Inference making in young children: the concurrent and longitudinal contributions of verbal working memory and vocabulary

Language and Reading Research Consortium

Currie, Nicola, K.

Muijselaar, Marloes M. L.

Accepted 2 January 2019

Author Note:

This paper was prepared by a Task Force of the Language and Reading Research Consortium (LARRC) consisting of Kate Cain (Convener), Nicola K. Currie, and Marloes M. L. Muijselaar. LARRC project sites and investigators are as follows:

Ohio State University (Columbus, OH): Laura M. Justice (Site PI), Richard Lomax, Ann O'Connell, Jill Pentimonti[1], Stephen A. Petrill[2], Shayne B. Piasta

Arizona State University (Tempe, AZ): Shelley Gray (Site PI), Maria Adelaida Restrepo.

Lancaster University (Lancaster, UK): Kate Cain (Site PI).

University of Kansas (Lawrence, KS): Hugh Catts[3] (Site PI), Mindy Bridges, Diane Nielsen.

University of Nebraska-Lincoln (Lincoln, NE): Tiffany Hogan (Site PI), Jim Bovaird, J. Ron Nelson.[4]

MGH Institute of Health Professions (Boston, MA): Tiffany Hogan (Site PI)

1. Jill Pentimonti is now at American Institutes for Research (AIR).

2. Stephen A. Petrill was a LARRC co-investigator from 2010-2013.

3. Hugh Catts is now at Florida State University.

4. J. Ron Nelson was a LARRC co-investigator from 2010-2012.

This work was supported by grant # R305F100002 of the Institute of Education Sciences' Reading for Understanding Initiative. We are deeply grateful to the numerous staff, research associates, school administrators, teachers, children, and families who participated. Key personnel at study sites include: Crystle Alonzo, Lisa Baldwin-Skinner, Lauren Barnes, Garey Berry, Beau Bevens, Jennifer Bostic, Shara Brinkley, Janet Capps, Tracy Centanni, Beth Chandler, Lori Chleborad, Emmanuel Cortez, Willa Cree, Dawn Davis, Kelsey Dickerhoof, Jaclyn Dynia, Michel

Eltschinger, Kelly Farquharson, Tamarine Foreman, Yvonne Fraser, Abraham Aldaco Gastelum, Rashaun Geter, Sara Gilliam, Alexandria Hamilton, Cindy Honnens, Miki Herman, Hui Jiang, Elaine Joy, Jaime Kubik, Trudy Kuo, Gustavo Lujan, Chi Luu, Junko Maekawa, Carol Mesa, Denise Meyer, Maria Moratto, Kimberly Murphy, Marcie Mutters, Amy Pratt, Trevor Rey, Lizeth Sanchez-Verduzco, Amber Sherman, Shannon Tierney, Stephanie Williams, and Gloria Yeomans-Maldonado.

The views presented in this work do not represent those of the federal government, nor do they endorse any products or findings presented herein. Correspondence concerning this work should be sent to Kate Cain, k.cain@lancaster.ac.uk, Department of Psychology, Lancaster University, LA1 4YF, UK.

Abstract

Inference making is fundamental to the construction of a coherent mental model of a text. We examined how vocabulary and verbal working memory relate to inference development concurrently and longitudinally in 4- to 9-year-olds. Four hundred and twenty pre-kindergartners completed oral assessments of inference making, vocabulary breadth, vocabulary depth, and verbal working memory each year until grade 3. Concurrently, hierarchical regressions revealed that a greater proportion of total variance in inference making was explained by vocabulary and verbal working memory for younger than older children. Vocabulary breadth was a stronger predictor of inference than verbal working memory but the opposite pattern was found for vocabulary depth and verbal working memory. The longitudinal relations between inference making, vocabulary and verbal working memory were investigated in two separate cross-lagged models: one with vocabulary breadth and a second with vocabulary depth. Both vocabulary breadth and depth explained subsequent inference making and verbal working memory throughout the early grades. Inference making also predicted subsequent vocabulary depth. The results highlight the critical role of vocabulary knowledge in the development of inference ability both within and across time, the importance of vocabulary in supporting the development of verbal working memory, and the changing dynamics between language and memory in early development.

Keywords: inference making, vocabulary breadth, vocabulary depth, verbal working memory, longitudinal

Educational Impact and Implications Statement

Inference making is essential for successful listening and reading comprehension. This study examined how vocabulary knowledge (breadth: number of words known, and depth: what is known about a word's meaning) and verbal working memory were related to 4- to 9-year-olds' ability to make the inferences necessary to understand spoken narratives. Vocabulary knowledge and verbal working memory were stronger predictors of concurrent inference making ability for younger than for older children. Across time, reciprocal relations were evident: inference making predicted subsequent vocabulary depth, and both aspects of vocabulary knowledge supported later inference making and working memory. Educators should be aware of the critical roles of vocabulary and verbal working memory to young children's inference making, and how good vocabulary skills support verbal memory. Of note are the reciprocal relations that exist between vocabulary and inference; both skills should be fostered in the classroom to mutually support each other.

Inference making in young children: the concurrent and longitudinal contributions of verbal working memory and vocabulary

Successful text comprehension results in a representation of the state of affairs described in the text, typically referred to as a mental model or a situation model (Johnson-Laird, 1983; Kintsch, 1998). The construction of a fully specified and coherent mental model involves going beyond the surface details in a text by combining information across sentences and integrating background knowledge with textual information. These coherence processes involve inference making. Our focus in this paper is the development of children's ability to generate inferences between 4 to 9 years, and how this ability is related to vocabulary and verbal working memory.

Children make inferences from narrative texts from a young age (Filiatrault-Veilleux, Bouchard, Trudeau, & Desmarais, 2016; Kendeou, Bohn-Gettler, White, & Van Den Broek, 2008; Lepola, Lynch, Laakkonen, Silvén, & Niemi, 2012; Silva & Cain, 2015; Tompkins, Guo, & Justice, 2013). These early inferential skills between 4 to 6 years predict both concurrent and subsequent listening and reading comprehension (Lepola, Lynch, Kiuru, Laakkonen & Niemi, 2016; Silva & Cain, 2015). In older children, there are significant improvements in inference skill, particularly between the ages of 6 to 10 years (Ackerman, 1986; Barnes, Dennis, & Haefele-Kalvaitis, 1996; Currie & Cain, 2015). However, to date we do not know what drives these developmental improvements. In this paper we examine concurrently and longitudinally the role of two factors critical to inference making: vocabulary and verbal working memory.

Vocabulary is strongly linked to children's listening and reading comprehension in general (Cain, Oakhill & Bryant, 2004; Cromley & Azevedo, 2007; Florit, Roch, Altoe, & Levorato, 2009; Lynch et al., 2008). Such measures of the broad construct of text comprehension (of either spoken or written text) often tap the ability to generate targeted inferences, in addition to memory for details explicitly stated in the text. As a result, there is typically a strong relationship between measures designed to tap inference making ability specifically and standardized measures of reading and listening comprehension (e.g., Cain et al., 2004; Elleman, 2017; Language and Reading Research Consortium (LARRC) & Muijselaar 2018; Silva & Cain, 2015).

Vocabulary is also related to children's performance on measures designed specifically to assess inference making (Currie & Cain, 2015; Daugaard, Cain & Elbro, 2017; Silva & Cain, 2015) and inference has also been found to mediate the relationship between vocabulary and measures of general reading and listening comprehension (e.g., see Cromley & Azevedo, 2007; Daugaard et al., 2017 and Kim, 2017 for further discussion). The simple explanation for this close relationship is that vocabulary enables inference making. When combining information across different clauses or sentences in a text to achieve local coherence, vocabulary knowledge may be critical to make that link: for example, knowledge of synonyms, pronouns and category exemplar pairings (e.g., the bird – the robin) helps to aid integration. Likewise, readers and listeners often have to draw on their vocabulary and background knowledge to understand elements of a text such as the setting and a protagonist's motivations and responses to an event to achieve global coherence (Cain & Oakhill, 1999; Long & Chong, 2001). Indeed, vocabulary is a key predictor of inference making between 6 to 11 years (Cain, Oakhill, & Bryant, 2004; Chrysochoou, Bablekou, & Tsigilis, 2011; Currie & Cain, 2015). However, we do not know what role vocabulary plays in the developmental improvements seen in inference skill over time and whether a reciprocal relationship exists between vocabulary and inference.

In a longitudinal study of 7- to 11-year-olds, Oakhill and Cain (2012) found that vocabulary at 7 to 8 years predicted inference one year later, raising the possibility that vocabulary has a direct influence on growth in inference making ability. Studies of the relationship between reading comprehension and vocabulary are also relevant, because measures of reading comprehension often tap inference making ability (e.g., Cain et al., 2004; Elleman, 2017; Daugaard et al., 2017; Kim, 2017; LARRC, 2017). One such relevant study investigated growth in children's reading comprehension between 6 to 11 years (Quinn, Wagner, Petscher & Lopez, 2015). This study found that reading comprehension was predicted by earlier vocabulary knowledge, but the opposite relationship (i.e. reading comprehension predicting later vocabulary) was not found. Together these studies support the viewpoint that vocabulary facilitates the development of inference making, because it provides the information necessary to generate many inferences.

In addition, other work demonstrates the reciprocal relationship between vocabulary and comprehension skill whereby earlier reading comprehension and inference skills predict later vocabulary performance (Lepola, Lynch, Laakkonen, Silven, & Niemi, 2012; Verhoeven & Van Leeuwe, 2008; Verhoeven et al., 2011). For example, Verhoeven et al. (2011) found that grade 2 reading comprehension predicted grade 3 vocabulary and Lepola et al. (2012) found that inference at age 4 predicted vocabulary at age 5. The mechanism for this effect could be the role that reading comprehension skills, such as inference making, play in supporting children's ability to derive the meanings of new and unfamiliar words from context (Cain, Oakhill, & Lemmon, 2004). Cain et al. (2004) found that 9- to 10-year-olds with poor comprehension skills were less able to infer the meaning of a novel word from text than those with good comprehension skills, particularly when the working memory demands of the task were high and involved integrating information across several filler sentences. Therefore inference skill could impact word learning and vocabulary growth, and working memory could influence this effect. To date, this relationship has not been explored longitudinally.

An important factor examined in this study is the nature of vocabulary knowledge. Breadth of vocabulary knowledge is considered the number of words known, and is typically measured by single-word receptive and expressive vocabulary tasks that determine whether or not a lexical entry for a given word exists. In contrast, vocabulary depth is conceptualised as what is known about words, and has been measured by tasks that require word definitions, synonyms, and similarities (e.g., Cain & Oakhill, 2014; Ouellette, 2006; Swart et al., 2017).

