Psychology of Language and Communication 2021, Vol. 25, No. 1

DOI: 10.2478/plc-2021-0008

Lucas Federico Sterpin[1], Sofía Soledad Ortiz[1], Jésica Formoso[1, 2], Juan Pablo Barreyro[1, 2]
[1] School of Psychology, Universidad de Buenos Aires, Argentina
[2] National Scientific and Technical Research Council (CONICET), Ministry of Science, Technology and Innovation, Argentina

The role of vocabulary knowledge on inference generation: A meta-analysis

Successful text comprehension results in a coherent mental model of the situation being described. To achieve this, the reader has to infer certain information by connecting parts of the text to their prior knowledge. An important construct involved in this process is vocabulary knowledge, usually divided into breadth and depth. We conducted a meta-analysis on 23 studies, and explored the fit of five different models to establish an effect size of both dimensions of vocabulary on inference making, as well as its developmental trajectory in children aged 3-12. We found a significant and moderate effect of vocabulary knowledge of both modalities. Vocabulary type was not a significant moderator, but age was, meaning that there was a similar effect for both breadth and depth and that the strength of the correlations decreased with age. Heterogeneity was high overall, meaning that more moderators should be assessed in future studies.

Key words: children, meta-analysis, vocabulary, inference, moderator

Address for correspondence: Lucas Federico Sterpin, School of Psychology, Universidad de Buenos Aires, Buenos Aires, Argentina.
E-mail: sterpinlucas@gmail.com

Text or discourse comprehension depends on the execution and integration of many cognitive processes (van den Broek et al., 2005; van den Broek & Espin, 2012). The cognitive processes of reading comprehension can be divided into two categories: (a) lower level processes that involve translating the written code into meaningful language units and (b) higher level processes that involve combining these units into a meaningful and coherent mental representation (Kendeou et al., 2014). Both lower and higher level processes are needed for a successful comprehension.

A good reader uses their prior knowledge to interpret the text or discourse information to construct a coherent mental representation of what the text is about (Kendeou et al., 2014; Kintsch, 1988). This representation is the foundation from which the reader can retell the story, apply knowledge that has been acquired from the text, identify the theme, and so forth.

One source of reading comprehension problems concerns the ability to generate inferences. Inferences allow the reader or listener to construct meaningful connections between text elements and relevant background knowledge (Oakhill et al., 2003; van den Broek, 1990), and for this reason, inferences are crucial for comprehension. In fact, research shows that good readers generate more and better inferences than those of lesser skill-level (Bowyer-Crane & Snowling, 2005; Singer & Ritchot, 1996), and the role of inference generation in text comprehension has been widely corroborated (Florit et al., 2011; Oakhill & Cain, 2013).

Inferences establish local and global connections throughout the text (Currie & Cain, 2015; Graesser et al., 1994). Local connections are necessary for local coherence and typically involve integrating separate but related statements within the text. Global connections involve inferring goals that motivate particular actions or establishing the overall theme of a text, and this often relies on the ability to connect ideas that can be distributed throughout the text and are not explicitly signaled by a single word (Currie & Cain, 2015). Even though Cain and Oakhill (2014) report that vocabulary is a stronger predictor of elaborative rather than connective inferences in reading, there is statistical evidence that suggests that the construct is fairly unidimensional and that the separate types of inferences cannot be measured in a consistent and dependable way (Muijselaar, 2018). For this reason, we will not distinguish between connective or elaborative inferences in the following analyses.

Inference making requires a solid knowledge of the vocabulary involved in the text. Vocabulary knowledge is divided into two core components: breadth and depth. Most researchers agree that vocabulary breadth refers to the number of words a person knows, but there is less consensus over what vocabulary depth entails. Initially, it referred to the ability to define words, considering its multiple possible meanings and where in a sentence it would be used best (Cronbach, 1942; Li & Kirby, 2015). However, as psycholinguistic research advanced, scientists added more caveats to this construct. For example, Perfetti's lexical quality hypothesis (Perfetti, 2007; Perfetti & Hart, 2002) claims that a good performance on language related processes depends, at least in part, on the quality of the representation of the orthography, phonetics, grammar, pragmatics, and semantics of the vocabulary that the reader has on their mental lexicon. If the reader lacks quality in some of these aspects, then the respective process would be hindered, and a complete comprehension of the passage would be at risk.

Another influential model is that by Proctor et al. (2012), which includes both morphological and syntactic awareness as components of vocabulary depth, arguing that knowing how to interpret or derive different morphological forms of words from their roots helps to grow vocabulary knowledge. In turn, syntactic awareness could allow the reader to understand the syntactic structure to which the newly formed word fits best.

Vocabulary breadth and depth are associated with different outcomes. Vocabulary breadth is linked to reading decoding, because phonological representations map onto orthographic representations (Ouellette, 2006; Wise et al., 2007). At the same time, larger vocabulary knowledge facilitates efficient word retrieval and faster word identification (Wise et al., 2007). Larger depth of vocabulary is associated with stronger lexical–semantic representations, which leads to more efficient semantic access and better reading comprehension (Nation & Snowling, 1999; Ouellette, 2006; Paul & Gustafson, 1991). This last point, coupled with the lexical quality hypothesis, allows us to think about a deep connection between vocabulary depth and inference-making ability; without a clear knowledge of a word's meaning and possible conceptual relations, one would not have anything to activate during reading and would not be able to generate a good inference.

Following this idea, Currie and Cain (2015) tested 130 children aged 5-10 on both breadth and depth of vocabulary and both elaborative and connective inferences. They found an association between composite vocabulary score and global coherence inferences in children aged 6, 8, and 10, but found an association between this measure and local coherence inferences only in the 6- and 8-year-old groups. In the same vein, Ouellette (2006) found an association between both

Table 1. Vocabulary dimensions.

Breadth	Depth
It refers to the number of words a person knows, regardless of whether or not they know its meaning.	It refers to the knowledge that a person has about a word in terms of its semantic richness (polysemy, synonyms, semantic neighbours) and in terms of its morphological, phonological and syntactic qualities.
It is usually measured using naming tasks.	It is usually measured using tasks that require providing or recognizing the definition of a word, or by tasks that require manipulating some aspect of its meaning, such as a semantic fluency task.
It is linked to decoding, because phonological representations map onto orthographic representations.	It is linked to faster and more efficient word retrieval and identification, as well as to reading comprehension.

types of vocabulary and text comprehension, with a bigger effect size for depth rather than breadth. In contrast, Tannenbaum et al. (2006) reported a bigger effect size for breadth. It is important to note that the measure of vocabulary breadth used in this last experiment was critiqued by Oakhill et al. (2015), who considered it more appropriate as a measure of vocabulary depth, so more research in this area is necessary.

Regarding the development of this relation, different studies report a drop in the effect size of vocabulary on the ability to generate inferences as the child grows. For example, Lynch et al. (2008) found an association between vocabulary breadth and a comprehension measure that included inferential questions in 4-year-olds, but not in 6-year-olds. This is not fully congruent with the results of Currie and Cain (2015), who reported this drop at a later point in the child's development.

