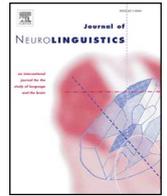


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Situation model building ability uniquely predicts first and second language reading comprehension



Henriette Raudszus*, Eliane Segers, Ludo Verhoeven

Behavioural Science Institute, Radboud University, Montessorilaan 3, P.P. Box 9104, 6500 HE, Nijmegen, the Netherlands

ABSTRACT

We examined the unique role of textbase memory and situation model building ability in first (L1) and second (L2) language reading comprehension. Participants were 76 monolingual and 102 bilingual children in 4th grade. A pathfinder network approach was used to assess textbase memory and situation model building ability, on top of other well-known cognitive and linguistic predictors of reading comprehension (working memory, nonverbal reasoning, decoding, vocabulary, and grammar). Reading comprehension was assessed by a standardized task unrelated to the textbase and situation model building task. The results showed that there was no difference between L1 and L2 readers in nonverbal reasoning, working memory, textbase memory and situation model building. L2 readers were more efficient decoders than L1 readers, but lagged behind on vocabulary, grammar, and reading comprehension. Situation model building ability was found to predict reading comprehension over and above the other cognitive and linguistic predictors to the same extent in both groups.

1. Introduction

Reading comprehension is a prerequisite for academic success and societal participation. For many children, the transition from ‘learning to read’ to ‘reading to learn’ in 4th grade is a stumbling block in their reading development (Hirsch, 2003), and this is even more so for children learning to read in their second language (L2) (e.g., Droop & Verhoeven, 2003). Learning from text requires readers to not only to build up a literal memory of propositions in a text (i.e., the *textbase*), but also to build a coherent *situation model* of concepts in the text and their interrelations, integrated with prior knowledge (van Dijk & Kintsch, 1983). In building such a situation model, vocabulary is crucial, as it forms the bridge between word decoding and higher-level integration processes, such as inferencing and comprehension monitoring (Perfetti & Stafura, 2014). Previous research has shown that L2 readers usually do not struggle with word decoding, while their smaller target language vocabulary clearly plays a role in their reading comprehension problems (for a review, see Melby-Lervåg & Lervåg, 2014). However, whether and how L2 readers’ higher-level integration processes contribute to their reading comprehension outcomes is still unclear. This is problematic given the central role of integration processes in proficient reading.

A seminal model of reading comprehension is the reading systems framework (Perfetti & Stafura, 2014), which posits that reading comprehension entails interactive cognitive processing, builds on various knowledge sources (such as orthographic, linguistic, and world knowledge), and is constrained by limited cognitive resources. In this view, well-defined and richly connected mental representations of words (high lexical quality) allow for efficient word identification, which frees up attentional resources for higher-level integration processes. While there is an abundance of research showing the importance of lexical quality for reading comprehension outcomes in L2 readers (e.g., Babayiğit, 2015; Melby-Lervåg & Lervåg, 2014; Proctor, Silverman, Haring, & Montecillo, 2012), much less is known about how integration processes at the text level contribute to L2 reading comprehension outcomes. From the perspective of the reading systems framework, lower lexical quality in the L2 is likely to impact higher-level processing

* Corresponding author.

E-mail address: h.raudszus@bsi.ru.nl (H. Raudszus).

<https://doi.org/10.1016/j.jneuroling.2018.11.003>

Received 6 April 2018; Received in revised form 17 October 2018; Accepted 26 November 2018

Available online 11 January 2019

0911-6044/ © 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

negatively. On the other hand, L2 readers might be able to “patch” their mental representation of the text with L1 knowledge resources and increased recruitment of cognitive control (Raudszus, Segers, & Verhoeven, 2018; Li & Clariana, 2018), leading to a relatively good mental representation despite lower-level linguistic problems.

Assessing the mental representation of a text is difficult. Simple reading comprehension questions often focus directly on small parts of the text itself, and not on the mental model that has been built of the entire text, in which a connection with long-term memory has been made. Relationship judgements scaled by a pathfinder algorithm are a promising method of assessment of text-level integration abilities. Building on the assumption that the mental representation of a text can be approximated by a network of concepts and their relations, properties of a reader's representation of the text can be examined (Clariana, 2010). The similarity of a child's text representation to the literal text can be seen as an indicator of textbase memory, while similarity to an expert model can be seen as a proxy for situation model quality (Fesel, Segers, Clariana, & Verhoeven, 2015). In the current study, we employed a pathfinder approach to measure textbase memory and situation model building ability in 4th grade L1 and L2 readers, in order to investigate the impact of higher-order integration abilities on general reading comprehension outcomes in a linguistically diverse sample.

1.1. Levels of processing in reading comprehension

Definitions about what exactly reading comprehension entails differ, but there is general agreement that the outcome of successful reading comprehension is a mental representation of concepts from the text and their interrelations (Kendeou, Van den Broek, White, & Lynch, 2009; Li & Clariana, 2018; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007; van den Broek & Helder, 2017). The construction-integration model of reading comprehension (van Dijk & Kintsch, 1983) distinguishes two levels of representation: The *textbase*, which is a propositional representation of the text, and the *situation model*, which is a more elaborate representation of the text integrated with prior knowledge. In the construction-integration framework, comprehension is a result of bottom-up processes such as word recognition and spreading activation to associated concepts, and top-down processes such as comprehension monitoring and inhibition of irrelevant associations (Kintsch, 2005). This is consistent with Perfetti and Stafura's (2014) reading systems framework. According to the reading systems framework, word decoding activates linguistic and general knowledge, which then needs to be integrated through inferencing and monitoring processes. Word-to-text integration takes place at the levels of the sentence, textbase, and situation model. These models from reading research are also consistent with neurobiological models of language processing, such as the memory, unification, and control paradigm (Hagoort, 2005), according to which word representations are retrieved from long-term memory and unified into larger units, a process that is guided by cognitive control. These models share the assumption that word recognition needs to be complemented with higher-level cognitive processes such as inferencing and comprehension monitoring, and that the outcome of successful comprehension is a mental representation of relationships between concepts in a text.

The theoretical distinction between word decoding and higher-level comprehension processes is well-supported by a considerable body of research showing that in children learning to read, word decoding and language comprehension skills develop independently, at least to some extent, and that both contribute to reading comprehension (e.g., Kendeou et al., 2009; Lervåg, Hulme, & Melby-Lervåg, 2017; Storch & Whitehurst, 2002; van Viersen et al., 2018). However, language comprehension as operationalized in this line of research subsumes a wide variety of skills, such as verbal working memory, vocabulary, and inferencing (e.g., Lervåg, Hulme, & Melby-Lervåg, 2017), which map onto different processing levels theoretically. For example, basic vocabulary and syntax might be sufficient to build a representation of the textbase, but for an integrated situation model, higher-level inferencing and comprehension monitoring are necessary.

For children learning to read in their L1, there is evidence for a distinction between language and general knowledge on the one hand, and the ability to utilize both to construct a coherent mental representation from a text on the other hand. Research on inferencing, for example, has shown that inferencing is an important contributor to reading comprehension (Kendeou, McMaster, & Christ, 2016), which is separable from basic reading skills and vocabulary (Kendeou, Bohn-Gettler, White, & van den Broek, 2008) as well as from background knowledge (Cain, Oakhill, Barnes, & Bryant, 2001). This does not mean that inferencing is entirely independent from lower-level language skills such as vocabulary, however. For children aged five to ten years, Currie and Cain (2015) found that vocabulary predicted the ability to make global coherence inferences. Adding to this, Cain and Oakhill (2014) reported that in ten- to eleven-year-olds, vocabulary depth in particular was related to higher-level inferencing skill. Segers and Verhoeven (2016) found that both easy and more complex logical inferences were predicted by lexical quality, against their original hypothesis that only easier text-based inferences would be related to vocabulary. Thus, while inferencing makes an independent contribution to reading comprehension, vocabulary and inferencing are clearly related.

1.2. Higher-level processing in second language reading comprehension

For children who learn to read in their L2, low reading comprehension outcomes are a pervasive problem. Previous research has clearly shown that lower comprehension scores of L2 readers in primary school are linked to lower target language vocabulary (for a review, see Melby-Lervåg & Lervåg, 2014). This concerns both the number of words known (vocabulary breadth) and the quality of their representations (vocabulary depth) (Proctor et al., 2012), as well as the accessibility of semantic information (Cremer & Schoonen, 2013). L2 readers have also been found to have weaker target language syntactic skills (e.g., Geva & Farnia, 2012; Lesaux, Lipka, & Siegel, 2006). It is conceivable that problems in the lower-level processes of lexical access and syntactic integration also lead to poorer higher-level integration (see e.g., Segers & Verhoeven, 2016).