Both aspects of vocabulary knowledge predict reading comprehension (Ouellette, 2006; Richter, Isberner, Naumann & Neeb, 2013; Swart et al., 2017; Tannenbaum et al., 2006) and inference making specifically (Cain & Oakhill, 2014). For example, Swart et al. (2017) found that separate measures of vocabulary breadth and depth accounted for unique variance (8% and 2% respectively) in 9-year-olds' reading comprehension after controlling for non-verbal reasoning, short-term memory and decoding. In this case vocabulary breadth had a greater influence on reading comprehension performance than vocabulary depth. A study of 10- to 11-year-olds found that both vocabulary breadth and depth were more important predictors of inference making, than of literal memory for the text (Cain & Oakhill, 2014). Furthermore, while both aspects of vocabulary explained additional variance in inference making after age and word reading accuracy were controlled, vocabulary depth was a stronger predictor than vocabulary breadth. Ouellette (2006) also found stronger relations between vocabulary depth and reading comprehension, than vocabulary breadth. However, due to differences between studies in the measurement of vocabulary breadth and depth it is difficult to ascertain whether there are developmental changes in the relative importance of each. An advantage of the current study is that we use the same instruments to assess vocabulary and inference at each time point, enabling us to assess the relative influence of vocabulary breadth and depth at different ages.

The importance of both vocabulary breadth and vocabulary depth in relation to inference (Cain & Oakhill, 2014) can be interpreted within Perfetti's (2007) lexical quality hypothesis. The lexical quality hypothesis proposes that more precise knowledge about words, including their semantic representations and interconnections with other related words, helps to support efficient comprehension of text. In terms of vocabulary breadth, an underspecified representation of a critical inference-promoting word limits the likelihood of that inference being made. In terms of vocabulary depth, inferences rely heavily on accurate and robust lexical representations and inter-related, semantic networks that enable links to be made between words that are thematically related. Thus, both breadth and depth of vocabulary knowledge are important for accurate inference making. However, measures of vocabulary depth, tapping well-specified knowledge of individual words and their interrelations with other known words, might be a better predictor of inference development than vocabulary breadth, because it taps the ease with which related concepts for a topic can be activated. In this study, we included a comprehensive assessment of vocabulary, which enabled us to carry out analyses that compared the relations between inference making and these two aspects of vocabulary (depth and breadth), concurrently and longitudinally, to enable if one was a better predictor of inference development than another within and across development.

The other factor that we consider in relation to inference making is verbal working memory. Working memory refers to the memory systems involved in simultaneously storing information while processing new incoming information (Baddeley & Hitch, 1974). Verbal working memory is related to reading and listening comprehension (Cain et al., 2004; Florit et al., 2009; Leather & Henry, 1994; Seigneuric & Ehrlich, 2005) and also inference making specifically (Chrysochoou et al., 2011; Currie & Cain, 2015). Verbal working memory is proposed to support inference making because readers and listeners must access just presented information while they process and integrate the current piece of text or their background knowledge into the unfolding mental model.

There are few studies investigating the relations between children's verbal working memory, vocabulary, and their inference making. One exception is Currie and Cain (2015; see also Chrysochoou et al., 2011) who examined the relative influence of verbal working memory and vocabulary on the generation of coherence inferences in cross-sectional comparisons of 6-, 8- and 10-year-olds. Where verbal working memory was related to inference making, its effect was consistently mediated by vocabulary knowledge. Similarly, when considering broader measures of reading comprehension, some studies report that verbal working memory does not explain unique variance in reading comprehension once factors such as vocabulary are controlled (Muijselaar & de Jong, 2015). However, as with the literature on the relations between vocabulary and inference making/reading comprehension, there are also exceptions here: verbal working memory has been shown to explain unique variance (over and above vocabulary) in children's concurrent reading and listening comprehension (Cain, Oakhill, & Bryant, 2004; Florit et al., 2009; Seigneuric & Ehrlich, 2005), indicating that it could make a unique contribution to inference making. If working memory is important for vocabulary learning, relations between working memory and vocabulary learning may be stronger when examining the longitudinal development of these skills: the concurrent modelling reported in previous research may have obscured developmental effects. We examined these relations over time in our study.

Chronological age may be an important factor to consider when examining the relative importance of verbal working memory to inference (Chrysochoou & Bablekou, 2011; Currie & Cain, 2015). In a cross-sectional study Chrysochoou and Bablekou (2011) found that verbal working memory predicted inference skill over and above vocabulary in 5-year-olds but its effect decreased with age. One explanation for this finding is that older children may have richer background knowledge to draw upon and make use of comprehension strategies to a greater extent than younger children, therefore relying less on working memory resources (Elleman & Compton, 2017; van den Broek, Bohn-Gettler, Kendeou, Carlson, & White, 2011).

The current study

We examined the relations between inference making, vocabulary, and verbal working memory in children aged 4-5 through to 8-9 years in a longitudinal study. We used structural regression modeling to be able to investigate unique developmental relations between those variables. The longitudinal design in the current study enabled us to examine the nature of the developmental relationships between inference, vocabulary, and working memory over time.

To provide a comprehensive assessment of the construct of inference making, we included two types of coherence inferences in this study: local coherence inferences, which involve the integration of information within text, enabling the reader or listener to go beyond a surface level interpretation, and global coherence inferences, which involve inferring goals and motivations of characters within a narrative or establishing an overall theme. A recent study confirmed the empirical distinction of these two types of inference, but critically demonstrated that they shared more variance in common, arguably because both are necessary for adequate text comprehension (LARRC & Muijselaar, 2018). For that reason, we treat coherence inferences as a single construct in this paper.

We addressed the following research questions:

1) What are the concurrent relations between vocabulary breadth, vocabulary depth, verbal working memory, and inference making in children aged 4-5 years through to 8-9 years, and do the relations change with age? Based on the findings of Currie and Cain (2015) and Chrysochoou et al. (2011) we predicted that vocabulary would explain significant and unique variance in inference making at each age. Given the importance of vocabulary depth to general comprehension skills and inference (Cain & Oakhill, 2014; Ouelette, 2006; Richter et al., 2013), we expected vocabulary depth to be a stronger predictor of inference. We predicted verbal working memory to be related to inference to a lesser extent, in line with the research outlined above. Following Chrysochoou and Bablekou (2011), we expected verbal working memory to have a greater influence in the youngest age groups.

2) Do vocabulary breadth, vocabulary depth and verbal working memory predict unique variance in subsequent inference making? This is the first study to examine directly the predictors of inference longitudinally. However, given the key role of vocabulary in reading comprehension development and its concurrent prediction of inference at different ages, we were able to make informed predictions.

We expected both vocabulary breadth and depth to predict subsequent inference throughout the age range (Currie & Cain, 2015; Seigneuric & Ehrlich, 2005; Verhoeven & Van Leeuwe, 2008). For verbal working memory, we predicted a direct influence of verbal working memory on subsequent inference skill only in the youngest age groups (Currie & Cain, 2015; Chrysochoou & Bablekou, 2011).

3) Are there reciprocal relations between inference, vocabulary and verbal working memory? Given the findings of Lepola et al. (2012) and Verhoeven and colleagues (Verhoeven & van Leeuwe, 2008; Verhoeven et al. 2011) we expected to find that inference skill predicted vocabulary development, in addition to the prediction of inference from vocabulary. Our predictions about the relations between verbal working memory, inference, and vocabulary were more speculative. Concurrently, there is evidence both that vocabulary supports verbal working memory (Nation et al., 1999; Van Dyke & Johns, 2012) and that verbal working memory (in addition to inference) supports vocabulary learning, but only when the new word and defining context are not adjacent (Cain et al., 2004). For that reason, we expected stronger predictive relations from vocabulary to working memory, than vice versa.

Method

Participants

Participants were 420 pre-kindergartners (*M* = 60 months, *SD* = 4.35; 58% boys) who were followed for 5 consecutive years from pre-kindergarten onwards (thus in prekindergarten (P), kindergarten (K), grade 1 (G1), grade 2 (G2), and grade 3 (G3)). Of these children, 305 stayed in the study and progressed through to grade 3, whilst 38 progressed only to grades 1 (N=2) or 2 (N=36) because they were held back. These children took part in a larger longitudinal study on listening and reading comprehension. For children who repeated a class, the data from the first year was used. For more information about the study and the participants, we refer to Language and Reading Research Consortium (LARRC), Farquharson, and Murphy (2016). We also refer readers to published work with this sample that reports on the structure of the inference task specifically (LARRC & Muijselaar, 2018).

Instruments

For the present study, measures of inference making, vocabulary, and verbal working memory were selected. All standardized measures were administered according to the procedures described in the manual.

Inference making. To assess inference making, a researcher-developed measure, the Inference Task, based on work by Cain and Oakhill (1999), Oakhill and Cain (2012), and Currie and Cain (2015), was used. Children listened to two short narratives and answered eight open-ended questions after each one: 4 questions assessed local coherence inferences, and 4 questions assessed global coherence inferences. At grades K and above, one story was new and one was from the pair of stories administered in the previous grade. Length and format of the stories across grades was similar, with stories selected to be suitable for different grades by content (to align with general knowledge) and earlier pilot testing. See Appendix A for an example of a narrative and inference questions. The questions were scored by trained assessors. Each inference question was scored on a 0 to 2 point scale. Reliability for the entire task was sufficient to good: Cronbach's alphas ranged between .74 to .84, depending on grade. The average score of the questions corresponding to the two narratives at each grade level was used in the analyses.

Vocabulary breadth. To assess vocabulary breadth, the Peabody Picture Vocabulary Test – Fourth Edition (PPVT-4; Dunn & Dunn, 2007) and the Expressive Vocabulary Test – Second Edition (EVT-2; Williams, 1997-2007) were used. These are single-word tasks that tap vocabulary breadth because they determine whether or not a lexical entry exists for a given word; knowledge of the broader semantics of the target word (and also distractors for the PPVT-2) are not required to perform the task. For each, the starting item is determined by the child's age (and requires a basal to be established) and testing ceases when a prescribed number of incorrect responses has been made. For the PPVT, children were asked to point to the picture, out of four options, that corresponded to a verbally presented word. For the EVT, children were required to provide a single word to label a picture, or to provide a synonym for a target word. The internal consistency was calculated for this sample and found to be high for both measures: Cronbach's alphas were between .93 and .96, depending on grade. Total raw scores were used in the analyses.

Vocabulary depth. Vocabulary depth was assessed with the Clinical Evaluation of Language Fundamentals – Fourth edition: Subtest Word Classes 1 and 2 (CELF-WC; Semel, Wiig, & Secord, 2003). CELF-WC 1 was administered to children in pre-kindergarten to Grade 2. For children in Grade 3, CELF-WC 2 was used. The same start point is used for all children for each version and testing ceases when a prescribed number of incorrect responses has been made. Children were asked to select two words that belonged to each other from a series of three or four words. This assesses knowledge of interrelations between word meanings and, therefore, depth of vocabulary knowledge. In order to measure deeper understanding of each word's meaning, children were then required to explain how the two words go together. This results in separate scores for receptive and expressive vocabulary. Reliabilities for this sample were sufficient to high: Cronbach's alphas were .75 to .93 for receptive, .69 to .85 for expressive, depending on grade. Raw scores for receptive and expressive vocabulary were used in the analyses.

