ADDRESSING READING COMPREHENSION IN ELEMENTARY-SCHOOL READERS: HOW DO DECODING AND ORAL LANGUAGE SKILLS PREDICT PERFORMANCE ON TWO MEASURES, AND ARE THE SAME STUDENTS IDENTIFIED AS HAVING A COMPREHENSION DIFFICULTY?

by

Sarah Campbell

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in School Psychology

at

Mount Saint Vincent University
Halifax, Nova Scotia
September 2021

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

Reading Comprehension draws on multiple skills and knowledge but is frequently assessed as if it were a single skill. There are questions about the validity and differences between measures, with few studies addressing these issues directly. In first through third-grade students, pseudoword decoding accounted for similar amounts of variance in two measures of reading comprehension, while oral language skills contributed more to a multiple-choice test than to a cloze test. The prediction by each oral language skill was similar across grades, except for syntactic awareness which predicted more variance in cloze test scores with each increasing grade. Of students identified as below the 16th and 25th percentile by the multiple-choice test, 34.2% and 36.9% were not identified by the cloze test. This highlights inconsistencies between who is identified between these two measures. Based on these results and previous findings, I propose a change in the assessment of reading comprehension.

Keywords: reading comprehension, oral language, decoding, measurement, identification, early elementary

Introduction

The ability to comprehend text is the primary goal of reading and it is positively correlated with academic achievement (e.g., Elosúa et al., 2012). Reading comprehension is a complex cognitive activity that draws on multiple skills and knowledge. However, it is often studied, assessed, and taught as if it were a single skill or ability (e.g., Cain & Bignell, 2014, Catts et al., 2008; Clark & Kamhi, 2014). The complex nature of reading comprehension makes it difficult to assess, as it is hard to have a single measure that adequately draws upon all that comprehension entails (Cain, 2016; Francis, et al., 2005; Kintsch & van Dijk, 1978). One goal has been to accurately assess and detect reading comprehension difficulties in children, as this is important to facilitate early and effective intervention (Cain & Oakhill, 2006).

In this study, I provide an overview of several key areas of research related to reading comprehension, that provides a context for my narrower research questions. First, I briefly introduce several theoretical models of reading comprehension. Next, I discuss important factors that contribute to a reader's ability to comprehend text, such as language skills and background knowledge. A further section will discuss the cognitive strategies that have been found to help improve reading comprehension. I then narrow in on current methods that are used to measure reading comprehension. My research examines the measurement of reading comprehension in young elementary-school children. The breadth of topics briefly reviewed will contribute to illuminating the complexity of reading comprehension and set the context for my examination of the consistency and breadth of the two measures examined in this study.

Theoretical Models of Reading Comprehension

The Simple View of Reading states that reading comprehension is the product of two general constructs or skill sets: word-level decoding and language comprehension (Gough &

Tunmer, 1986; Hoover & Gough, 1990). Decoding is defined broadly, as the ability to recognize printed words accurately and quickly in order to gain access to the meaning of words contained in the text (Hoover & Tunmer, 2018). Language comprehension refers to the ability to extract and construct both literal and inferred meanings from language (Lonigan et al., 2018). Language comprehension involves a broad range of oral language skills such as one's knowledge of vocabulary and syntax (Lonigan et al., 2018). Despite its name, The Simple View of Reading does not suggest that the reading process is simplistic, rather than the complex process of reading can be organized into two parts, each important and neither one sufficient on its own (Gough & Tunmer, 1986; Hoover & Gough, 1990).

For skilled reading comprehension, both decoding and language comprehension must be sufficiently developed. This suggests that reading comprehension will be impaired for anyone who has difficulty in either of these two broadly defined cognitive activities (Hoover & Tummer, 2018). The Simple View of Reading also proposes that developmentally, word reading is the primary source of individual differences for young readers. Language comprehension becomes more dominant across development which was confirmed in one meta-analysis (Garcia & Cain, 2014). In general, for a younger child who is learning to decode, reading comprehension is limited by word recognition skills. Due to limited ability to decode printed words, at this point in development, a child is better able to comprehend language through speech than through text. As skills in decoding improve, these constraints are lessened, and the reader's comprehension of spoken and written texts becomes more similar. Language comprehension thus becomes the limiting factor for successful reading comprehension with continued development. As a caveat, Garcia and Cain (2014) also found that the way decoding and reading comprehension are assessed in any given study impacts the observed relationship between these constructs over time.

Catts and his colleagues' (2006) findings supported the notion that the two central constructs in the Simple View of Reading determine one's ability to comprehend text. In their study, Grade 8 students who were poor comprehenders (with adequate word reading) were compared to those who were poor decoders, as well as to typically achieving readers. The poor comprehenders had deficits in oral language comprehension and average abilities in phonological processing. Poor decoders had deficits in phonological processing and average language comprehension abilities. Additionally, this same pattern of findings was observed in their retrospective analysis of the students in Kindergarten, Grade 2, and Grade 4. Catts and colleagues (2006) thus argued that these skill sets function largely independently of one another, as predicted by the Simple View of Reading.

