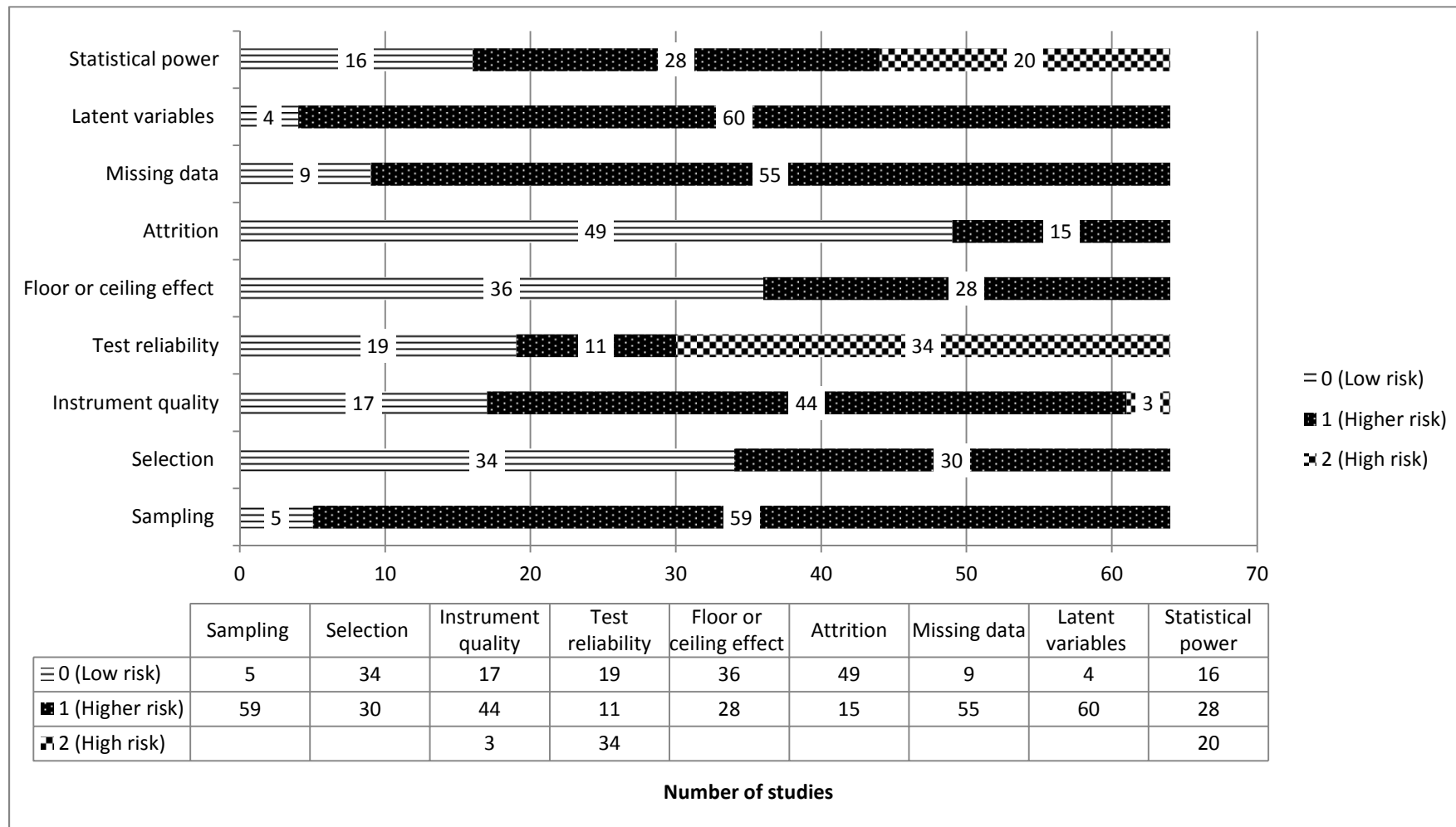


Figure 6. Study quality in the included studies.