Verbal working memory. Verbal working memory was measured with the Woodcock Johnson Auditory Memory Test (WJAM; Woodcock, McGrew, & Mather, 2001) and an experimenter-developed Memory Updating Test (MU; based on Belacchi, Carretti, & Cornoldi, 2010). For the WJAM, children listened to the names of a series of objects and digits. They were asked to reorder the series, saying the objects in sequential order first, followed by the digits sequentially ordered. Testing started at the same item for all age groups and was discontinued when a prescribed number of incorrect responses were made. Cronbach's alphas for this sample were all good (.78 to .82). In the Memory Updating task, children were asked to recall the one/two/three/four smallest things from a spoken list of objects that differed in physical size. Testing started at the same item for all age groups and was discontinued when a prescribed number of incorrect responses were made. The reliability of the updating task was sufficient to good (Cronbach's alphas = .74 to .84). The total raw scores of both memory measures were used in the analyses.

Procedure

All measures were administered to the children, individually, in multiple sessions within a 5- or 6-month time frame (January to May/June). Breaks were scheduled if testing sessions took place on the same day, so that individual sessions did not last longer than 60 minutes. These sessions were carried out by trained research staff in a quiet room within the child's school, home, local university site or community center.

Analyses

We first investigated the fit of the measurement models for P, K, G1, G2, and G3. The measurement models consisted of latent variables for inference making with local and global coherence inference making as indicators, vocabulary breadth with

PPVT and EVT as indicators, vocabulary depth with word classes expressive and receptive vocabulary as indicators, and verbal working memory with WJAM and Memory Updating as indicators. From these measurement models, factor scores were extracted for further analyses.

The concurrent predictors of inference making at each age were investigated with hierarchical regression analyses. Two different models were estimated, all predicting inference making. In the first regression model vocabulary breadth was entered as a first step, and verbal working memory as a second step. In the second model, vocabulary depth rather than breadth was entered in the first step.

The longitudinal relations between inference making, and vocabulary breadth, vocabulary depth, and verbal working memory were investigated in separate cross-lagged models. Two cross-lagged models with three variables were constructed (see Figure 1: inference making, vocabulary breadth, verbal working memory; inference making, vocabulary depth, verbal working memory). We assessed vocabulary depth and breadth in separate analyses, each competing against verbal working memory. These two aspects of vocabulary share variance (Cain & Oakhill, 2014; Tannenbaum et al., 2006), but were assessed independently against verbal working memory to test the different predictions for each.

Mplus Version 7.11 (Muthén & Muthén, 2012) was used to conduct the analyses. Full Maximum Likelihood Estimation was used to obtain parameter estimates in the face of missing data. To evaluate the fit of the models, we inspected the chi-square goodness-of-fit test-statistic, the root mean square error of approximation (RMSEA) and its corresponding confidence interval, and the comparative fit index (CFI) (Kline, 2011). A nonsignificant chi-square value was taken as good model fit (Hayduck, 1996). RMSEA values between .05 and .08

indicated satisfactory approximate fit, and an RMSEA below .05 was taken as good approximate fit (Browne & Cudeck, 1993). A model with a CFI between .90 and .95 was considered acceptable, and a CFI above .95 was taken as good incremental model fit (Hu & Bentler, 1999).

Results

Descriptive Statistics and Preliminary Analyses

Data were first screened for missing data and outliers. Outliers (scores that were more than three standard deviations above or below the mean) were replaced by the lowest or highest non-outlying score (z = -3 to z = 3). The number of outliers was between 0.2% and 0.7% per grade (number of outliers / total number of data points). In total 5% of the scores were missing in P, 14% in K, 21% in Grade 1, 25% in Grade 2, and 32% in Grade 3 due to sample attrition.

The descriptive statistics of all measures (inference making, vocabulary, and verbal working memory) by grade are presented in Table 1. Data were normally distributed with skewness ranging from -1.53 to 0.81 and kurtosis ranging from -0.86 to 2.08 (Kline, 2011). The drop in scores for the WC test between grades 2 and 3 is likely due to different tests administered to different grades, as recommended in the manual (see Methods). The correlations among the inference making, vocabulary, and verbal working memory measures were significant (see Table 2).

Factor scores were extracted from the measurement models in all grades. These models consisted of latent variables for inference making, vocabulary breadth, vocabulary depth, and verbal working memory. Because the factor loadings of expressive vocabulary on the latent vocabulary depth factor were very high and resulted in estimation problems, these factor loadings were fixed at .9 and the corresponding residual variances at .19. The fit of the measurement models was adequate to good (see Table 3), so the factor scores from these models were used in further analyses. For children with complete missing data at a certain grade, factor scores were not generated. Thus in the following regression analyses, from the 420 pre-kindergartners, 369 of the sample was left in K, 339 in grade 1, 319 in grade 2, and 305 in grade 3. In the longitudinal analyses, the missing data was estimated with FIML.

Concurrent Predictions of Inference Making by Vocabulary and Verbal Working Memory

Hierarchical regression analyses were carried out to determine how vocabulary breadth and depth, and verbal working memory predicted inference making in each grade (see Table 4). The first set of models, with vocabulary breadth and verbal working memory predicting inference making, revealed that less total variance in inference making was explained in grade 3 (51%) than in grades 1 and 2 (67 and 71% respectively) than in P and K (82 and 85% respectively).

Vocabulary breadth explained a sizeable and significant additional amount of variance in inference making in PK through to grade 2 (76 – 81%) and slightly less in grade 3 (50%). When entered after this variable, the variance explained by verbal working memory was small but significant in P, K, and grade 3 (1 – 4 %), but memory did not explain significant variance in inference making in grades 1 and 2 when entered after vocabulary breadth. The unique effects of vocabulary breadth on inference making were large (β = .58 – β = .90). Verbal working memory had a moderate effect on inference making in P and K (β = .36; β = .38) and a small effect in grade 3 (β = .14), but there were no significant unique effects of verbal working memory on inference making in grades 1 and 2.

The second set of models, with vocabulary depth instead of vocabulary breadth as a predictor are also reported in Table 4. As for vocabulary breadth, less total variance in inference making was explained by vocabulary depth in grade 3 than in the earlier grades. As before, vocabulary depth predicted significant and unique variance, but the amount explained was less than by vocabulary breadth (19 - 46%). In contrast to the previous analysis, verbal working memory predicted sizeable and significant variance at each grade (14 – 38%), although the contribution reduced with increasing age. The unique effects of vocabulary depth on inference making were small in P (β = .16), and even negative in K (β = -.09). The latter may be explained by suppression because of the high correlation between vocabulary depth and verbal working memory factor scores (r = .75); the correlation between the vocabulary depth and inference factor was .61. The effect of vocabulary depth on inference making was moderate in grades 1 and 2 (β = .31; β = .35), and not significant in grade 3. With respect to verbal working memory, high unique effects were found in P and K (β = .76; β = .94), and moderate effects in grades 1 to 3 (β = .45 – β = .53).

In sum, vocabulary breadth made a significant contribution to inference making from P through to grade 1, with its contributions decreasing with increasing age. Vocabulary depth affected inference making from P through to grade 2, but there was no effect from grade 2 to grade 3. In general, vocabulary breadth was a stronger predictor than was vocabulary depth. In contrast, verbal working memory did not make a consistent contribution to inference making across grades and the strength of its contribution was dependent on the type of vocabulary included in the model. The contribution of vocabulary depth to working memory was small in all grades, and not significant from grade 2 till 3. Vocabulary breadth also had a small effect on working memory in K and these effects were moderate from K through to Grade 2.

Longitudinal Relations between Inference Making, Vocabulary, and Verbal Working Memory.

To examine the longitudinal relations between inference, vocabulary, and verbal working memory we ran cross-lagged path models with three variables. The first model concerned inference making, vocabulary breadth and verbal working memory. This model had a good fit to the data, χ^2 (36) = 45.77, p = .128, RMSEA = .025 (90% CI [.000, .045]), CFI = 1.00. The stability paths between all variables are not displayed (see Table 5 for values). The stability of inference making was moderate (.15 - .34), verbal working memory had a higher stability (.35 - .47), and the stability of vocabulary breadth was high (.55 - .71).

In Figure 2, the significant cross-lagged paths are presented. This model shows that vocabulary breadth has a strong relation to later inference making and verbal working memory in the early grades. Vocabulary breadth in P, K and grade 1 explained inference and also verbal working memory in subsequent grades. The negative effect of inference making in K on verbal working memory in grade 1 may be explained by suppression, because the correlation between the inference making and vocabulary breadth factor in K was very high (r = .90), and the correlation between inference making in K and verbal working memory in grade 1 was also high (r = .66).

Additionally, we investigated the longitudinal relations in a model with inference making, vocabulary depth, and verbal working memory. This model had a good fit to the data, χ^2 (36) = 49.67, p = .064, RMSEA = .030 (90% CI [.000, .049]), CFI = 1.00. The stability paths are not presented; the stability of vocabulary depth was low to moderate (.03 - .42). The significant cross-lagged paths are displayed in Figure 3. Similar to vocabulary breadth, vocabulary depth influenced the development of inference making and verbal working memory, early in development. Notably, in P and K vocabulary depth predicted later inference making, and in P, K, and grade 1 vocabulary depth predicted later verbal working memory. Note however that these effects are much smaller than for vocabulary breadth. Note also a difference in the relations between inference, vocabulary, and verbal working memory between the two analyses. Inference making predicted subsequent performance in vocabulary depth between K and G1, G1 and G2, and G2 and G3: such an effect was not evident for vocabulary breadth. Additional significant relations were found between inference in P and verbal working memory in K, and verbal working memory in P and vocabulary depth in K. For both path models, we investigated whether the parameter estimates changed after removing the nonsignificant cross-lagged paths. Removing the nonsignificant paths slightly changed the parameter estimates (M_{change} = .07), but it did not affect the results and conclusions.

Discussion

In this study, the concurrent and longitudinal relations between inference making, vocabulary breadth and depth, and verbal working memory were investigated in children from pre-kindergarten through to grade 3. Key findings from the concurrent prediction of inference making were: a greater proportion of variance in inference making was explained by verbal working memory and vocabulary in younger than in older children; vocabulary breadth had a stronger influence on inference making than did verbal working memory, whilst the opposite was found for vocabulary depth, which was less strongly predictive of concurrent inference making than was verbal working memory. Key findings from our longitudinal analyses were: both vocabulary breadth and depth were significant enablers of subsequent inference making and verbal working memory in the early grades; inference predicted subsequent vocabulary depth but not breadth. We discuss these findings in turn, together with implications and suggestions for future research.