Therefore, the aim of the current meta-analysis was to evaluate the effect size of vocabulary on inference-making ability in children. We hypothesized that both breadth and depth of vocabulary would be positively correlated with inference-making ability, but we expected a greater effect size of vocabulary depth. We also analyzed the developmental trajectory of this relationship. Even though most researchers agree with a multi-component view of vocabulary depth, such as the ones described above, only a small number of studies actually use multiple tests to measure it, with definitions and synonym evaluations usually being the only aspects tested (Tran et al., 2020). Additionally, given that there is no consensus over what a test of vocabulary depth should be, we considered standardized tests or tasks that measure a child's ability to provide the definition or another semantic manipulation of a word as tasks of vocabulary depth, whereas we considered tests or tasks that do not require a definition as tasks of vocabulary breadth. An example of the former is the Vocabulary subtest of the Wechsler Intelligence Scale for Children (Wechsler, 2012) while an example of the latter is the British Picture Vocabulary Scale (Dunn et al., 2009). Focusing only on the semantic aspect would paint an incomplete picture of the contribution of vocabulary depth to inference making, but it would be one that is consistent with the literature. Future work could expand ours by analyzing the role of morphological, syntactic, and/or phonological awareness on inference making. Besides that, this strategy resulted in a more homogeneous sample of research methodologies, making our results more reliable.

Another distinction used by researchers when studying vocabulary learning (Milton, 2009) is the difference between receptive and productive vocabulary (Nation, 1990; Shahov, 2012, Fan, 2000; Laufer, 1998; Laufer & Paribakht, 1998; Henriksen, 1999; Nation, 2001; Schmitt, 2014). This difference is widely accepted in research of vocabulary acquisition (Shahov, 2012) and is linked to the memory recovery modality. In this field, receptive vocabulary is often defined as the ability to recognize the form of a word (Laufer et al., 2004), perceive its meaning (Webb, 2008), or provide its synonym (Webb, 2009). On the other hand, productive vocabulary is often defined as the ability to retrieve the form and meaning (Laufer et al., 2004; Webb, 2008), or to produce the word according to its context (Webb, 2009).

Methodology

To summarize the literature on the association between vocabulary breadth and depth and inference making, we followed the PRISMA standards (Moher et al., 2009). We performed a total of 12 searches in the PubMed, APA PsycNet, SciELO, and Wiley Online Library databases of articles published between January 1, 2000, and June 30, 2020, using the following terms: "vocabulary" and "inference making" (*n* = 197), "verbal ability" and "inference making" (*n* = 95), "prior knowledge" and "inference making" (*n* = 160), "vocabulario" and "inferencias" (*n* = 2), "habilidad verbal" and "inferencias" (*n* = 0), "conocimiento previo" and "inferencias" (*n* = 8). Eleven additional studies were identified from the reference lists of the articles deemed eligible. Of the identified records 167 were duplicated, leaving a total of 296 articles to screen.

For the studies to be included in the analysis:

- The sample had to consist of children aged between 3 and 12 years, with no developmental, learning, psychiatric, or neuropsychological conditions and/or sensory loss reported by the parents or the schools.
- The results had to include at least one correlation between a measure of vocabulary and a measure of inference-making ability.

Based on these criteria, 200 articles were excluded after reading their abstracts, leaving 96 full texts to be assessed for eligibility. Of the remaining studies, 72 were excluded for not meeting the inclusion criteria and, after further analysis, one additional record was excluded because the vocabulary measure was not clearly described. The 23 studies that were included in the analysis are detailed in Table 2.

Data Analysis

As we were assessing the association between two numerical variables, we used the Pearson correlation coefficient as a measure of the effect size (Cheung, 2014, 2019; Fernández-Castilla et al., 2020). We extracted all reported coefficients between vocabulary and inference making from each study and transformed them using Fisher's transformation.

Many studies reported more than one correlation coefficient, either because they included different age groups or because they assessed both vocabulary depth and breadth. This resulted in multiple dependent effect sizes nested within studies. To address this dependency, we conducted multi-level analyses which allowed

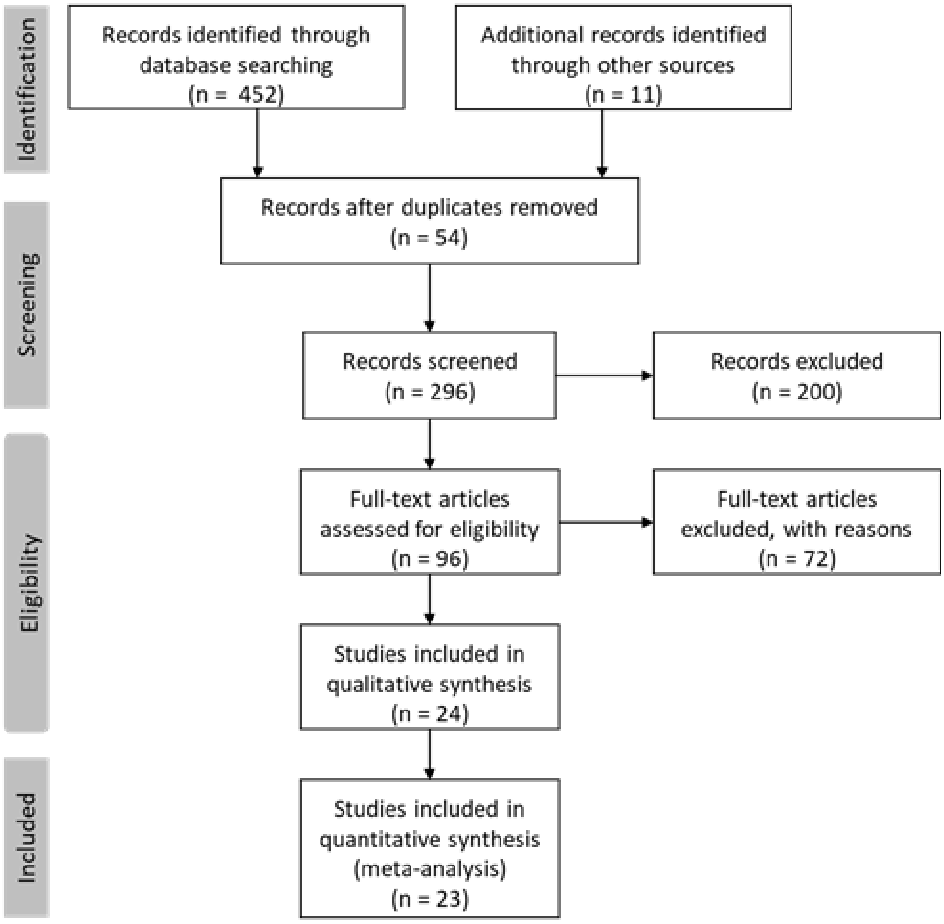

Figure 1. Meta-analysis flow diagram for study selection.

us to incorporate sampling variation for each coefficient (Level 1), within-study variation (Level 2), and between-study variation (Level 3; Moeyaert et al., 2017).