PRESCHOOL PREDICTORS OF READING COMPREHENSION

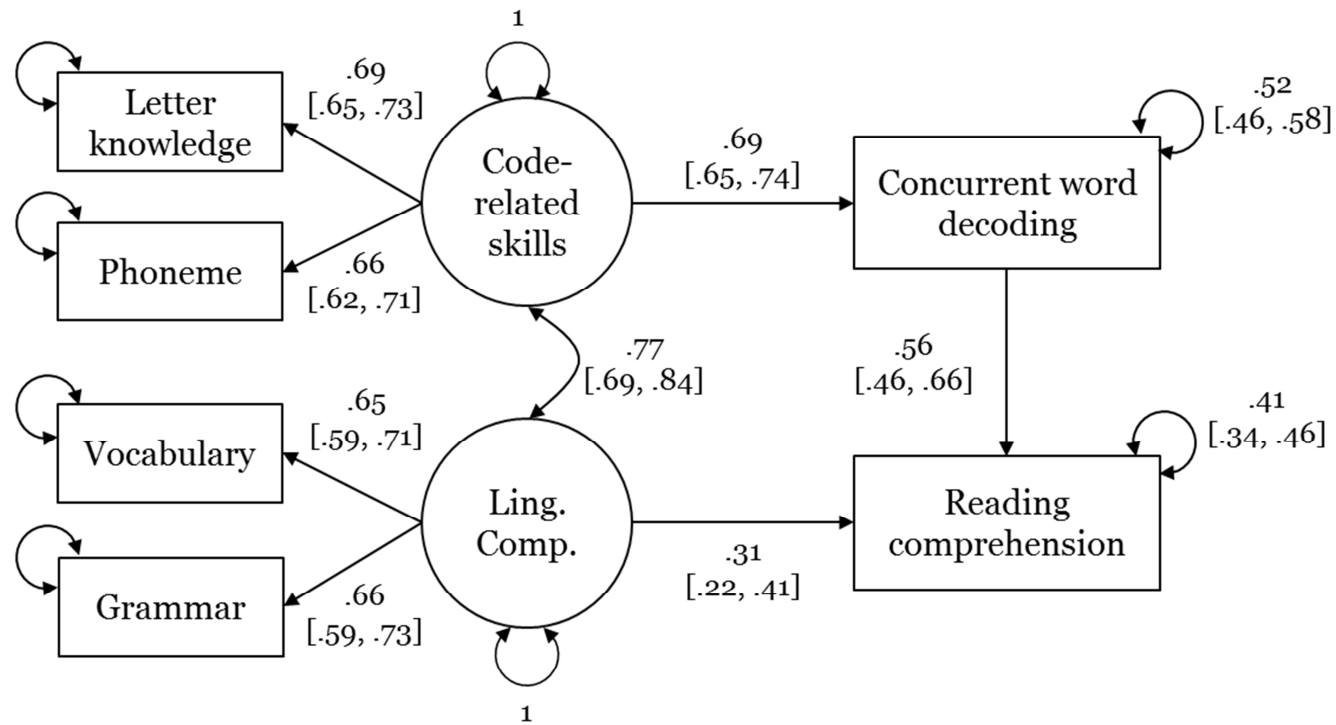
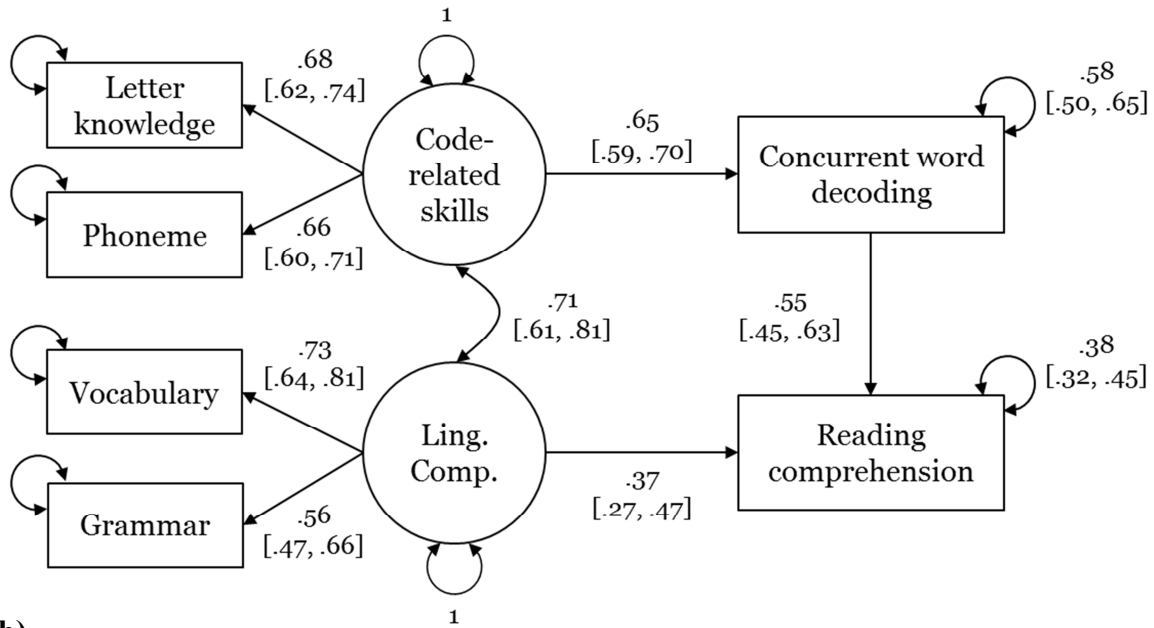


