INTERNATIONAL LITERACY ASSOCIATION READING RESEARCH QUARTERLY

Learning to Read Connections—Sensitivity to Collocation Frequency Links Vocabulary Size and Reading Comprehension in Middle Childhood

Alexandra M. A. Schmitterer
DIPF, Leibniz Institute for Research and Information in Education, Frankfurt am Main, Germany

IDeA Center for Individual Development and Adaptive Education, Frankfurt am Main, Germany

Caterina Gawrilow
IDeA Center for Individual Development and Adaptive Education, Frankfurt am Main, Germany

University of Tübingen, Tübingen, Germany

Claudia Friedrich
University of Tübingen, Tübingen, Germany

ABSTRACT

The collocation frequency of words in the language environment contributes to early vocabulary development. Vocabulary size, in turn, predicts children's reading comprehension skills later in development. Both collocation frequency and reading comprehension have been connected to inferential reasoning at different time points in development. Here, it was hypothesized that 8-year-old children's ($N = 147$; 76 female) sensitivity to collocation frequency would be related to vocabulary size and reading comprehension skills of varying complexity. Participants completed an auditory thematic judgment task to assess their sensitivity to collocation frequency (response accuracy or speed). In the task, children were presented with a short sentence containing a reference word (e.g., "John sees the cloud.") and asked to judge which of two subsequent words best fit the sentence (e.g., "rain" or "lip"). Semantic relatedness between reference words and test words was operationalized in three levels (strong, weak, and distant) based on a corpus-based analysis of collocation frequency. Multilevel and mediation analyses confirmed that thematic judgment responses were related to corpus-based measures of collocation frequency and were associated with vocabulary size and reading comprehension skills at the sentence and text level. Furthermore, thematic judgment predicted vocabulary size and reading comprehension when the relation of decoding and reading comprehension was taken into account. The study highlights sensitivity to collocation frequency as a link between early language comprehension development (i.e., lexical retrieval and inferential reasoning) and reading comprehension in middle childhood. It also integrates theoretical approaches from computational network or distributional semantics studies and behavioral experimental studies.

Reading comprehension plays an important role in everyday life. The ability to read and comprehend text fluently involves many complex cognitive processes that develop in early and middle childhood. Longitudinal studies in multiple alphabetic languages broadly distinguish two underlying cognitive components: decoding and language comprehension (Caravolas et al., 2019; Florit & Cain, 2011; Hjetland et al., 2020; Hoover & Tunmer, 2018; Nation, 2019). Together, they have been shown to explain more than 85% of the variability in reading comprehension abilities in middle childhood (Chiu, 2018; Lervåg et al., 2018; Lonigan et al., 2018).

Decoding skills (i.e., word recognition) are explained by underlying phonological and phoneme-grapheme convergence processes (e.g., phonological awareness, letter-sound knowledge, rapid naming; Caravolas et al., 2019; Moll et al., 2014; Schmitterer & Schroeder, 2019a; Zugarramurdi et al., 2022). However, disentangling the cognitive processes underlying the relation between early language comprehension

INTERNATIONAL LITERACY ASSOCIATION

Reading Research Quarterly, 0(0)
pp. 1–20 | doi:10.1002/rrq.548

development and reading comprehension skills in middle childhood is still considered a complex problem (Castles et al., 2018; Duke & Cartwright, 2021). Furthermore, while decoding skills explain more variability in reading comprehension in early reading development, the contribution of language comprehension increases throughout reading development (Hjetland et al., 2019, 2020; Hulme et al., 2015). Thus, understanding how early language comprehension development relates to reading comprehension is relevant for reading acquisition in the long term. In this study, we aim to investigate the role that the early acquired cognitive architecture of semantic knowledge based on co-occurrences might play in the relation between vocabulary size and reading comprehension abilities in middle childhood.

Lexical Structure, Vocabulary Development and Collocation Frequency

Computational models of semantic representations assume that lexical knowledge (i.e., vocabulary) is stored in networks or vector spaces (Collins & Loftus, 1975; Firth, 1957; Harris, 1954; Landauer & Dumais, 1997; Mikolov et al., 2013; Pennington et al., 2014; McClelland & Rogers, 2003; Stella et al., 2017; Steyvers & Tenenbaum, 2005; Turney & Pantel, 2010). In such structures, words are represented by nodes, which are connected by links (or vectors) that represent specific relations between the words. An underlying assumption is that the processing and meaning representation of any given word is specified by the words it relates to.

Links between words in these models can be semantic relations based on features (e.g., cucumber—is edible; cucumber—is a vegetable, McRae et al., 2005), free association norms (e.g., dog—bone; Nelson et al., 1998), or co-occurrence statistics (e.g., cloud—rain; Fourtassi et al., 2020; Fourtassi, 2020). Because words often co-occur within specific thematic contexts (e.g., clouds and rain during a walk), semantic relations based on co-occurrence are also referred to as thematic relations (Mirman et al., 2017). Particularly in distributional semantic models (mainly vector-based), co-occurrence measures are used as the basic statistical information to specify the models (Flores et al., 2020; Fourtassi, 2020; Pennington et al., 2014; Unger et al., 2023). The underlying assumption is that words with similar meanings or roles tend to appear in similar linguistic contexts. In this study, we use the term collocation frequency to highlight that sensitivity to co-occurrence measures is connected to the frequency with which two words co-occur (e.g., within 5 words, within a sentence; Evert, 2008; Unger et al., 2023). In this study, we hypothesize that sensitivity to collocation frequency in the input language links vocabulary size and reading comprehension.

Only few computational studies have studied early vocabulary development in other languages than English. Here, we introduce studies using network growth model approaches that have been replicated across multiple languages. These studies have investigated learning principles that aim to explain early vocabulary development (Fourtassi et al., 2020; Hills et al., 2009). In these studies, semantic networks were fitted for 130 high-frequency nouns typically acquired before 30 months of age, and vocabulary growth was modeled based on Age-of-Acquisition norms (AoA; MCDI; Dale & Fenson, 1996). AoA norms provide information from parent questionnaires that qualify the age at which each noun was acquired for at least 50% of children (Braginsky et al., 2019). The structure of semantic networks was fitted using feature-based semantic relations (McRae et al., 2005) and association norms (Nelson et al., 1998). In addition, it was investigated whether the strength of relatedness between nodes and the number of nodes connected to each other (i.e., connectivity) were consistent with co-occurrence statistics from corpora of child-directed speech (CHILDES; MacWhinney, 2014; Wordbank; Frank et al., 2017). The networks followed these principles: A new word is more likely to be learned if it is semantically related to other words. For example, the word "rain" is learned faster if the word "cloud" is already known or appears together with "rain" in the child's environment. Furthermore, a new word is more likely to be learned if it is semantically related to words that are connected to many other words (i.e., high connectivity and similar contexts). For example, the word "rain" will be learned more quickly if it is semantically related to a hub including "cloud," "thunder," "mountain," "lightning," and "water".

Seminal network growth studies by Hills et al. (2009) and Fourtassi et al. (2020) found that young children's lexical structure in 10 languages represented associative rather than feature-based relations. Since association norms have been explained by co-occurrence statistics (Fourtassi, 2020; Lund et al., 1996; Spence & Owens, 1990; Unger et al., 2023), this points to the importance of collocation frequency in early vocabulary development. In addition, the growth of the networks has been explained by the number of co-occurring words with the target nouns (i.e., connectivity) in the language environment (i.e., preferential acquisition). Thus, in these models words that were linked to many co-occurring words in child-directed speech were acquired earlier than words that were linked to few co-occurring words. Studies with adults show similar growth effects that highlight the importance of semantic relatedness and connectivity for lexical growth (Steyvers & Tenenbaum, 2005). Based on this evidence for lexical growth in early childhood and adulthood, we assume that for middle childhood, computational models would show that collocation frequency in the language environment and the structure of children's individual lexica contribute to vocabulary growth.

The assumption that semantic connectivity based on collocation frequency plays an important role in early vocabulary development is corroborated by computational studies focusing on interindividual differences in early vocabulary development. Beckage et al. (2011) conducted semantic network analyses based on AoA data from typically developing children and children with significant delays in vocabulary development (late talkers). Individual networks were constructed by linking words in each child's vocabulary to co-occurrence statistics from child-directed speech (CHILDES; MacWhinney, 2014). For each individual network, various measures of connectivity were computed, and mean connectivity was compared between typically developing children and late talkers. Results indicated that, while controlling for vocabulary size, late talkers' semantic networks had lower connectivity and greater distances between nodes than typically developing children's semantic networks. The increased network connectivity of typically developing children compared to late talkers may indicate either a higher sensitivity to the connectivity of co-occurring words in their language environment or higher connectivity in their own lexicon. Taken together, the results of these studies indicate that children's word representations are structured according to collocation frequency and children may use their sensitivity to thematic connectivity to increase vocabulary size.

Sensitivity to Collocation Frequency across Development

Sensitivity to thematic relations has been investigated in behavioral studies in childhood and adulthood (Lany & Saffran, 2013; Savic et al., 2022, 2023; Unger et al., 2023; Wojcik & Kandhadai, 2020). At the behavioral level, cognitive structures are often investigated using priming paradigms or tasks that tap into semantic matching processes (e.g., triad paradigms). In priming paradigms, participants are often asked to decide whether a word exists or not (i.e., lexical decision). In triad paradigms, participants are presented with three words and asked to judge the difference between word relations or to identify a target word among all three words. Stimuli in both paradigms include unrelated control items and primes or distractors that are semantically related to the target. Judgment latencies in semantic priming paradigms are typically faster on primed trials than on control trials. This is explained by the parallel activation of related items during lexical retrieval (McClelland & Rogers, 2003; Rogers & McClelland, 2004) In priming paradigms, a related prime is presented just before the target. Thus, the prime co-activates the target and facilitates subsequent retrieval. However, in task designs where distractors must be inhibited, the effects of relatedness can also be inhibitory (see, e.g., Friedrich et al., 2013). For example, in a triad paradigm, when participants must decide which of two semantic relations is stronger (i.e., a semantic matching process), the decision will be more difficult when two relations are similar in strength than when they are dissimilar in strength.

In line with the findings of computational studies, behavioral studies suggest that adults, toddlers, younger children, and older children are sensitive to semantic relations between words (Lany & Saffran, 2013; Savic et al., 2023; Unger et al., 2023). Semantic relations in this context include relatedness based on association norms or co-occurrence measures (Arias-Trejo & Plunkett, 2013; Buchanan et al., 2001; Landrigan & Mirman, 2018; Mirman et al., 2017; Savic et al., 2022, 2023; Unger et al., 2020; Wojcik & Kandadhai, 2020). For example, in a triad paradigm, Vales and Fisher (2019) asked 3- to 9-year-old children to judge whether a panel of three pictures contained a reference to a spoken target (e.g., "bone"). In addition to the target picture, panels included either two unrelated distractors (e.g., a flower and a broom) or one unrelated and one related distractor (e.g., a flower and a puppy). The target-distractor association was based on adult association ratings. Children took longer to decide whether the target was part of the triad when a related word was included. The inhibitory effect varied as a function of the strength of the relatedness between the target and the related distractor: The more related the target and distractor were, the longer it took children to make their judgment. Thus, because children had to inhibit the distractor to make their judgment, an inhibitory effect was observed. In the context of semantic networks, this is explained by the path length between nodes: The shorter the path length, the stronger the relatedness, the more effortful the decision, and the stronger the inhibitory effect.