Previous research on L2 reading comprehension in children provides evidence for the view that low L2 vocabulary hinders higher-level processing in L2 reading comprehension. For example, [Kieffer \(2012\)](#) found that in Spanish-speaking English language learners, vocabulary was a better predictor of reading comprehension attainment than higher-level language skills such as story retell, contrary to findings for L1 peers. In a similar vein, [Droop and Verhoeven \(1998\)](#) found that minority language children could only use larger background knowledge to their advantage in linguistically less complex texts. Similarly, [Burgoyne, Whiteley, and Hutchinson \(2013\)](#) reported that in third grade, children learning English as an additional language showed lower English reading comprehension scores, and that this was not only related to vocabulary, but also to poorer integration of background knowledge with information from the text. These findings are in line with the lexical quality hypothesis, according to which readers are hindered by their smaller vocabularies, which propagates to less effective higher-level integration processes.

For adults, there are also indications that situation model building might be poorer in an L2 than in the L1. For example, [Morishima \(2013\)](#) reported that in contrast to L1 English readers, adult Japanese learners of English failed to detect inconsistencies in a text when there was an intervening sentence between the prior information and the inconsistency. This was ascribed to difficulty in attention allocation when L2 lower-level processing is demanding. Also consistent with this reasoning are the results of an event-related potential study by [Romero-Rivas et al. \(2017\)](#). They showed that, compared to native Spanish readers, L2 Spanish readers (with diverse L1s) showed an extended late negativity when integrating information that was inconsistent with prior world knowledge into the discourse context. This was interpreted as an indication of increased effort for L2 discourse integration.

There is evidence that increased cognitive effort in L2 discourse processing is related to lower lexical quality in the L2. For example, in a longitudinal fMRI study with adult foreign language learners, [Grant, Fang, and Li \(2015\)](#) found that as L2 proficiency grew, cognitive control engagement decreased, and semantic processing increased in a language membership judgement task. This is in line with findings by [Prat and Just \(2010\)](#), who reported that readers with larger vocabularies showed less activation in frontal regions than readers with poorer vocabulary when reading sentences of varying complexity. In all, it seems that L2 reader's situation model building ability might mainly be hampered by lower lexical quality, which leads to an increased need for attentional resource allocation to lower-level linguistic processes.

It has been suggested that, even if L2 lower-level processing is less efficient, L2 readers might be able to (partly) compensate for poorer lower-level L2 processes by increasingly engaging cognitive control, and partly relying on L1 knowledge to compensate for L2 knowledge gaps ([Hacquebord, 1999](#); [Kim & Clariana, 2015](#); [Welie, Schoonen, & Kuiken, 2017](#)). For example, [Welie, et al. \(2017\)](#) found that eight-grade bilingual students did not show lower text-structure inferring skills than their monolingual peers, despite lower scores on basic language measures and on a standardized reading comprehension test. They argued that a compensatory mechanism might be at play, with bilinguals recruiting more other resources such as general reasoning ability and executive control. Furthermore, [Kim and Clariana \(2015\)](#) showed that when L2 readers were allowed to use their L1 in an L2 reading task, the resulting mental model of the text in the L2 was improved compared to a condition in which only the L2 was used. This suggests that L2 readers might improve their comprehension by engaging L1 resources. L1 engagement, however, is a process that is likely to demand additional cognitive resources ([Li & Clariana, 2018](#)). This raises the question what the impact of higher-order integration abilities is on reading comprehension outcomes in children reading in their L2.

1.3. Assessment of textbase memory and situation model building

Research on reading comprehension usually investigates either the predictive skills that contribute to reading comprehension, or the component processes that result in the final mental representation ([Kendeou et al., 2016](#)). We argue that assessing the outcome of processes at the text level, such as textbase memory and situation model quality, also provides information about component skills, which can then be related to performance on standardized reading comprehension tests. Unfortunately, many of the measures traditionally used to assess the mental representation of text add considerable cognitive load to reading (such as think-aloud protocols), or rely on additional production skills (such as summary writing or retelling). Constructing a coherent representation of a whole text is a taxing task – additional processing demands during assessment could obfuscate the process. Other measures make use of very short texts (such as in self-paced reading or event-related potential studies). Assessing shorter stretches of discourse might not be representative of the representation that is built of longer stretches of text, because of the higher cognitive load of processing longer text. Both the addition of cognitive demands and the use of too short texts are problematic for the assessment of children, for whom higher-level processing is a skill still in development.

A task that offers global text-level assessment with minimal additional cognitive and linguistic demands is thus needed. Building on the assumption that the mental model of a text consists of concepts and the relationships between them ([Rapp et al., 2007](#); [van den Broek & Helder, 2017](#); [Zwaan, 2015](#)), a promising method is asking readers how related they think concepts from a text are, after reading the text. The relationship judgments can be translated to proximity data, and scaled by a pathfinder algorithm. The pathfinder algorithm searches for the shortest paths between concepts, which reflect salient concept associations ([Schvaneveldt, Durso, & Dearholt, 1989](#)). The resulting network representation can be thought of as an approximation of the reader's mental model of a text. The network can then be compared to a referent network ([Clariana, 2010](#); [Clariana & Wallace, 2009](#)). Assuming that expert readers converge in their mental models of a text, comparing a student's network to the expert reference provides information about the quality of the mental model. The mental representation that experienced readers build of a text can be tapped by deriving an average pathfinder network from expert readers' relationship judgements of concepts in the text. Similarity of a student's network to the average expert network can be quantified by dividing the number of links in common between the two networks by the total number of links of both networks combined (intersection divided by the union, [Clariana & Wallace, 2009](#)). Similarity calculated in this way captures how many conceptual relations from the text are extracted by expert readers and students.

Previous research has successfully applied the pathfinder network approach to investigate reading comprehension. For example, Davis and Guthrie (2015) showed that similarity between high school students' pathfinder networks and expert networks was correlated with a standardized reading comprehension measure. Similarity of student and expert networks of concepts in a text has also been found to reflect text coherence (Britton & Gülgöz, 1991). Fesel et al. (2015) extended this line of research by investigating different levels of mental text representation: They assessed similarity to a sequential text network (reflecting proximity of terms in a text) as a representation of linear text knowledge, and similarity to a referent expert network as a measure of knowledge network quality, and used both measures to predict reading comprehension of the same text in monolingual Dutch 6th grade readers. They found that similarity to the sequential model predicted reading comprehension, and that similarity to the expert model only predicted reading comprehension in a more complicated hypertext reading condition.

The two similarity measures used by Fesel et al. (2015) map onto theoretically different levels of text comprehension: Similarity to sequential text reflects the propositions that are contained in the text – mapping onto textbase memory –, while similarity to an expert network also captures elaborative inferencing and integration of knowledge across text sections, mapping onto situation model building. While Fesel et al. (2015) used pathfinder network measures to predict scores on reading comprehension questions on the same text, it should also be possible to relate children's ability to remember the textbase, and to build a situation model of the text, to their general reading comprehension outcomes. Assessing children's text representation, relating this to comprehension as measured on a different text, and controlling for other cognitive and linguistic skills, would allow to better disentangle the contribution of abilities at different levels of comprehension.

1.4. Current study

In the current study, we employed a pathfinder network approach to measure textbase memory and situation model building ability in a sample of linguistically diverse 4th grade students in the Netherlands. Participants were also assessed on nonverbal reasoning, working memory, decoding efficiency, vocabulary, grammar, and general reading comprehension. The standardized general reading comprehension outcome measure was unrelated to the materials used to assess textbase memory and situation model building ability. This allowed us to address the following questions:

- 1.) Do L1 and L2 readers in primary school differ in their textbase memory and situation model building ability, in addition to differences in vocabulary, grammar, and general reading comprehension?
- 2.) Do textbase memory and situation model building ability predict general reading comprehension to the same extent in L1 and L2 readers, over and above nonverbal reasoning, working memory, decoding, vocabulary, and grammar?

Concerning the first research question, we expected L2 learners to score lower than L1 learners on vocabulary, grammar, and general reading comprehension, in line with previous research. Regarding textbase memory and situation model building ability, on the one hand, it could be expected that L2 learners score lower on these measures too, as lower-level linguistic processing propagate higher-level integration processes. On the other hand, L2 readers may rely on L1 knowledge and more cognitive control, leading to similar levels of higher-level processing. With respect to the second research question, we expected textbase memory and situation model building ability to predict general reading comprehension to the same extent in L1 and L2 readers, on top of cognitive and linguistic factors.