First, we note that concurrent relations between inference, vocabulary, and verbal working memory were stronger in younger children (P and K) than in older children (grades 1, 2, and 3). This is in keeping with other research; stronger relations between similar measures of inference, vocabulary, and verbal working memory have been reported for younger relative to older children (Currie & Cain, 2015; Chrysochoou, & Bablekou, 2011). Of note, when examining the dimensionality of language and cognitive skills, we find that memory and oral language (vocabulary, grammar, discourse-level skills) form a unidimensional construct in prekindergarten children, but separable factors from around kindergarten age (LARRC & Nnanatu, in preparation). In the current study, we found that less total variance in inference making was explained for the older age groups compared to the younger children in the concurrent analyses. Again, this pattern is echoed by other work; a recent examination of the concurrent prediction of reading comprehension found that language and memory skills explained more variance in poorer comprehenders than in better comprehenders (LARRC & Logan, 2017). Our inference materials were comparable in terms of length and structure for the different grades, which may have allowed for greater concurrent prediction by vocabulary and memory in the younger grades. Together, this body of work suggests that variables additional to vocabulary and verbal working memory may influence inference making performance in older and/or better comprehenders, than in younger children or less skilled comprehenders.

One such variable may be background knowledge. Prior knowledge of a topic is critical to successful reading and listening comprehension (e.g., Elleman & Compton, 2017; Kendeou & van den Broek, 2005; Schneider, Korkel, & Weinert, 1989), enabling comprehenders to make inferences and build coherent representations of the text's meaning (Kintsch & Rawson, 2005). In addition, older children and better comprehenders may set a more appropriate 'standard of coherence' when reading and listening to texts, resulting in more targeted inference making to ensure coherence (van den Broek, et al., 2011). In support of this, both background knowledge and knowledge and use of reading strategies predict unique variance in reading comprehension and inference making (Ahmed et al., 2016; Cain, 1999; Cromley & Azevedo, 2007). Although background knowledge overlaps with vocabulary knowledge (Ahmed et al., 2016), our vocabulary measures tapped knowledge at the word level and therefore differ from the type of world/conceptual knowledge questions typically used in measures of background knowledge (e.g., see Cromley & Azevedo, 2007 and Ahmed et al., 2016). Future research should include a measure of background knowledge, in addition to measures of vocabulary and verbal working memory, to identify the most sensitive predictors of inference making in different age groups.

A second key finding from our concurrent analyses of the prediction of inference making was how the pattern of prediction differed by vocabulary type: vocabulary breadth was a consistent and strong predictor of inference across grades, whereas the contribution made by vocabulary depth was much weaker. In our introduction, we argued for the opposite pattern of prediction based on previous research (Cain & Oakhill, 2014), namely that vocabulary depth would be more important for inference making, because inference making often depends on knowledge of relations between words. We might point to practical reasons for this difference, such as the different age groups participating in the two studies and our use in this study of latent constructs to minimize measurement error. But we believe a consideration of these findings offers the potential for theoretical insights, as follows.

Accurate and precise representations of word meanings are critical for text comprehension, in addition to rich semantic networks (Perfetti, 2007). Our analyses indicate that for children aged between 4 to 9 years, the ability to answer inference-tapping questions is more strongly related to vocabulary breadth than depth; if children do not know the meaning of critical words in a text they cannot make the inference from that word. Related to this, our findings are not completely at odds with the extant literature. Tannenbaum et al.'s (2006) study of third graders found that vocabulary breadth was more strongly correlated with reading comprehension performance, for which inference making is integral, compared with vocabulary depth. In line with this, a recent study of adult readers, also found that vocabulary breadth was the more critical predictor of reading comprehension (Binder, Cote, Lee, Bessette, & Vu, 2017).

Different researchers have operationalized vocabulary depth and breadth in different ways (see Cain & Oakhill, 2014; Kieffer & Lesaux, 2012; Ouellette, 2006; Swart et al., 2017; Tannenbaum et al., 2006), which can make it hard to integrate findings across studies. Thus, there is a clear need to refine our definition of the construct of word knowledge to study further the relations between vocabulary knowledge and its role in inference making, as well as text comprehension more generally. Other researchers have sought to expand the construct to include morphological and grammatical knowledge, in addition to semantics (Foorman et al., 2015; Proctor, Silverman, Harring, & Montecillo, 2012; Spencer et al., 2015), and have also included speeded measures (Tannenbaum et al., 2006). Specific to inference making, we propose that a task that assesses the speed or automaticity with which semantic meanings are accessed may be more strongly predictive of inference making, particularly when considering performance on a measure of inference that assesses the ease or automaticity of inference making itself.

There is empirical support for this position. Perfetti, Yang and Schmalhofer (2008) found that adults with poor comprehension skills were slower at identifying the meaning of critical words in short passages (e.g., hospital) and linking them to related words given earlier in the passage (e.g., emergency room). Individuals with well-specified knowledge of words and, critically, their interrelations with other known words in various contexts would be able to carry out this task more efficiently. For children, there is preliminary evidence that speed of access to semantic knowledge is a stronger predictor of general reading comprehension skill in 9- to 10-year-olds, than is accuracy of knowledge (Oakhill, Cain, & McCarthy, 2015), lending support to this idea. We therefore recommend that future research on this topic includes speed/automaticity measures of vocabulary knowledge and also inference making.

A final discussion point about the findings from our concurrent analyses was that vocabulary breadth shared more overlapping variance with verbal working memory than did vocabulary depth. Vocabulary knowledge and the quality of lexical representations is proposed to enhance not only reading comprehension, but also verbal working memory because precise and stable semantic and phonological representations of words will support the maintenance of verbal information in verbal working memory tasks (Nation, Adams, Bowyer-Crane, & Snowling, 1999; Van Dyke, Johns, & Kukona, 2014). Our findings lend weight to this argument.

Turning to the analyses that explored the longitudinal relations between inference making, vocabulary, and verbal working memory, we found consistency in the relations between vocabulary and inference; in the early grades, both aspects of vocabulary were significant enablers of subsequent inference making (although the relation between vocabulary depth in grade 1 and inference in grade 2 was not significant, $p = .13$). It should be noted, however, that the strengths of the relations from early vocabulary to later inference were stronger for vocabulary breadth than depth, which is at odds with the prediction that deeper vocabulary knowledge would be a more critical factor in inference making, as discussed earlier in relation to our findings from the concurrent analyses. We believe that the reasons proposed for those concurrent findings are also valid here.

A striking difference between the analyses involving vocabulary depth and breadth was that inference making skill in kindergarten, grade 1, and grade 2 predicted subsequent vocabulary depth, but not breadth. Other studies of early language development have also found that early inference making skills predict subsequent vocabulary depth (Lepola, Lynch, Laakkonen, Silven, & Niemi, 2012). Our finding of reciprocal relations between vocabulary and inference are in line with some (e.g., Lepola et al., 2012; Verhoeven et al., 2011) but not all (e.g., Quinn et al., 2015) studies of the relations between vocabulary and reading or listening comprehension. We believe that these differences in findings are due to the types of task used to measure reading comprehension. For example, Quinn et al (2015) used a cloze task where children provided a missing word at the end of a short passage, a task that has been shown to be more highly dependent on word decoding skills than listening comprehension (Keenan, Betjemann, & Olson, 2008). In contrast, Verhoeven and colleagues (Verhoeven & Van Leeuwe, 2008; Verhoeven et al., 2011) assessed reading comprehension with explicit and implicit questions at the end of texts and Lepola et al. (2012) assessed inference making ability by asking inference-tapping questions while children viewed a picture book. Our findings converge with other studies that have used comprehension tasks that include direct assessment of inference.

Examination of the literature provides a possible mechanism for this relation. Derivation of new word meanings from context using inference is considered a crucial mechanism for vocabulary learning (Cain, Oakhill, & Elbro, 2003; Jenkins, Matlock, & Slocum, 1989; Keiffer & Lesaux, 2012; McKeown, 1985; Nagy, Herman, & Anderson, 1985). Context rarely provides a definition alongside a new word, instead information about related words is present that narrows down the range of possible definitions. Thus, inference making ability may tap knowledge about semantic relations between words and the ability to select appropriate meanings of words from context, resulting in a relationship between inference making ability and vocabulary depth over time (Daugaard et al., 2017). We note, however, that across time, the stability of vocabulary breadth was high, whereas the stability of vocabulary depth was low to moderate, allowing for greater variance to be explained by other variables.

Whether stability is an artefact of the tasks used or a critical difference in the nature and development of these two aspects of vocabulary is a topic for future research. A consideration of the cognitive processes involved in the assessments used to measure these two aspects of vocabulary and how that might impact on longitudinal relations is necessary in relation to this. For example, the receptive and expressive measures of vocabulary breadth used in this study were essentially matching tasks, whereas the measures of depth may tap not only word knowledge but also processes critical to inference making, such as explaining how two things are related. Future research might include experimental measures of vocabulary knowledge for which the processing requirements of breadth and depth measures were more similar. Overall, what is clear from the patterns of relations between variables across time, is that we need to refine our models of reading and listening comprehension development to include dynamic relations between language and cognitive skills in the course of development (Quinn et al., 2015).

There are some additional findings and limitations to this study that warrant comment. First, the negative relation between inference making in kindergarten and later verbal working memory in the analyses of vocabulary breadth. As noted in the results section, the first finding is likely a suppression effect: there were high concurrent correlations between inference making and vocabulary breadth in kindergarten, and also between inference making and verbal working memory in grade 1. Second, there was a positive relation between inference making in prekindergarten and verbal working memory in kindergarten, but only in the analysis that included vocabulary depth. Given the presence of this relationship for only one grade and aspect of vocabulary and the strong interrelatedness between early oral language and memory (LARRC & Nnanatu, in preparation), we reason that this is a 'spurious' finding. Third, we estimated the factor structure for each time point, rather than in one large model with all five time points included (Widaman, Ferrer & Conger, 2010). The latter approach resulted in estimation problems. Additionally, we did not check longitudinal measurement invariance; it is perhaps not surprising that the factor structure should be invariant, given that complex constructs such as inference making, vocabulary, and verbal working memory, are unconstrained skills, informed by other skills and knowledge, as well as metacognitive strategies, and should be expected to change over time. Lastly, future studies that use cross-lagged panel modeling (CLPM) may consider random intercept cross-lagged panel modeling

(RI-CLPM) as proposed by Hamaker, Kuiper, and Grasman (2015). In the former models, the stability paths (or autoregressive effects) represent both a between-person and a within-person component. In RI-CLPM, random intercepts are added to account for between-person variance, so that the variance explained by those two components can be interpreted.

A final limitation of note is that the words tested in our standardized measures of vocabulary breadth and depth were not the same items. In addition, our measure of depth considered the relations between two words that were related by category or were synonyms (the latter for grade 3 only). Experimental measures of vocabulary depth and breadth that assessed knowledge for the same words, and which were designed to assess depth across a range of relations (e.g., shared features), would prove useful in future research design to extend our understanding of how vocabulary depth and breadth relate to inference making.