Sensitivity to thematic relatedness in children was also found in a longitudinal study (Schmitterer & Schroeder, 2019b) with a similar design. Notably, this study used co-occurrence measures from a corpus of child-directed literature to operationalize semantic relatedness (Schroeder et al., 2015). In their thematic judgment task, 4- to 7-year-olds were presented with a sentence containing a target word (e.g., "Jan is looking at the cloud."). Children were then presented with two words that were either strongly (strong: "rain"), weakly (weak: "mountain"), or distantly related (control: "lip") to the target word ("cloud"). The children were asked to judge which word would better fit the sentence. The compared relations were either close to each other in strength of relatedness (strong-weak: "cloud-rain" vs. "cloud-mountain") or distant to each other in strength of relatedness (strong-distant: "cloud-rain" vs. "cloud-lip"). Across development, children were more likely to choose the most strongly related word ("rain") in the condition where the relatedness distance between the word pairs was greater than in the condition where the relatedness strength was closer. Furthermore, because the material in the study was based on a corpus of child-directed literature rather than child-directed speech, the results suggest a relation between sensitivity to collocation frequency in verbal input and exposure to written language.

Sensitivity to Collocation Frequency and Reading Comprehension

The suggestion that thematic judgments are related to reading comprehension was further supported by an additional analysis of the longitudinal study by Schmitterer and Schroeder (2019b): thematic judgments in 4- and 5-year-olds predicted early reading comprehension skills (i.e., word-picture matching) in the same children 14 months later at the end of first grade. Responses on the thematic judgment task predicted children's reading abilities beyond children's results in phonological working memory, letter-sound knowledge, and syntactic comprehension tasks.

To date, a relation between thematic judgment and reading has only been found for beginning word-level reading skills (i.e., word-picture matching). However, reading theories suggest that there may also be a link to more complex reading comprehension abilities. For example, the reading system framework (Perfetti & Stafura, 2014), which posits cascading processing, argues that lexical processing at the word level is a pressure point for more complex comprehension processes at the sentence or text level. Thus, if lexical processes at the word level, such as the retrieval of semantic information from the lexicon, are impaired, higher-level processes at the sentence and text level will also be impaired.

This link between sensitivity to collocation frequency at the word-level and reading comprehension skills is illustrated by a study focusing on interindividual differences in reading comprehension in 10-year-old children (Nation & Snowling, 1999). Typically developing readers and children with a specific reading comprehension deficit (i.e., poor comprehenders) were presented with recordings of strongly associated prime-target pairs (e.g., "dog"—"cat") or weakly associated pairs (e.g., "airplane"—"train"). Children were asked to decide whether the target was a real word or not (i.e., lexical decision task). Associations were based on co-occurrence measures of child-directed literature (Carroll et al., 1971). Children's reaction times were faster in conditions with strong associations than in conditions with weak associations. For poor comprehenders, the facilitatory effect of association strength was weaker than for children without reading difficulties. This suggests that children with reading comprehension difficulties may co-activate fewer strongly related items than typically developing children. In consequence, this could indicate less connectivity or fewer entries in the lexica of children with reading comprehension difficulties.

Nation and Snowling's (1999) study shows an association between reading comprehension difficulties at the sentence or text level and reduced sensitivity to the connectivity of thematic relations at the word level. However, even when considering these findings it remains unclear whether thematic judgment would predict reading comprehension at the sentence or text level in addition to other semantic measures. For example, because measures of vocabulary size and thematic judgments should be related, and both measure lexical retrieval at the word level, thematic judgments may not add to the well-documented relation between vocabulary size and reading comprehension (Duke & Cartwright, 2021; Hjetland et al., 2020; Lervåg et al., 2018; Verhoeven et al., 2011). However, there is another cognitive process that may link complex reading comprehension skills and thematic judgment: inferential reasoning.

Sensitivity to Collocation Frequency, Reading Comprehension, and Inferential Reasoning

Inferential reasoning has been found to be relevant to reading comprehension abilities (Duke & Cartwright, 2021; Elleman, 2017) and, in early development, has been linked to collocation frequency (Fisher et al., 2011). In reading, an example of inferential reasoning is the ability to perceive and use indicators of local coherence within or between sentences to infer syntactic or semantic relations (Ahmed et al., 2016; Kintsch, 1998). One example would be identifying referents (e.g., animal in boat) to a main character (e.g., tiger) throughout a text or sentence (e.g., Life of Pi). Co-occurrences and a sensitivity thereof might aid the ability to identify referents that are associated. If poor comprehenders have less entries or less connectivity in their individual lexica, they might also be less likely to use statistical information from co-occurrences to infer semantic relatedness in reading.

In young children, sensitivity to collocation frequency has been linked to inferential reasoning (Fisher et al., 2011; Godwin et al., 2013; Matlen et al., 2015). For example, Fisher et al. (2011) asked 4-year-olds to judge whether word pairs share features. Half of the word pairs regularly occurred together in child-directed speech (e.g., "bunny-rabbit"), and half of them did not (e.g., "alligator-crocodile"). Children were more likely to recognize that the word pairs shared common features when the words also co-occurred regularly. This suggests that sensitivity to collocation frequency aids inferential reasoning. However, whether inferential reasoning as indicated by thematic judgment responses is related to higher-level reading skills in middle childhood remains to be investigated.

This Study

In this study, we focused on the relation between thematic judgment, vocabulary size, and various reading comprehension skills in a group of older children. We collected data from 147 German speaking third-grade students. Our sample was diverse regarding their reading comprehension abilities. We obtained teacher ratings for about half of the sample. These ratings indicated that teachers would definitely or potentially recommend reading intervention for approximately 68% of the participants. We anticipated a high degree of variability in decoding and comprehension skills that would allow us to identify underlying cognitive processes in typically developing and struggling readers.

We used an adapted version of Schmitterer and Schroeder's (2019b) thematic judgment task. The item selection process is shown in Figure 1. First, we analyzed whether responses to this task measured children's ability to discriminate between different strengths of thematic relatedness in our item set. Second, we examined the relation between interindividual differences in thematic judgment responses and interindividual differences in vocabulary size, and reading comprehension at the word, sentence, and text level. Third, we analyzed whether thematic judgment mediated the relation between vocabulary size and complex reading comprehension abilities. Thus, thematic judgment would represent a link between word-level lexical retrieval and inferential reasoning required for higher-level comprehension skills.In addition, we examined whether these results hold, when we add decoding. With this approach, we aimed to investigate whether thematic judgment would emerge as a relevant mediator in a common modeling approach in reading research (i.e., Hjetland et al., 2020). Furthermore, with our model we explored the relation of decoding to vocabulary size and thematic judgment: Decoding relies strongly on phonological processes which also play an important role in the built up of semantic networks in early vocabulary development (Fourtassi et al., 2020). Thus, decoding may be related to both thematic judgment and vocabulary size and may weaken either contribution to reading comprehension. Moreover, recent reading models treat vocabulary as a bridging process between decoding and language comprehension skills (i.e., inferential reasoning; Duke & Cartwright, 2021). Therefore, if the results of the first model variant would not hold, we sought to explore a model in which vocabulary mediates the relation between thematic judgment and reading comprehension. This would indicate that a component representing inferential reasoning within thematic judgment would directly contribute to reading comprehension and would predict vocabulary size.

Regarding the first research question, we expected that children would be sensitive to differences in the strength of thematic relations between word pairs and this sensitivity to differences would be connected to collocation frequency. Our specific hypotheses were as follows:

1a Children will be more likely to identify a word pair as a "better fit" than a distractor word pair (mean accuracy) when the difference in collocation frequency between the contrasting word pairs in the input language is large versus small.

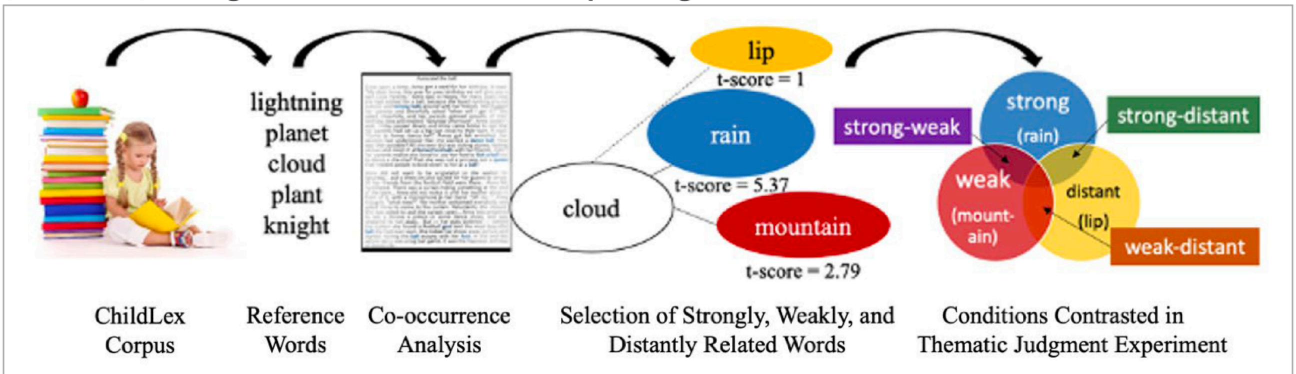
1b Children are more likely to show fast accurate responses (mean latency of accurate decisions) when the maximum collocation frequency between the contrasting word pairs is high (i.e., short path) versus low (i.e., long path).

Second, we expected to find a relation between overall thematic judgment accuracy and vocabulary size. In addition, we examined whether there would be a relation between accuracy or speed of thematic judgment responses and various reading skills. Our specific hypotheses were as follows:

2a Children with large vocabulary sizes will be more likely to respond more accurately in the thematic judgment task overall than children with small vocabulary sizes.

2b We examine whether children with good decoding or reading comprehension skills at the word, sentence, or text level respond overall faster in accurate responses or more accurately

FIGURE 1
Schematic Description of Stimuli Selection in Thematic Judgment Task. *Note.* Co-occurrence measure = t-score (Evert, 2008); The higher the t-score, the shorter the path length.



in the thematic judgment task than children with poor decoding or reading comprehension skills.

Finally, we conducted a mediation analysis to determine whether thematic judgment responses would mediate the relation between vocabulary size and reading comprehension. In addition, we examined whether these results hold, when we add decoding. If the results would not hold, we aimed to explore a model where vocabulary mediates the relation between thematic judgment and reading comprehension. Our hypotheses were as follows:

3a Children's vocabulary size (X) predicts children's thematic judgment responses (M), which in turn predicts reading comprehension skills (Y).

3b The effects in (3a) will remain stable when we add decoding ability as a covariate predicting reading comprehension (W).

3c Depending on the results, we may explore a variant where thematic judgment (X) predicts vocabulary size (M), which in turn predicts reading comprehension skills (Y).