The Reading Systems Framework (Perfetti & Stafura, 2014) is another model that informs current research on the development of reading comprehension. One principle of the Reading Systems Framework is that reading builds on three key sources of knowledge: orthographic knowledge, linguistic knowledge, and general knowledge (Raudszus et al., 2019). Orthographic knowledge refers to internal representations of the written form of spoken words that allow quick and automatic word reading (Apel, 2011). This contributes to successful reading comprehension. Linguistic knowledge refers to one's knowledge of aspects of the language system (e.g., phonology, vocabulary, syntax, and morphology). Phonology and orthographic knowledge are key in developing and supporting skilled word reading. Knowing the meanings of individual words, vocabulary knowledge is important to understand sentences and the meaning of the text as a whole. Some researchers propose that vocabulary knowledge is a central driver of reading comprehension, even in younger children (Spencer et al., 201 memory, integration of this knowledge within the context of the text, and inhibitory control (Perfetti & Stafura, 2014).

The Reading Systems Framework incorporates aspects of model-construction theories, originating with Kintsch's work (e.g., Kintsch, 1988). These theories posit three levels of text representation: surface, textbase, and situation models (Kintsch, 1988; Zwaan & Radvansky, 1998). The surface representation is made up of words and clauses that appear in the text. Building an accurate surface representation of a text is needed to create an accurate representation at the textbase and situation level. If a reader struggles with aspects of oral language such as syntactic awareness, morphological awareness, or vocabulary, it will interfere with their ability to create an accurate representation at the textbase and situation levels. The textbase model involves a representation of the propositions and the relationships among these in a text. A proposition is a basic unit of meaning, made up of a single relational term and the concepts and arguments that it relates (Keenan & Brown, 1984). To then build a working model, or what is called a situation model of the text, the reader must fully integrate the concepts and propositions activated by the text with their prior knowledge of these and related concepts (van Dijk & Kintsch, 1983). Prior to van Dijk and Kintsch's (1983) description of situation models, text comprehension was largely viewed as the creation of mental representations of the text itself, rather than a representation of the situation described by the text. van Dijk and Kintsch (1983) did not abandon the concept of mental representations of a text, rather they proposed that readers construct both text-based representations and situation models while reading (Zwaan & Radvansky, 1998).

During text comprehension, readers build a situation model by monitoring several key aspects of the text such as protagonists, time, space, causation, and intentionality (Zwaan et al., 1995a; Zwaan et al., 1995b). By monitoring and integrating information from key dimensions of the text, readers gradually build and update their representation of the situation being explained in the text and put together a situation model. In addition to the information in the text, readers extend their situation models by incorporating their own knowledge of the world. This level of text representation is proposed to be associated with deep processing or better comprehension as the reader is able to go beyond the textual information to develop a more in-depth understanding of what the text is about (Schoot et al., 2010).

Research has shown that task demands can influence the extent to which readers are able to construct situation models while reading (Stine-Morrow et al., 2001; Zwaan et al., 1995). Zwaan and colleagues (1995) examined the extent to which readers monitored the temporal, spatial, and causal dimensions of a story under two conditions. In a “normal reading” condition, participants were told to read the story as they would normally read a short story for pleasure. In a second condition, “memory focused,” participants were instructed to read so that they would be able to give a detailed account of what happened in the story. The researchers hypothesized that this condition would cause readers to focus more on the text itself, rather than on the situation described within the text. Findings confirmed their hypothesis, indicating that the instruction to read for memory was disruptive to the monitoring of situational dimensions in the story. Thus, the task at hand can, like the memory-focused condition, affect one’s ability to construct an in-depth understanding of the text (Zwaan et al. 1995).

Researchers have also investigated the types of instruction that can encourage readers to build a situation model (e.g., Radvansky et al., 2001). Schoot and colleagues (2010) conducted a study with elementary students, to determine whether situation-focused instructions could encourage the formation of a situation model. Children were assigned to one of two conditions.

In a “standard instruction” condition, participants were told to *understand* what the text was about and in a situational instruction condition, participants were told to *imagine* the events and developments of the text. These latter directions were thought to encourage readers to position themselves in the story, thus encouraging the construction of a situation model. After reading several fictional narrative texts, readers who heard the “situational” instructions were better able to answer comprehension questions than those in the comparison condition. Focused instructions to help the reader situate themselves in the text may aid comprehension and highlight the potential benefits for situation model construction when teaching reading comprehension (Schoot et al., 2010).

Theoretical models of reading comprehension help to inform our understanding of the construct and provide insight concerning the measurement of reading comprehension. These models make clear that reading comprehension is not a single ability, rather it is facilitated by the development of a variety of skills and knowledge. Knowledge of these many essential components of reading comprehension helps to evaluate the popular measures of reading comprehension examined in the current study and can help determine if the measures are adequately assessing a reader’s ability to comprehend a text as would be the case in reading tasks in the classroom and real-world, outside of the context of the tests.