2. Method

2.1. Participants

The participants were 178 4th-grade children in the Netherlands (90 girls, 88 boys; mean age 10 years, SD = 5 months). In the seven participating schools (ten classrooms), all 4th-grade children without language or reading disorders who were born in the Netherlands took part in the study, unless their parents or guardians declined consent. The current sample is identical to the sample in Raudszus et al. (2018). For 76 participants, Dutch was the only language they used (L1 group); 102 used Dutch and another language (L2 group). The L2 children had diverse language backgrounds. Languages represented most in the bilingual group were Turkish (21%), Moroccan Arabic (19%), and Berber (15%). Participants' bilingualism status was assessed by means of a language background questionnaire. Children were asked if they knew any language other than Dutch, and if so, what language(s) they spoke with their mother, father, siblings, friends, and extended family, respectively. Answers were coded as 1 'only Dutch', 2 'mostly Dutch, sometimes other language', 3 'mostly other language, sometimes Dutch', 4 'only other language'. Any answer other than 'only Dutch' on the questionnaire meant the child was categorized as bilingual (following Cremer & Schoonen, 2013; Melby-Lervåg & Lervåg, 2014).

2.2. Materials

2.2.1. Standardized measures of precursors of reading comprehension

Nonverbal reasoning. Nonverbal reasoning was assessed by the Raven Standard Progressive Matrices (Raven, 2006). In this task, children were shown 60 patterns with a missing piece, and were asked to indicate which puzzle piece would complete the pattern. Patterns were grouped into five sets of increasing difficulty. Each correct answer was awarded one point, resulting in a maximum of 60 with higher scores indicating better nonverbal reasoning ability. Test-retest reliability is reported to be 0.88 for 13-year-olds

(Raven, 2006).

Working memory. Working memory was assessed by the backward digit span of the Wechsler Intelligence Scale for Children III, Dutch version (WISC-III-NL, Kort et al., 2005). The experimenter named sequences of digits that the child had to repeat in reverse order, starting with a sequence of two digits. Every two trials, the length of the sequence was increased by one digit, until the child failed to correctly reverse two sequences of the same length. Each sequence that was repeated backwards correctly was awarded one point, with a maximum of 16 points. Cronbach's α and split-half reliability indices range from 0.50 to 0.66 (Kort et al., 2005).

Decoding. Word decoding was assessed by the Eén-MinuuT-Test [One Minute Test] Version B (Brus & Voeten, 1999). In this task, children were given one minute to read aloud as many Dutch words as possible, which increased in orthographic complexity. The score on the task was the number of words read correctly within one minute, with a maximum of 116. Parallel test reliability ranges from 0.89 to 0.92 (Brus & Voeten, 1999).

Pseudoword decoding skills were assessed by the Klepel Version B (van den Bos, Lutje Spelberg, Scheepstra, & De Vries, 1994), in which the child was given two minutes to read out as many pseudowords as possible, with the pseudowords increasing in complexity. The score was the number of pseudowords read correctly in two minutes, with a maximum score of 116. Parallel test reliability is 0.89 (van den Bos et al., 1994).

For subsequent analyses, word and pseudoword decoding scores were standardized and averaged into one 'decoding' score.

Vocabulary. Vocabulary depth in Dutch was assessed by the Word Definition Task of the Taaltoets Allochtone Kinderen Bovenbouw [Language Test for Minority Children Grades 4–6] (Verhoeven & Vermeer, 1993). In this task, the child was asked to define words that the experimenter named. Based on the scoring guidelines, complete formal definitions were awarded 2 points, and functional definitions were awarded 1 point. The test had 25 items, resulting in a maximum score of 50. Cronbach's α for this test is 0.90 (Verhoeven & Vermeer, 1996).

Dutch vocabulary breadth was assessed by the Peabody Picture Vocabulary Test (PPVT-III-NL, Schlichting, 2005). As the experimenter named a word, the child was shown four pictures and had to point to the one depicting the word. Items were grouped in sets of 12, and after 9 incorrect answers in one set, the task was discontinued. Each correct item was awarded one point. In accordance with the testing manual, assessment started with a particular set. Lower-numbered sets were assumed to be known and given a full score. Because we expected the children in our sample to have relatively small vocabularies, assessment was started with set 6 for all participants, originally intended for ages 5;6-6;5. Reliability (lambda2-coefficient) ranges between 0.96 and 0.95 for children between 9 and 11 years old (Schlichting, 2005).

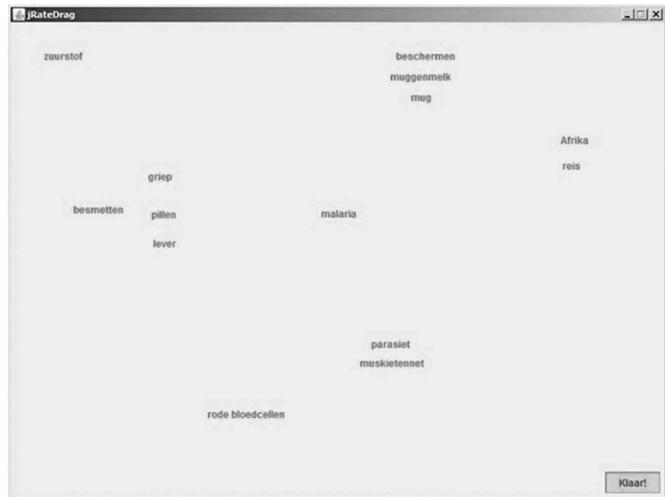
For subsequent analyses, one 'vocabulary' score was computed by standardizing and averaging the vocabulary depth and breadth scores.

Grammar. Knowledge of Dutch grammar was assessed by the Sentence Reading Task of the Taaltoets Allochtone Kinderen Bovenbouw [Language Test for Minority Children Grades 4–6] (Verhoeven & Vermeer, 1993). In this task, children read sets of three sentences and had to indicate which sentence was incorrect, or whether all sentences were correct. Incorrect sentences contained violations of difficult features of Dutch grammar, such as verbal morphology, word order, and gender agreement. Each correct answer was awarded one point, with a maximum of 40 points. Cronbach's α for this test is 0.86 (Verhoeven & Vermeer, 1996).

General reading comprehension. General reading comprehension was assessed by the Text Reading Task of the Taaltoets Allochtone Kinderen Bovenbouw [Language Test for Minority Children Grades 4–6] (Verhoeven & Vermeer, 1993). Children read two expository texts of approximately 300 words, which contained gaps. For each gap, children were asked which of three options fit the gap best. In order to choose the correct answer, children had to combine information from the surrounding context and world knowledge. Each correct answer was awarded one point, with a maximum of 40 points. Cronbach's α for this test is 0.75 (Verhoeven & Vermeer, 1996).

2.2.2. Pathfinder networks as a measure of text-level abilities

Pathfinder networks. To assess children's text-level abilities, a pathfinder network approach was used (Clariana & Wallace, 2009). This approach is based on the assumption that knowledge consists of associations between concepts, and that this knowledge structure can be elicited by similarity ratings (Schvaneveldt et al., 1989). We collected similarity ratings using a computerized sorting task (jRateDrag v.2.0, Schuelke (n.d.)). Children first read a 500-word text on an unfamiliar topic (malaria), taken from the materials of de Leeuw, Segers, and Verhoeven (2014), who adapted the text from Bouckaert (2007). After participants read the text, the text was taken away. Participants were then instructed that they would see terms on a computer screen, and that the task was to 'drag words that are more related closer together and words that are less related further apart'. The terms were 14 core concepts from the text selected by the first author. Children could try out the task with three practice terms not included in the text, to get used to the interface. After the practice, they completed the actual sorting task on their own: Fourteen core terms from the malaria text written appeared distributed over the task window randomly, and were to be sorted. To reduce cognitive load, terms that were moved already turned from black to green. Children could not refer back to the text during the sorting task. When the child said they were ready, the experimenter asked whether they were happy with the location of all terms on the screen, and closed the interface. To ensure children had read the text, they were asked five yes-/no-questions about the contents. Monolinguals and bilinguals did not differ on this task ($M_{L1} = 3.7$, $SD_{L1} = 0.9$; $M_{L2} = 3.5$, $SD_{L2} = 1.0$; $p = .34$, $d = 0.14$). For each child, a similarity matrix was then computed from pixel distances in the final sorted map, containing one distance value for each pair of terms in the task. Similarity matrices were transformed into network structures by a pathfinder scaling algorithm as implemented in JPathfinder v 1.0, with Minkowski's r set to 3 and $q = n-1$. These parameter settings were found to result in networks which capture both local relations and more global inferences (Tang & Clariana, 2017b; Taricani & Clariana, 2006). Fig. 1 shows an example sorting task screen, proximity matrix, and pathfinder network for one participant.



| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|-----------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----|
| 1. malaria | - | | | | | | | | | | | | | |
| 2. mosquito | 99 | - | | | | | | | | | | | | |
| 3. parasite | 71 | 30 | - | | | | | | | | | | | |
| 4. mosquito net | 320 | 340 | 340 | - | | | | | | | | | | |
| 5. oxygen | 210 | 285 | 267 | 200 | - | | | | | | | | | |
| 6. mosquito repellent | 363 | 374 | 377 | 50 | 251 | - | | | | | | | | |
| 7. red blood cells | 254 | 332 | 313 | 215 | 47 | 264 | - | | | | | | | |
| 8. Africa | 236 | 197 | 196 | 536 | 446 | 572 | 488 | - | | | | | | |
| 9. infect | 36 | 136 | 107 | 321 | 189 | 367 | 230 | 258 | - | | | | | |
| 10. liver | 305 | 373 | 358 | 180 | 97 | 223 | 66 | 542 | 286 | - | | | | |
| 11. travel | 219 | 200 | 192 | 530 | 425 | 569 | 465 | 44 | 236 | 522 | - | | | |
| 12. protect | 284 | 293 | 296 | 54 | 203 | 81 | 229 | 490 | 290 | 208 | 487 | - | | |
| 13. flu | 79 | 170 | 146 | 278 | 134 | 326 | 176 | 312 | 54 | 231 | 290 | 253 | - | |
| 14. pills | 202 | 215 | 215 | 125 | 165 | 163 | 203 | 411 | 210 | 208 | 406 | 82 | 178 | - |

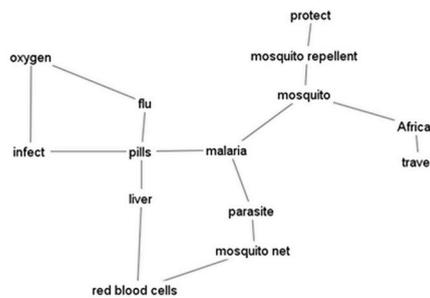


Fig. 1. Example sorting task and pathfinder network from one participant. From top to bottom: Terms as sorted by the participant, proximity matrix (distance in pixels between terms), and pathfinder network ($r = 3, q = n-1$) derived from the proximity matrix. Figure is available under <https://doi.org/10.6084/m9.figshare.5437432.v2> under CC license.

Textbase memory. To assess children's textbase memory, a sequential model of the terms as they appear in the text was created, following the approach described in Clariana (2010). All terms used in the sorting task were extracted from the text in the order in which they appeared in the text. Adjacent words received a value of 1 in a proximity matrix, non-adjacent terms a value of 0, and a pathfinder network was computed from the matrix ($r = 3$ and $q = n-1$). Fig. 2 shows the sequential pathfinder network. For each child, the similarity of their pathfinder network to the sequential network was calculated. Network similarity was defined as the number of links in common between the networks divided by the total number of links of both networks combined (intersection divided by the union, Clariana & Wallace, 2009). The sequential text model has 21 links. Thus, by way of illustration, if a child's model has 16 links, of which 9 are the same as in the sequential text model, this would result in a similarity of $9/(21 + 16) = 0.24$. Similarity calculated in this way can range between 0 and 1, with higher values indicating better textbase memory.

Situation model building ability. To assess situation model building ability, children's networks were compared to a referent expert network (following, e.g., Fesel et al., 2015). For the referent network, ten literacy researchers from our department completed the same task as the children, and average distances were calculated from their sorting task data. The pathfinder network derived

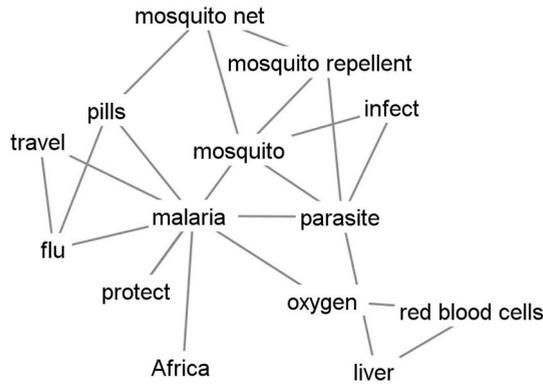


Fig. 2. Sequential model. Similarity to this model was used to assess textbase memory. Figure is available under <https://doi.org/10.6084/m9.figshare.5437462.v2> under CC license.

from the expert average ($r = 3, q = n-1$) is shown in Fig. 3. The expert network clearly depicts the gist of the text, which can be paraphrased as follows (core concepts italicized): *Malaria infection* is spread by *mosquitoes*, which carry a *parasite*. The parasite goes into the *liver*, where it affects *red blood cells*, so they cannot transport *oxygen* well anymore. To protect oneself against malaria, one can use *mosquito repellent*, *mosquito nets*, and *pills*. Malaria symptoms resemble the *flu*, and *malaria* is a risk when *travelling* to regions of *Africa*. This indicates that pathfinder networks derived from relationship judgements do indeed capture meaningful aspects of the text. To further check the validity of the referent network, we estimated the correlations (square root of similarity) between each expert's pathfinder network and the referent network. Correlations ranged from $r = .59-.71$, which is comparable to the values found by Fesel et al. (2015). The similarity between each child's network and the referent expert network was computed by calculating the intersection of networks divided by the union, resulting in a value between 0 and 1, with higher similarity indicating better situation model building ability. For example, the average expert network has 20 links. If a child's network has 16 links, of which 9 are the

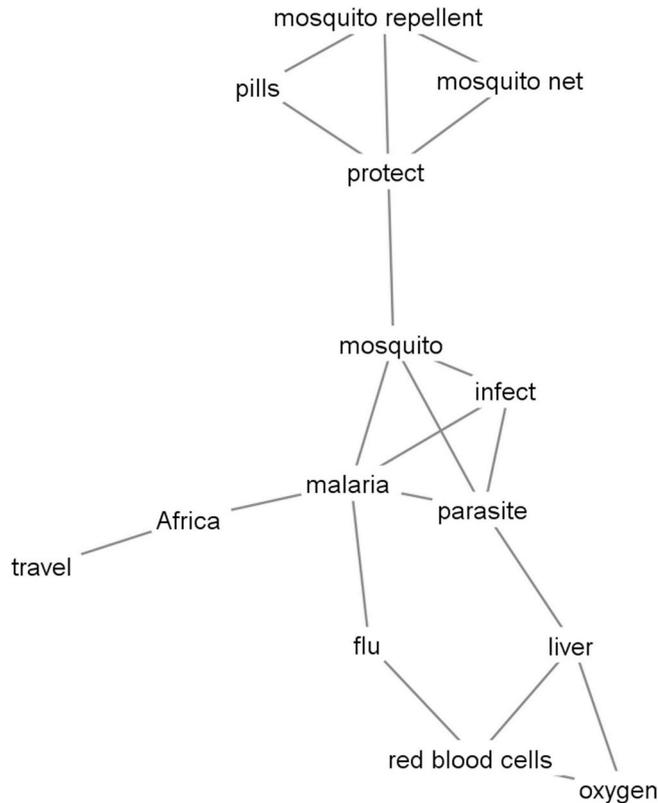


Fig. 3. Average expert pathfinder network. This model was used as a benchmark to assess situation model building ability. Figure is available under <https://doi.org/10.6084/m9.figshare.5437465.v2> under CC license.

same as in the expert model, this would result in a similarity of $9/(20 + 16) = 0.25$.

2.3. Procedure

After the study was approved by the Ethics Committee of our university, the first author contacted schools with a high population of bilingual students. Seven schools agreed to participate. They received information about the findings and individual results in exchange for their collaboration. Parents or guardians of the participant received a letter about the study, and could decline the participation of their child.

Children were tested in the second half of 4th grade. The nonverbal reasoning task was administered in a classroom session of 40 min. The general reading comprehension and grammar tasks were administered in one classroom session of two times 30 min with a break in between. All other tasks were administered in two individual sessions of 30–45 min each. The individual sessions also included two computerized reading tasks used for another study. All tests were conducted by the first author and seven trained undergraduate students of Educational Science.

2.4. Analyses

Nine children (three monolingual, six bilingual) had missing scores on the nonverbal reasoning task due to absence during the classroom session. Five of these children (two bilingual, four monolingual) also had missing scores due to absence during the grammar and general reading comprehension assessment. Missing values were deleted casewise for t-tests and listwise for the regression analyses. All statistical analyses were conducted with R (version 3.4.1, R Core Team, 2017). Differences between L1 and L2 readers were investigated by means of independent samples t-tests. The contribution of text-level abilities to general reading comprehension was investigated by means of hierarchical regression and commonality analysis. Model assumptions were tested using the *gvlma* package version 1.0.0.2 (Peña & Slate, 2006). The commonality analysis was carried out using *Nimon, Oswald, and Roberts'* (2013) *yhat* package (version 2.0–0). The visualization of commonality coefficients was inspired by *Lex and Gehlenborg* (2014).

3. Results

3.1. Differences between L1 and L2 readers

To investigate mean differences between monolingual and bilingual readers, independent sample t-tests were conducted. Table 1 shows the descriptive statistics for all quantitative measures included in this study. L2 readers did not differ from L1 readers with respect to nonverbal reasoning ($p = .07$, $d = 0.29$) and working memory ($p = .96$, $d = 0.01$). L2 readers were more efficient than L1 readers at word decoding ($p < .01$, $d = -0.45$) and pseudoword decoding ($p < .001$, $d = -0.57$), but L2 readers had lower scores than L1 readers on vocabulary depth ($p < .001$, $d = 0.62$), vocabulary breadth ($p < .001$, $d = 0.67$), grammar ($p < .05$, $d = 0.31$), and general reading comprehension ($p < .01$, $d = 0.47$). L1 and L2 readers did not differ significantly in their textbase memory ($p = .85$, $d = 0.03$) or situation model building ability ($p = .24$, $d = 0.18$).