We conclude with two notable educational implications. The first stems from our finding of strong relations between vocabulary and inference making, which suggests reciprocal relations across the course of development. For vocabulary growth, both direct instruction of words and also strategies such as inference from context are needed to develop these skills; in turn, this supports inference making by providing the relevant vocabulary knowledge and also developing inference itself. Second, our tasks were all delivered aurally; children were not required to read the texts. If children were required to read the stories in our inference task, we would expect performance for younger and poorer readers to be strongly predicted by their word reading ability, rather than vocabulary and working memory (Language and Reading Research Consortium, 2015). We recommend that tasks such as ours are used to assess children's higher-level language skills without the confound of decoding.

In sum, we have demonstrated that vocabulary and verbal working memory are strong predictors of concurrent inference making and that vocabulary in particular supports the development of inference making longitudinally. In addition, our longitudinal findings also contribute to a different line of research that proposes that verbal ability supports verbal working memory (Nation et al., 1999; Van Dyke & Johns, 2012). A key message for educators is that strong vocabulary skills support inference making, and that both inference making and vocabulary support each other reciprocally across development. A related message for researchers is that we need to be mindful of the dynamics of development when we consider the relations between language and cognitive skills.

References

Ackerman, B. P. (1986). Referential and causal coherence in the story comprehension of children and adults. *Journal of Experimental Child Psychology, 41*, 336-366. doi: 10.1016/0022-0965(86)90044-5

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*, 66-82. doi: 10.1016/j.cedpsych.2016.02.002

Baddeley, A.D., & Hitch, G. (1974). Working Memory. In G. Bower (Ed.), *The Psychology of learning and motivation* (*Vol. 8,* pp 47-90). New York: Academic Press.

Barnes, M.A., Dennis, M., & Haefele- Kalvatis, J. (1996). The effects of knowledge availability and knowledge accessibility on coherence and elaborative inferencing in children from six to fifteen years of age. *Journal of Experimental Child Psychology, 61*, 216-241. doi: 10.1006/jecp.1996.0015

Belacchi, C., Carretti, B., & Cornoldi, C. (2010). The role of working memory andupdating in Coloured Raven Matrices performance in typically developing children. *European Journal of Cognitive Psychology, 7,* 1010-1020. doi: 10.1080/09541440903184617

Binder, K. S., Cote, N. G., Lee, C., Bessette, E., & Vu, H. (2017). Beyond breadth: the contributions of vocabulary depth to reading comprehension among skilled readers. *Journal of Research in Reading*, *40*, 333-343. doi: 10.1111/1467-9817.12069

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136-162). Newbury Park, CA: Sage.

Cain, K. (1999). Ways of reading: How knowledge and use of strategies are related to reading comprehension. *British Journal of Developmental Psychology, 17*, 293-309. doi: 10.1348/026151099165285

Cain, K., & Oakhill, J. V. (1999). Inference making and its relation to comprehension failure. *Reading and Writing. An Interdisciplinary Journal*, *11*, 489-503. doi:10.1023/A:1008084120205

Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Annee Psychologique, 114*, 647-662. doi:10.4074/S0003503314004035

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

Cain, K., Oakhill, J. V., & Elbro, C. (2003). The ability to learn new word meanings from context by school-age children with and without language comprehension difficulties. *Journal of Child Language, 30*, 681-694. doi: 10.1017/S0305000903005713

Cain, K., Oakhill, J., & Lemmon, K. (2004). Individual differences in the inference of word meanings from context: the influence of reading comprehension, vocabulary knowledge, and memory capacity. *Journal of Educational Psychology, 96*, 671-681. doi: 10.1037/0022-0663.96.4.671

Chrysochoou, E., & Bablekou, Z. (2011). Phonological loop and central executive contributions to oral comprehension skills of 5.5 to 9.5 years old children. *Applied Cognitive Psychology, 25*, 576-583. doi: 10.1002/acp.1723

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

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

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

Daugaard, H., Cain, K., & Elbro, C. (2017). From words to text: Inference making mediates the role of vocabulary in children's reading comprehension. *Reading and Writing*, *30*, 1773-1788. doi: 10.1007/s11145-017-9752-2

Dunn, L. M. & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test-Fourth Edition*. MN: Pearson Assessment.

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*, 761-781. doi:10.1037/edu0000180

Elleman, A. M., & Compton, D. L. (2017). Beyond comprehension strategy instruction: What's next? *Language, Speech, and Hearing Services in Schools*, *48*, 84-91. doi: 10.1044/2017_LSHSS-16-0036

Filiatrault-Veilleux, P., Bouchard, C., Trudeau, N., & Desmarais, C. (2016). Comprehension of inferences in a narrative in 3- to 6-year-old children. *Journal of speech, language, and hearing research, 59*, 1099-1110. doi: 10.1044/2016_JSLHR-L-15-0252

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

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

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*, 102-116. doi:10.1037/a0038889

Hayduck, L. A. (1996). *LISREL issues, debates, and strategies.* Baltimore, MD: Johns Hopkins University Press.

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

Jenkins, J. R., Matlock, B., & Slocum, T. A. (1989). Two approaches to vocabulary instruction: The teaching of individual word meanings and practice in deriving word meaning from context. *Reading Research Quarterly, 24*, 215-235. doi: 10.2307/747865.

Johnson-Laird, P. N. (1983). *Mental models : towards a cognitive science of language, inference and consciousness*: Cambridge: Harvard University Press.

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

Kendeou, P., & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology, 97*, 235-245. doi: 10.1037/0022-0663.97.2.235

Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading, 31*, 259-272. doi: 10.1111/j.1467-9817.2008.00370.x

Kieffer, M. J., & Lesaux, N. K. (2012). Knowledge of words, knowledge about words: dimensions of vocabulary in first and second language learners in sixth grade. *Reading and Writing*, *25*, 347-373. doi:10.1007/s11145-010-9272-9

Kim, Y.-S. G. (2017). Why the simple view of reading is not simplistic: Unpacking the simple view of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading, 21*, 310-333. doi:10.1080/10888438.2017.1291643

Kintsch, W. (1998). *Comprehension: a paradigm for cognition*: Cambridge New York, NY : Cambridge University Press.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 209-226). Malden, MA: Blackwell.

Kline, R. B. (2011). *Principles and practice of structural equation modeling (3rd ed.)*. New York, NY: Guilford Press.

Language and Reading Research Consortium (LARRC) (2017). Oral language and listening comprehension: Same or different constructs? *Journal of Speech, Language, and Hearing Research, 60*, 1273-1284. doi: 10.1044/2017_JSLHR-L-16-0039

Language and Reading Research Consortium (LARRC), & Logan, J. (2017). Pressure points in reading comprehension: A quantile multiple regression analysis. *Journal of Educational Psychology, 109*, 451-464. doi: 10.1037/edu0000150

Language and Reading Research Consortium (LARRC), Farquharson, K., & Murphy, K. A. (2016). Ten steps to conducting a large, multi-site, longitudinal investigation of language and reading in young children. *Frontiers in Psychology*, *7*, 1-16. doi:10.3389/fpsyg.2016.00419

Language and Reading Research Consortium (LARRC), & Muijselaar, M. (2018). The dimensionality of inference making: are local and global inferences distiniguishable? *Scientific Studies of Reading, 22,* 117-136. doi: 10.1080/10888438.2017.1371179

Language and Reading Research Consortium (LARRC), & Nnanatu, C. (in prep). The dimensionality of language: emergent separation of vocabulary, grammar, discourse and memory over time.

Leather, C.V., & Henry, L.A. (1994). Working memory span and phonological awareness tasks as predictors of early reading ability. *Journal of Experimental Child Psychology, 58,* 88-111. doi: 10.10006/jecp.1994.1027

Lepola, J., Lynch, J. S., Kiuru, N., Laakkonen, E., & Niemi, P. (2016). Task orientation, and foundational reading skills as predictors of grade 3 reading comprehension. *Reading Research Quarterley, 51*, 373-390. doi: 10.1002/rrq.145

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

Long, D. L., & Chong, J. L. (2001). Comprehension skill and global coherence: a paradoxical picture of poor comprehenders' abilities. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1424-1429. doi: 10.1037//0278-7393.27.6.1424

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

McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20*, 482-496. doi: 10.2307/747855.

Muijselaar, M. M. L., & de Jong, P. F. (2015). The effects of updating ability and knowledge of reading strategies on reading comprehension. *Learning and Individual Differences, 43*, 111-117. doi: 10.1016/j.lindif.2015.08.011

Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide*. *Seventh edition*. Los Angeles, CA: Muthén & Muthén.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*, 233-253. doi:10.2307/747758

Nation, K., Adams, J. W., Bowyer-Crane, C. A., & Snowling, M. J. (1999). Working memory deficits in poor comprehenders reflect underlying language impairments. *Journal of Experimental Child Psychology, 73*, 139-158. doi:10.1006/jecp.1999.2498

Oakhill, J., & Cain, K. (2012). The precursors of reading comprehension and word reading in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading, 16*, 91-121, doi: 0.1080/10888438.2010.529219

Oakhill, J., Cain, K., & McCarthy, D. (2015). Inference processing in children: the contributions of depth and breadth of vocabulary knowledge. In O'Brien, E. J., Cook, A. E., & Lorch, R. F. (Eds.), *Inferences during reading* (pp. 140-159). Cambridge, UK: Cambridge University Press.

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

Perfetti, C. (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357-383. doi: 10.1080/10888430701530730

Perfetti, C., Yang, C., & Schmalhofer, F. (2008). Comprehension skill and word to text integration processes. *Applied Cognitive Psychology, 22,* 303-318, doi: 10.1002/acp.1419

Proctor, C.P., Silverman, R.D., Harring, J.R. & Montecillo, C. (2012). The role of vocabulary depth in predicting reading comprehension among English monolingual and Spanish-English bilingual children in elementary school. *Reading and Writing*, *24*, 1635-1664. doi: 10.1007/s11145-011-9336-5

Quinn, J. M., Wagner, R. K., Petscher, Y., & Lopez, D. (2015). Developmental relations between vocabulary knowledge and reading comprehension: a latent change score modeling study. *Child Development*, *86*, 159-175. doi:10.1111/cdev.12292

Richter, T., Isberner, M., Naumann, J., & Neeb, Y. (2014). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading, 17*, 415-434. doi: 10.1080/10888438.2013.764879

Schneider, W., Körkel, J., & Weinert, F. E. (1989). Domain-specific knowledge and memory performance: A comparison of high- and low-aptitude children. *Journal of Educational Psychology*, *81*, 306-312. doi:10.1037/0022-0663.81.3.306

Seigneuric, A., & Ehrlich, M.-F. (2005). Contribution of working memory capacity to children's reading comprehension: a longitudinal investigation. *Reading and Writing: An Interdisciplinary Journal, 18*, 617-656. doi: 10.1007/s11145-005-2038-0

Semel, E., Wiig, E. H. & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals–Fourth Edition; Examiner's Manual.* Bloomington, MN: Pearson, Inc.

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

Spencer, M., Muse, A., Wagner, R. K., Foorman, B., Petscher, Y., Schatschneider, C., . . . Bishop, M. D. (2015). Examining the underlying dimensions of morphological awareness and vocabulary knowledge. *Reading and Writing, 28*, 959-988. doi: 10.1007/s11145-015-9557-0.