Language Skills

Congruent with the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), language comprehension skills cannot be maximally engaged to allow for effective comprehension until adequate skill in decoding develops (Lervag et al., 2018). Once decoding becomes more efficient, language comprehension gains importance for reading development (Garcia & Cain, 2014). Oral language comprehension is thought to include one’s vocabulary knowledge and understanding of the structures of language – syntax and morphology knowledge (e.g., Lonigan et al. 2018; American Speech-Language-Hearing Association n.d.).

Vocabulary refers to the stored knowledge one has about the meanings of individual words. A strong vocabulary is correlated with successful reading comprehension (Carroll, 1993). Two dimensions of vocabulary knowledge are vocabulary breadth and depth. Vocabulary breadth describes the number of words *known* by the reader and vocabulary depth refers to the *quality* of the meaning that the learner knows (Chen & Lui, 2020). The lexical quality hypothesis (Perfetti, 2007) proposes that the quality of word knowledge is as important for comprehending spoken and written language as the number of words known.

Research on children's reading has found that both breadth and depth of vocabulary knowledge are predictive of reading comprehension (e.g., Oakhill & Cain, 2012). For example, Ouellette (2006) found that fourth-grade students' reading comprehension (measured with a passage comprehension test) was positively correlated with both breadth and depth of vocabulary knowledge (measured with the Test of Word Knowledge; Wiig & Secord, 1992). However, Ouellette (2006) found that only depth of vocabulary knowledge predicted unique variance in reading comprehension when decoding and word reading was controlled. Other research, however, has pointed to vocabulary breadth as having a strong correlation with understanding text (Tannenbaum et al., 2006). Thus, both the number of words in a reader's vocabulary and the extent of knowledge associated with word items appear important for reading comprehension.

Morphological awareness is another important component of oral language in the context of reading and its development. A morpheme is the smallest meaningful unit of language (Apel, 2014), and these meaningful units make up words. Morphological awareness refers to one's understanding that words are built by combining prefixes, roots, and suffixes and one's ability to recognize and manipulate these units (Carlisle, 1995). Within morphology, a distinction can be made between inflectional and derivational morphology. Inflections change the grammatical function of a word (such as when *-ed* is added to the word *play* to create *played*) and derivations change the meaning of a word (adding *-ful* to the word *play* to form *playful;* Kirby et al. 2012).

Research has shown that there is a link between morphological awareness and success in reading comprehension. For example, Deacon and Kirby (2004) found that second-grade students' morphological awareness contributed unique variance to their concurrent reading comprehension and their later, Grade 5 reading comprehension. Other studies have supported this unique relationship between morphological awareness and reading comprehension (e.g., Metsala et al., 2021). It has been proposed that being able to recognize the morphological structure of a word helps a reader make sense of unfamiliar words (Deacon et al., 2014). For example, a reader could infer the meaning of the word *hopeful* from its component parts: *hope* and *-ful*.

Morphological processing has also been shown to be impaired in students who have poor reading comprehension, but adequate word reading (Nation et al., 2004; 2005). Nation and colleagues (2005) found that eight-year-olds who were poor comprehenders (but with average decoding skills) had difficulty compared to typically developing readers with past tense production for regular and irregular verbs. Findings implicating deficits in morphological awareness for students with specific comprehension difficulties have also been supported in other research (e.g., Kirby et al., 2012).

Syntactic awareness has similarly been shown to predict variance in performance on reading comprehension measures. Syntactic awareness refers to a reader's understanding of grammatical rules and sentence structure (Layton et al., 1998). One proposal has been that the effect of syntactic awareness on reading comprehension is mediated by vocabulary skills. Syntactic awareness leads to increased breadth and depth of vocabulary, which in turn influences reading comprehension (Guo et al., 2011). Having well-developed syntactic skills may also support reading comprehension by enabling readers to monitor their comprehension while reading (Guo et al., 2011). Understanding how word order works has been shown to contribute to reading comprehension beyond vocabulary and morphological awareness (e.g., Metsala et al., 2021).

Comprehending text has thus been shown to be associated with individual oral language skills. It appears that well-developed vocabulary knowledge, syntactic awareness, and morphological awareness contribute to skilled text comprehension (e.g., Nation & Snowling, 2000).

Background Knowledge

Knowledge, or the information that is stored in long-term memory, has been shown to influence the amount of information a reader retains during reading and how well they understand the text (Cook & Guerard, 2005; Recht & Leslie, 1988). Knowledge includes a reader's store of concepts, ideas, and the relationships among them (Cabell & Hwang, 2020). In reading research, knowledge is often referred to as either *prior* or *background knowledge* and can be categorized based on the breadth of knowledge under discussion. Domain knowledge, or content knowledge, is related to a particular field of study or discipline (e.g., Alexander & Kulikowich, 1991). Knowledge that is related to more than one field of study or discipline is referred to as general knowledge (e.g., Kozminsky & Kozminsky, 2001). While foundational word reading and language skills are essential for successful reading comprehension, having adequate domain knowledge in the topic area of a text is also critical for an in-depth understanding of a text (Cabell & Hwang, 2020).