Figure 7. Meta-analytic structural equation model describing the pathways of reading comprehension for the overall sample of studies.

Note. 95% confidence intervals are shown in brackets.

(a)



(b)

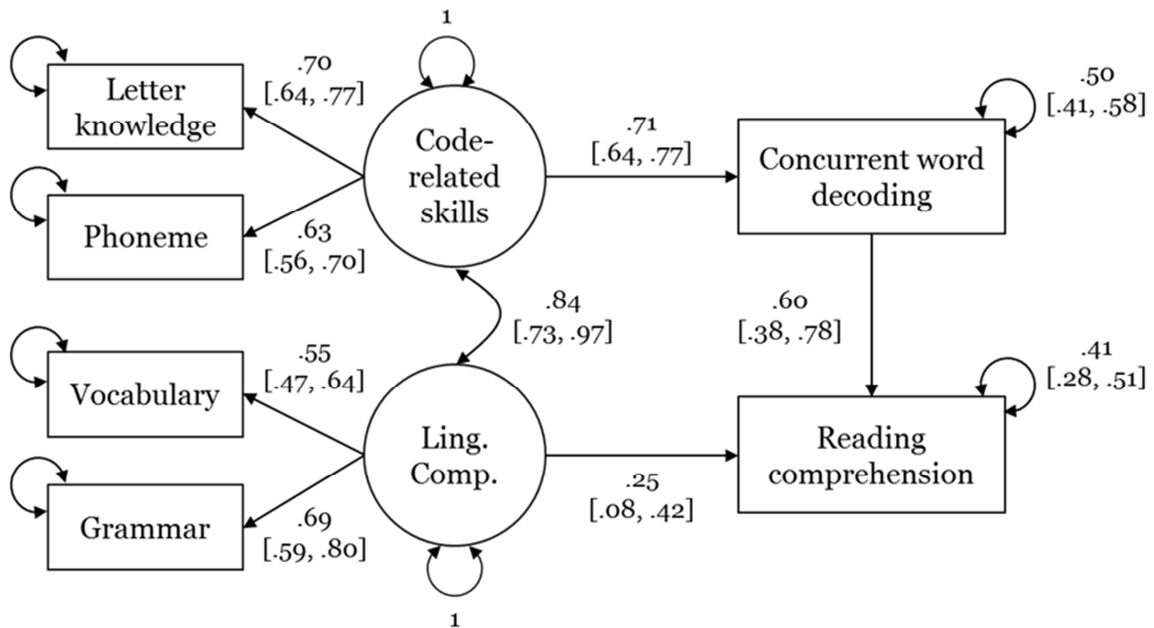


Figure 8. Multi-group, meta-analytic structural equation model describing the pathways of reading comprehension for (a) early reading, and (b) later reading.

Note. 95% confidence intervals are shown in brackets.