We adjusted five multi-level models. The first one included only the intercept to obtain the overall effect. This was the estimation of the true correlation coefficient assuming all studies measure the same variables. The second model included type of vocabulary (breadth and depth) as a moderator to assess if the estimated correlation coefficient varied within each level of that variable. The third model included receptive and productive vocabulary type as a moderator. A fourth model was adjusted using the mean age of the participants as a moderator to test if the association between vocabulary and inference making varied with it. Finally, a fifth model included vocabulary type (breadth and depth), receptive and productive vocabulary type, and age to assess the effect of these variables on the estimated coefficient when controlling for the other one. To estimate the parameters in the models, we used the restricted maximum likelihood method.

Table 2. Characteristics of the studies included in the analysis.

Authors	N	Sample age	Number of coefficients	Vocabulary task	Type of vocabulary assessed	Inference task	Type of inference assessed	Presentation	Type of text	Other variables
Cain & Oakhill (2014)	83	10 - 11	4	British Picture Vocabulary Scales – BPVS (Dunn et al., 1997) and two measures of vocabulary and similarities of Wechsler Intelligence Test for Children - III (Wechsler, 1994)	Breadth and depth	Inference tasks developed from Cain & Oakhill (1999)	Connective and elaborative	Oral and written	Narrative	Word reading, general comprehension and literal information comprehension
Cain et al. (2004)	271	8-11	6	Vocabulary subtest of Gates – MacGinitie, level 2 (MacGinitie et al., 2001) and British Picture Vocabulary Scale (Dunn et al., 1997)	Breadth	At time 1, The constructive inference task from Oakhill (1982). At time 2 and 3, stories with open, inferential questions extracted from Cain & Oakhill (1999)	No difference	Oral and written	Narrative	Reading ability, Verbal ability, working memory and comprehension (comprehension monitoring, knowledge of the structure of the story)
Chrysochoou et al. (2011)	92	8 - 10	3	Peabody Picture Vocabulary Test, 4th Ed (Dunn & Dunn, 2007).	Breadth	Stories based on Oakhill (1984) and Cain & Oakhill (1999)	Connective and elaborative	Oral and written	Narrative	Reading fluency, literal comprehension, comprehension control, backwards digit recall, counting recall, listening recall, nonword list recall, word list recall, matching word

Currie & Cain (2015)	130	5 - 10	6	The British Picture Vocabulary Scale–Third Edition (Dunn et al., 1997)	Breadth	Based on Language and Reading Research Consortium (2015) The dimensionality of language ability in young children. Submitted for publication.	Connective and elaborative	Oral	Narrative	Working memory
Currie & Muijselaar (2019)	420	4 - 9	20	Peabody Picture Vocabulary Test, 4th Ed. (Dunn & Dunn, 2007), Expressive vocabulary test (Williams, 2007) and Clinical evaluation of language fundamentals - CELF (Semel et al., 2006)	Breadth and depth	The Inference task based on Cain & Oakhill (1999); Oakhill & Cain, (2012) and Currie & Cain (2015)	Connective and elaborative	Oral	Narrative	Verbal working memory
Dale et al. (2010)	4.892	12	1	Subtest de Multiple choice del Wechsler Intelligence Scale for Children (Wechsler, 1994)	Depth	Level 2 of subtest Making Inferences of Test of Language Competence (Wiig et al., 1989)	Connective	Oral and written	Narrative	Listening Grammar y Figurative Language
Davies et al. (2020)	100	4	1	Language Content index del Clinical evaluation of language fundamentals - CELF (Wiig et al., 2004)	Breadth	Stories based on Currie & Cain (2015)	No difference	Oral	Narrative	-
Florit et al. (2011)	221	4 - 5	1	Peabody Picture Vocabulary Test, 4th Ed. (Dunn & Dunn, 2007)	Breadth	Five simple paragraphs, consisting of three or four sentences followed by two inference questions (ad hoc)	Elaborative	Oral	Narrative	Listening comprehension, working memory, receptive vocabulary, Verbal intelligence and inferential skills

Jones et al. (2016)	67	6 - 11	1	The expressive one word picture vocabulary test – EOWPVT (Brownell, 2000)	Breadth	Inference questions that were used as a test to assess the understanding of the actions and intentions of the characters and were scored according to the answer (ad-hoc)	Elaborative	Visual	Narrative	Macro-level narrative skills (Narrative content and narrative structure) and Micro-level narrative skills (cohesion, grammatical morphemes and narrative and evaluative devices). Age and non-verbal ability were investigated as predictors of performance on the narrative skills.
Kendeou et al. (2008)	113 of 4 years and 108 of 6 years	4-6 and 6-8 (longitudinal study)	6	Peabody Picture Vocabulary Test, 4th Ed. (Dunn & Dunn, 2007)	Breadth	Open questions about the causes of central events in the story, peripheral events, character goals, and the central theme of the story. In addition, there was a free recall task. (ad hoc)	Elaborative	Oral	Narrative	Phonological awareness, letter, knowledge of letters and decoding of words.
Kim (2017)	262	4.88 - 5.78	1	Picture Vocabulary Test of the Woodcock–Johnson III (Woodcock, R. W., McGrew, K. S., & Mather, 2001)	Breadth	Based on Inference subtest of Comprehensive Assessment of Spoken Language - CASL (Carrow-Woolfolk, 1999)	Elaborative	Oral	Narrative	Attention, working memory, morphosyntactic knowledge

Kim (2020)	201	6.54 – 7.14	2	The Korean Test of Language and Literacy Skills (Y. S. G. Kim et al., n.d.)	Breadth	Based on Inference subtest of Comprehensive Assessment of Spoken Language (Carrow-Woolfolk, 1999)	No difference	Oral	Narrative	Reading comprehension, text reading fluency, word reading, listening comprehension, Theory of mind, comprehension monitoring, grammatical knowledge and working memory
Kim (2020)	165	7-10	1	The Picture Vocabulary subtest of the WJ-III (Woodcock, R. W., McGrew, K. S., & Mather, 2001)	Breadth	The Picture Vocabulary subtest of the WJ-III (Woodcock, R. W., McGrew, K. S., & Mather, 2001)	Elaborative	Oral	Narrative	Reading comprehension, word reading, listening comprehension, attention, working memory, theory of mind, grammatical knowledge and comprehension monitoring
Lepola et al., (2012)	130	4, 5 and 6 (longitudinal study)	3	The word definition test (Silvén & Rubinov, 2010)	Depth	Questions and materials developed by Paris & Paris, (2003)	No difference	Oral	Narrative	Oral comprehension, working memory, phonological awareness
Lepola et al. (2016)	90	4, 6 and 9 (longitudinal study)	3	The word definition test (Silvén & Rubinov, 2010)	Depth	4 and 6 years = questions and materials developed by Paris & Paris (2003). 9 years = The picture book A Boy, a Dog and a Frog by Mercer Mayer (1967) was used at time 3 to measure inference making.	No difference	Oral and written	Narrative	Word knowledge, phonological awareness, listening comprehension, reading speed, reading fluency, and reading comprehension