The role that background knowledge plays in reading comprehension has been well established in research, with studies demonstrating that a reader's domain knowledge of the text's topic has an impact on what they comprehend and learn as they read (e.g., Cervetti et al., 2016). For example, Recht and Leslie (1988) examined seventh and eighth-grade student's general reading comprehension and their knowledge about specific texts. They aimed to determine whether each influenced the students' ability to recall the details of a passage. Participants were identified based on a general measure of reading comprehension, as either "good readers" (above the 70^{th} percentile) or "poor readers" (below the 30^{th} percentile). Students silently read passages about a half-inning of a baseball game. The authors found that children's understanding of the passage largely depended on their pre-existing knowledge of the topic. Those who had more baseball knowledge recalled more than children with less domain knowledge regardless of their measured comprehension ability. This research suggests that even students who have low scores on a comprehension measure can compensate when they have strong domain knowledge of the topic in the text (Recht & Leslie, 1988).

Moravcsik and Kintsch (1993) also showed that domain knowledge facilitates text comprehension. The authors had university students read a passage about an unidentified, complex procedure and manipulated the amount of context that was provided. When readers were given a passage that included a title, they were able to use their domain knowledge about the topic (e.g., doing laundry) and they were better at recalling details of the text, compared to those who were not provided with a title. That is, the title became a cue that prompted readers to activate background knowledge about the topic which improved recall of the passage. The researchers also manipulated the coherence of the passage. This was done to determine whether the quality of writing of the text could compensate for a lack of domain knowledge. They found that while reading a more coherent passage helped with comprehension, it did not entirely compensate for a reader's lack of domain knowledge. Good writing has been proposed to allow the reader to form a coherent textbase. However, having domain knowledge allows the reader to embed that textbase into an accurate representation of the situation described by the text which supports a richer understanding (Morvacski & Kintsch, 1993).

Research on background knowledge, including experimental studies that have manipulated or taught the relevant knowledge, may have implications for instruction and intervention in the school setting. The evidence supports the claim that building knowledge supports a reader's ability to comprehend new material (Cervetti et al., 2016). This research would appear to also have implications for measuring general reading comprehension ability.

Reading Comprehension Strategies

Good readers use a variety of reading comprehension strategies while reading (Reutzel et al., 2005). Poor readers use fewer strategies and are less likely to learn them unless they are given explicit strategy instruction (Harris & Pressley, 1991; National Reading Panel, 2000). Strategies can be defined as deliberate, goal-directed attempts to control and modify one's own efforts to understand and construct meaning from text (Afflerbach et al., 2017). It is important that the reader intentionally and consciously applies the strategies, particularly when comprehension falters. It is also important that they have conditional knowledge so that they know why and when to apply different strategies (Magnusson, 2019).

Both intentionality and conditional knowledge involve metacognitive awareness insofar as students monitor their own reading process. Metacognitive strategies involve using higher-order cognitive processes to regulate one's own learning by planning, monitoring, and evaluating (Zhang & Seepho, 2013). Good readers use metacognitive strategies to focus on both the process and the product of reading by monitoring and evaluating their comprehension as they read. Students benefit from engaging in this metacognitive activity during reading because they can monitor their comprehension, clarify difficulties, and restore the process when it fails (Muhid et al., 2020).

For a long time, reading research was somewhat unbalanced, focusing more on the development of word reading and the related skills of phonological awareness and fluency (e.g., Boulineau et al., 2004). It is often assumed that once students are proficient at decoding and can read fluently, their ability to comprehend text will automatically follow (Hagaman et al., 2012). While there is a correlation between these factors, there are skills, strategies, and additional foundational reading skills that play a role in reading comprehension. Cognitive strategies that have accumulated evidence to support their effectiveness when used by readers include activating background knowledge, generating questions related to the text, summarizing text, organizing information graphically, learning the structures of stories, and monitoring comprehension during reading (Guthrie et al, 2004; Harris & Pressley, 1991; National Reading Panel, 2000). Research has shown that explicit and systematic instruction in reading comprehension strategies improves students' ability to comprehend text (e.g., Graham & Bellert, 2004; Guthrie et al., 2004).

While instruction of metacognitive and cognitive strategies has been examined in isolation, research has also examined the instruction of reading comprehension strategies in combination. The National Reading Panel (2000) endorsed multiple cognitive strategy instruction, reasoning that good readers do not use one single comprehension strategy at a time when they read. Rather, they coordinate a set of strategies when interacting with a text (Reutzel et al., 2005). Reutzel and colleagues (2005) investigated multiple strategy instruction to determine if this leads to improved comprehension outcomes when compared to single strategy instruction. They compared Single Strategy Instruction (SSI; n=38) in which reading comprehension strategies were taught one at a time, with Transactional Strategy Instruction (TSI; n=42), for which students were taught a set of comprehension strategies embedded in an engaging routine. The strategies included activating background knowledge, examining text structure, predicting, visualizing, monitoring, questioning, and summarizing. On the comprehension subtest of the Gates-MacGinitie reading test (MacGinitie et al., 2000) there was no significant difference between the SSI and TSI groups after 16 weeks of instruction. This research suggests that explicitly teaching strategies one at a time (SSI) or in combination (TSI) may not differ in the effectiveness of increasing students' reading comprehension. Whether a single strategy is taught at once or a set of strategies, students should ultimately be explicitly taught several strategies and should be given plenty of opportunities to practice (Shanahan et al., 2010).