3.2. Textbase memory and situation model building ability as predictor of general reading comprehension

Bivariate correlations of all variables included in subsequent analyses are shown in Table 2, with correlations for the L1 group below the diagonal, and correlations for the L2 group above the diagonal. For both groups, general reading comprehension was positively correlated with vocabulary, grammar, textbase memory, and situation model building ability. For the L1 group, nonverbal reasoning and working memory were also positively correlated with general reading comprehension, whereas in the L2 group, a positive correlation between general reading comprehension and decoding was found.

A hierarchical multiple regression analysis was conducted to evaluate the contribution of situation model building ability to

Table 1
Mean scores on cognitive and linguistic measures for L1 and L2 children.

| | L1 | | | L2 | | |
|-----------------------------|----------|------------------------|---------|----------|------------------------|---------|
| | <i>n</i> | <i>M</i> (<i>SD</i>) | Range | <i>n</i> | <i>M</i> (<i>SD</i>) | Range |
| Nonverbal reasoning | 73 | 39.9 (7.2) | 20–53 | 96 | 37.9 (7.1) | 20–57 |
| Working memory | 76 | 4.3 (1.1) | 2–7 | 102 | 4.3 (1.4) | 1–8 |
| Word decoding | 76 | 61.9 (11.3) | 37–87 | 102 | 67.7 (13.8) | 34–102 |
| Pseudoword decoding | 76 | 50.4 (14.1) | 19–84 | 102 | 59.2 (16.4) | 22–105 |
| Vocabulary depth | 76 | 24.7 (7.0) | 9–39 | 102 | 20.1 (7.6) | 4–37 |
| Vocabulary breadth | 76 | 116.7 (10.2) | 95–144 | 102 | 109.8 (10.5) | 81–135 |
| Grammar | 75 | 21.0 (6.7) | 6–35 | 98 | 19.0 (6.0) | 4–31 |
| Textbase memory | 76 | .14 (.05) | .06–0.3 | 102 | .14 (.06) | .03–.30 |
| Sit. model building ability | 76 | .21 (.08) | .07–.42 | 102 | .20 (.08) | .03–.42 |
| Reading comprehension | 75 | 25.6 (5.8) | 9–36 | 98 | 23.0 (5.0) | 7–33 |

Table 2

Correlations between variables used in the regression analysis for L1 (n = 75–76) and L2 (n = 98–102).

| | L2 | | | | | | | |
|--------------------------------|--------|--------|-------|--------|--------|--------|--------|--------|
| | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
| 1. Nonverbal reasoning | – | .10 | -.14 | .19 | .31** | .12 | .23* | .14 |
| 2. Working memory | .45*** | – | .25* | .08 | .35*** | -.17 | .02 | .13 |
| 3. Decoding | .17 | .18 | – | .03 | .27** | .07 | .04 | .30** |
| 4. Vocabulary | .50*** | .27* | .04 | – | .33*** | .29** | .30** | .50*** |
| 5. Grammar | .55*** | .45*** | .37** | .44*** | – | .18 | .29** | .54*** |
| 6. Textbase memory | .19 | .15 | -.13 | .18 | .23* | – | .36*** | .30** |
| 7. Sit. model building ability | .27* | .12 | .10 | .26* | .34** | .46*** | – | .34*** |
| 8. Reading comprehension | .51*** | .28* | .20 | .61*** | .62*** | .24* | .34*** | – |

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

general reading comprehension (see Table 3). In the first step, bilingualism status was entered as a known predictor of general reading comprehension. In the second step, two general cognitive ability measures, nonverbal reasoning and working memory, were added. In the third step, all linguistic predictors were entered. Measures of higher-level processing were added next, in separate steps to investigate their independent contribution: In the fourth step, textbase memory was added to the model, not contributing additional explained variance. In the fifth step, situation model building ability was added to the model and found to explain a significant amount of extra variance. In the sixth step, the interaction between textbase memory and bilingualism and situation model building ability and bilingualism were not significant predictors and did not explain additional variance. Effects of bilingualism (step 1) and nonverbal reasoning (step 2) disappeared after adding linguistic predictors (step 3). In the final model, vocabulary ($\beta = 0.35$), grammar ($\beta = 0.34$), and situation model building ability ($\beta = 0.26$) emerged as significant predictors. The model's residuals were examined and not found to violate any assumptions, using the procedure described in Peña and Slate (2006) as implemented in Peña and Slate (2014).

3.3. Unique and interrelated contributions to general reading comprehension

While hierarchical regression makes it possible to investigate the unique contribution of a predictor to a model's explained variance, commonality analysis offers a more complete picture of regression effects, giving insight into the interrelations between predictors (Nimon & Reio, 2011). For commonality analysis, first an all possible subset (APS) regression is conducted. From the APS regression, for each predictor, shared and unique variance can be computed. We employed Nimon, Oswald, and Roberts' (2013) yhat package (version 2.0–0) to examine the commonality coefficients of the above regression model. Fig. 4 illustrates the results. Considering unique prediction, vocabulary and grammar explained the largest amount of general reading comprehension variance (8.6% and 7%, respectively), followed by situation model building ability (1.8%). A lot of variance was shared between predictors. Vocabulary explained 22.6% of general reading comprehension variance in concert with other predictors; for grammar this was 26%. The largest shared explained variance was contributed by grammar and vocabulary (3.7%), decoding and grammar (2.2%), and nonverbal reasoning, grammar and vocabulary (2.1%). Situation model building ability had 14% explained variance shared with other predictors, mainly vocabulary, grammar, textbase memory, and nonverbal reasoning. Textbase memory explained virtually no variance uniquely (0.03%), but it contributed 8% of shared variance, mainly with grammar, vocabulary, situation model building ability, and nonverbal reasoning.

4. Discussion

The present study extended previous research on the predictors of reading comprehension, by investigating textbase memory and situation model building ability in 4th grade L1 and L2 readers, and relating these to their general reading comprehension, as assessed by an unrelated standardized measure. We found that assessing children's mental model of a text by deriving pathfinder networks from relationship judgements was a fruitful strategy. Textbase memory and situation model building ability did not differ between L1 and L2 children, while there was an L2 disadvantage for vocabulary, grammar, and general reading comprehension. Situation model building ability, but not textbase memory, predicted general reading comprehension uniquely in both groups, along with vocabulary and grammar. Commonality analysis showed that situation model building explained 1.8% of general reading comprehension variance uniquely and 14% in concert with other predictors. Textbase memory contributed no unique variance, but had 8% of explained variance shared with other predictors. Situation model building ability and textbase memory shared variance with vocabulary, grammar, nonverbal reasoning, and each other.

The children in our sample were similar in their language usage patterns and scores on standardized measures to reports from the previous literature. The L2 children reported to use their L1 most with their parents, both L1 and L2 with their siblings, and mostly L2 with their friends, a pattern that is often found in multilingual homes (Unsworth, 2016). As has been found before, L2 and L1 readers did not differ in working memory and nonverbal reasoning, but L2 readers had weaker target language vocabulary, grammar, and general reading comprehension, while decoding was a relative strength (Melby-Lervåg & Lervåg, 2014). In accordance with the

Table 3
Hierarchical regression analysis predicting reading comprehension (n = 169).