López-Escribano et al. (2013)	33	8,1 – 9,4	1	Kaufman brief intelligence test Kaufman, K-BIT (Kaufman & Kaufman, 2000)	Breadth	Reading comprehension subtest of Evaluation of Reading Processes for Children – Revised Edition - PROLEC (Cuetos et al., 2007)	No difference	Visual	Narrative	Reading comprehension, spelling choice task rapid word segmentation task, rapid automatized naming task; word reading precision, pseudoword reading precision, reading span task.
Nash & Heath (2011)	26	8	1	The British Picture Vocabulary Scale-II (Dunn & Dunn, 2007)	Breadth	Literal and inferential reading comprehension (ad hoc)	No difference	Oral (vocabulary) - Inference (written)	Narrative	Literal information, text recall, narrative comprehension, verbal working memory, reading accuracy, word reading and decoding
Oakhill, Cain, & Bryant (2003)	102	7-8 and 10-11 (longitudinal study)	4	The Gates–MacGinitie Vocabulary subtest (MacGinitie et al., 2001) vocabulary and similarities of Wechsler Intelligence Test for Children - III (Wechsler,1994)and British Picture Vocabulary Scale – BPVS (Dunn & Dunn, 2007)	Breadth and depth	The constructive inference task from Oakhill's (1982) study and stories with open-ended questions that tapped inferential information, which were taken from a study by Cain & Oakhill (1999)	No difference	Oral and written	Narrative	Working memory, general comprehension, text structure, monitoring, literal comprehension, macrostructure, knowledge of mathematics, reading

Oakhill & Cain (2012b)	102	7–11 (longitudinal study)	6	The Gates–MacGinitie Vocabulary subtest, Level 2, Form K (MacGinitie et al., 2001), vocabulary and similarities of Wechsler Intelligence Test for Children - III (Wechsler, 1994) and British Picture Vocabulary Scale (Dunn et al., 1997).	Breadth and depth	The constructive inference task from Oakhill's (1982) study and stories with open-ended inferential questions, which were taken from Cain & Oakhill (1999)	No difference	Oral and written	Narrative	Text structure, comprehension monitoring, comprehension, grammar knowledge, working memory, phonological awareness, reading ability
Silva & Cain (2015)	82	4-6	1	The British Picture Vocabulary Scale-II (Dunn et al., 1997)	Breadth	Experimental assessment of inference making and literal comprehension (ad-hoc)	No difference	Oral and images	Narrative	Working memory, knowledge of grammar and general non-verbal intelligence
Silverman et al. (2014)	274	8-11	1	WMLS–R, picture vocabulary subtest (Woodcock et al., n.d.) and semantic and the Clinical evaluation of language fundamentals - CELF (Wiig et al., 2004)	Breadth and depth	The Gates–MacGinitie Reading Test–Fourth Edition (MacGinitie et al., 2001) and Test of Silent Reading Efficiency and Comprehension - TOSREC (Wagner et al., 2010)	No difference	Oral and written	Narrative	Phonological awareness, global comprehension
Strasser & del Río (2014)	254	4 - 6	2	WPPSI (Wechsler, 1998) and Peabody Picture Vocabulary Test (Dunn & Dunn, 2007)	Breadth and depth	Adaptation of a task developed by Oakhill (1982)	No difference	Oral	Narrative	Comprehension monitoring, working memory, theory of mind, inhibition and attention control
Villalonga Penna et al., 2014	94	8-10	1	Adaptation of the Vocabulary Subtest of the Wechsler Intelligence Scale for Children	Depth	Inference tasks developed from Cain & Oakhill (1999)	No difference	Oral	Narrative	Decoding and reading comprehension

We used likelihood ratio tests to test the variation in effect sizes between and within studies when comparing the deviance scores of two and three level models. Because the correlation coefficients were converted to Fisher's z scale and all analyses were performed using the transformed values, the estimates were later converted back to Pearson's r.

We used a forest plot to report effect sizes and the Egger test to assess the presence of publication bias (Sterne & Egger, 2005).

Results

Characteristics of the Sample

The data set comprises 76 effect sizes nested within 23 studies published between 2004 and 2020. No records were included for the years 2000-2002, 2006, and 2018. The number of effect sizes per study varies from one to nine, except for Currie & Muijselaar (2019), which included 20 effect sizes. Effect size and their distribution according to year of publication, type of vocabulary, and mean age can be found in Figures 2 and 3. In total, the information gathered corresponds to 8334 participants. Of those, 4892 are from Hayiou-Thomas et al. (2010), who administered Web-based measures of vocabulary and inference making to 12-year-olds participating in a larger study of development in twins. The percentage of female participants ranged from 33% to 77%. One effect size (Silverman, et al., 2014) was excluded from the analyses as it did not define the type of vocabulary assessed. Egger's test was not significant (t = 0.19, p = .85) indicating no funnel plot asymmetry.

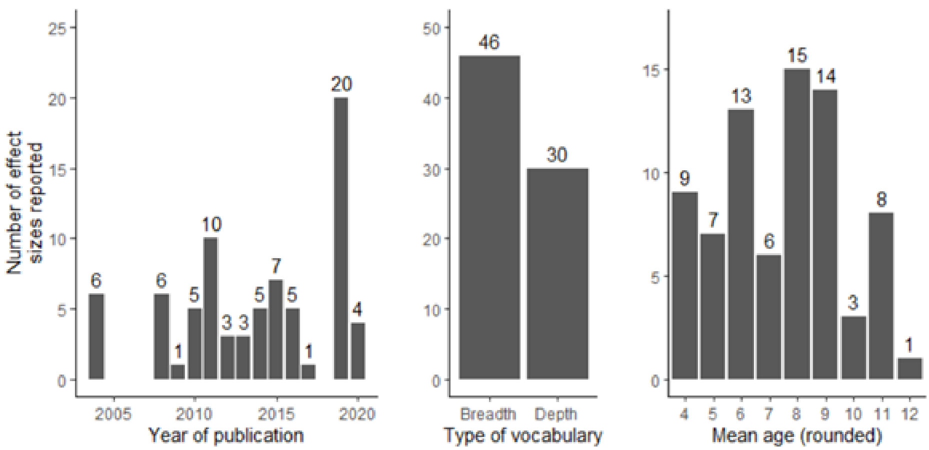

Figure 2. Distribution of effect sizes according to year of publication, type of vocabulary and mean age rounded to integers.