Measurement of Reading Comprehension

Although research has demonstrated the complexity of reading comprehension for real-life reading tasks, investigations concerning how best to measure reading comprehension within this context are somewhat sparse. Tests of reading comprehension that are commonly used in practice and research were created by test-developers, who do not necessarily share the same conceptualization of reading comprehension as the researchers. Therefore, there is potential variation in the conceptualization of reading comprehension across tests and in what exactly different standardized tests are measuring. Furthermore, some researchers do create their own measures of reading comprehension to get at more aspects and complexity of real-life reading activities (e.g., Guthrie et al., 2004; Clemens & Fuchs, 2021). In doing so, these researchers are acknowledging the limitations of standardized reading comprehension tests. In addition to differences in conceptual underpinnings, tests of reading comprehension appear to vary in the demands they place on the reader through differing formats and administration procedures, and on the reading-related skills and knowledge that the tests may draw most heavily upon (e.g., Cutting & Scarborough, 2006). Although research is not abundant, across the last several decades, there have been examinations and questions concerning the validity of, and differences between, measures of reading comprehension (e.g., Nation & Snowling, 1997; Keenan & Meenan, 2014). I next provide a brief summary of this research.

Nation and Snowling (1997) compared the association of scores from a comprehension test using a multiple-choice sentence completion procedure (also known as a cloze task; The Suffolk Reading Scale; Hagley, 1987) to one that required the reader to read aloud a series of short passages and then respond to orally presented comprehension questions (Neale Analysis of Reading Ability-Revised; Neale, 1989). The scores from the two tests were strongly correlated ($r = .75$) for a group of 7 - 9 years olds, with each test accounting for about 56% of the variance in the other test. They also identified groups of students with good versus poor listening comprehension (answering both literal and inferential questions following each of a series of orally presented stories). The group of children who had low performance on the listening comprehension test performed more poorly on the Neale test of reading comprehension than the good listening comprehenders; however, the two groups were comparable on the cloze test (i.e., the Suffolk scale). The researchers argued that the test with the cloze format did not correctly identify 'poor comprehenders' (low listening comprehension; Nation & Snowling, 1997).

Nation and Snowling (1997) formally examined the extent to which decoding and listening comprehension each contributed to these two separate tests. Decoding ability and listening comprehension each showed differences in the relative contribution to the two tests. A factor analysis showed that the measure that employed answering orally presented questions about each passage loaded similarly onto each of the decoding and listening comprehension factors to comparable extents (.62 and .67 respectively). The cloze test loaded higher on the decoding factor compared to the comprehension factor (.88 vs .26, respectively). A regression analysis confirmed that reading comprehension measures with a cloze format rely largely on word reading skills, rather than on one's ability to understand concepts and relationships in language and stories. Nation and Snowling (1997) emphasized the importance of considering the underlying skills being tapped by reading comprehension tests to avoid overlooking language comprehension deficits that will influence real-world reading. This study highlights concerns that the nature of a test's format influences what contributing skills are mostly measured and may affect which students would be identified as having difficulties in comprehending text and should have access to interventions.

In 2014, Keenan and Meenan's results called into question Nation and Snowling's (1997) conclusions that scores on reading comprehension tests with differing formats (a cloze format test and an open-ended question format test) were highly correlated. In their large sample of 8-18-year-olds, Keenan and Meenan (2014) found that the average correlation between common tests of reading comprehension was moderate (r = .45 - .68). When tests of similar formats to those compared by Nation and Snowling (1997) were compared (a cloze and an open-ended question format test), the correlation was .56 (Keenan & Meenan, 2014). The current study will seek to contribute to this research by examining the correlations of different tests in a younger sample.

To further examine the notion that reading comprehension measures using a cloze format are largely influenced by decoding skill, Keenan and colleagues (2008) observed performance across additional measures, with a larger sample (*n* = 510), and a larger age range of students (8-18 years). Four measures of reading comprehension were examined: a Passage Comprehension test for which students provide the missing word in a sentence (WJ-PC; Woodcock et al. 2001); a test which required students to read a passage aloud and answer short questions (QRI; Leslie & Caldwell, 2001); one that required students to read a series of passages aloud and answer orally presented multiple-choice questions (GORT; Wiederholt & Bryant, 1992); and the fourth test for which participants silently read sentences and choose which of four photos best expresses the sentence's meaning (PIAT; Dunn & Markwardt, 1970). In general, the correlations among tests were weak to moderate (0.31-0.48), with the exception of the stronger correlation (0.70) between the cloze format test (WJ-PC) and the picture-matching test (PIAT). A series of regression analyses showed that decoding accounted for a high proportion of unique variance in these two tests after the variance accounted for by listening comprehension was controlled (34 % and 32%, for the PIAT and WJPC, respectively). At the same time, listening comprehension accounted for a smaller amount of unique variance in these two tests (5% and 7%, respectively) once decoding was controlled. On the other hand, for the two tests that had students answer questions about a text, decoding accounted for a smaller amount of unique variance after accounting for listening comprehension (8% and 3% for the GORT and QRI, respectively), and listening comprehension accounted for the same or a greater proportion of unique variance overall (10% and 17%).