| Predictor | ΔR^2 | B | SE B | β |
|--|--------------|--------|-------|---------|
| Step 1 | | | | |
| Constant | 0.036** | 25.41 | 0.61 | – |
| Bilingualism | | –2.19 | 0.81 | –0.20** |
| Step 2 | | | | |
| Constant | 0.107*** | 14.46 | 2.36 | – |
| Bilingualism | | –1.76 | 0.78 | –0.16* |
| Nonverbal reasoning | | 0.21 | 0.05 | 0.29*** |
| Working memory | | 0.57 | 0.30 | 0.14 |
| Step 3 | | | | |
| Constant | 0.317*** | 17.52 | 1.98 | – |
| Bilingualism | | –0.64 | 0.67 | –0.06 |
| Nonverbal reasoning | | 0.04 | 0.05 | 0.05 |
| Working memory | | –0.22 | 0.26 | –0.05 |
| Decoding | | 0.68 | 0.35 | 0.12 |
| Vocabulary | | 2.38 | 0.40 | 0.39*** |
| Grammar | | 0.33 | 0.06 | 0.39*** |
| Step 4 | | | | |
| Constant | 0.008 | 24.45 | 2.10 | – |
| Bilingualism | | –0.72 | 0.66 | –0.07 |
| Nonverbal reasoning | | 0.03 | 0.05 | 0.04 |
| Working memory | | –0.14 | 0.26 | –0.03 |
| Decoding | | 0.68 | 0.35 | 0.12 |
| Vocabulary | | 2.22 | 0.40 | 0.36*** |
| Grammar | | 0.31 | 0.06 | 0.37*** |
| Textbase memory | | 10.83 | 5.85 | 0.11 |
| Step 5 | | | | |
| Constant | 0.015* | 15.57 | 2.08 | – |
| Bilingualism | | –0.74 | 0.65 | –0.07 |
| Nonverbal reasoning | | 0.02 | 0.05 | 0.03 |
| Working memory | | –0.11 | 0.26 | –0.03 |
| Decoding | | 0.67 | 0.34 | 0.12 |
| Vocabulary | | 2.11 | 0.40 | 0.35*** |
| Grammar | | 0.29 | 0.06 | 0.34*** |
| Textbase memory | | 5.96 | 6.11 | 0.06 |
| Situation model building ability | | 10.06 | 4.17 | 0.15* |
| Step 6 | | | | |
| Constant | –0.000 | 14.58 | 2.29 | – |
| Bilingualism | | 0.96 | 1.98 | 0.09 |
| Nonverbal reasoning | | 0.02 | 0.05 | 0.03 |
| Working memory | | –0.11 | 0.26 | –0.03 |
| Decoding | | 0.64 | 0.35 | 0.12 |
| Vocabulary | | 2.12 | 0.40 | 0.35*** |
| Grammar | | 0.29 | 0.06 | 0.34*** |
| Textbase memory | | 2.54 | 10.18 | 0.03 |
| Situation model building ability | | 17.15 | 6.72 | 0.26* |
| Bilingualism*Textbase memory | | 4.77 | 12.84 | 0.07 |
| Bilingualism*Sit. model building ability | | –11.28 | 8.29 | –0.25 |
| Total adjusted ΔR^2 | 0.483*** | | | |

Note. *** $p < .001$, ** $p < .01$, * $p < .05$.

previous literature, the effect of bilingualism on reading comprehension was mediated by linguistic predictors, and vocabulary and grammar emerged as the most important predictors of reading comprehension (Babayigit, 2015).

Concerning higher-level processing, we found no differences between L1 and L2 readers in their textbase memory or situation model building ability. This is in contrast to several previous studies on other higher-level skills in L2 readers (e.g., Burgoyne et al., 2013; Droop & Verhoeven, 1998; Kieffer, 2012), but in line with the recent study by Welie et al. (2017). Despite an L2 disadvantage in vocabulary, grammar, and a standardized reading comprehension measure (in line with our first hypothesis), L2 readers seem to be able to construct a mental model of a text to the same extent as L1 readers. This is in line with earlier findings suggesting that, when L2 readers are assessed on propositional content rather than on specific lexical or syntactic items, they are likely to score similarly to their L1 peers (Sorenson Duncan, 2017; Verhoeven & Vermeer, 1984). L2 readers might be able to compensate for less efficient processing at lower linguistic levels by increasingly engaging cognitive control, possibly patching gaps in their L2 knowledge by

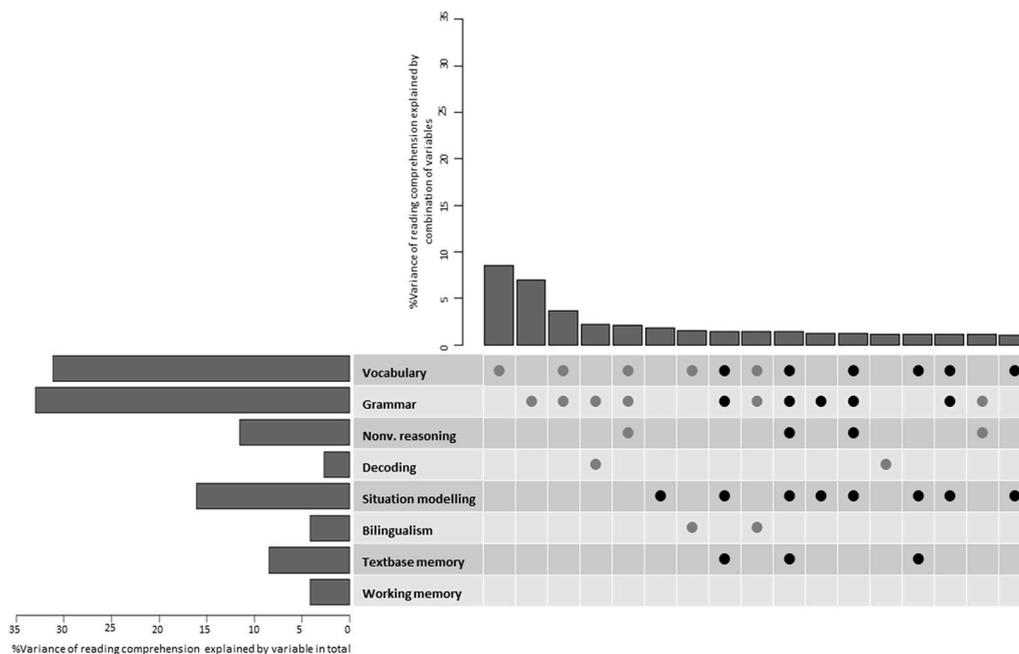


Fig. 4. Unique and shared variance of reading comprehension explained. The upper bar graph shows unique and shared variance explained by the combinations of variables, as indicated by dots in the matrix below. Only sets explaining more than 1% of variance are shown. Sets of variables containing situation model building ability are highlighted in black. The bar chart on the left shows the total amount of variance explained by a variable (unique plus shared).

recruiting L1 conceptual representations (Li & Clariana, 2018). A different explanation is that the expository text we used, while appropriate for the grade level, was quite difficult for the children in our sample. The presence of unfamiliar words might have levelled the playing field, as all children had to infer some words from the text. The same can be said for the unfamiliar topic: Both monolingual and bilingual children had to rely on information from the text, so differences in background knowledge that are often found between monolinguals and bilinguals (Droop & Verhoeven, 1998), are also less likely to play a role. Bilinguals might in fact have an advantage in that situation, as they are more used to making inferences based on incomplete knowledge (Hacquebord, 1999). The majority of bilinguals in our sample was not more likely than L1 participants to have previous knowledge of the text topic, malaria. Only ten participants spoke languages of countries with a high incidence of malaria, and removing those participants from the analysis did not change the pattern of results. Further research is needed to disentangle the effects of L1 and L2 vocabulary, background knowledge, and cognitive control engagement on situation model building ability in monolingual and bilingual readers.

Regarding the contribution of higher-level processing to general reading comprehension outcomes, we found that for both L1 and L2 readers, textbase memory shared explained variance of general reading comprehension with vocabulary, grammar, situation model building ability, and nonverbal reasoning. However, it did not contribute unique variance to general reading comprehension. As Fesel et al. (2015) did find a contribution of textbase memory to reading comprehension, our findings are, at first sight, surprising. However, Fesel et al. (2015) used similarity to a sequential text model to predict comprehension questions on the same text. It is conceivable that remembering the textbase of one text would predict comprehension scores for the same text, but it might not be a skill that predicts reading comprehension in general. This is even more so considering that our measure of textbase memory, like the one in Fesel et al. (2015), compared children's mental model to a sequential model of the text, in which terms occurring close together in the text were linked. In the construction-integration model, the textbase is constructed on the basis of argument overlap rather than proximity alone. It is possible that our text memory measure reflected literal memory of the order of concepts in the text, rather than of propositions and their interrelations. Future research should investigate how the similarity of a student's mental model to a network based on propositional analysis predicts reading comprehension. An even more fine-grained measure might be achieved by using connection strength as predicted by computational models such as the Landscape Model (Tzeng, van den Broek, Kendeou, & Lee, 2005; van den Broek, Young, Tzeng, & Linderholm, 1999) as a basis for the referential model.

Situation model building ability, on the other hand, did predict general reading comprehension, both considering shared variance with other predictors (especially nonverbal reasoning, grammar, vocabulary, and textbase memory), and as a unique predictor. It is important to note that situation model building ability was measured using a different text than the one used in the standardized general reading comprehension assessment, and that the most important known cognitive and linguistic precursors were controlled for. This suggests that situation model building ability is a measurable and transferrable skill, distinct from nonverbal reasoning and language proficiency. Our findings confirm theoretical predictions on the importance of situation model building for reading comprehension (Perfetti & Stafura, 2014; van Dijk & Kintsch, 1983), and extend previous research that has shown inferencing to be relatively independent of linguistic abilities (e.g., Kendeou, Bohn-Gettler, White, & van den Broek, 2008). It remains to be

investigated what constitutes the core of this situation model building ability, i.e., what elements of similarity between a reader's situation model and the referent expert network are most predictive of reading comprehension differences. Representing the relationship of complex clusters of concepts, relying on both linguistic knowledge and cognitive ability, is a likely constituent, considering the shared variance with nonverbal reasoning, vocabulary, and grammar.