Method

Participants

The data analyzed in this study were part of the iLearn pre- and post-test study (see Schmitterer & Brod, 2021). For this study, a cross-sectional subsample of children who participated at the beginning of the school year (in 2019) was used. Testing took place in group and individual sessions at the children's schools and was conducted by trained student experimenters. The children and a legal guardian consented to participate in the study. The study was approved by an institutional review board and the ethics committee of the authorized ministry of education prior to data collection.

We initially included data from 162 children from 17 primary schools and 29 classes from the region of Hesse in Germany. The mean age of the children was 8 years and 10 months (SD = 6.5 months), 50.6% were female and 49.4% were male (no other gender was reported). The children came from different socio-economic backgrounds (M_{HISEI} = 54.79, SD_{HISEI} = 16.75, $range_{HISEI}$ = 17–89; Ganzeboom, 2010). Fifty-seven percent of the participants spoke German as their first language, 17% spoke German as one of their first languages, and 26% spoke German as their second language. Parents of about 37% of the children reported that their children had an impairment (e.g., visual impairment, mental disorders). However, parents of only 10% of these children (n = 6) reported that their children did not receive any form of intervention.

The thematic judgment response was administered in individual sessions. As part of the project, teachers were asked to select students with reading difficulties and a total of up to eight students for individual testing. In addition, we asked them to rate the need for reading intervention for each student in three levels: "not necessary", "potentially necessary", or "definitely necessary". Teachers reported that in their ratings they orientated to the largest extent on their students' basic reading skills (i.e., decoding) and advanced reading skills (i.e., reading comprehension on a text level). They reported to orientate to a lesser extent on other linguistic prerequisites (i.e., phonological awareness or grammatical knowledge) or vocabulary knowledge (see Schmitterer & Brod, 2021, table 5). In addition, empirical data suggest that teachers' ratings were most strongly predicted by students' spelling skills followed by reading comprehension on a text-level (Schmitterer & Brod, 2021).

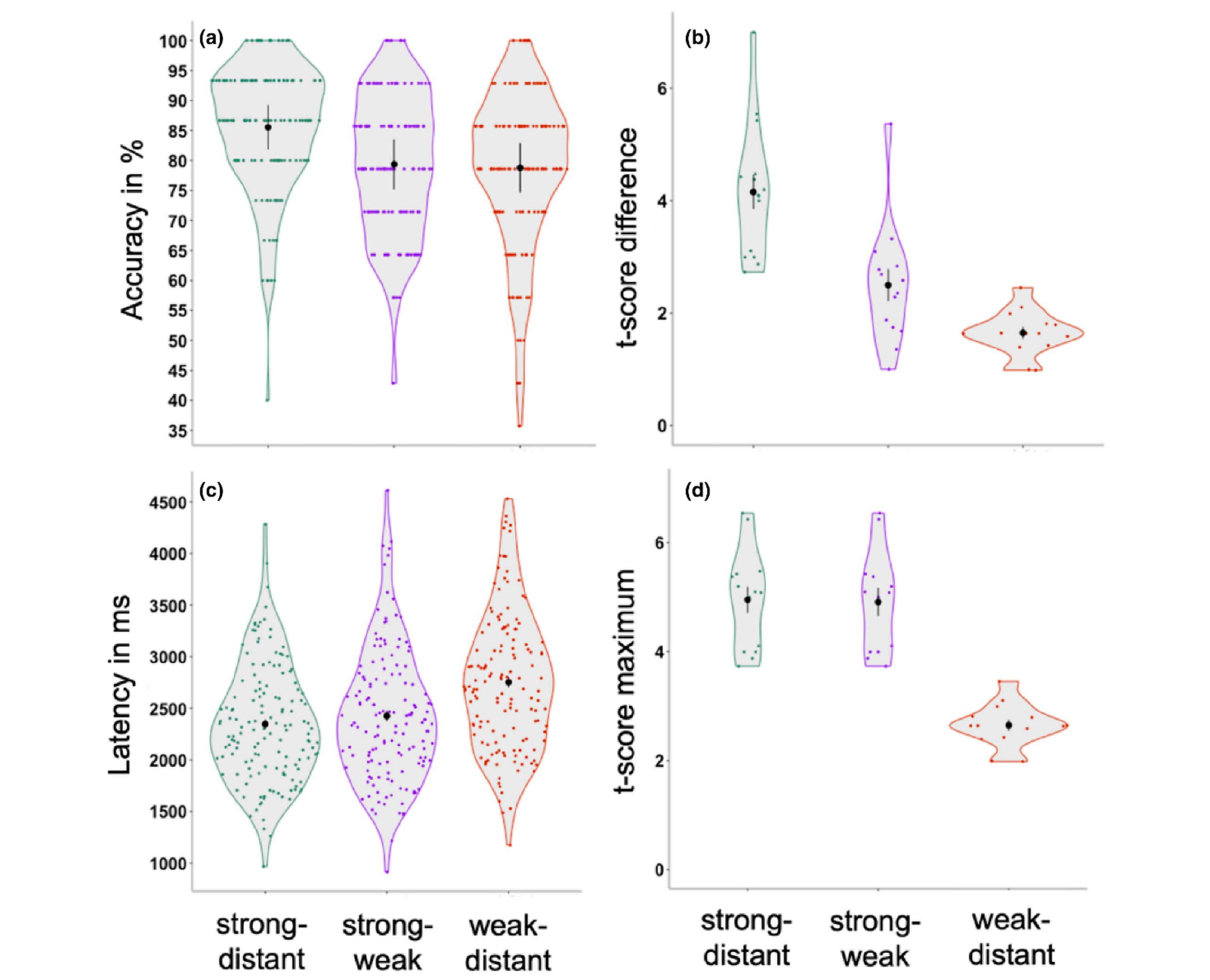
For the 162 students, we obtained teacher judgments for 52% of the students (n = 83). For these students, the teachers' judgments were that 68% definitely (n = 26) or potentially (n = 30) needed a reading intervention. From the 162 participants, 15 participants were excluded because they scored below the chance level on thematic judgment responses (n = 6), scored below clinical criteria for phonological working memory and nonverbal intelligence (n = 4), had a clinical condition that affected their responses (i.e., hearing impairment, ADHD, n = 2), or exhibited avoidant behavior during group or individual testing, resulting in data loss (n = 3).

Thematic Judgment Task Assessing Sensitivity to Collocation Frequency

Design and Procedure

The design, stimulus selection, and procedure of the thematic judgment experiment are described in more detail in the supplementary material and in the original study (Schmitterer & Schroeder, 2019b). The thematic judgment task used in this study was an adapted task. For this study, we excluded homonymous reference words (i.e., words with multiple meanings) and added a third condition: strong-distant [originally: unrelated], strong-weak [originally: association], and weak-distant. The task consisted of 15 context sentences that followed a subject-verb-object structure. The object of the sentence was one of 15 preselected high-frequency reference words. Based on these 15 reference words, we computed a co-occurrence frequency measure in a corpus of child-directed literature to determine the relationship between the reference word and the test words in the language input at three levels of relationship strength: strong, weak, or distant (Schroeder et al., 2015; t-score; Evert, 2008). Sentence conditions and verbs were manipulated based on t-score (see descriptions in supplementary material; Table S1; Figure 1; Figure 2, panels b and d) and controlled for lemma frequency, number of phonemes, phonological neighborhood density, and taxonomic relatedness (Tables S2 & S3). Familiarity was

FIGURE 2
Between Condition Effects of Participant Responses and Stimuli Statistics. *Note*. Panels (a) and (c) show responses of participants (*N* = 147), panels (b) and (d) show between condition effects of thematic relatedness measures per reference word (*N* = 15); arrows represent standard errors.



controlled for by parent AoA ratings in the original study.The experiment followed a mixed stimulus-within-condition design. Each child responded to all stimuli in each condition (45 stimuli; fully crossed). Each target sentence was presented three times (once for each condition) to each child (see Figure S1). Items were nested within target sentences and conditions (stimuli-within-condition). Participants were instructed as follows: "Soon you will hear a sentence followed by two words. Please judge which of the two words fits the sentence better. If you think the first word fits the sentence better, press 1. If you think the second word fits the sentence better, press 2." The children indicated their judgments by pressing the appropriate key on a laptop. In total, there were 90 pairs (45 counterbalanced pairs) organized into six lists of 15 trials, each containing each target sentence once, of which three lists (= blocks) were presented to each child in a Latin-square design. Stimuli presentation within blocks was random. Each block was followed by a short break. Prior to analyses, we conducted power analyses (Westfall et al., 2014) on stimuli-within-condition and fully crossed designs with an estimated 100 participants and 45 stimuli (d=.05) and found power estimates greater than 0.8.

Standardized Assessments

Vocabulary Size

Participants were asked to produce 40 nouns, verbs, or adjectives based on pictures on a computer screen (Wortschatz- und Wortfindungstest [WWT]; Glück & Glück, 2011). The raw score was the sum of all correct responses. The test provided a list of alternative correct responses for each item, derived from the normative sample, to increase the objectivity of assessment. Criteria for item selection were word length, word frequency, parts of speech, and semantic category (categories derived from Dornseiff, 2020). Semantic categories included nature, body parts (including health and exercise), school, leisure, kitchen (including food and cooking), and social interaction. In total, items from 15 semantic categories were included to index vocabulary breadth. Internal consistency was good (Cronbach's $\alpha = .87$). Convergent validity was reported with a receptive vocabulary subtest in the same normative sample ($r = .72$). In our sample, expressive vocabulary also had a strong correlation with a receptive test of syntactic comprehension ($r = .59$; TSVK; Siegmüller et al., 2011). The children's standardized scores on expressive vocabulary were low (see Table 1). We assume that this is explained by differences between the normative sample (monolingual only, different region) and our sample. The distribution of raw scores was acceptable (see Figure S3).

Decoding

Participants were asked to correctly read pseudowords (1 min) from a list of 156 items (Salzburger Lese- und Rechtschreibtest [SLRT-II]; Moll & Landerl, 2010). The raw score was the number of correctly read pseudowords. Parallel test reliability was reported as $r = .90$ for third grade. Standardized scores were reported for the middle of third grade.

Reading Comprehension

Reading comprehension was assessed on the word, sentence, and text level (ELFE-II, Lenhard et al., 2017). All three subtests were speeded (2×6 min, 1×7 min). For the word level, children had to choose the matching word to a picture from a list of five words. For the sentence level, children had to fill a gap in a sentence by choosing the correct word from a list of five words. For the text level, children were presented with short text passages describing familiar situations (e.g., play date) or natural phenomena (e.g., diamonds made of coal). Children answered multiple choice questions about the content of each text passage. The raw score for all three tests was the sum of correct responses. Split-half reliability for each subtest was reported to be acceptable to high (word level: $r = .98$; sentence level $r = .97$; text level: $r = .87$).