In both the PIAT and the WJ-PC, the researchers noted that the assessment of comprehension appears to rely heavily on the decoding of a single word. In the PIAT, the child reads a single sentence and chooses the photo that best represents the text. The differences between each photo are generally depictions of alternatives that would correspond to an incorrect decoding of one word in the sentence. These authors suggested that the WJ-PC is similar insofar as providing a correct response frequently depended on being able to decode a single word (Keenan et al., 2008). Overall, this study extends the findings of Nation and Snowing (1997) that different tests of reading comprehension rely to varying degrees on skills such as decoding and oral language comprehension.

Keenan and colleagues (2008) also found that decoding skills were more strongly associated with reading comprehension performance for younger, and less skilled readers, consistent with theoretical accounts of reading development (Catts et al., 2005; Hoover & Tunmer, 1993). Keenan and colleagues (2008) extended previous findings by showing that differences in the extent to which these developmental differences are observed depend on the type of reading comprehension test. The percentage of unique variance accounted for by the addition of an interaction term for age and decoding was much larger for the PIAT (31%) and the WJ-PC (27%) than it was for the other tests (all $\leq$ 6%). These findings suggest that in younger and less skilled readers, data from these two tests (and tests with similar formats) may be most representative of decoding ability rather than other cognitive aspects that are expected to influence reading comprehension. Thus, not only do different reading tests rely more on some skills than others but also the same test may rely on skills to a different extent, depending on the child's age and skill level.

In a similar study, Cutting and Scarborough (2006) examined this question in students in Grades 1 through 10, across three popular reading comprehension measures. One test requires the student to read a series of passages aloud and answer the orally presented multiple-choice questions after it is removed from view (GORT–3; Wiederholt & Bryant, 1992). The second test has students read passages silently and answer two orally presented open-ended questions (one literal and one inferential) while the text remains in view (Wechsler Individual Achievement Test; WIAT; Wechsler, 1992). The last test was different from the previous two in that it was a group-administered test that requires students to read passages silently and answer three to six written multiple-choice questions while the passage is still in view. Items gradually increase in difficulty, and students are given a 35-minute time limit to complete all items (Gates–MacGinitie Reading Test-Revised; G–M; MacGinitie et al., 2000). Students also completed three vocabulary measures, testing of each of receptive vocabulary, expressive vocabulary, and knowledge of word meanings (the student indicated which two words from a series are most closely related). Cutting and Scarborough (2006) reported that the contribution of word reading varied across measures, accounting for 6% of unique variance in G-M scores, 8.0% in GORT-3 scores, and 12.0% in WIAT scores. Vocabulary accounted for 15.0% of unique variance in G-M scores, and 9.0% in both GORT-3 and WIAT scores. Cutting and Scarborough (2006) also found that the amount of variance accounted for by word reading was not related to participants' skill level at reading words. This differs from Keenan and colleagues' (2008) findings that the same test relied on word reading more in younger, less skilled decoders.

From this small number of studies that have directly addressed this question, it is clear that different measures of reading comprehension tap contributing skills to varying extents. Measures that rely on cloze tasks seem to be more closely associated with word reading skills, perhaps not giving a good indication of language comprehension and the ability to understand and relate concepts in the text (e.g., Nation & Snowling, 1997). This may be a problem because measures of comprehension that employ cloze tasks do not identify students that are at-risk for reading comprehension difficulties due to poor language comprehension. We might expect these children to have difficulty in comprehending text in the classroom, especially with increasing grades and text difficulty.

Indeed, Keenan and Meenan (2014) reported considerable inconsistencies across three common tests of reading comprehension in terms of who, in their sample of 8-18-year-olds, was identified with a reading comprehension deficit. The authors identified the lowest 100 scoring participants on each test (about 10% of their sample) as having a comprehension deficit and compared the degree of overlap between tests. The percentage of students that were identified by any two of the three tests ranged from 32% to 56%. A child who was among the low-scoring group on the PIAT (multiple-choice format where photos are chosen to demonstrate comprehension) had only a 56% chance of being identified on the WJ-PC (cloze format comprehension test). Keenan and Meenan (2014) found that the GORT (multiple-choice format comprehension test) and the WJ-PC had 32% of poor comprehenders in common. There were only 20 children identified across all three tests.

There are several limitations in this small body of research comparing measures of reading comprehension. First, there has not been a focus on young children who are early in their development of reading skills. Second, studies that have included young readers in their samples have wide age ranges and have not examined whether associations between component skills and reading comprehension scores change within a more limited age range (e.g., Keenan & Meenan, 2014). Third, the measures of language comprehension have been constrained to listening comprehension or vocabulary. Finally, there have not been any examinations of the test items to further explain the measures and why they may vary in their outcomes.