With respect to the relative contributions of textbase memory and situation model building ability to general reading comprehension, in line with our expectations, we did not find differences between L1 and L2 readers. After general linguistic and cognitive skills are accounted for, the ability to build up a mental representation of a text seems to be equally important for L1 and L2 readers.

To our knowledge, ours was the one of the first studies to combine a sorting task with a pathfinder algorithm to assess children's situation model building ability. This method has considerable practical and methodological advantages compared to other assessment methods: Children can complete the task fast and intuitively. The task does not put demands on dual processing (as do think-aloud tasks), language production (as do open-ended questions), distractor effects (as do multiple choice tasks), or detailed questions with one single correct answer (as inference questions often do). Even when part of the text was not understood by the reader, or some of the terms in the sorting task were unknown, it was still possible to be credited for the part that was understood. A disadvantage of the method as implemented in this study is that the terms in the sorting task were pre-selected: Children were thus not challenged to decide which terms in the text contain relevant information. On the other hand, this allows for comparison of the networks with an expert reference network. Centrality of the terms in the pathfinder network still gives an indication of the relevance of terms in the participants' mental model. For example, inspection of the average child network shows that many children connected *flu* to all other 'health terms' (*infect, pills, red blood cells*), indicating that many children failed to notice its peripheral role in the text. The experts, on the other hand, linked *flu* to *malaria*, reflecting their understanding that the flu was only an example of something resembling malaria.

At this point, several limitations of the current study need to be acknowledged. Our sample of L2 children was very heterogeneous. Considering that the representation of knowledge structures varies across languages (Kim & Clariana, 2015), future research should investigate whether native language typology has an influence on situation model building in L2 reading. In addition, to further investigate the role of cognitive control and L1 engagement in compensating for L2 knowledge gaps, future research should aim to directly measure the allocation of cognitive resources and the activation of L1 concepts during L2 reading, and relate these to measures of situation model building. Another limitation of our study is that we did not collect relationship judgements to construct a baseline model before children read the text. However, previous research has shown that performing a sorting task can influence the mental model of the topic area (Tang & Clariana, 2017a), so administering the sorting task before reading could have influenced text processing during reading. Finally, we measured all constructs at one time point only, and therefore cannot make any claims on causality. Longitudinal investigations of the contribution of situation model building ability to reading comprehension outcomes would be valuable to extend and validate the current findings, as would the inclusion of both a standardized reading comprehension measure and a traditional measure of comprehension of the text used for the pathfinder task.

In conclusion, a sorting task combined with a pathfinder network approach is useful for assessing text-level processing skills in primary school readers, which allows us to expand our explorations of classical theories of reading comprehension. L1 and L2 readers did not differ in textbase memory nor situation model building ability despite L2 readers' weaker target language skills, indicating that higher-level processes might be a relative strength in L2 readers. However, this was not reflected in a larger contribution of these skills to general reading comprehension in L2 readers. For both L1 and L2 readers, situation model building ability uniquely contributes to general reading comprehension outcomes, over and above grammar and vocabulary. L2 reading comprehension problems seem to originate at the level of L2 linguistic knowledge, rather than through problems in higher-level unification.

Acknowledgements

We would like to thank the research assistants and the schools, staff, parents, and pupils involved in this project. We would also like to thank the reviewers, whose comments helped improve the manuscript. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Declaration of interest: None.

References

- Babayiğit, S. (2015). The relations between word reading, oral language, and reading comprehension in children who speak English as a first (L1) and second language (L2): a multigroup structural analysis. *Reading and Writing: An Interdisciplinary Journal*, 28, 527–544. <https://doi.org/10.1007/s11145-014-9536-x>.
- Bouckaert, G. (2007). *Monsterlijke microben: En hoe ga je ze te lijf? [Monstrous microbes: And how to deal with them]*. Tiel, Belgium: Uitgeverij Lannoo nv.
- Britton, B. K., & Gülgöz, S. (1991). Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology*, 83, 329. <https://doi.org/10.1037/0022-0663.83.3.329>.
- Brus, B. T., & Voeten, M. J. M. (1999). *Een-minuut-test: vorm A en B (EMT)*. Nijmegen, The Netherlands: Berkhout.
- Burgoyne, K., Whiteley, H. E., & Hutchinson, J. M. (2013). The role of background knowledge in text comprehension for children learning English as an additional language. *Journal of Research in Reading*, 36, 132–148. <https://doi.org/10.1111/j.1467-9817.2011.01493.x>.
- Cain, K., & Oakhill, J. (2014). Reading comprehension and vocabulary: Is vocabulary more important for some aspects of comprehension? *L'Année Psychologique*, 114, 647–662. <https://doi.org/10.4074/S0003503314004035>.
- Cain, K., Oakhill, J. V., Barnes, M. A., & Bryant, P. E. (2001). Comprehension skill, inference-making ability, and their relation to knowledge. *Memory & Cognition*, 29, 850–859. <https://doi.org/10.3758/BF03196414>.
- Clariana, R. B. (2010). Deriving individual and group knowledge structure from network diagrams and from essays. In D. Ifenthaler, P. Pirnay-Dummer, & N. M. Seel (Eds.). *Computer-based diagnostics and systematic analysis of knowledge* (pp. 117–130). New York, NY: Springer. https://doi.org/10.1007/978-1-4419-5662-0_7.
- Clariana, R. B., & Wallace, P. E. (2009). A comparison of pair-wise, list-wise, and clustering approaches for eliciting structural knowledge. *International Journal of Instructional Media*, 36(3), 287–302.