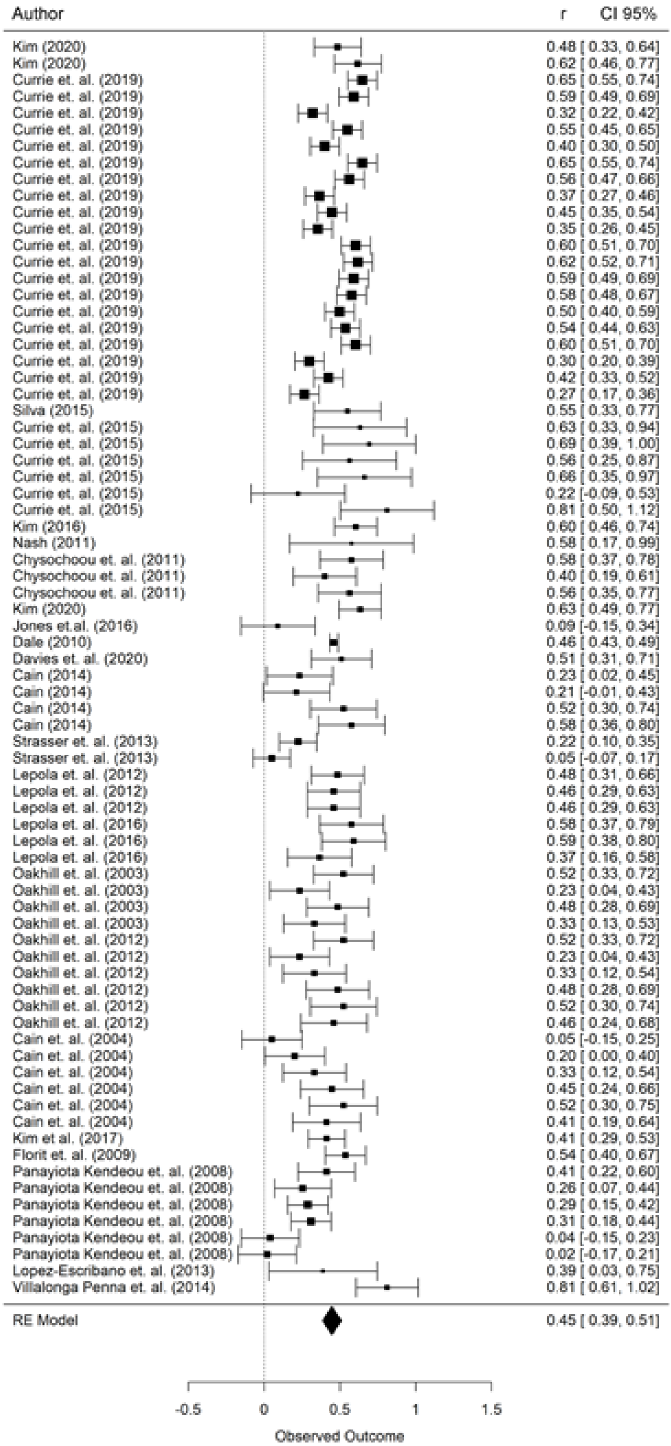

Figure 3. Fisher's z transformation of the correlation coefficients and their confidence intervals.

Association Between Vocabulary and Inference Making

The overall association between vocabulary and inference making was a moderate effect of 0.42 (Fisher's z = 0.45 [95% CI = 0.39, 0.51]) and it was statistically significant (t[75] = 14.37, p < .001). Within-study variance (σ^2 = 0.01, $\chi^2(1)$ = 77.38, p < .001) and between-study variance (σ^2 = 0.01, $\chi^2(1)$ = 10.37, p = .001) were also significant. To analyze the distribution of the variance, we used the formula described in Cheung (2014). Results showed that 14.28% (I^2 = 0.1428) of the total variance could be attributed to within-study sampling error, 42.82% (I^2 = 0.4282) corresponded to differences between effect sizes within studies (Level 2) and 42.91% (I^2 = 0.4291) corresponded to differences in effect sizes between studies (Level 3). The variability in effect sizes was more than what would be explained by sampling error alone, which suggests that other variables are moderating the association between vocabulary and inference making.

We adjusted a second multi-level analysis with type of vocabulary as a moderator. Results showed that the overall effect was not moderated by the type of vocabulary assessed by the studies (F[1, 74] = 0.07, p = .80) and that after including this variable, there was still significant unexplained variance between all effect sizes in the data set (Q[74] = 348.90, p < .001). We adjusted a third multi-level analysis with receptive and productive vocabulary type as a moderator. As with the previous model, the overall was not moderated by this variable (F[1, 77] = 1.24, p = .27) and the unexplained variance between effect sizes did not diminish (Q[77] = 414.36, p < .001).

A fourth multi-level analysis included mean age as a moderator. Results showed that this variable was a significant moderator (F[1, 74] = 1.15, p = .045) of the association between vocabulary and inference making. As the mean age reported in the studies increased, the strength of the correlation decreased (β = -0.02, t(74) = -2.04, p = .04). As with the previous model, after including this variable, there still was significant unexplained variance between all effect sizes in the data set (Q[74] = 342.34, p < .001). Within-study variance (σ^2 = 0.01, χ^2[1] = 45.99, p < .001) and between-study variance (σ^2 = 0.02, χ^2[1] = 12.39, p = 0.001) were also significant. Of the total variance, 13.51% (I^2 = 0.1351) could be attributed to within-study sampling error, 33.45% (I^2 = 0.3345) corresponded to differences between effect sizes within studies (Level 2) and 53.03% (I^2 = 0.53.03) corresponded to differences in effect sizes between studies (Level 3).

Finally, to assess the effect of mean age while controlling for vocabulary type and receptive and productive vocabulary type, we adjusted a fifth model including the three measures. The resulting model was no longer significant (F[3, 74] = 1.47, p = .23).

Discussion

Successful text comprehension results in a coherent memory-based representation of the state of affairs described in the text, typically referred to as a mental model or a situation model (Johnson-Laird, 1983; Kintsch, 1998; Silva & Cain, 2015), The construction of a fully specified and coherent mental model involves going beyond the surface details of a text by combining information across sentences and integrating background knowledge with textual information. These coherence processes involve inference making (Currie & Muijselaar, 2019). Inference comprehension is a complex ability that recruits distinct cognitive domains, such as vocabulary, memory, attention, and executive functions. Vocabulary knowledge is a critical component of language, literacy, and comprehension (Duncan et al., 2007; Marchman & Fernald, 2008). Many empirical studies have demonstrated that there is a close relationship between vocabulary knowledge and reading comprehension (Cain & Oakhill, 2014; Chrysochoou et al., 2011; Cromley & Azevedo, 2007), and between vocabulary knowledge and inference making (Currie & Cain, 2015; Lepola et al., 2016; Nash & Heath, 2011; Silva & Cain, 2015).

According to the lexical quality hypothesis (Perfetti, 2007; Perfetti & Hart, 2002), skilled reading depends on high-quality lexical representations, and therefore, vocabulary should be a powerful predictor of reading. Anderson and Freebody (1981) first made the distinction between the two dimensions of vocabulary knowledge: breadth and depth. They proposed that vocabulary breadth refers to the number of words that the person knows at least on some of the significant aspects of their meaning, while vocabulary depth is the quality or depth of understanding. In particular, breadth of vocabulary refers to how many words a person knows, whereas depth of vocabulary refers to how well a person knows these words.

Prior research has found a role of vocabulary in inference making (Currie & Cain, 2015; Lepola et al., 2016; Nash & Heath, 2011; Silva & Cain, 2015). The purpose of this study was to extend this research by examining the role of vocabulary knowledge, and type of vocabulary (breadth and depth), productive and receptive vocabulary on inferences, by adjusting five multi-level models in a meta-analysis.