Covariates

Phonological Working Memory

Participants completed a digit span task (Arbeitsgedächtnistestbatterie [AGTB]; von Hasselhorn et al., 2012) consisting of eight adaptive test trials: they would advance to the next level after correctly answering two consecutive items with the same digit span. The raw score was calculated as the mean of the last two correctly

TABLE 1
Descriptive Statistics of Experimental and Standardized Assessments

Variable	*M*	SD	Range	Max
Thematic judgment accuracy (%)	81.32	9.06	56–98	43[a]
Thematic judgment latency (ms)	2465	1151.74	451–6000	—
Vocabulary size	16.75	8.05	1–33	40
	32.57	19.5	0–62	
Decoding of pseudowords	40.55	14.97	8–88	156[b]
	41.48	8.85	27–66	
Reading comprehension word level	34.21	9.27	10–58	75[b]
	45.92	8.18	25–66	
Reading comprehension sentence level	12.11	4.98	0–24	36[b]
	43.11	8.14	25–62	
Reading comprehension text level	6.69	3.76	0–17	26[b]
	41.84	8.89	25–63	
Phonological working memory	3.91	0.62	2.4–5.5	—
	46.16	8.6	29–67	

Note. $N = 147$; for standardized assessments the first row depicts raw scores, the second row depicts standardized scores (T).
[a]Two items were excluded from the analysis.
[b]Speeded tests.

answered span sizes. The internal consistency was good (Cronbach's α = .68–.78, depending on age).

Analyses

All analyses were carried out in RStudio (2022.12.0) using R 4.1.0 (R Core Team, 2021). Codebooks, code, data, and supplementary material with additional descriptions of the material, tables, and figures are available on OSF (link provided in Data Availability Statement).

Thematic Judgment Experiment

Preprocessing

In the experimental analysis, affirmative responses to the stronger related words are referred to as "accurate" responses. Participants' accurate and inaccurate responses were distributed across all items. We identified six participants and two items that were outliers regarding response accuracy (below chance). The reported results are without outliers.

Latencies of inaccurate responses were significantly longer than latencies of accurate responses (see Figure S2). Inaccurate responses provided too few data points to draw significant conclusions from (e.g., Fazio, 1990). Thus, we only included latencies of correct responses (80.7%). For accurate responses, we excluded latencies below 300 and above 8000 ms. In addition, we excluded latencies that deviated more than 2.5 SD from the log-transformed participant or item mean. In total, appr. 1.3% of data points for correct responses were discarded.

Analyses

For hypotheses (1a) and (1b), we analyzed between-condition effects (i.e., strong-distant vs. strong-weak vs. weak-distant pairs, see Figure 1) in response accuracy and latency with (generalized) linear mixed-effects models using the {(g)lmer} function from the lme4 package (Bates et al., 2015). The response accuracy model was based on a binomial error distribution and used a logit link transformation (Cohen et al., 2013). Response latencies were log-transformed prior to analysis. Intercepts for participants and items were included as crossed random effects. Because item structure was crossed (target sentences crossed with participants) and nested (items nested within target sentences), we included target sentences as an additional random effect. To account for contextual information and because participants were nested in classes, we also included class as an additional random effect. We conducted an a priori analysis with null models that included the condition as a singular fixed effect using different random-effect variations and compared model fit with the {anova} function from the car package (Fox & Weisberg, 2019). We chose one of the most parsimonious random effect structures. All continuous fixed effects were standardized, and conditions were dummy-coded with the strong-distant condition as the reference level.

To validate our procedure, we then ran models with experimental lists, presentation blocks, or the thematic relation between the verb and the reference word in the target sentence, as a fixed effect (see code r1 on OSF). The results showed a list effect for response accuracy models but not for response latency models. However, because these list effects were not systematic, we concluded that this effect was random. There was also a presentation block effect, with children responding significantly more accurately and faster during the first presentation block than during the last presentation block. Since the thematic judgment experiment was the last task of the individual session and we assume that the children knew this, this could have influenced their concentration. Therefore, we controlled for block effects in the final models. Finally, we did not find any effects of the verb-reference word relation, suggesting that our target sentences provided plausible but not idiomatic contexts for the reference word.

The final models for hypotheses (1a) and (1b), included condition and phonological working memory as fixed effects. The final model structure was as follows: (g) lmer (av ~ condition + ph. working memory + block + (1 | participant) + (1 | reference word) + (1 | item) + (1|class), data). To check for multicollinearity, we calculated variance inflation factors for each model using the {vif} function from the car package (Fox & Weisberg, 2019). All variance inflation factors in these and all subsequent analyses were less than 2. Intra-class correlation coefficients (ICC), the explained variability for each random effect and conditional and marginal R^2 for each model were calculated using the {tab_model} function from the sjPlot package (Lüdecke et al., 2023). Post hoc analyses were conducted by applying single-degree-of-freedom contrasts based on the cell mean estimates in separate models with the intercept set to zero and using the same parameters as in the base models.

To validate the relation between our results and the collocation frequency measures, we ran models in which we replaced the condition factor with the t-score difference and the t-score maximum value. We entered both t-score measures into both models to see whether the measure we hypothesized (hypotheses 1a and 1b) would predict the respective thematic judgment response over the other measure for the respective response type. We assumed that the t-score difference measure would be a predictor of response accuracy because, in the response accuracy model, all responses (correct or incorrect) were entered into the model. Thus, the responses represented children's sensitivity to the relation between the reference word and the test words. In the latency analysis, only correct responses were included, and responses represented children's speed in choosing the more strongly related word. Therefore, we assumed that the t-score maximum would be the stronger predictor here. The results are summarized in Table 2. Figure 2 highlights the relation between collocation measures and behavioral responses per condition.

For hypotheses (2a) and (2b) we added vocabulary size, decoding, and several reading comprehension tasks of varying complexity as fixed effect in the model. The final model structure was: (g)lmer (av ~ condition +

TABLE 2
Model Summaries for Effects of Condition and Collocation Statistics

	Accuracy			Latency		
Fixed effects	**Odds ratio**	**SE**	***p***	**Estimate**	**SE**	***p***
Intercept	7.92	1.51	**<.001**	2132.15	63.34	**<.001**
Condition effects						
Strong—Weak	0.58	0.14	**.020**	1.01	0.02	.684
Weak—Distant	0.55	0.13	**.010**	1.18	0.03	**<.001**
Main effects covariates						
Ph. working memory	1.18	0.06	**.002**	1.01	0.02	.532
Presentation block	0.88	0.03	**<.001**	0.94	0.01	**<.001**
Random effects						
σ^2	3.290			0.146		
$\tau_{00\ participants}$	0.229			0.046		
$\tau_{00\ items}$	0.329			0.003		
$\tau_{00\ class}$	0.007			0.002		
$\tau_{00\ reference\ words}$	0.111			0.003		
ICC	0.171			0.271		
Marginal R^2/Conditional R^2	0.029/0.195			0.045/0.304		
Fixed effects	**Odds ratio**	**SE**	***p***	**Estimate**	**SE**	***p***
Intercept	5.47	0.73	**<.001**	2255.83	63.93	**<.001**
Effects of collocation statistics						
t-score difference	1.52	0.29	**.029**	1.01	0.02	.795
t-score maximum	0.80	0.15	.232	0.92	0.02	**<.001**
Main effects covariates						
Ph. working memory	1.18	0.06	**.002**	1.01	0.02	.532
Presentation block	0.88	0.03	**<.001**	0.94	0.01	**<.001**
Random effects						
σ^2	3.290			0.146		
$\tau_{00\ participants}$	0.229			0.046		
$\tau_{00\ items}$	0.355			0.004		
$\tau_{00\ class}$	0.007			0.002		
$\tau_{00\ reference\ words}$	0.098			0.004		
ICC	0.173			0.278		
Marginal R^2/conditional R^2	0.027/ 0.195			0.044/0.310		

Note. Accuracy models: 6321 observations; Latency models: 4673 observations; all models: 147 participants, 29 classes, 43 items (trials) and 15 reference words. Bold values represent significant effects.

vocabulary size/decoding/reading comprehension + ph. working memory + (1 | participant) + (1 | reference word) + (1 | item) + (1| class), data). We again checked for unobtrusive multicollinearity. Furthermore, we used the {anova} function from the car package (Fox & Weisberg, 2019) to compare model fit between the base models (Table 2) and the model with the added fixed effects to determine whether a significant amount of additional variance was explained by the added variable. Results are reported in Tables S5–S7 and shown in Figure 3.

Mediation Analysis

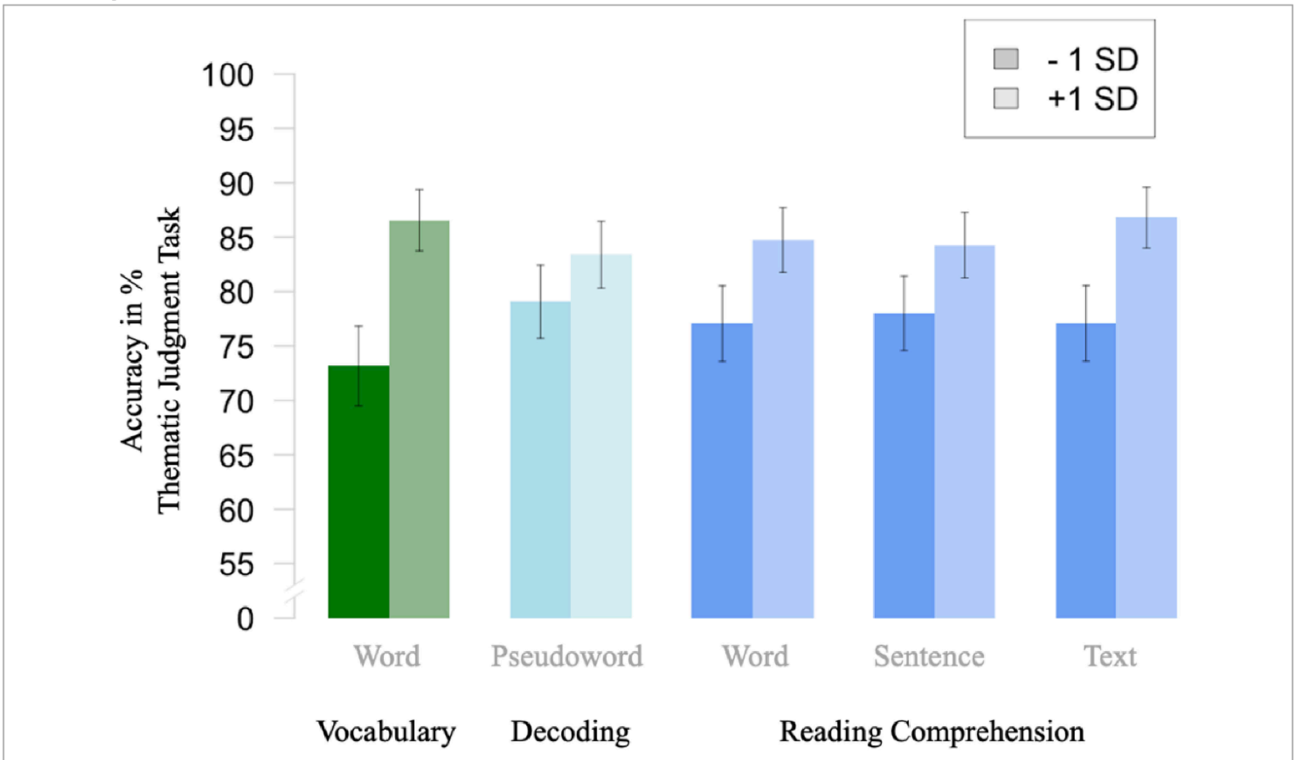
Mediation analyses were conducted using the lavaan package (Rosseel et al., 2017). Prior to our analyses, we conducted a power analysis for two models on the pwrSEM application (Wang & Rhemtulla, 2021). In the first model, we tested whether thematic judgment (M) mediated the relation between vocabulary size (X) and reading comprehension (Y). In the second model, we added a direct effect from decoding (W) on reading comprehension. In an additional model, we explored a variant, where vocabulary size mediated the relation between thematic judgment and reading comprehension. We selected the smallest effect sizes of interest ($a = .35$; $b = .35$, $c = .35$, and $c' = .3$; factor loadings = .8; Lakens et al., 2018) based on effects reported in previous literature (e.g., Hjetland et al., 2020; Verhoeven et al., 2011) and conducted the analyses for different sample sizes (100, 120, 150, 162). The results showed power estimates of above .8 for all effects of interest when the sample size was <= 120. We did not calculate the mediation analysis as multilevel models. The explained variability at the class level for thematic judgment was very small (less than 2%) which contributed to convergence issues in multi-level models.