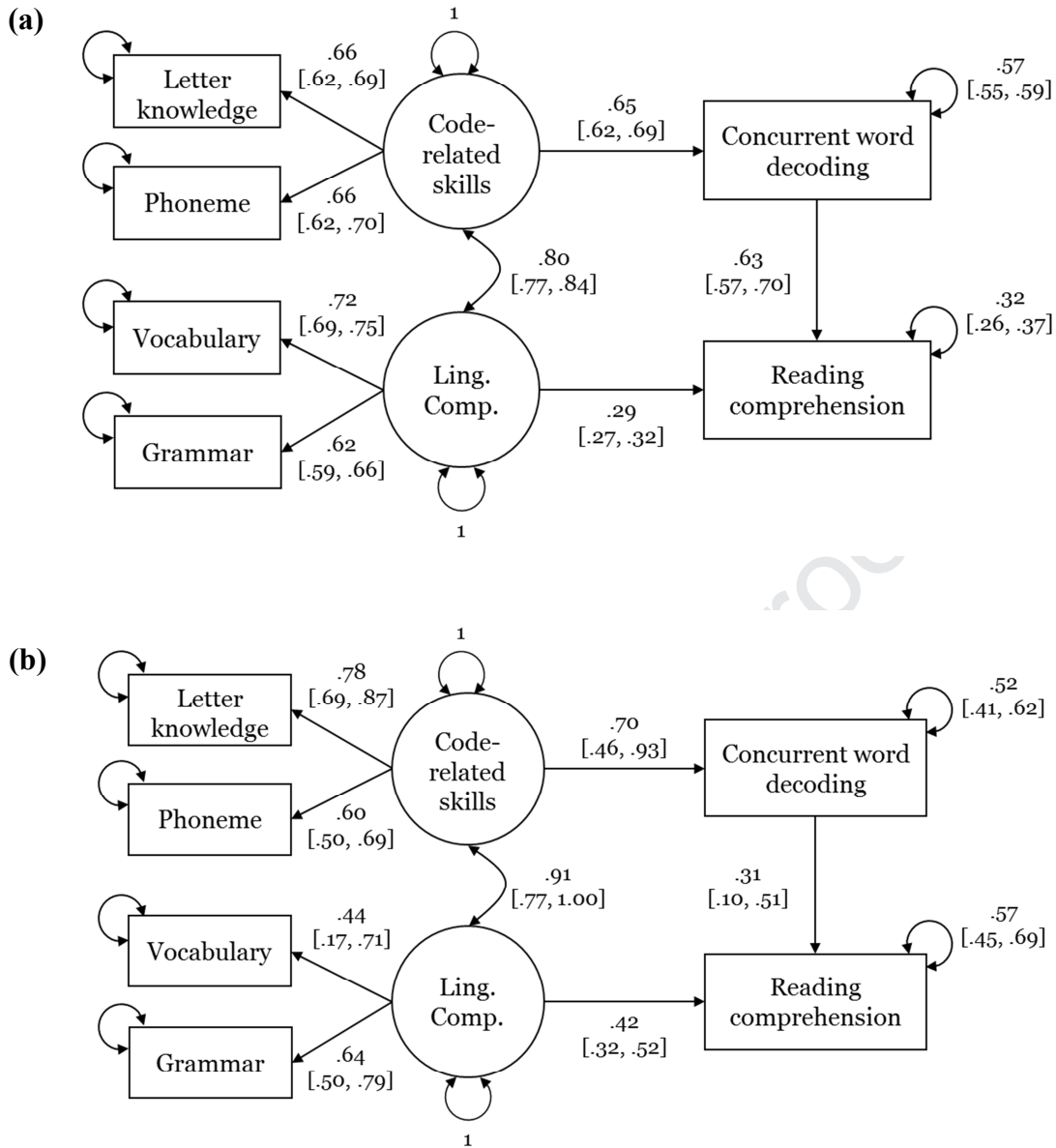


Figure 9. Multi-group, meta-analytic structural equation model describing the pathways of reading comprehension for (a) multiple-choice, and (b) open-ended assessments of reading comprehension.

Note. 95% confidence intervals are shown in brackets.

Highlights

- Moderate longitudinal correlations between skills in preschool and later reading comprehension
- Code-related skills and linguistic comprehension are separable predictors of later reading comprehension
- Two distinct but related pathways from preschool language to reading comprehension in school
- Linguistic comprehension has a key role in the development of reading comprehension