The first model analyzed the global effect of vocabulary on inferences, without differentiating the type of vocabulary. The results showed a moderate and significant association (.42) between vocabulary and inference making. This means that vocabulary knowledge is relevant for children's inference making, as research in the area has highlighted. The analysis also suggests, taking into account the heterogeneity index, that other variables are moderating the association between vocabulary and inference.

The second proposed model was intended to analyze the overall effect of vocabulary on inferences with type of vocabulary (depth and breadth) as a moderator. This model evaluated the hypothesis that the depth of the vocabulary has a greater effect on inferences than the breadth of vocabulary. The results indicate that the association between vocabulary and inference making is not moderated by the type of vocabulary assessed by the studies. This means that there are no differences in the correlation patterns between breadth and depth of vocabulary as we measured it and inference making, at least in relation to text and discourse comprehension. Taking into account other components of vocabulary depth reviewed in the Introduction - morphological, syntactic, and phonological awareness - could change this conclusion, but more research is necessary.

The third model analyzed the moderator effect of productive and receptive vocabulary. The results suggest that there are no differences in the correlation patterns between receptive and productive vocabulary in relation to the making of inferences. Taking into account the results of the two previous models, the differentiation between vocabulary breadth and depth could be linked to the differentiation between receptive and productive vocabulary.

In order to study other possible moderators in the relation between vocabulary and inference making, a fourth multi-level analysis was performed including the mean age as a moderator. The findings showed that mean age is a significant moderator; this means that younger children have a stronger relation between vocabulary knowledge and inference making than older ones. In this way, as the mean age reported in the studies decreases, the strength of the correlation increases. Although the pattern of correlation obtained was not consistent across ages, the analysis indicated that there is still significant unexplained variance between all effect sizes in the data set. For these reasons, a fifth model was proposed.

The fifth model was adjusted using the mean age, vocabulary type (breadth and depth), and receptive and productive vocabulary to assess the effect of these variables on the estimated coefficient when controlling for the other one. The data analysis from this model did not show a significant fit to the data. Our results are in agreement with cross-sectional and longitudinal studies that show similar strength correlations of both types of vocabulary with the making of inferences (e.g., Cain et al., 2004; Kendeou et al., 2008; Oakhill & Cain, 2012). On the other hand, the mean age moderator effect indicates that younger children's inference skills are, at least, supported more strongly by vocabulary knowledge, and this strength decreases with age. The idea that vocabulary knowledge is an early predictor of reading comprehension skills (Oakhill & Cain, 2012) agrees with this result.

Although most researchers in the development of reading comprehension skills agree with a multi-component view of vocabulary, only a small number of studies actually use multiple tests to measure it. Although the complexity of the vocabulary is greater than that observed or measured with a single test, more research will be required to more precisely describe the relationship between different vocabulary modalities and different comprehension skills. A limitation observed by this study is that in general, a majority of the studies have used a task that requires a definition to assess vocabulary depth, and a task that does not require a definition to assess vocabulary breadth. Our results agree with the importance of assessing vocabulary with a greater range and variety of tests in order to understand the dimensionality of this concept.

The significant unexplained variance between all effect sizes in the data set suggests that further research is needed to understand which variables are moderating the association between vocabulary and inference making in addition to age. One kind of variable that could influence this relationship is related to development itself, like working memory, sustained attention, and other reading skills. Another group could be linguistic or psycholinguistic variables, like text gender or type of inference. A third kind of moderator could be related to methodological procedures, like input modality (oral or written), type of response (open response or multiple choice), or scoring modality. In the end, different sets of skills are implicated in the development of inference making, and vocabulary is just an important one of them.

Conclusion

Vocabulary knowledge is usually divided into different modalities, like breadth and depth, and receptive and productive vocabulary. The results indicate a strong relation between vocabulary and inference making during text comprehension that is independent of the modality. Children's age showed a moderator effect in the relation between vocabulary and inference making. This association is stronger in younger children than in the older ones.

Acknowledgments

The authors acknowledge the personnel of the Department of Mental Health and Experimental Psychology of the School of Psychology of the University of Buenos Aires.

Conflict of Interest Disclosure

The authors declare no conflicts of interest.

Funding

This research was funded by UBACyT 2020 n°20020190100029BA granted to the fourth author.

Research Ethics Statement

The authors declare that no Research Ethics Statement was required.

References

Anderson, R. C., & Freebody, P. (1981). Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 77–117). International Reading Association.

Bowyer-Crane, C., & Snowling, M. J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Educational Psychology, 75*(2), 189–201. https://doi.org/10.1348/000709904X22674

Brownell, R. (2000). *Expressive one-word picture vocabulary test manual*. Academic Therapy.

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

Cain, K., & Oakhill, J. V. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing, 11*(5/6), 489–503. https://doi.org/10.1023/A:1008084120205

*Cain, K, & Oakhill, J. V. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? In *Annee Psychologique* (Vol. 114, Issue 4, pp. 647–662). Editions NecPlus. https://doi.org/10.4074/S0003503314004035

Carrow-Woolfolk, E. (1999). *Comprehensive assessment of spoken language*. Pearson.

Cheung, M. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods, 19*(2), 211–229. https://doi.org/10.1037/a0032968

Cheung, M. (2019). Guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review, 29*, 387–399. https://doi.org/10.1007/s11065-019-09415-6

*Chrysochoou, E., Bablekou, Z., & Tsigilis, N. (2011). Working memory contributions to reading comprehension components in middle childhood children. *American Journal of Psychology, 124*(3), 275–289. https://doi.org/10.5406/amerjpsyc.124.3.0275

Cromley, J. G., & Azevedo, R. (2007). Testing and refining the direct and inferential mediation model of reading comprehension. *Journal of Educational Psychology, 99*(2), 311–325. https://doi.org/10.1037/0022-0663.99.2.311

Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *Journal of Educational Research, 36*(3), 206–217. https://doi.org/10.1080/00220671.1942.10881160

Cuetos, F., Rodríguez, B., Ruano, E., & Arribas, D. (2007). *PROLEC-R: Batería de evaluación de los procesos lectores, revisada* [Evaluation of reading processes for children, revised edition]. TEA Ediciones.

*Currie, N. K, & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology, 137*, 57–75. https://doi.org/10.1016/j.jecp.2015.03.005

*Currie, N. K., & Muijselaar, M. M. L. (2019). Inference making in young children: The concurrent and longitudinal contributions of verbal working memory and vocabulary. *Journal of Educational Psychology, 111*(8), 1416–1431. https://doi.org/10.1037/edu0000342

*Dale, P. S., Harlaar, N., Hayiou-Thomas, M. E., & Plomin, R. (2010). The etiology of diverse receptive language skills at 12 years. *Journal of Speech, Language, and Hearing Research, 53*(4), 982–992. https://doi.org/10.1044/1092-4388(2009/09-0108)

*Davies, C., McGillion, M., Rowland, C., & Matthews, D. (2020). Can inferencing be trained in preschoolers using shared book-reading? A randomised controlled trial of parents' inference-eliciting questions on oral inferencing ability. *Journal of Child Language, 47*(3), 655–679. https://doi.org/10.1017/S0305000919000801

Duncan, G., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1443. https://doi.org/10.1037/0012-1649.44.1.217

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, fourth edition.* Pearson.