Results

Thematic Judgment Task in Middle Childhood

For hypothesis (1a), we hypothesized that children would be more likely to identify a word pair as "better matched" (mean accuracy) than a distractor word pair, if the difference in collocation frequency between the contrasting word pairs in the input language is large (i.e., low competition with the distractor) versus small (i.e., high competition with the distractor). Therefore, the difference in thematic relations (i.e., the t-score difference) between conditions is the strongest indicator of our hypothesis. The results were consistent with this hypothesis. Children responded more accurately in the strong-distant condition ($M = 86\%$, $SE = 1\%$) compared to the weak-distant

FIGURE 3
Effects of Vocabulary Size and Reading Abilities on Thematic Judgment Responses. *Note*. Plot based on raw data; arrows represent standard errors.



condition (M=79%, SE=1%, Δ=7%, β=.60, SE=.23, t=2.59, p<.01) and in the strong-distant condition compared to the strong-weak condition (M=79%, SE=1% Δ=7%, β=.54, SE=.23, t=2.32, p<.05). However, response accuracy was not significantly different between the strong-weak and the weak-distant conditions (Δ=0.1%, β=.06, SE=.23, t=.26, p>.05). This was a similar pattern to the between-condition effects shown by the t-score difference values, indicating that the operationalization of the task was consistent with children's response patterns (Figure 2, panels a and b). Furthermore, we ran an additional model with a standardized value of the t-score difference as a continuous fixed effect. while controlling for the t-score maximum and found a main effect for t-score difference (see Table 2). Thus, the larger the difference, the more likely children were to choose the more strongly related test word.

In hypothesis (1b), we hypothesized that children would be more likely to show a fast accurate response (mean latency of accurate decisions) when the maximum collocation frequency between the contrasting word pairs was high (short path) versus low (long path). This hypothesis was also supported. Children were faster in the strong-distant (M=2327 ms, SE=26 ms) and the strong-weak (M= 2391 ms, SE=29 ms) condition compared to the weak-distant condition (M=2712 ms, SE=33 ms; ($\Delta_{\text{strong-distant}}$ = 385 ms, β=.17, SE=.03, t=6.9, p<.001; $\Delta_{\text{strong-weak}}$= 321 ms, β=.16, SE=.03, t=6.33, p<.001)). Response latencies of accurate responses did not differ between the strong-distant and strong-weak conditions (Δ=64 ms, β=.01, SE=.02, t=0.41, p>.05). This pattern also matched the between-condition pattern of the maximum t-score (Figure 2, panels c and d). Finally, in line with our assumptions, the analysis with continuous collocation frequency measures found that children gave faster accurate responses on trials with higher t-score maximums than on trials with lower t-score maximums (Table 2).

Finally, we found a significant main effect for phonological working memory in the accuracy models (Table 2). To rule out the possibility that this finding affected between-condition effects, we conducted additional analyses with data that excluded children with phonological working memory problems (T<35). These analyses yielded the same between-condition pattern of results without a main effect of phonological working memory (see code r1 on OSF).

Thematic Judgment and Vocabulary Size, Decoding and Reading Comprehension

All results for this section are reported in Tables S5–S7 and are summarized in Figure 3. In line with our hypotheses, we found that children with larger vocabulary sizes also gave more accurate responses in the thematic judgment task. The same was true for children's reading comprehension on sentence, and text level, as well as in tendency on the word level. Among reading comprehension abilities, the size of the effect increased with task complexity. Thus, we were able to affirm hypotheses (2a) and refine assumptions regarding hypothesis (2b).

Thematic Judgment as Link between Vocabulary Size and Reading Comprehension

Based on our findings, we selected two reading measures for mediation analysis. First, a reading comprehension composite score based on the scores of the reading comprehension subtests (i.e., word, sentence, and text level). The mean accuracy score of thematic judgment correlated significantly with the outcome of all subtests (see Figure S4). Second, we chose reading comprehension on the text-level as a manifest variable because it had the strongest correlation with thematic judgment and correlated significantly with vocabulary size. We focused on response accuracy because response latency or a response/accuracy measure (LISAS; Vandierendonck, 2017) did not correlate significantly with vocabulary size or reading comprehension (see code r^2 on OSF). We ran two models for each dependent reading variable. One for the basic mediation analysis to see if thematic judgment (M) mediated the relation of vocabulary size (X) and reading comprehension (Y; Figure 4a) and the second one with decoding to see if the mediation effect would remain stable (Figure 4b). Because effects changed between the first and the second model, we added another exploratory analysis (Figure 4c) to clarify the relations between the variables. All variables were entered into the models as z-standardized raw scores.

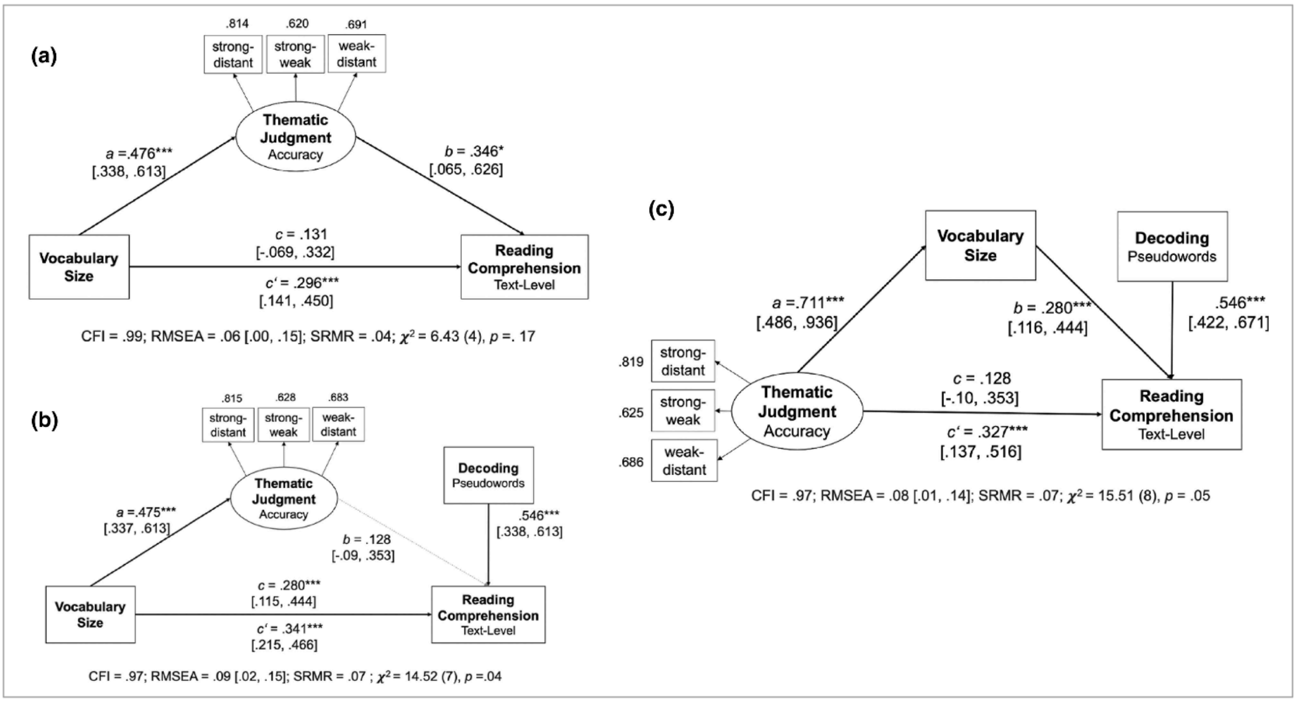
Accuracy of Thematic Judgment as Mediator for Reading Comprehension

Thematic judgment responses were not standardized. Therefore, we performed internal consistency analyses and a cluster analysis using the {alpha} and {iclust} functions of the *psych* package (Revelle, 2022). We removed 15 items that were either negatively correlated with the scale or showed item-scale-correlations of below 0.2. The final item set (29 items; strong-distant: 11; strong-weak: 8; weak-distant: 10) had an acceptable internal consistency (Cronbach's α=.73). The cluster analysis indicated that internal consistency was highest if one cluster was assumed (Table S4). We entered thematic judgment as latent construct into the models. The latent variables represented the conditions (see Figure 4).

Composite Reading Comprehension Score

For the composite reading score, the word and sentence level tasks had a good fit (>.8) while the text level task

FIGURE 4
Mediation Models for Reading Comprehension on the Text Level and Reading Fluency. ***Note.*** ***p*** **< .05*;** ***p*** **< .01 **;** ***p*** **< .001.**



had a lower but acceptable fit (>.6). Both models for this score are shown in Figure S5. The mediation model of the composite reading score had an acceptable fit (CFI = .96; RMSEA = .09 [.04, .14]; SRMR = .07; χ^2 = 25.72 (12), *p* =. 01). Vocabulary size did not significantly predict reading comprehension, either before or after accounting for thematic judgment responses (c: β = −.02, *SE* = .09, *z* = −.21, *p* > .05; c': β = 12, *SE* = .07, *z* = 1.731, *p* = .08). However, vocabulary size predicted thematic judgment (a: β = .47, *SE* = .07, *z* = 6.78, *p* < .001), and thematic judgment in turn predicted reading comprehension (b: β = .30, *SE* = .13, *z* = 2.27, *p* < .05), leading to a significant mediation effect (a*b: β = .14, *SE* = .06, *z* = 2.20, *p* < .05).

The second model included decoding. The model also had an acceptable fit (CFI = .97; RMSEA = .08 [.04, .12]; SRMR = .10; χ^2 = 34.11 (17), *p* =. 01). However, including decoding changed the pattern of results. Vocabulary size now significantly predicted reading comprehension both before and after considering thematic judgments (c: β = 18, *SE* = .05, *z* = 3.29, *p* < .01; c': β = 19, *SE* = .04, *z* = 4.5, *p* < .001) and still predicted thematic judgment (a: β = .47, *SE* = .07, *z* = 6.76, *p* < .001). The mediation effect was not significant (a*b: β = .01, *SE* = .03, *z* = .26, *p* > .05) as thematic judgment no longer predicted reading comprehension (a: β = .02, *SE* = .07, *z* = .26, *p* > .05). Expectedly, decoding strongly predicted reading comprehension (β = .73, *SE* = .05, *z* = 13.51, *p* < .001).