The Present Study

School psychologists use norm-referenced, reading comprehension measures in school settings to assess students' abilities to understand text. Results from these tests are interpreted to represent a unitary reading comprehension construct or ability and are used to inform expectations for classroom performance (Catts & Kamhi, 2014). The purpose of the present study is to examine the measurement of reading comprehension across two common reading comprehension measures for early elementary students. In this study, I ask three research questions for a group of students in first through third grade concerning the measurement of reading comprehension with the Passage Comprehension subtest of the Woodcock Reading Mastery Tests, Third Edition (Woodcock et al., 2001) and the comprehension subtest of the Gates-MacGinitie Reading Tests—Second Canadian Edition (MacGinitie et al., 2000; G-M). As previously described, the WJ-PC has children read short passages and provide a missing word to demonstrate their comprehension. The G-M, a group-administered test, has children read passages and answer three to six multiple-choice questions about each passage (see methods section for more detail).

The first research question I address is what is the strength of the relationship between the reading comprehension scores on the two tests, and does the strength of this association differ across the first through third-grade participants? Keenan and Meenan (2014) found the average correlation between common tests of reading comprehension to be weak to moderate (r = .45-.68), but their students spanned a large age range (8-18 years of age). On the other hand, Nation & Snowling (1997) examined the relationship between a cloze task and a more open-ended task and found a strong correlation for those between 7-9 years of age ($r = .75$). I hypothesize that the correlation may be similarly strong for my sample of students.

My second research question is whether the relative contributions of pseudoword decoding and oral language to each of these two reading comprehension measures differ in young elementary students? Based on previous research, I expect that decoding skills may be more strongly predictive of the cloze format comprehension test (WJ-PC) than the multiple-choice format comprehension test (G-M). Furthermore, oral language may predict more unique variance in students' performance on the multiple-choice format comprehension test (G-M) than the cloze format comprehension test (WJ-PC). Related to this question, I will examine whether this relationship appears to change with age. In Keenan and colleagues' (2008) study amongst students from 8-18 years of age, decoding predicted more variance in the comprehension of younger and less skilled readers. The current study will investigate whether this is true for participants in the early elementary school years.

My third research question asks, will the same group of students be identified as poor comprehenders on both reading comprehension measures in this study? Based on comparisons of categorizing readers in previous studies (Keenan & Meenan, 2014; Nation & Snowling, 1997), I do not expect a high degree of agreement in identifying the poorest performers on the two measures in this study. In the current study, I examine this agreement in younger students than in this previous research. In general, younger children's decoding skills have been proposed to account for the bulk of the variance in reading comprehension and limit the influence from oral language skills (e.g., Catts et al., 2005). That is, skilled reading comprehension is limited without the development of decoding. Thus, the two tests in this study may identify largely different groups of children as poor comprehenders. The cloze task (WJ-PC) may primarily identify those poor in pseudoword decoding, while the multiple-choice format-test may also identify those who are poorer on other components that contribute to reading comprehension, such as oral language or knowledge skills.

One further purpose of the present study is to provide an initial, cursory examination and comparison of the vocabulary and syntax in the reading passages for both of these tests. This examination will be undertaken in order to compare two aspects of the language demands placed on young students across these two tasks. Across test items on both the measures, I will do a comparison of the word frequency in the passages for each grade, as well as compare the sentence structure, in terms of clauses, in each test. This preliminary analysis will help support an understanding of potential differences between these tests that have not been previously studied.

Examining these research questions will help to better understand similarities and differences in these two comprehension measures. This will include the ways in which component skills contribute to overall performance on each test and potentially differ across these early grades and the congruency between which students are identified as poor comprehenders. The findings will also allow me to begin to describe the vocabulary and syntax that make up the passages of these two tests. Overall, this study will help to inform my ongoing thinking and use of reading comprehension measures in my future role as a school psychologist.

Method

Participants

Data for the present study was collected as part of a larger study with Principal Investigator, Dr. Jamie Metsala. All children were recruited from a medium-size, mostly suburban school district in Eastern Canada. Two samples were included in this study. The first sample consists of 64 students in Grade 1. For this Grade 1 sample, the decoding and oral language measures (vocabulary, syntactic and morphological awareness) were completed in the early winter of Grade 1. The two reading comprehension measures were collected one year later, in the early winter of the students' Grade 2 year. This context needs to be kept in mind for the grade-based comparisons, as the Grade 1 students completed their reading comprehension tests in Grade 2.

The second sample consists of 57 students in Grade 2 and 59 students in Grade 3. The vocabulary, syntactic and morphological awareness, and pseudoword decoding measures were completed in the late fall of the academic year. The two reading comprehension measures were completed in the spring of that year, about 6 months later.

Measures

Pseudoword Decoding. Students completed the Word Attack subtest of the Woodcock Reading Mastery Tests, Third Edition (Woodcock et al., 2001). In this subtest, the student was presented with and asked to read aloud a list of pronounceable pseudowords that progressively increased in difficulty. This task assesses a student's ability to sound out unfamiliar words. The manual reported value for Cronbach's Alpha reliability coefficient for this measure was $\alpha = .94$ indicating a high degree of consistency (Bland & Altman, 1997).