- Cremer, M., & Schoonen, R. (2013). The role of accessibility of semantic word knowledge in monolingual and bilingual fifth-grade reading. *Applied Psycholinguistics*, 34, 1195–1217. <https://doi.org/10.1017/S0142716412000203>.
- Currie, N. K., & Cain, K. (2015). Children's inference generation: The role of vocabulary and working memory. *Journal of Experimental Child Psychology*, 137, 57–75. <https://doi.org/10.1016/j.jecp.2015.03.005>.
- Davis, M. H., & Guthrie, J. T. (2015). Measuring reading comprehension of content area texts using an assessment of knowledge organization. *The Journal of Educational Research*, 108, 148–164. <https://doi.org/10.1080/00220671.2013.863749>.
- de Leeuw, L., Segers, E., & Verhoeven, L. (2014). Context, task, and reader effects in children's incidental word learning from text. *International Journal of Disability, Development and Education*, 61, 275–287. <https://doi.org/10.1080/1034912X.2014.934010>.
- Droop, M., & Verhoeven, L. (1998). Background knowledge, linguistic complexity, and second-language reading comprehension. *Journal of Literacy Research*, 30, 253–271. <https://doi.org/10.1080/10862969809547998>.
- Droop, M., & Verhoeven, L. (2003). Language proficiency and reading ability in first-and second-language learners. *Reading Research Quarterly*, 38(1), 78–103. <https://doi.org/10.1598/RRQ.38.1.4>.
- Fesel, S. S., Segers, E., Clariana, R. B., & Verhoeven, L. (2015). Quality of children's knowledge representations in digital text comprehension: Evidence from pathfinder networks. *Computers in Human Behavior*, 48, 135–146. <https://doi.org/10.1016/j.chb.2015.01.014>.
- Geva, E., & Farnia, F. (2012). Developmental changes in the nature of language proficiency and reading fluency paint a more complex view of reading comprehension in ELL and EL1. *Reading and Writing: An Interdisciplinary Journal*, 25, 1819–1845. <https://doi.org/10.1007/s11145-011-9333-8>.
- Grant, A. M., Fang, S. Y., & Li, P. (2015). Second language lexical development and cognitive control: A longitudinal fMRI study. *Brain and Language*, 144, 35–47. <https://doi.org/10.1016/j.bandl.2015.03.010>.
- Hacquebord, H. (1999). A Dutch comprehension test for identifying reading problems in L1 and L2 students (Research Note). *Journal of Research in Reading*, 22, 299–303. <https://doi.org/10.1111/1467-9817.00093>.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9, 416–423. <https://doi.org/10.1016/j.tics.2005.07.004>.
- Hirsch, E. (2003). Reading comprehension requires knowledge-of words and of the world: Scientific insights into the fourth-grade slump and the nation's stagnant comprehension scores. *American Educator*, 27(1), 10–29.
- JPathfinder v1.0, available from: interlinkinc.net.
- 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. <https://doi.org/10.1111/j.1467-9817.2008.00370.x>.
- Kendeou, P., McMaster, K. L., & Christ, T. (2016). Reading comprehension: core components and processes. *Policy Insights from the Behavioral and Brain Sciences*, 3, 62–69. <https://doi.org/10.1177/2372732215624707>.
- Kendeou, P., Van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, 101, 765–778. <https://doi.org/10.1037/a0015956>.
- Kieffer, M. J. (2012). Early oral language and later reading development in Spanish-speaking English language learners: Evidence from a nine-year longitudinal study. *Journal of Applied Developmental Psychology*, 33, 146–157. <https://doi.org/10.1016/j.appdev.2012.02.003>.
- Kim, K., & Clariana, R. B. (2015). Knowledge structure measures of reader's situation models across languages: Translation engenders richer structure. *Technology, Knowledge and Learning*, 20, 249–268. <https://doi.org/10.1007/s10758-015-9246-8>.
- Kintsch, W. (2005). An overview of top-down and bottom-up effects in comprehension: The CI perspective. *Discourse Processes*, 39, 125–128. <https://doi.org/10.1080/0163853X.2005.9651676>.
- Kort, W., Schittekatte, M., Dekker, P. H., Verhaeghe, P., Compaan, E. L., Bosmans, M., et al. (2005). *WISC-III/NL Wechsler Intelligence Scale for Children. David Wechsler. Derde editie NL. Handleiding en verantwoording*. Amsterdam, The Netherlands: Harcourt Test Publishers/Nederlands Instituut voor Psychologen.
- Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2017). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*. <https://doi.org/10.1111/cdev.12861>.
- Li, P., & Clariana, R. B. (2018). Reading comprehension in L1 and L2: An integrative approach. *Journal of Neurolinguistics*. Advance online publication <https://doi.org/10.1016/j.jneuroling.2018.03.005>.
- Lesaux, N. K., Lipka, O., & Siegel, L. S. (2006). Investigating cognitive and linguistic abilities that influence the reading comprehension skills of children from diverse linguistic backgrounds. *Reading and Writing: An Interdisciplinary Journal*, 19, 99–131. <https://doi.org/10.1007/s11145-005-4713-6>.
- Lex, A., & Gehlenborg, N. (2014). Points of view: Sets and intersections. *Nature Methods*, 11, 779–779. <https://doi.org/10.1038/nmeth.3033>.
- Melby-Lervåg, M., & Lervåg, A. (2014). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin*, 140, 409. <https://doi.org/10.1037/a0033890>.
- Morishima, Y. (2013). Allocation of limited cognitive resources during text comprehension in a second language. *Discourse Processes*, 50, 577–597. <https://doi.org/10.1080/0163853X.2013.846964>.
- Nimon, K., Oswald, F., & Roberts, J. K. (2013). *yhat: Interpreting regression effects. R package version 2.0-0*. <https://CRAN.R-project.org/package=yhat>.
- Nimon, K., & Reio, T. G. (2011). Regression commonality analysis: a technique for quantitative theory building. *Human Resource Development Review*, 10, 329–340. <https://doi.org/10.1177/1534484311411077>.
- Peña, E. A., & Slate, E. H. (2006). Global validation of linear model assumptions. *Journal of the American Statistical Association*, 101, 341–354. <https://doi.org/10.1198/016214505000000637>.
- Peña, E. A., & Slate, E. H. (2014). *gvlma: Global Validation of Linear Models Assumptions. R package version 1.0.0.2*. <https://CRAN.R-project.org/package=gvlma>.
- Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, 18, 22–37. <https://doi.org/10.1080/10888438.2013.827687>.
- Prat, C. S., & Just, M. A. (2010). Exploring the neural dynamics underpinning individual differences in sentence comprehension. *Cerebral Cortex*, 21(8), 1747–1760. <https://doi.org/10.1093/cercor/bhq241>.
- 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: An Interdisciplinary Journal*, 25, 1635–1664. <https://doi.org/10.1007/s11145-011-9336-5>.
- R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, 11(4), 289–312. <https://doi.org/10.1080/10888430701530417>.
- Raudszus, H., Segers, E., & Verhoeven, L. (2018). Lexical quality and executive control predict children's first and second language reading comprehension. *Reading and Writing: An Interdisciplinary Journal*, 31, 405–424. <https://doi.org/10.1007/s11145-017-9791-8>.
- Raven, J. C. (2006). *Handleiding RAVEN Standard Progressive Matrices (Nederlandse vertaling)*. USA: Texas: Harcourt Assessment, Inc.
- Romero-Rivas, C., Corey, J. D., Garcia, X., Thierry, G., Martin, C. D., & Costa, A. (2017). World knowledge and novel information integration during L2 speech comprehension. *Bilingualism: Language and Cognition*, 20, 576–587. <https://doi.org/10.1017/S1366728915000905>.
- Schlichting, L. (2005). *Peabody Picture Vocabulary Test-III-NL*. Amsterdam, The Netherlands: Harcourt Assessment BV.
- Schuelke, M. (n.d.). JrateDrag, v2.0. Available from: <http://endaemon.com/>.
- Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. In G. Bower (Vol. Ed.), *The psychology of learning and motivation: Advances in research & theory: Vol. 24*, (pp. 249–284). New York: Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60539-3](https://doi.org/10.1016/S0079-7421(08)60539-3).
- Segers, E., & Verhoeven, L. (2016). How logical reasoning mediates the relation between lexical quality and reading comprehension. *Reading and Writing: An Interdisciplinary Journal*, 29, 577–590. <https://doi.org/10.1007/s11145-015-9613-9>.
- Sorenson Duncan, T. (2017). *How does maternal education influence language acquisition? Interdependencies between environment and input in the bilingual development of immigrant and refugee children* Doctoral dissertation. University of Alberta.
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental*

- Psychology*, 38, 934. <https://doi.org/10.1037/0012-1649.38.6.934>.
- Tang, H., & Clariana, R. (2017a). *The influence of sorting tasks on second language reading comprehension: From the knowledge structure perspective*. Paper presented at the AECT International Convention, Jacksonville, FL.
- Tang, H., & Clariana, R. (2017b). Leveraging a sorting task as a measure of knowledge structure in bilingual settings. *Technology, Knowledge and Learning*, 22(1), 23–35. <https://doi.org/10.1007/s10758-016-9290-z>.
- Taricani, E. M., & Clariana, R. B. (2006). A technique for automatically scoring open-ended concept maps. *Educational Technology Research & Development*, 54, 65–82. <https://doi.org/10.1007/s11423-006-6497-z>.
- Tzeng, Y., van den Broek, P., Kendeou, P., & Lee, C. (2005). The computational implementation of the Landscape Model: Modeling inferential processes and memory representations of text comprehension. *Behavior Research Methods, Instruments, & Computers*, 37, 277–286.
- Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis, & S. Montanari (Eds.). *In Lifespan perspectives on bilingualism* (pp. 136–196). Berlin, Germany: Mouton and De Gruyter. <https://doi.org/10.1515/9783110341249-008>.
- van den Bos, K. P., Lutje Spelberg, H. C., Scheepstra, A. J. M., & De Vries, J. R. (1994). *De Klepel: Vorm A en B. Een test voor de leesvaardigheid van pseudoworden [The Klepel: Form A and B. A test of the ability to read pseudowords]*. Nijmegen, The Netherlands: Berkhout.
- van den Broek, P., & Helder, A. (2017). Cognitive processes in discourse comprehension: passive processes, reader-initiated processes, and evolving mental representations. *Discourse Processes*, 54, 360–372. <https://doi.org/10.1080/0163853X.2017.1306677>.
- van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading. In H. van Oostendorp, & S. R. Goldman (Eds.). *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Erlbaum.
- van Dijk, T. A., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York, NY: Academic Press.
- Verhoeven, L., & Vermeer, A. (1984). Assessment procedures of oral second language proficiency: spontaneous speech vs. elicitation data. *Interlanguage Studies Bulletin*, 50–68.
- Verhoeven, L., & Vermeer, A. (1993). *Taaltoets Allochtone Kinderen-Bovenbouw [Language Test for Minority Children-Grades 4-6]*. Tilburg, The Netherlands: Zwijsen.
- Verhoeven, L., & Vermeer, A. (1996). *Taalvaardigheid in de bovenbouw [Language ability in the upper grades of primary school]*. Tilburg, The Netherlands: Tilburg University Press.
- van Viersen, S., de Bree, E. H., Zee, M., Maassen, B., van der Leij, A., & de Jong, P. F. (2018). Pathways into literacy: The role of early oral language abilities and family risk for dyslexia. *Psychological Science*. <https://doi.org/10.1177/0956797617736886>.
- Welie, C., Schoonen, R., & Kuiken, F. (2017). The role text structure inference skill plays for eighth graders' expository text comprehension. *Reading and Writing: An Interdisciplinary Journal*, 1–30. <https://doi.org/10.1007/s11145-017-9801-x>.
- Zwaan, R. A. (2015). Situation models, mental simulations, and abstract concepts in discourse comprehension. *Psychonomic Bulletin & Review*, 23, 1028–1034. <https://doi.org/10.3758/s13423-015-0864-x>.