Swart, N., Muijselaar, M., Steenbeek-Planting, E., Droop, M., Jong, P., & Verhoeven, L. (2017). Differential lexical predictors of reading comprehension in fourth graders. *Reading and Writing, 30*, 489-507. doi: 10.1007/s11145-016-9686-0

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

Tompkins, V., Guo, Y., & Justice, L. M. (2013). Inference generation, story comprehension, and language skills in the preschool years. *Reading and Writing: An Interdisciplinary Journal, 26*, 403-429. doi: 10.1007/s11145-012-9374-7

Van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M.J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. T. McCrudden, J. Magliano & G.Schraw (Eds.), *Relevance instructions and goal focusing in text learning* (pp.123-140). Greenwich, CT: Information Age Publishing.

Van Dyke, J. A., & Johns, C. L. (2012). Memory interference as a determinant of language comprehension. *Language and Linguistics Compass, 6*, 193-211. doi: 10.1002/lnc3.330

Van Dyke, J. A., Johns, C. L., & Kukona, A. (2014). Low working memory capacity is only spuriously related to poor reading comprehension. *Cognition, 131*, 373-403. doi: 10.1016/j.cognition.2014.01.007

Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: a longitudinal study. *Applied Cognitive Psychology, 22*, 407-423. doi: 10.1002/acp.1414

Verhoeven, L., Van Leeuwe, J., Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading, 15*, 8-25. doi: 10.1080/10888438.2011.536125

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives, 4,* 10-18. doi: 10.1111/j.1750-8606.2009.00110.x

Williams, K. T. (2007). *Expressive Vocabulary Test, Second Edition.* Bloomington, MN: Publisher: Pearson.

Woodcock, R.W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Test of Cognitive Abilities*. Itasca, IL: Riverside Publishing.

Appendix A

Example of a text and questions from the Inference Task (* = local coherence inference). Please see LARRC and Muijselaar (2018) for the full set of stories and questions.

A NEW PET (P & K)

Tim had a new pet called Sparky. Sparky was soft, furry, and very playful. At first, Sparky slept indoors in a cardboard box with a nice soft blanket. Sparky soon grew very big. Tim decided to build a kennel and a tall wooden fence around the back yard.

Tim went to the store. He already had a hammer and a saw, but he needed some wood and some nails. Tim built the kennel first. His friend Jack helped him to build the fence. Jack held the wood and Tim banged in the nails. The fence was soon finished. Even though Tim's thumb was bruised and sore, he was smiling. He put the hammer that had caused the pain away in his toolbox. He was very pleased with his hard work.

That evening, Tim moved Sparky into his new home. But, Sparky did not like his new home. His old cardboard box was still indoors and Sparky missed his nice soft blanket.

Questions

1. What sort of animal was Sparky?

Answer: dog

2. What did Tim buy at the store? *

Answer: wood and nails

3. Who put up the fence? *

Answer: Tim & Jack, Tim & his friend, the man & his friend

4. Why did Tim need a tall fence?

Answer: because Sparky could jump/ so Sparky didn't run away

5. Why did Tim have a sore thumb?

Answer: banged/hit his thumb with hammer etc.

6. Where was Sparky's kennel? *

Answer: in yard, outside in back yard

7. Why did Sparky no longer sleep in the cardboard box?

Answer: he was too big, he had grown too big, outgrown it

8. Where was Sparky's blanket? *

Answer: (still) in his box, in the house

Table 1

Descriptive Statistics for Inference Making, Vocabulary, and Verbal Working Memory by Grade

	Pre-kindergarten				Kindergarten				Grade 1				Grade 2				Grade 3			
	M	*SD*	S	K	*M*	*SD*	S	K	*M*	*SD*	S	K	*M*	*SD*	S	K	*M*	*SD*	S	K
IT local	0.88	0.49	0.14	-0.80	1.07	0.44	-0.39	-0.39	1.17	0.41	-0.61	0.08	1.16	0.43	-0.50	-0.41	1.38	0.45	-0.63	-0.37
IT global	0.76	0.40	0.15	-0.35	1.18	0.39	-0.38	-0.15	1.24	0.40	-0.55	0.02	1.21	0.43	-0.35	-0.58	1.57	0.36	-0.88	0.33
PPVT	94.06	18.94	-0.37	0.09	113.60	16.86	-0.45	0.45	128.12	15.04	-0.31	0.41	140.01	15.28	-0.39	0.55	152.54	14.61	-0.17	0.67
EVT	70.07	13.74	0.03	-0.23	85.09	12.63	-0.15	-0.45	95.77	13.31	-0.12	-0.20	106.40	12.78	-0.22	0.17	116.28	13.87	0.08	-0.10
WC: rec	14.30	4.43	-0.82	0.07	17.71	3.05	-1.53	2.08	19.22	1.67	-1.30	1.75	19.88	1.15	-0.89	-0.02	10.76	2.86	-0.29	-0.04
WC: exp	7.81	4.54	-0.02	-0.82	12.54	3.97	-0.86	0.49	15.09	2.42	-0.73	0.48	16.38	2.19	-0.68	0.39	6.99	2.76	0.03	-0.86
WJAM	5.81	4.54	0.70	-0.37	11.45	4.88	-0.08	-0.45	15.64	5.09	-0.23	-0.06	17.69	5.39	-0.45	-0.22	19.76	6.20	-0.41	-0.08
MU	4.21	2.65	0.81	0.17	6.56	3.17	0.58	0.15	8.80	3.91	0.57	0.17	11.45	4.93	0.66	0.58	13.77	5.46	0.55	-0.13

Note. M = Mean; SD = Standard Deviation; S = Skewness; K = Kurtosis. IT: Inference Task; Vocabulary measures: PPVT: Peabody Picture Vocabulary Test – Fourth Edition; EVT: Expressive Vocabulary Test; WC: CELF-4, Subtest Word Classes. Rec: receptive; Exp: expressive; Verbal Working memory measures: WJAM: Woodcock Johnson Auditory Memory; MU: Memory Updating.

CELF-WC 1 was administered in P, K, G1, G2; CELF-WC 2 was administered in G3.

Table 2

Correlations among all Measures by Grade

	IT local	IT global	PPVT	EVT	WC: receptive	WC: expressive	WJAM	MU
IT local	1							
IT global	.67/.60/.47 / .58/.59	1						
PPVT	.57/.53/.31 / .50/.38	.54/.55/.53 / .52/.46	1					
EVT	.57/.51/.35 / .42/.34	.49/.54/.54 / .48/.46	.72/.71/.74/ .77/.72	1				
WC: receptive	.36/.32/.17 / .18/.25	.35/.32/.31 / .17/.32	.42/.46/.54/ .52/.58	.45/.49/.53/ .49/.61	1			
WC: expressive	.50/.40/.29 / .40/.26	.47/.42/.43 / .37/.33	.51/.56/.37/ .20/.61	.58/.61/.37/ .32/.66	.78/.80/.67/ .56/.76	1		

WJAM	.37/.26/.22 / .27/.25	.39/.35/.31 / .30/.27	.35/.27/.41/ .49/.33	.39/.35/.42/ .47/.42	.27/.28/.33/ .36/.35	.38/.32/.27/ .22/.39	1	
MU	.34/.25/.17 / .17/.22	.38/.36/.22 / .31/.25	.40/.32/.30/ .36/.25	.40/.35/.34/ .33/.36	.22/.27/.27/ .28/.25	.30/.37/.22/ .20/.26	.38/.33/.40/ .46/.43	1

Note. Correlations in Pre-kindergarten/Kindergarten/Grade 1/Grade 2/Grade 3 [1]

For all correlations: $p < .01$.

IT: Inference Task; Vocabulary measures: PPVT: Peabody Picture Vocabulary Test – Fourth Edition; EVT: Expressive Vocabulary Test; WC: CELF-4, Subtest Word Classes. Verbal working memory measures: WJAM: Woodcock Johnson Auditory Memory; MU: Memory Updating.

CELF-WC 1 was administered in P, K, G1, G2; CELF-WC 2 was administered in G3.

Longitudinal correlations can be found in Appendix B.

Table 3

Fit Indices for the Measurement Models for Inference Making, Vocabulary Breadth and Depth, and Verbal Working Memory

Grade	χ^2	*df*	RMSEA	90% CI	CFI
P	20.82	15	.030	.000 - .059	1.00
K	14.32	15	.000	.000 - .047	1.00
G1	7.02	15	.000	.000 - .000	1.00
G2	33.78**	15	.063	.034 - .091	.98
G3	31.51**	15	.061	.031 - .091	.98

Note. P: Pre-kindergarten; K: Kindergarten; G1: Grade 1; G2: Grade 2; G3: Grade 3.

Table 4

Hierarchical Regression Analyses for the Concurrent Predictions of Inference Making by Vocabulary Breadth and Depth, and Verbal Working Memory

	P		K		G1		G2		G3	
	ΔR^2	β[a]	ΔR^2	β[a]	ΔR^2	β[a]	ΔR^2	β[a]	ΔR^2	β[a]
1. Vocabulary breadth	.79**	.58**	.81**	.58**	.67**	.87**	.70**	.90**	.50**	.60**
2. Verbal working memory	.03**	.36**	.04**	.38**	.00	-.06	.00	-.08	.01*	.14*
Total R^2	.82		.85		.67		.71		.51	
1. Vocabulary depth	.46**	.16**	.38**	-.09*	.34**	.31**	.36**	.35**	.19**	.11
2. Verbal working memory	.30**	.76**	.38**	.94**	.14**	.47**	.14**	.45**	.17**	.53**
Total R^2	.76		.76		.48		.50		.36	

[a] Standardized regression coefficients are provided for the final models.

* $p < .05$. ** $p < .01$.

Note. P: Pre-kindergarten; K: Kindergarten; G1: Grade 1; G2: Grade 2; G3: Grade 3.