Reading Comprehension on Text Level

The mediation model for text-level reading comprehension had a good fit (CFI = .99; RMSEA = .06 [.00, .15]; SRMR = .04; χ^2 = 6.43 (4), *p* =. 17; see Figure 4a). Vocabulary size did not significantly predict reading comprehension before, but after thematic judgment responses were considered (c: β = .13, *SE* = .10, *z* = 1.28, *p* > .05; c': β = .30, *SE* = .08, *z* = 3.76, *p* < .001). Vocabulary size predicted thematic judgment (a: β = .48, *SE* = .07, *z* = 6.79, *p* < .001) and thematic judgment in turn predicted reading comprehension (b: β = .35, *SE* = .14, *z* = 2.41, *p* < .05). The mediation effect was significant, indicating that thematic judgment contributed to the significant relation between vocabulary size and reading comprehension (a*b: β = .16, *SE* = .07, *z* = 2.32, *p* < .05).

The model including decoding had an acceptable fit (CFI = .97; RMSEA = .09 [.02, .15]; SRMR = .07; χ^2 = 14.52 (7), *p* = .04). However, the effect of thematic judgment on reading comprehension and the mediation effect were not significant (b: β = .13, *SE* = .12, *z* = 1.11, *p* > .05; a*b: β = .06, *SE* = .05, *z* = 1.10, *p* > .05; see Figure 4b). Vocabulary size predicted thematic judgment (a: β = .48, *SE* = .07, *z* = 6.77, *p* < .001) and reading comprehension (c: β = 28, *SE* = .08, *z* = 3.34, *p* < .01; c': β = .34, *SE* = .06, *z* = 5.31, *p* < .001). In addition, decoding strongly predicted reading comprehension (β = .55, *SE* = .06, *z* = 8.56, *p* < .001).

Despite a moderate correlation with text-level reading comprehension (*r* = .31), thematic judgment no longer contributed to the explanation of reading comprehension or the relation between vocabulary size and reading comprehension, when decoding was added. Thus, we went on to explore a variant where vocabulary size would mediate the relation between thematic judgment and reading comprehension.

We computed an additional model in which vocabulary size was entered as a mediator of thematic judgment and reading comprehension and included decoding (see Figure 4c). This model had an acceptable fit (CFI = .97; RMSEA = .08 [.01, .14]; SRMR = .07; $\chi^2 = 15.51$ (8), $p = .05$). Thematic judgment did not initially predict text-level reading comprehension (c: $\beta = 13$, $SE = .11$, $z = 1.11$, $p > .05$). However, thematic judgment strongly predicted vocabulary size (a: $\beta = 71$, $SE = .11$, $z = 6.1$, $p < .001$) and after accounting for vocabulary size, significantly predicted reading comprehension (c': $\beta = .33$, $SE = .10$, $z = 3.38$, $p < .01$). The mediation effect of vocabulary size on the relation between thematic judgment and reading comprehension was significant ($\beta = .20$, $SE = .06$, $z = 3.00$, $p < .01$). Vocabulary size (b: $\beta = .28$, $SE = .08$, $z = 3.34$, $p < .01$) and decoding ($\beta = .55$, $SE = .06$, $z = 8.59$, $p < .001$) both also significantly predicted reading comprehension.

Discussion

This study focused on the relation between sensitivity to collocation frequency, vocabulary size, and reading comprehension skills in third-grade students with low reading comprehension skills. Our results support the hypothesis that sensitivity to collocation frequency is related to both vocabulary size and reading comprehension. Thus, the ability to use collocation frequency as a source of semantic information is positively related to the acquisition of comprehension skills throughout development. We also conclude that collocation sensitivity connects comprehension abilities on the word level (i.e., requiring lexical retrieval) and more complex comprehension tasks (i.e., requiring inferential reasoning) on the sentence or text level. Overall, the results clearly indicate that thematic judgment (i.e., a measure of sensitivity to collocation frequency), in addition to vocabulary size contributed to explaining variance in reading comprehension ability, even when other relevant cognitive processes (i.e., decoding) are controlled.

Thematic Judgment as Indicator for Semantic Processing Throughout Childhood

Our results confirm the results of previous studies showing young children, older children, and adults are sensitive to thematic relations (Lany & Saffran, 2013; Savic et al., 2023; Schmitterer & Schroeder, 2019b; Unger et al., 2023, 2020; Unger & Fisher, 2021; Vales & Fisher, 2019; Wojcik & Kandhadai, 2020). Our study adds to previous studies by directly linking measures of co-occurrence to children's responses between conditions (t-score; Evert, 2008; Unger et al., 2023). Thus, we confirm and expand previous findings that find co-occurrence statistics to be directly related to behavioral outcomes in word knowledge (Unger et al., 2023). Of note for future studies, we used an item subset of a study with 4- to 6-year-olds (Schmitterer & Schroeder, 2019b). In the present study with 8-year-olds, we found similar results. We used thematic relatedness of high frequent words with an early age of acquisition onset (e.g., Braginsky et al., 2019). This group of words has been found to be similarly represented in semantic networks of toddlers and adults (Hills et al., 2008) and could serve as an indicator for sensitivity to collocation frequency across development.

The Relation of Thematic Judgment Responses and Vocabulary Size

Our results also suggest that vocabulary size and thematic judgments are linked. This finding is consistent with previous work linking the structure and growth of semantic networks based on co-occurrence measures to early vocabulary development (Flores et al., 2020; Fourtassi, 2020; Fourtassi et al., 2020; Hills et al., 2009). Children were more likely to respond in the direction of the strongest related word in the thematic judgment task when they had a larger compared to a smaller vocabulary. This is consistent with our assumption that children with higher compared to lower sensitivity to collocation frequency in the input language are also more likely to form highly interconnected semantic networks and thus learn more new words in the same amount of time. At the same time, high connectivity may facilitate lexical retrieval of word-level conceptual information or semantic relatedness information. The average path length of links in the network of children with highly interconnected vocabularies should be shorter than in children with sparsely interconnected semantic networks (see e.g., Beckage et al., 2011). This should make retrieval less effortful.

Thematic Judgment, Inferential Reasoning, and Reading Comprehension

Our results indicate that thematic judgments in middle childhood are related to reading comprehension skills of varying complexity (see Figure 3). These findings are consistent with a cascading understanding of comprehension processes in reading. For example, the reading system framework (Perfetti & Stafura, 2014) suggests that word-level comprehension processes are a pressure point for higher-level reading processes. In the case of thematic judgments, information about word relations is retrieved.

Thus, if retrieval of information about word relations is impaired, reading comprehension at the word, sentence, and text level will also be low.

In addition, our results show that word-level thematic judgment responses in 8-year-olds tended to be better explained by individual differences in complex reading comprehension tasks on the text-level than by a word-level reading comprehension task. We suggest that this is explained by the fact that thematic judgments measure at least two cognitive processes. First, thematic judgments foster lexical retrieval of semantic information. This ability is relevant in vocabulary and reading comprehension tasks of any complexity. Second, thematic judgments require inferential reasoning. Inferential reasoning is relevant for complex reading comprehension tasks (Ahmed et al., 2016; Duke & Cartwright, 2021) and has been linked to collocation frequency in young children (Fisher et al., 2011). In line with our assumptions, our findings suggest that children who are more sensitive to collocation frequency, as indicated by thematic judgment responses, are also better in reading tasks that require inferential reasoning than children who are less sensitive to collocation frequency.

Thematic Judgment linking Vocabulary Size and Reading Comprehension

Our mediation analysis initially indicated that response accuracy in thematic judgments mediated the relation between vocabulary size and reading comprehension (Figure 4). However, when we added decoding as an additional predictor of reading comprehension skills, the effect of thematic judgment on reading comprehension became insignificant while the effect of vocabulary size on reading comprehension increased. In line with reading research, children's decoding skills explained a large amount of variance in reading comprehension (Hjetland et al., 2020; Hoover & Tunmer, 2018). Because decoding relies on phoneme-grapheme conversion (see e.g., Castles et al., 2018; Hjetland et al., 2020), we conclude that a large amount of variance in reading comprehension that is explained by decoding, represents phonological processes. Our productive vocabulary size measure included phonological processes, but we controlled phonological processes in the thematic judgment task. Thus, we assume that the inclusion of decoding shifted the shared variance that was explained by vocabulary size in reading comprehension. Before including decoding, vocabulary size explained some phonological and some semantic variance in reading comprehension. In addition, thematic judgment responses then mediated semantic lexical retrieval processes between vocabulary size and reading comprehension. However, after the addition of decoding skills accounted for a large portion of the phonological variance in reading comprehension, vocabulary size shared primarily semantic processes with reading comprehension. This led to an attenuated mediation effect of the thematic judgment on the relation of vocabulary size and reading comprehension because the relation between vocabulary and reading comprehension then explained semantic lexical retrieval processes to a larger extend.

Because we hypothesized that the thematic judgment task measures not only lexical retrieval but also inferential reasoning, and thus should explain an additional amount of variance in reading comprehension, we tried an alternative approach. We examined whether the relation between thematic judgment and reading comprehension was mediated by vocabulary size. In this scenario, vocabulary size would bridge semantic and phonological processes in lexical retrieval at the word level. In addition, thematic judgment would predict vocabulary size because sensitivity to collocation frequency may precede (productive) vocabulary growth in development (Fourtassi et al., 2020; Hills et al., 2009). Furthermore, thematic judgment would be related to reading comprehension based on shared variance explained by inferential reasoning that does not overlap with variance explained by lexical retrieval of concepts (i.e., vocabulary size) or phonological processing (i.e., decoding). This exploratory model (Figure 4c) then showed that all three variables contributed significantly to explaining variance in text-level reading comprehension, and that vocabulary size mediated the relation between thematic judgment and reading comprehension.

We concluded that this exploratory approach fits well with research on early language comprehension development (Fisher et al., 2011; Fourtassi et al., 2020; Savic et al., 2023; Unger & Fisher, 2021). In addition, this approach supports middle childhood reading research in several ways. First, it shows that both language comprehension and decoding contribute to reading comprehension (Hjetland et al., 2020; Hoover & Tunmer, 2018). Second, it shows that word-level lexical processes predict higher-level reading comprehension skills (Perfetti & Stafura, 2014). Third, it confirms that cognitive processes behind language comprehension are complex (Castles et al., 2018), and finally, that vocabulary may bridge cognitive processes in reading comprehension (Duke & Cartwright, 2021).

Limitations

Our study had several limitations. First, performance on our measure of vocabulary size was lower than expected (although with an acceptable distribution; see Figure S3), and vocabulary size did not correlate with some typically related measures (i.e., reading comprehension on word or sentence level; see Figure S4). We suspect that this is due to differences between the normative sample and our sample (i.e., region and language background). Furthermore, our vocabulary measure was a measure of productive vocabulary, whereas the reading comprehension measures were receptive. However, most of our results that included vocabulary size were in line with previous literature. We would recommend replicating our findings with other vocabulary measures.