Dunn, L. M., Dunn, D. W., Sewell, J., Styles, B., & Shamsan, Y. (2009). *The*

British Picture Vocabulary Scale (3rd ed.). GL Assessment.
Dunn, L. M., Dunn, L. M., Whetton, C., & Burley, J. (1997). *British Picture Vocabulary Scale II*. NFER-Nelson.
Fan, M. (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal, 31*, 105–119. https://doi.org/10.1177/003368820003100205
Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, N., Onghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods, 52*, 2031–2052. https://doi.org/10.3758/s13428-020-01373-9
*Florit, E., Roch, M., & Levorato, M. C. (2011). Listening text comprehension of explicit and implicit information in preschoolers: The role of verbal and inferential skills. *Discourse Processes, 48*(2), 119–138. https://doi.org/10.1080/0163853X.2010.494244
Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371–395. https://doi.org/10.1037/0033-295X.101.3.371
Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition, 21*(2), 303–317. https://doi.org/10.1017/S0272263199002089
Hayiou-Thomas, M. E., Harlaar, N., Dale, P. S., & Plomin, R. (2010). Preschool speech, language skills, and reading at 7, 9, and 10 years: Etiology of the relationship. *Journal of Speech, Language, and Hearing Research, 53*(2), 311–332. https://doi.org/10.1044/1092-4388(2009/07-0145)
Johnson-Laird, P. N. (1983). *Mental models*. Cambridge University Press.
*Jones, A. C., Toscano, E., Botting, N., Marshall, C. R., Atkinson, J. R., Denmark, T., Herman, R., & Morgan, G. (2016). Narrative skills in deaf children who use spoken English: Dissociations between macro and microstructural devices. *Research in Developmental Disabilities, 59*, 268–282. https://doi.org/10.1016/j.ridd.2016.09.010
Kaufman, A. S., & Kaufman, N. L. (2000). *K-Bit, Test Breve de Inteligencia de Kaufman.* Pearson.
*Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading, 31*(3), 259–272. https://doi.org/10.1111/j.1467-9817.2008.00370.x
Kendeou, P., Van Den Broek, P., Helder, A., & Karlsson, J. (2014). A cognitive view of reading comprehension: Implications for reading difficulties. *Learning Disabilities Research and Practice, 29*(1), 10–16. https://doi.org/10.1111/ldrp.12025
*Kim, Y.-S. G. (2020). Toward integrative reading science: The direct and indirect effects model of reading. *Journal of Learning Disabilities, 53*(6), 469–491. https://doi.org/10.1177/0022219420908239

*Kim, Y. S. G. (2017). Multicomponent view of vocabulary acquisition: An investigation with primary grade children. *Journal of Experimental Child Psychology, 162*, 120–133. https://doi.org/10.1016/j.jecp.2017.05.004

*Kim, Y. S. G. (2020). Hierarchical and dynamic relations of language and cognitive skills to reading comprehension: Testing the direct and indirect effects model of reading (DIER). *Journal of Educational Psychology, 112*(4), 667–684. https://doi.org/10.1037/edu0000407

Kim, Y. S. G., Cho, J. R., & Park, S. G. (n.d.). *The Korean Test Of Language And Literacy Skills (KOLLS)*. Korea Guidance.

Kintsch, W. (1998). *Comprehension. A paradigm for cognition.* Cambridge University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182. https://doi.org/10.1037/0033-295X.95.2.163

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or different? *Applied Linguistics, 19*(2), 255–271. https://doi:10.1093/applin/19.2.255

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing, 21*(2), 202–226. https://doi.org/10.1191/0265532204lt277oa

Laufer, B., & Paribakht, T. (1998). The relationship between passive and active vocabularies: effects of language learning context. *Language Learning, 48*, 365–391. https://doi.org/10.1111/0023-8333.00046

*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*(4), 373–390. https://doi.org/10.1002/rrq.145

*Lepola, J., Lynch, J., Laakkonen, E., Silvén, M., & Niemi, P. (2012). The role of inference making and other language skills in the development of narrative listening comprehension in 4-6-year-old children. *Reading Research Quarterly, 47*(3), 259–282. https://doi.org/10.1002/RRQ.020

Li, M., & Kirby, J. R. (2015). The effects of vocabulary breadth and depth on English reading. *Applied Linguistics, 36*(5), 611–634. https://doi.org/10.1093/applin/amu007

*López-Escribano, C., de Juan, M. R. E., Gómez-Veiga, I., & García-Madruga, J. A. (2013). Estudio predictivo de la comprensión lectora en estudiantes españoles de tercero de Primaria. *Psicothema, 25*(2), 199–205. https://doi.org/10.7334/psicothema2012.175

Lynch, J. S., van den Broek, P., Kremer, K. E., Kendeou, P., White, M. J., & Lorch, E. P. (2008). The development of narrative comprehension and its relation to other early reading skills. *Reading Psychology, 29*(4), 327–365. https://doi.org/10.1080/02702710802165416

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2001). *Gates-*

MacGinitie Reading Tests, Level 7/9, Form S (4th ed.). Riverside Publishing.

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood: Fast-Track Report. *Developmental Science, 11*(3), F9–F16. https://doi.org/10.1111/j.1467-7687.2008.00671.x

Milton, J. (2009). Measuring second language vocabulary acquisition. *Multilingual Matters.*

Moeyaert, M., Ugille, M., Natasha Beretvas, S., Ferron, J., Bunuan, R., & Van den Noortgate, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology, 20*(6), 559–572. https://doi.org/10.1080/13645579.2016.1252189

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., Altman, D., Antes, G., Atkins, D., Barbour, V., Barrowman, N., Berlin, J. A., Clark, J., Clarke, M., Cook, D., D'Amico, R., Deeks, J. J., Devereaux, P. J., Dickersin, K., Egger, M., Ernst, E., … Tugwell, P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine, 6*(7). https://doi.org/10.1371/journal.pmed.1000097

Muijselaar, M. M. L. (2018). The dimensionality of inference making: Are local and global inferences distinguishable? *Scientific Studies of Reading, 22*(2), 117–136. https://doi.org/10.1080/10888438.2017.1371179

*Nash, H., & Heath, J. (2011). The role of vocabulary, working memory and inference making ability in reading comprehension in Down syndrome. *Research in Developmental Disabilities, 32*(5), 1782–1791. https://doi.org/10.1016/j.ridd.2011.03.007

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press.

Nation, I.S.P. (1990). *Teaching and learning vocabulary.* Heinle & Heinle Publishers.