Table 5

Longitudinal Relations between Inference Making, Vocabulary Breadth/Depth, and Verbal Working Memory

Stability paths / autoregressive effects of models with vocabulary breadth				
Inference making				
ON	G2	G1	K	P
G3	.35**	.12	.03	.20**
G2	-	.15*	.26**	.07
G1	-	-	.26**	.20**
K	-	-	-	.33**
Vocabulary breadth				
ON	G2	G1	K	P
G3	.64**	.21**	.07	.06
G2	-	.64**	.23**	.12**
G1	-	-	.73**	.28**
K	-	-	-	.71**
Verbal working memory				

ON	G2	G1	K	P
G3	.43**	.23**	.10	.14*
G2	-	.33**	.23**	.09
G1	-	-	.27**	.31**
K	-	-	-	.49**

Stability paths / autoregressive effects of models with vocabulary depth

Inference making

ON	G2	G1	K	P
G3	.42**	.11	.06	.23**
G2	-	.31**	.32**	.11*
G1	-	-	.45**	.30**
K	-	-	-	.51**

Vocabulary breadth

ON	G2	G1	K	P
G3	.03	.03	.19**	.05
G2	-	.23**	.23**	.11

G1	-	-	.42**	.12*
K	-	-	-	.37**

Verbal working memory

ON	G2	G1	K	P
G3	.43**	.22**	.13*	.12*
G2	-	.41**	.25**	.14*
G1	-	-	.21*	.43**
K	-	-	-	.50**

Covariances

	IT	VOC_b		IT	VOC_d
IT	1	-	IT	1	-
VOC_b	.89/.71/.50/.56/.38	1	VOC_d	.68/.29/.27/.34/-.04	1
VWM	.87/.69/.22/.24/.33	.89/.61/.54/.51/.49	VWM	.87/.69/.25/.27.33	.69/.54/.28/.26/.25

Indirect effects of models with vocabulary breadth

IT_G3←VOCb_G2←VWM_G1	-.00	IT_G2←VOCb_G1←VWM_K	-.01	IT_G1←VOCb_K←VWM_P	.03
IT_G3←VWM_G2←VOCb_G1	-.01	IT_G2←VWM_G1←VOCb_K	-.05	IT_G1←VWM_K←VOCb_P	.00

VOCb_G3←IT_G2←VWM_G1	-.00	VOCb_G2←IT_G1←VWM_K	-.00	VOCb_G1←IT_K←VWM_P	-.01
VOCb_G3←VWM_G2←IT_G1	.01	VOCb_G2←VWM_G1←IT_K	.00	VOCb_G1←VWM_K←IT_P	-.00
VWM_G3←IT_G2←VOCb_G1	.01	VWM_G2←IT_G1←VOCb_K	-.03	VWM_G1←IT_K←VOCb_P	-.09*
VWM_G3←VOCb_G2←IT_G1	.00	VWM_G2←VOCb_G1←IT_K	-.02	VWM_G1←VOCb_K←IT_P	.04
Indirect effects of models with vocabulary depth					
IT_G3←VOCd_G2←VWM_G1	-.00	IT_G2←VOCd_G1←VWM_K	-.01	IT_G1←VOCd_K←VWM_P	.04
IT_G3←VWM_G2←VOCd_G1	.00	IT_G2←VWM_G1←VOCd_K	.01	IT_G1←VWM_K←VOCd_P	-.00
VOCd_G3←IT_G2←VWM_G1	.02	VOCd_G2←IT_G1←VWM_K	-.00	VOCd_G1←IT_K←VWM_P	.10**
VOCd_G3←VWM_G2←IT_G1	.00	VOCd_G2←VWM_G1←IT_K	.00	VOCd_G1←VWM_K←IT_P	-.03
VWM_G3←IT_G2←VOCd_G1	.00	VWM_G2←IT_G1←VOCd_K	.00	VWM_G1←IT_K←VOCd_P	.01
VWM_G3←VOCd_G2←IT_G1	.00	VWM_G2←VOCd_G1←IT_K	.04*	VWM_G1←VOCd_K←IT_P	.01

Note. *p < .05. **p < .01.

Covariances in Pre-kindergarten/Kindergarten/Grade 1/Grade 2/Grade 3

For all covariances: $p < .01$.

Note. P: Pre-kindergarten; K: Kindergarten; G1: Grade 1; G2: Grade 2; G3: Grade 3.

IT: Inference Task; VOCb: Vocabulary Breadth; VOCd: Vocabulary Depth; VWM: Verbal Working Memory.

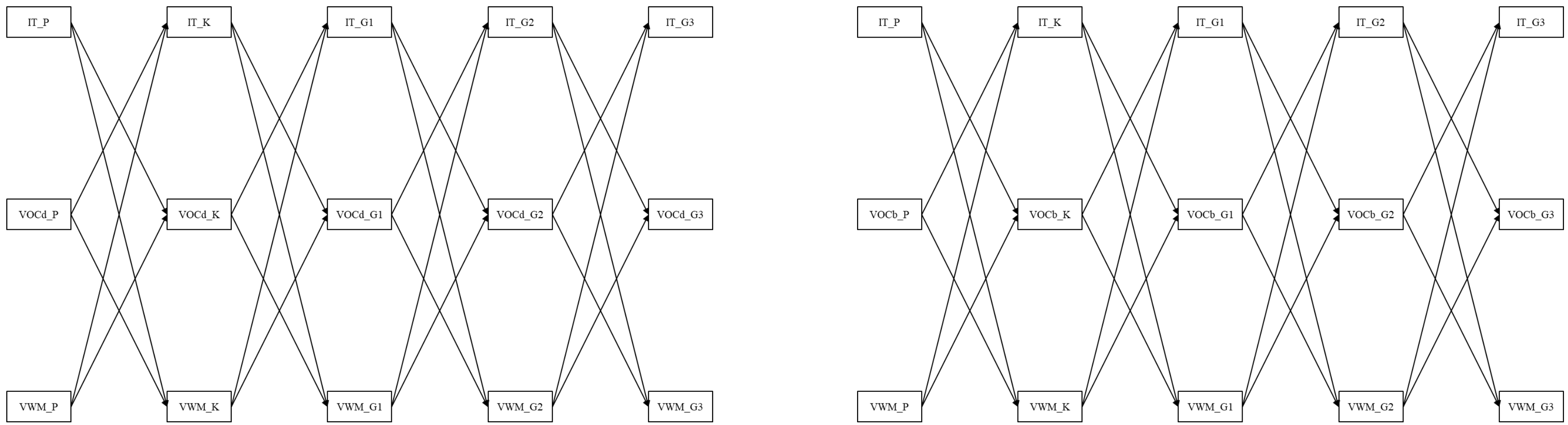

Figure 1. Proposed cross-lagged models

Note. Covariances and stability paths are not displayed. P: Pre-kindergarten; K: Kindergarten; G1: Grade 1; G2: Grade 2; G3: Grade 3. IT: Inference Task; VOCb: Vocabulary Breadth; VOCd: Vocabulary Depth; WM: Verbal Working Memory.

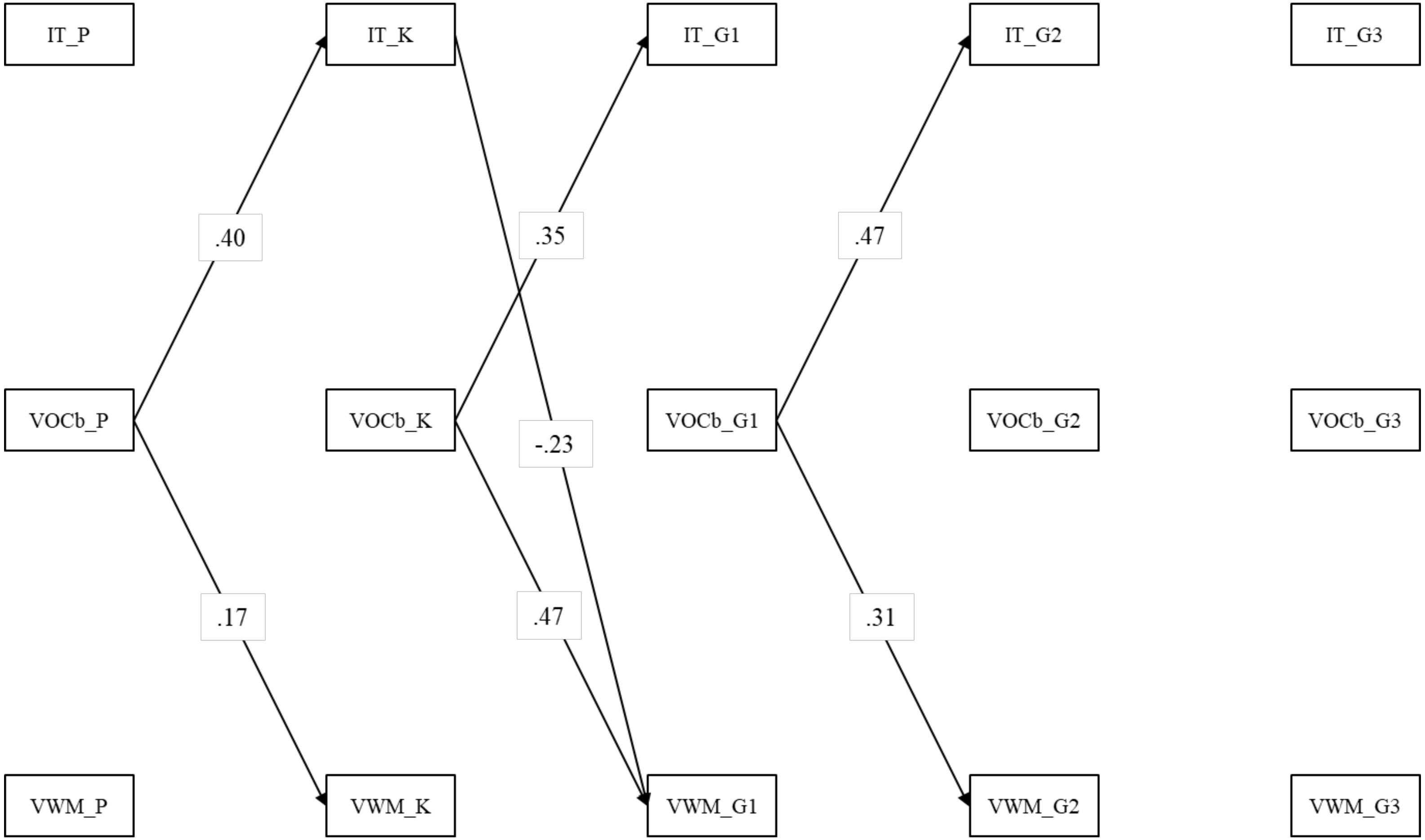

Figure 2. Longitudinal Relations between Inference Making, Vocabulary Breadth, and Verbal Working Memory

Note. Only significant cross-lagged paths are depicted. Covariances and stability paths are not displayed. P: Pre-kindergarten; K: Kindergarten; G1: Grade 1; G2: Grade 2; G3: Grade 3. IT: Inference Task; VOCb: Vocabulary Breadth; VOCd: Vocabulary Depth; WM: Verbal Working Memory.