Second, the item set for the thematic judgment task was not standardized, and after reducing the item set, the internal consistency and the loadings for the latent variable did not go beyond the interpretation of an acceptable fit. This is not uncommon for meaning-based measures (see, e.g., the measure for morphological awareness in Caravolas et al., 2019). However, even though the item set was highly controlled for frequency, phonological and orthographic features, and taxonomic relatedness, some additional semantic features like idiomatic or compositional multiword expressions may have also influenced lexical processing in some items (Jiang et al., 2020; Kessler et al., 2021; Ulicheva et al., 2021). Thus, more, and different item sets for thematic judgments should be developed and tested.

Third, our experimental analysis showed that the largest amount of explained variability was between participants or between items, while second-level effects of variability between reference words or between classes or schools were very small. These results suggest that our assessment of thematic judgment captured a sensitivity to linguistic contexts that was relatively independent of social contexts. This may be partly explained by our selection of high-frequency words. We are not aware of any other study that has reported class 1 and class 2 level effects for both items and participants. This makes comparisons difficult. However, it would be interesting to see more studies reporting similar random effects structures in the future. Regarding the small amount of variability explained by class or school level effects, we suspect that the fact that children were pre-selected by teachers for individual sessions based on their reading ability may have contributed to this finding. In addition, only up to eight children per class participated in individual sessions, which meant that we had a small number of children per class in the analysis. Future studies could address contextual variability by administering similar tasks to whole classes.

Future Directions

The results of our study, together with previous literature suggest that thematic judgment could be an indicator linking comprehension processes in early language development and early reading development (e.g., Schmitterer & Schroeder, 2019b; Unger et al., 2020; Unger & Fisher, 2021; Vales & Fisher, 2019). Longitudinal studies that focus on the role of sensitivity to collocation frequency across development could further explore whether the relation between this ability and vocabulary size and various reading comprehension skills changes across development and what external influences (i.e., access to books, quality of explanations) are associated with these relations.

Second, we found that children who are likely to benefit from reading intervention have better reading comprehension skills and larger vocabularies if they are also more accurate in their thematic judgments. In addition, laboratory-based training studies that rely on co-occurrence statistics as a baseline for their design already yielded positive results (Savic et al., 2022, 2023). Moreover, some training approaches for special needs populations (e.g., autistic children) use network visualization to make the shared origin of semantic relatedness palpable (i.e., Thinking Maps; Mashal & Kasirer, 2011). These approaches could be combined with existing reading comprehension intervention programs (see, e.g., Duke et al., 2021).

Third, experts agree that the comprehension processes in reading models are more complex than decoding processes (Castles et al., 2018). Some studies of reading development in recent years have expanded the language component in reading models by including separate meaning-related lexical components at the word-level, such as morphological awareness (e.g., Caravolas et al., 2019; Lervåg et al., 2018). Thematic judgment could be an additional semantic component that provides an informative contribution to language comprehension.

Finally, the influence of collocation frequency on children's language processing has been studied more in oral language than in written language. However, some studies focusing on multiword expressions already indicate that collocation frequency affects children's reading also in written language processing (Jiang et al., 2020). Thematic judgment could be used as an operationalization to examine the effects of collocation frequency in written material as well.

Summary

In this study, we examined the relation between a measure of sensitivity to collocation frequency in spoken language (i.e., thematic judgments) with vocabulary size and reading comprehension measures of varying complexity in 8-year-old children ($N = 147$). Our results indicate that responses in thematic judgments are related to co-occurrence statistics from child-directed literature and are associated with individual differences in vocabulary size and reading comprehension skills of varying complexity. Moreover, interindividual differences in thematic judgments predicted vocabulary size and reading comprehension even when decoding was controlled. The results link studies of lexical growth and inferential reasoning in early development with comprehension

processes during reading acquisition in middle childhood. We conclude that thematic judgments are based on a sensitivity to collocation frequency and tap into semantic processes of lexical retrieval and inferential reasoning. Finally, sensitivities to collocation frequency have yet to be addressed in combined language and reading comprehension interventions.

Author Contributions

Alexandra Schmitterer was responsible for conceptualization, methodology, formal analysis, resources, data curation, project administration, writing the original draft, and visualization. Caterina Gawrilow and Claudia Friedrich contributed through writing, reviewing, editing, and supervision.

Acknowledgments

We thank our student research assistants for their support in data collection. Furthermore, we thank Marcus Hasselhorn, Garvin Brod, Leonard Tetzlaff, and all collaborators for their role in supporting the iLearn project. Open Access funding enabled and organized by Projekt DEAL.

Funding Information

This research was funded by the German Federal Ministry of Education and Research, as part of the LONDI research program and its subproject iLearn (grant number: 01GJ1710B).

Ethics Approval and Conflict of Interest Disclosure

This study was approved by the Institutional Review Board (Ethics Committee) of the DIPF | Leibniz Institute for Research and Information in Education (reference number: iLearn) and the Ethics Committee of the Hessian Ministry for Culture, Education and Opportunities (reference number: GWU 695). Parents or legal guardians gave written consent for their children to participate prior to data collection. Children could also agree to participate in the study with a simplified version of the consent form and were informed before data collection about their right to refuse or stop participation at any time without negative consequences. In addition, teachers gave written consent for their own participation in the study prior to data collection. The consent forms included permission to use the collected data in scientific research projects, including publications, in an anonymized form. The authors have no conflicts of interest to declare.

Data Availability Statement

The analyses presented here were preregistered: https://osf.io/eajfc/. Data, codebook, code, and supplementary material are available here: https://osf.io/m9fbw/.

REFERENCES

Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology*, *44*, 68–82. https://doi.org/10.1016/j.cedpsych.2016.02.002

Arias-Trejo, N., & Plunkett, K. (2013). What's in a link: Associative and taxonomic priming effects in the infant lexicon. *Cognition*, *128*(2), 214–227. https://doi.org/10.1016/j.cognition.2013.03.008

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS One*, *6*(5), e19348. https://doi.org/10.1371/journal.pone.0019348

Braginsky, M., Yurovsky, D., Marchman, V. A., & Frank, M. C. (2019). Consistency and variability in children's word learning across languages. *Open Mind*, *3*, 52–67. https://doi.org/10.1162/opmi_a_00026

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review*, *8*, 531–544. https://doi.org/10.3758/BF03196189

Caravolas, M., Lervåg, A., Mikulajová, M., Defior, S., Seidlová-Málková, G., & Hulme, C. (2019). A cross-linguistic, longitudinal study of the foundations of decoding and reading comprehension ability. *Scientific Studies of Reading*, *23*(5), 386–402. https://doi.org/10.1080/10888438.2019.1580284

Carroll, J. B., Davies, P., & Richman, B. (1971). *Word frequency book*. American Heritage.

Castles, A., Rastle, K., & Nation, K. (2018). Ending the Reading wars: Reading acquisition from novice to expert. *Psychological Science in the Public Interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

Chiu, Y. D. (2018). The simple view of Reading across development: Prediction of grade 3 Reading comprehension from prekindergarten skills. *Remedial and Special Education*, *39*(5), 289–303. https://doi.org/10.1177/0741932518762055

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, *28*, 125–127. https://doi.org/10.3758/BF03203646

Dornseiff, F. (2020). *Der deutsche Wortschatz nach Sachgruppen*. De Gruyter. https://doi.org/10.1515/9783110457742

Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, *56*, S25–S44. https://doi.org/10.1002/rrq.411

Duke, N. K., Ward, A. E., & Pearson, P. D. (2021). The science of reading comprehension instruction. *The Reading Teacher*, *74*(6), 663–672. https://doi.org/10.1002/trtr.1993

Elleman, A. M. (2017). Examining the impact of inference instruction on the literal and inferential comprehension of skilled and less skilled readers: A meta-analytic review. *Journal of Educational Psychology*, *109*(6), 761–781. https://doi.org/10.1037/edu0000180