Nation, K., & Snowling, M. J. (1999). Developmental differences in sensitivity to semantic relations among good and poor comprehenders: Evidence from semantic priming. *Cognition, 70*(1), B1–B13. https://doi.org/10.1016/S0010-0277(99)00004-9

Oakhill, J. (1982). Constructive processes in skilled and less skilled comprehenders' memory for sentences. *British Journal of Psychology, 73*(1), 13–20. https://doi.org/10.1111/j.2044-8295.1982.tb01785.x

*Oakhill, J. V., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*(2), 91–121. https://doi.org/10.1080/10888438.2010.529219

Oakhill, J V. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology, 54*(1), 31–39. https://doi.org/10.1111/j.2044-8279.1984.tb00842.x

Oakhill, J V, & Cain, K. (2013). The precursors of reading ability in young readers: evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*(2), 91–121. https://doi.org/10.1080/10888438.2010.529219

*Oakhill, J V, Cain, K., & Bryant, P. (2003). Dissociation of single-word reading and text comprehension skills. *Language and Cognitive Processes, 18*, 443–468. https://doi.org/10.1080/01690960344000008

Oakhill, J V, Cain, K., & McCarthy, D. (2015). Inference processing in children: The contributions of depth and breadth of vocabulary knowledge. In E. J. O'Brien, A. E. Cook, & R. F. Lorch, Jr. (Eds.), *Inferences during reading* (pp. 140–159). Cambridge University Press. https://doi.org/10.1017/cbo9781107279186.008

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554–566. https://doi.org/10.1037/0022-0663.98.3.554

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly, 38*(1), 36–76. https://doi.org/10.1598/RRQ.38.1.3

Paul, P. V., & Gustafson, G. (1991). Comprehension of high-frequency multimeaning words by students with hearing impairment. *Remedial and Special Education, 12*(4), 52–61. https://doi.org/10.1177/074193259101200408

Perfetti, C, & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (pp. 189–213). John Benjamins. https://doi.org/10.1075/swll.11.14per

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357–383. https://doi.org/10.1080/10888430701530730

Proctor, C. P., Silverman, R. D., Harring, J. R., & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing, 25*(7), 1635–1664. https://doi.org/10.1007/S11145-011-9336-5

Schmitt, N. (2014). Size and depth of vocabulary knowledge: What the research shows. *Language Learning, 64*(4), 913–951. https://doi.org/10.1111/lang.12077

Semel, E. M., Wiig, E. H., & Secord, W. (2006). *Clinical evaluation of language fundamentals* (4th ed.). Harcourt Assessment.

Shahov, V. (2012). Measuring L2 receptive and productive vocabulary knowledge, *University of Reading Language Studies Working Papers, 4*, 37–45.

*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*(2), 321–331. https://doi.org/10.1037/a0037769

Silvén, M., & Rubinov, E. (2010). Language and preliteracy skills in bilinguals and monolinguals at preschool age: Effects of exposure to richly inflected speech from birth. *Reading and Writing, 23*(3), 385–414. https://doi.org/10.1007/s11145-009-9206-6

Silverman, R. D., Proctor, C. P., Harring, J. R., Doyle, B., Mitchell, M. A., & Meyer, A. G. (2014). Teachers' instruction and students' vocabulary and comprehension: An exploratory study with English monolingual and Spanish-English bilingual students in grades 3-5. *Reading Research Quarterly, 49*(1), 31–60. https://doi.org/10.1002/rrq.63

Singer, M., & Ritchot, K. F. M. (1996). The role of working memory capacity and knowledge access in text inference processing. *Memory & Cognition, 24*(6), 733–743. https://doi.org/10.3758/BF03201098

Sterne, J. A., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and analysis* (pp. 90–110). John Wiley & Sons, Ltd.

*Strasser, K., & del Río, F. (2014). The role of comprehension monitoring, theory of mind, and vocabulary depth in predicting story comprehension and recall of kindergarten children. *Reading Research Quarterly, 49*(2), 169–187. https://doi.org/10.1002/rrq.68

Tannenbaum, K. R., Torgesen, J. K., & Wagner, R. K. (2006). Relationships between word knowledge and reading comprehension in third-grade children. *Scientific Studies of Reading, 10*(4), 381–398. https://doi.org/10.1207/s1532799xssr1004_3

Tran, A. H., Tremblay, K. A., & Binder, K. S. (2020). The factor structure of vocabulary: An investigation of breadth and depth of adults with low literacy skills. *Journal of Psycholinguistic Research, 49*(2), 335–350. https://doi.org/10.1007/s10936-020-09694-8

Van den Broek, P. (1990). The causal inference maker: Towards a process model of inference generation in text comprehension. In D. Balota, G. Flores D'Arcais, & K. Rayner (Eds.), *Comprehension processes in reading* (pp. 423–446). Lawrence Erlbaum Associates.

Van Den Broek, P, Rapp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionist processes in accounts of reading comprehension. *Discourse Processes, 39*(2–3), 299–316. https://doi.org/10.1207/s15326950dp3902&3_11

Van Den Broek, P, & Espin, C. A. (2012). Connecting cognitive theory and assessment: Measuring individual differences in reading comprehension. *School Psychology Review, 41*(3), 315–325. https://doi.org/10.1080/02796015.2012.12087512

*Villalonga Penna, M. M., Padilla Sabaté, C., & Burin, D. (2014). Relaciones entre decodificación, conocimiento léxico-semántico e inferencias en niños de escolaridad primaria. *Interdisciplinaria, 31*(2), 259–274. https://doi.org/10.16888/interd.2014.31.2.5

Wagner, R. C., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *Test of Silent Reading Efficiency and Comprehension.* ProEd.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second language acquisition, 30*(1), 79–95. https://doi.org/10.1017/S0272263108080042

Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal, 40*(3), 360–376. https://doi.org/10.1177/0033688209343854

Wechsler, D. (1994). T*est de inteligencia para niños WISC-III, Manual.* Paidós.

Wechsler, D. (1998). *Test de Inteligencia para Preescolares (WPPSI-R).* Paidos.

Wechsler, D. (2012). *Test de inteligencia para niños (WISC-IV) (4°)*. Paidós.

Wiig, E. H., Secord, W., & Sabers, D. (1989). *Tests of Language Competence (Extended Edition).* Psychological Corporation.

Wiig, E. H., Secord, W., & Semel, E. M. (2004). C*ELF-Preschool 2: Clinical evaluation of language fundamentals-preschool.* Pearson/Psychological Corporation.

Williams, K. T. (2007). *Expressive Vocabulary Test (2nd ed.).* Pearson.

Wise, J. C., Sevcik, R. A., Morris, R. D., Lovett, M. W., & Wolf, M. (2007). The relationship among receptive and expressive vocabulary, listening comprehension, pre-reading skills, word identification skills, and reading comprehension by children with reading disabilities. *Journal of Speech, Language, and Hearing Research, 50*(4), 1093–1109. https://doi.org/10.1044/1092-4388(2007/076)

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock– Johnson III.* Riverside Publishing.

Woodcock, R. W., Muñoz-Sandoval, A. F., & Ruef, M. L. (n.d.). *Woodcock–Muñoz Language Survey–revised*. Riverside Publishing.