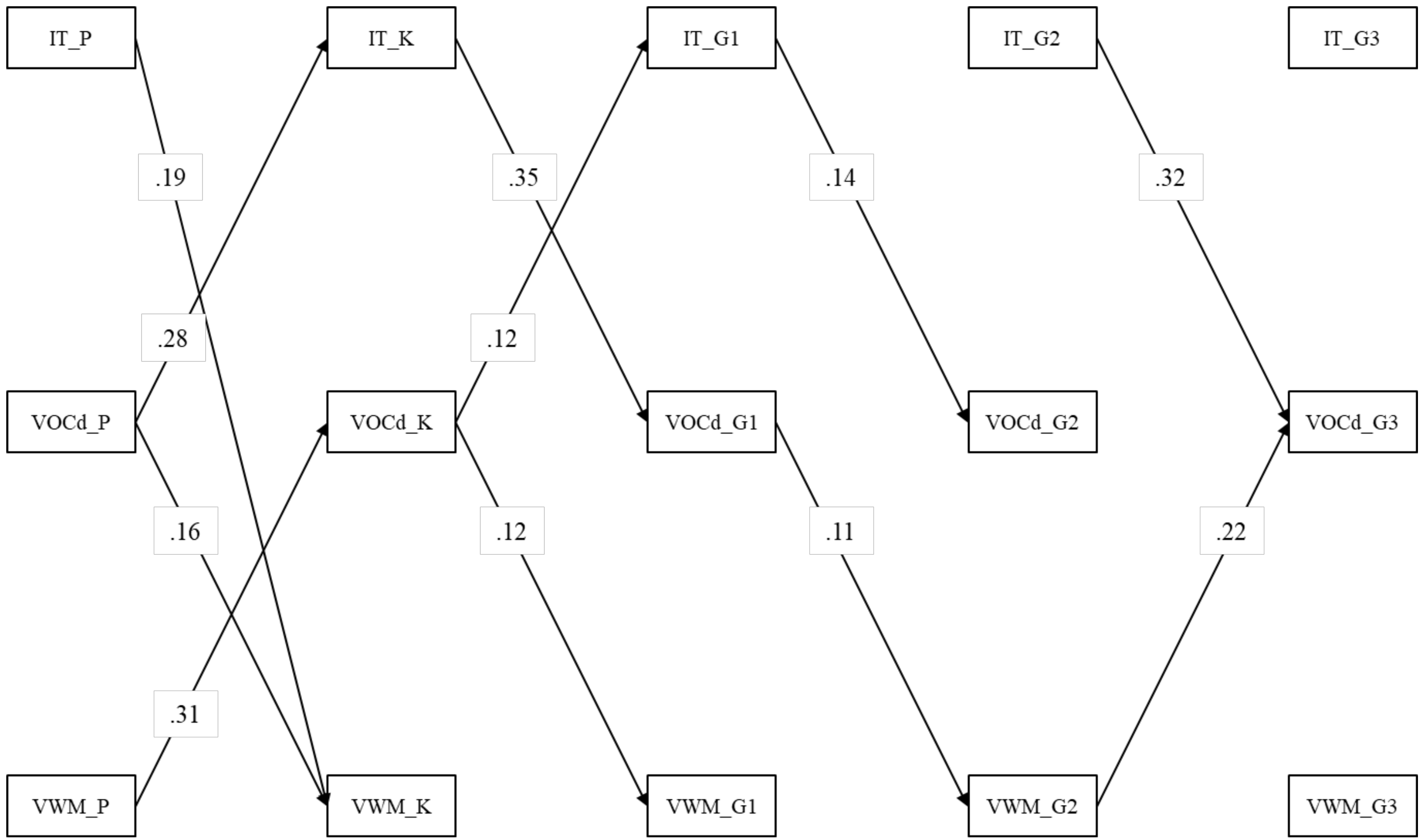

Figure 3. Longitudinal Relations between Inference Making, Vocabulary Depth, and Verbal Working Memory

Note. Only significant cross-lagged paths are depicted. Covariances and stability paths are not displayed. P: Pre-kindergarten; K: Kindergarten; G1: Grade 1; G2: Grade 2; G3: Grade 3. IT: Inference Task; VOCb: Vocabulary Breadth; VOCd: Vocabulary Depth; WM: Verbal Working Memory.

Appendix B

Longitudinal Correlations among All Measures

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1	IT local P	1																																							
2	IT global P	.67	1																																						
3	PPVT P	.57	.54	1																																					
4	EVT P	.57	.49	.72	1																																				
5	WC1 rec P	.36	.35	.42	.45	1																																			
6	WC1 exp P	.50	.47	.51	.58	.78	1																																		
7	WJAM P	.37	.39	.35	.39	.27	.38	1																																	
8	MU P	.34	.38	.40	.40	.22	.30	.38	1																																
9	IT local K	.56	.47	.49	.47	.26	.37	.31	.27	1																															
10	IT global K	.58	.55	.53	.56	.33	.44	.32	.34	.60	1																														
11	PPVT K	.54	.52	.71	.66	.39	.53	.36	.36	.53	.55	1																													
12	EVT K	.55	.51	.66	.75	.45	.61	.41	.37	.51	.54	.70	1																												
13	WC1 rec K	.29	.26	.37	.40	.52	.50	.30	.26	.32	.32	.45	.49	1																											
14	WC1 exp K	.38	.39	.49	.54	.51	.60	.38	.30	.39	.42	.56	.61	.80	1																										
15	WJAM K	.34	.33	.27	.32	.24	.36	.35	.32	.26	.35	.27	.35	.28	.32	1																									
16	MU K	.30	.27	.31	.27	.28	.32	.32	.44	.25	.36	.32	.35	.27	.37	.33	1																								
17	IT local G1	.38	.27	.36	.33	.25	.33	.16	.17	.38	.37	.33	.33	.25	.35	.23	.17	1																							
18	IT global G1	.53	.39	.55	.50	.33	.45	.22	.31	.47	.54	.51	.53	.36	.42	.27	.31	.47	1																						
19	PPVT G1	.53	.56	.71	.69	.43	.55	.41	.34	.50	.53	.77	.74	.45	.56	.28	.33	.31	.53	1																					
20	EVT G1	.55	.49	.64	.71	.43	.56	.40	.36	.49	.51	.66	.79	.43	.53	.32	.30	.34	.54	.74	1																				
21	WC1 rec G1	.23	.23	.25	.28	.40	.37	.22	.24	.29	.34	.36	.36	.52	.42	.23	.20	.17	.31	.37	.37	1																			
22	WC1 exp G1	.36	.28	.42	.47	.52	.53	.31	.22	.39	.41	.52	.57	.57	.59	.25	.23	.28	.43	.54	.53	.67	1																		
23	WJAM G1	.38	.33	.36	.39	.29	.43	.37	.39	.26	.32	.36	.43	.28	.40	.35	.33	.21	.31	.41	.41	.27	.33	1																	
24	MU G1	.23	.25	.26	.28	.24	.31	.28	.36	.22	.24	.28	.35	.29	.33	.27	.34	.17	.21	.30	.34	.22	.27	.40	1																
25	IT local G2	.40	.37	.42	.40	.23	.30	.23	.19	.43	.41	.40	.41	.25	.34	.23	.25	.37	.45	.47	.43	.20	.34	.21	.14	1															
26	IT global G2	.48	.41	.48	.46	.25	.34	.19	.25	.52	.52	.46	.47	.31	.38	.24	.22	.39	.50	.51	.50	.23	.32	.24	.21	.58	1														
27	PPVT G2	.50	.49	.65	.68	.41	.56	.43	.33	.47	.52	.72	.68	.43	.57	.36	.31	.36	.52	.82	.71	.38	.52	.41	.29	.50	.52	1													
28	EVT G2	.52	.47	.65	.71	.48	.62	.40	.34	.51	.52	.68	.78	.49	.58	.33	.30	.33	.49	.72	.82	.40	.56	.40	.31	.42	.48	.77	1												
29	WC1 rec G2	.26	.17	.19	.18	.30	.30	.22	.18	.16	.21	.23	.27	.35	.35	.18	.18	.08	.20	.18	.28	.38	.41	.19	.16	.18	.17	.20	.32	1											

30	WC1 exp G2	.33	.26	.37	.43	.33	.45	.29	.26	.34	.31	.46	.48	.40	.52	.29	.22	.20	.37	.48	.48	.35	.49	.26	.23	.40	.37	.52	.48	.56	1										
31	WJAM G2	.42	.38	.36	.40	.29	.42	.38	.35	.31	.35	.39	.43	.29	.37	.43	.35	.23	.29	.44	.45	.28	.37	.49	.39	.27	.30	.49	.47	.22	.36	1									
32	MU G2	.30	.26	.28	.33	.24	.38	.29	.27	.26	.31	.37	.37	.26	.34	.35	.34	.22	.25	.38	.36	.25	.37	.42	.37	.17	.31	.35	.33	.20	.28	.45	1								
33	IT local G3	.45	.29	.40	.31	.14	.25	.23	.19	.34	.34	.33	.33	.18	.21	.13	.17	.36	.34	.38	.35	.11	.20	.19	.12	.46	.50	.35	.31	.07	.13	.18	.18	1							
34	IT global G3	.43	.40	.40	.42	.20	.32	.26	.26	.40	.35	.45	.44	.28	.36	.14	.19	.33	.43	.47	.45	.22	.30	.25	.21	.46	.49	.46	.45	.17	.29	.26	.24	.59	1						
35	PPVT G3	.42	.43	.60	.59	.34	.43	.35	.35	.41	.44	.66	.63	.38	.48	.31	.30	.23	.39	.76	.66	.34	.45	.39	.28	.45	.44	.82	.73	.23	.45	.41	.27	.38	.46	1					
36	EVT G3	.50	.47	.61	.67	.38	.54	.43	.35	.46	.45	.61	.75	.43	.54	.32	.33	.30	.45	.70	.79	.32	.47	.43	.28	.44	.48	.75	.84	.28	.46	.43	.33	.33	.46	.72	1				
37	WC2 rec G3	.39	.39	.50	.51	.32	.42	.28	.33	.35	.40	.50	.52	.38	.43	.26	.26	.20	.43	.57	.58	.31	.40	.31	.28	.32	.41	.62	.62	.22	.40	.38	.33	.25	.32	.58	.61	1			
38	WC2 exp G3	.37	.42	.46	.54	.33	.43	.35	.36	.37	.39	.53	.57	.43	.50	.34	.28	.28	.40	.57	.60	.29	.41	.35	.28	.41	.42	.63	.62	.26	.44	.40	.34	.26	.33	.61	.66	.76	1		
39	WJAM G3	.34	.30	.33	.35	.25	.43	.40	.41	.26	.24	.34	.41	.26	.34	.44	.38	.25	.24	.38	.38	.21	.31	.52	.37	.19	.24	.42	.40	.17	.31	.64	.51	.25	.27	.33	.42	.35	.39	1	
40	MU G3	.29	.17	.25	.22	.15	.22	.27	.30	.25	.21	.23	.34	.25	.27	.34	.26	.21	.28	.29	.30	.22	.27	.39	.38	.15	.24	.26	.32	.27	.32	.35	.38	.22	.25	.25	.36	.25	.26	.43	1

Note. For all correlations: $p < .01$.

IT: Inference Task; Vocabulary measures: PPVT: Peabody Picture Vocabulary Test – Fourth Edition; EVT: Expressive Vocabulary Test; WC: CELF-4, Subtest Word Classes. Rec: receptive; Exp: expressive; Verbal working memory measures: WJAM: Woodcock Johnson Auditory Memory; MU: Memory Updating.

CELF-WC 1 was administered in P, K, G1, G2; CELF-WC 2 was administered in G3.