Evert, S. (2008). Corpora and collocations. In A. Lüdeling & M. Kytö (Eds.), *Corpus linguistics. An international handbook, 1212–1248*. Mouton de Gruyter. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.159.6220&rep=rep1&type=pdf retrieved from January 01, 2021
Fazio, R. H. (1990). A practical guide to the use of response latency in social psychological research. In C. Hendrick & M. S. Clark (Eds.), *Research methods in personality and social psychology* (pp. 74–97). Sage Publications, Inc.
Firth, J. R. (1957). A synopsis of linguistic theory 1930–55. In *Studies in linguistic analysis* (pp. 1–32). The Philological Society. Reprinted in Palmer (1968), pp. 168–205.
Fisher, A. V., Matlen, B. J., & Godwin, K. E. (2011). Semantic similarity of labels and inductive generalization: Taking a second look. *Cognition*, *118*(3), 432–438. https://doi.org/10.1016/j.cognition.2010.12.008
Flores, A. Z., Montag, J., & Willits, J. (2020). Using known words to learn more words: A distributional analysis of child vocabulary development. *arXiv preprint arXiv:2009.06810*.
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. https://doi.org/10.1007/S10648-011-9175-6
Fourtassi, A. (2020). Word co-occurrence in child-directed speech predicts children's free word associations. In *Proceedings of the workshop on cognitive modeling and computational linguistics* (pp. 49–53). Association for Computational Linguistics. https://doi.org/10.18653/v1/P17
Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children's semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, *44*(7), e12847. https://doi.org/10.1111/cogs.12847
Fox, J., & Weisberg, S. (2019). *An {R} companion to applied regression* (Third ed.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/
Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. https://doi.org/10.1017/S0305000916000209
Friedrich, C. K., Felder, V., Lahiri, A., & Eulitz, C. (2013). Activation of words with phonological overlap. *Frontiers in Psychology*, *4*, 556. https://doi.org/10.3389/fpsyg.2013.00556
Ganzeboom, H. B. G. (2010). A new international socio-economic index (ISEI) of occupational status for the international standard classification of occupation 2008 (ISCO-08) constructed with data from the ISSP 2002-2007. In *Paper presented at the annual conference of international social survey Programme*, Lisbon, Portugal. http://www.harryganzeboom.nl/Pdf/2010%20-%20Ganzeboom%20-%20A%20New%20International%20Socio-Economic%20Index%20ISEI%20of%20occupational%20status%20for%20the%20International%20Standard%20Classification%20of%20Occupation.pdf
Glück, C. W., & Glück, C. W. (2011). *Wortschatz-und Wortfindungstest für 6-bis 10-Jährige (WWT 6–10)*. Urban & Fischer.
Godwin, K. E., Matlen, B. J., & Fisher, A. V. (2013). Development of category-based reasoning in 4-to 7-year-old children: The influence of label co-occurrence and kinship knowledge. *Journal of Experimental Child Psychology*, *115*(1), 74–90. https://doi.org/10.1016/j.jecp.2012.11.008
Harris, Z. S. (1954). Distributional structure. *Word*, *10*(3), 146–162. https://doi.org/10.1080/00437956.1954.11659520
Hills, T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–739. https://doi.org/10.1111/j.1467-9280.2009.02365.x
Hills, T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. B. (2008). Is there preferential attachment in the growth of early semantic noun networks? *Proceedings of the 30th Annual Conference of the* 57 *Cognitive Science Society*, *October*, 1–6. papers://c941067e-36da-4589-acf1-2f5738fdb5a1/Paper/p406.
Hjetland, H., Brinchmann, E., Scherer, R., Hulme, C., & Melby-Lervåg, M. (2020). Preschool pathways to reading comprehension: A systematic meta-analytic review. *Educational Research Review*, *30*, 100323. https://doi.org/10.1016/j.edurev.2020.100323
Hjetland, H., 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*(5), 751–763. https://doi.org/10.1037/edu0000321
Hoover, W. A., & Tunmer, W. E. (2018). The simple view of reading: Three assessments of its adequacy. *Remedial and Special Education*, *39*(5), 304–312. https://doi.org/10.1177/0741932518773154
Hulme, C., Nash, H., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Journals.Sagepub.Com*, *26*(12), 1877–1886. https://doi.org/10.1177/0956797615603702
Jiang, S., Jiang, X., & Siyanova-Chanturia, A. (2020). The processing of multiword expressions in children and adults: An eye-tracking study of Chinese. *Applied PsychoLinguistics*, *41*(4), 901–931. https://doi.org/10.1017/S0142716420000296
Kessler, R., Weber, A., & Friedrich, C. K. (2021). Activation of literal word meanings in idioms: Evidence from eye-tracking and ERP experiments. *Language and Speech*, *64*(3), 594–624. https://doi.org/10.1177/0023830920943625
Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.
Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963
Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211
Landrigan, J. F., & Mirman, D. (2018). The cost of switching between taxonomic and thematic semantics. *Memory and Cognition*, *46*, 191–203. https://doi.org/10.3758/s13421-017-0757-5
Lany, J., & Saffran, J. (2013). Statistical learning mechanisms in infancy. *Comprehensive Developmental Neuroscience: Neural Circuit Development and Function in the Brain*, *3*, 231–248.
Lenhard, W., Lenhard, A., & Schneider, W. (2017). *ELFE II—Ein Leseverständnistest für Erst- bis Siebtklässler*. Hogrefe.
Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, *89*(5), 1821–1838. https://doi.org/10.1111/cdev.12861
Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018). Examining the simple view of Reading with elementary school children: Still simple after all these years. *Remedial and Special Education*, *39*(5), 260–273. https://doi.org/10.1177/0741932518764833
Lüdecke, D., Bartel, A., Schwemmer, C., Powell, C., Djalovski, A., & Titz, J. (2023). _sjPlot: Data Visualization for Statistics in Social Science. R package version 2.8.14. https://CRAN.R-project.org/package=sjPlot
Lund, K., Burgess, C., & Audet, C. (1996). Dissociating semantic and associative word relationships using high-dimensional semantic space. In G. W. Cottrell (Ed.), *Proceedings of the 18th annual conference of the cognitive science society* (pp. 603–608). Erlbaum.
MacWhinney, B. (2014). *The CHILDES project: Tools for analyzing talk*, Volume II. Psychology Press.
Mashal, N., & Kasirer, A. (2011). Thinking maps enhance metaphoric competence in children with autism and learning disabilities. *Research in Developmental Disabilities*, *32*(6), 2045–2054. https://doi.org/10.1016/j.ridd.2011.08.012
Matlen, B. J., Fisher, A. V., & Godwin, K. E. (2015). The influence of label co-occurrence and semantic similarity on children's inductive

generalization. *Frontiers in Psychology*, *6*, 1146. https://doi.org/10.3389/fpsyg.2015.01146
McClelland, J., & Rogers, T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews. Neuroscience*, *4*, 310–322. https://doi.org/10.1038/nrn1076
McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and non-living things. *Behavior Research Methods*, *37*(4), 547–559.
Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
Mirman, D., Landrigan, J.-F., & Britt, A. E. (2017). Taxonomic and thematic semantic systems. *Psychological Bulletin*, *143*(5), 499–520. https://doi.org/10.1037/bul0000092
Moll, K., & Landerl, K. (2010). *SLRT-II: Lese-und Rechtschreibtest.* Weiterentwicklung des Salzburger Lese-und Rechtschreibtests (SLRT). Huber.
Moll, K., Ramus, F., Bartling, J., Bruder, J., Kunze, S., Neuhoff, N., Streiftau, S., Lyytinen, H., Leppänen, P. H. T., Lohvansuu, K., Tóth, D., Honbolygó, F., Csépe, V., Bogliotti, C., Iannuzzi, S., Démonet, J. F., Longeras, E., Valdois, S., George, F., … Landerl, K. (2014). Cognitive mechanisms underlying reading and spelling development in five European orthographies. *Learning and Instruction*, *29*, 65–77. https://doi.org/10.1016/J.LEARNINSTRUC.2013.09.003
Nation, K. (2019). Children's reading difficulties, language, and reflections on the simple view of reading. *Australian Journal of Learning Difficulties*, *24*(1), 47–73. https://doi.org/10.1080/19404158.2019.1609272
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), 4–9. https://doi.org/10.1016/S0010-0277(99)00004-9
Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from http://w3.usf.edu/FreeAssociation/
Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532–1543). Association for Computational Linguistics. https://aclanthology.org/D14-1162.pdf
Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687
R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/
Revelle, W. (2022). Package "psych." *Cran.Rstudio.Org*. https://cran.rstudio.org/web/packages/psych/psych.pdf
Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.
Rosseel, Y., Yorgensen, D. T., De Wilde, L., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., Hallquist, M., Rhemtulla, M., Katsikatsou, M., Barendse, M., Rockwood, N., Scharf, F., Du, H., & Jamil, H. (2017). Package 'lavaan'.
Savic, O., Unger, L., & Sloutsky, V. M. (2022). Exposure to co-occurrence regularities in language drives semantic integration of new words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *48*(7), 1064–1081. https://doi.org/10.1037/xlm0001122
Savic, O., Unger, L., & Sloutsky, V. M. (2023). Experience and maturation: The contribution of co-occurrence regularities in language to the development of semantic organization. *Child Development*, *94*(1), 142–158. https://doi.org/10.1111/cdev.13844
Schmitterer, A. M. A., & Brod, G. (2021). Which data do elementary school teachers use to determine reading difficulties in their students? *Journal of Learning Disabilities*, *54*, 349–364. https://doi.org/10.1177/0022219420981990
Schmitterer, A. M. A., & Schroeder, S. (2019a). Effects of reading and spelling predictors before and after school entry: Evidence from a German longitudinal study. *Learning and Instruction*, *59*, 46–53. https://doi.org/10.1016/j.learninstruc.2018.09.005
Schmitterer, A. M. A., & Schroeder, S. (2019b). Young Children's semantic judgment of thematic relations predicts word Reading: Evidence from a longitudinal study. *Cognitive Development*, *50*, 22–35. https://doi.org/10.1016/j.cogdev.2019.01.002
Schroeder, S., Würzner, K. M., Heister, J., et al. (2015). childLex: A lexical database of German read by children. *Behav Res*, *47*, 1085–1094. https://doi.org/10.3758/s13428-014-0528-1
Siegmüller, J., Kauschke, C., von Minnen, S., & Bittner, D. (2011). TSVK. Test zum Satzverstehen von Kindern. In *Eine profilorientierte Diagnostik der Syntax*. Elsevier.
Spence, D. P., & Owens, K. C. (1990). Lexical co-occurrence and association strength. *Journal of Psycholinguistic Research*, *19*, 317–330. https://doi.org/10.1007/BF01074363
Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, *7*(1), 46730. https://doi.org/10.1038/srep46730
Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, *29*(1), 41–78. https://doi.org/10.1207/s15516709cog2901_3
Turney, P. D., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, *37*, 141–188. https://doi.org/10.1613/jair.2934
Ulicheva, A., Marelli, M., & Rastle, K. (2021). Sensitivity to meaningful regularities acquired through experience. *Morphology*, *31*(3), 275–296. https://doi.org/10.1007/S11525-020-09363-5
Unger, L., & Fisher, A. V. (2021). The emergence of richly organized semantic knowledge from simple statistics: A synthetic review. *Developmental Review*, *60*, 100949. https://doi.org/10.1016/j.dr.2021.100949
Unger, L., Vales, C., & Fisher, A. V. (2020). The role of Co-occurrence statistics in developing semantic knowledge. *Cognitive Science*, *44*(9), e12894. https://doi.org/10.1111/cogs.12894
Unger, L., Yim, H., Savic, O., Dennis, S., & Sloutsky, V. M. (2023). No frills: Simple regularities in language can go a long way in the development of word knowledge. *Developmental Science*, *26*, e13373. https://doi.org/10.1111/desc.13373
Vales, C., & Fisher, A. V. (2019). When stronger knowledge slows you down: Semantic relatedness predicts children's co-activation of related items in a visual search paradigm. *Cognitive Science*, *43*(6), e12746. https://doi.org/10.1111/cogs.12746
Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5
Verhoeven, L., Van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading*, *15*(1), 8–25. https://doi.org/10.1080/10888438.2011.536125
von Hasselhorn, M., Schumann-Hengsteler, R., Gronauer, J., Grube, D., Mähler, C., Schmid, I., Seitz-Stein, K., & Zoelch, C. (2012). *Arbeitsgedächtnistestbatterie für Kinder von fünf bis zwölf Jahren:(AGTB 5–12)*. Hogrefe.
Wang, Y. A., & Rhemtulla, M. (2021). Power analysis for parameter estimation in structural equation modeling: A discussion and tutorial. *Advances in Methods and Practices in Psychological Science*, *4*(1), 2515245920918253. https://doi.org/10.1177/2515245920918253
Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020–2045. https://doi.org/10.1037/xge0000014

Wojcik, E. H., & Kandhadai, P. (2020). Paradigmatic associations and individual variability in early lexical–semantic networks: Evidence from a free association task. *Developmental Psychology*, *56*(1), 53–69. https://doi.org/10.1037/dev0000844

Zugarramurdi, C., Fernández, L., Lallier, M., Valle-Lisboa, J. C., & Carreiras, M. (2022). Mind the orthography: Revisiting the contribution of Prereading phonological awareness to Reading acquisition. *Developmental Psychology*, *58*(6), 1003–1016. https://doi.org/10.1037/dev0001341

Submitted April 25, 2023
Final revision received April 26, 2024
Accepted May 9, 2024

Alexandra Schmitterer is an Associated Postdoc Researcher at the DIPF|Leibniz Institute for Research and Information in Education. She studies how language is processed in the mind and how language processes are connected to literacy and learning development. Email: a.schmitterer@dipf.de

Caterina Gawrilow is Professor of School Psychology at the University of Tübingen. Her main areas of interest are ADHD in children, adolescents, and adults, as well as experimental and applied research on self-regulation and executive functions. Email: caterina.gawrilow@uni-tuebingen.de

Claudia Friedrich is Professor of Developmental Psychology at the University of Tübingen. Her main area of interest is language acquisition with a focus on the developing lexicon and its cognitive foundations. Email: claudia.friedrich@psycho.uni-tuebingen.de

Supporting Information

Additional supporting information may be found in the online version of this article on the publisher's website:10.1002/rrq.548/suppinfo