Oral Language Skills. Oral language was measured with 3 tasks. Students' receptive vocabulary was measured with a modified version of the Peabody Picture Vocabulary Test (PPVT-IV; Dunn & Dunn, 2007). In this modified version, every third item was administered. All standardized procedures were otherwise followed. Similarly modified administrations of the PPVT have been used previously (e.g., Deacon et al., 2014; Sparks & Deacon, 2015) and validated with young children (Deacon et al., 2013). The manual-reported values for Cronbach's

Alpha were α =.85 for the full PPVT-IV, indicating satisfactory internal consistency (Bland & Altman, 1997).

Students' syntactic awareness was measured on a task with 15 items that required the student to make a judgment and/or correction to the syntax of sentences. In the first 10 items, students judged whether a sentence was syntactically correct (e.g., '*The cats meow and the dog*

) or incorrect (e.g., ') with errors being related to word order. In the remaining five items, students were asked to correct sentences with word order errors (e.g., '*The buses stop and go* can be corrected to '*The buses stop and the trains*

). For further detail on this measure, see Metsala et al. (2021). The Cronbach's Alpha reliability coefficient for this measure was α = .78 indicating a satisfactory level of internal consistency (Bland & Altman, 1997).

A morphological awareness measure examined inflectional morphology, which is well developed in students of this age (Robertson & Deacon, 2019). There were 20 items that made up this scale. In the first 13 items from the CELF-5 (Wiig et al. 2013), students were required to complete a partially spoken sentence stem that was associated with a provided picture (e.g., *Here is a boot. Here are two* ____). The remaining seven items required students to listen to a spoken sentence and correct the errors. Errors were in relation to subject-verb agreement and the incorrect use of morphemes (e.g., *The dogs plays and Matt sleep.* Correct response: *The dog plays and Matt sleeps* or *The dogs play and Matt sleeps*). For further detail on this measure, see Metsala et al. (2021). The Cronbach alpha reliability coefficient for this measure was α = .78 indicating an acceptable level of internal consistency (Bland & Altman, 1997).

Reading Comprehension. Students completed the Passage Comprehension subtest of the Woodcock Reading Mastery Tests, Third Edition (Woodcock et al., 2001). In this test, students read short passages to themselves and were asked to provide a missing word. The task was administered following the test's standardization procedures and administration continued until six consecutive errors occurred. For this measure, age-based standard scores were used in data analyses (*M* = 96.59, *SD* = 14.98). For reference, an average score on this task would be a standard score of 100. The Cronbach's alpha reliability coefficient provided by the manual was α = .96, indicating a high degree of internal consistency (Bland & Altman, 1997).

The Comprehension subtest of the Gates-MacGinitie Reading Tests—Second Canadian Edition (MacGinitie & MacGinitie, 1992), which is a group-administered measure, was also administered to assess reading comprehension. Students from the Grade 1 sample were given Level B, Form 3, and students from the Grade 2-3 sample were given Level C, Form 3. In the Level B measure, students were given a booklet containing passages ranging in length from 1-6 sentences followed by one comprehension question. In the Level C measure, students were given a booklet containing passages ranging from 1-13 sentences in length and were asked to complete 2-5 multiple-choice questions following the reading of each passage. Students were given 35 minutes to read these passages and answer the multiple-choice questions. T-scores, the age-based normative score provided by the Gates-MacGinitie, were used in data analyses for this measure (*M* = 47.27, *SD* = 9.53). For reference, an average score on this task would be a *t*-score of 50. The Cronbach's alpha reliability coefficient provided by the manual was α = .93, indicating a high degree of internal consistency (Bland & Altman, 1997).

Procedure

As mentioned, this was part of a larger study in Dr. Metsala's research lab. Testing was conducted by trained research assistants. Students were given a battery of measures at two different time points, and this study examines a portion of those tasks. At Time 1, students completed the three oral language measures (vocabulary, syntactic awareness, and morphological awareness) and a decoding measure (i.e., pseudoword decoding). At Time 2, reading comprehension was assessed using the two tests that are the focus of this study.

Results

Descriptive Statistics

Initial examination of the data showed that all reading and oral language measures were normally distributed. Previous analyses showed that there were no multivariate outliers for either of the two samples (Metsala et al., 2021; Sparks & Metsala, 2021). Across all measures, 2.1% of the data were missing. Missing data were replaced with multiple imputations using SPSS for all regression and correlation analyses. Pooled estimates were not available for R^2 and standardized regression coefficients in SPSS. Examination revealed that results from each of the five imputations produced similar R^2 values and maintained the same pattern of statistical significance. Therefore, statistical values from imputation 1 were chosen arbitrarily to report for each regression and correlation consistently. This reporting procedure avoids the possibility of choosing which imputation to report based on statistics that most strongly support any of my hypotheses and is used throughout all data analyses to maintain consistency.

Table 1 provides descriptive data, including mean and standard deviations for all raw and standardized scores used in this study. As can be seen in Table 1, the mean standard scores for both reading comprehension measures fell within the average range across all age groups. Zero-order correlations among raw scores are across all participants are presented in Table 2. Table 3 shows the correlations separately for each grade.

Correlations

Zero-order correlations among raw scores are presented in Table 2, first across all participants in this study and then separately for each grade (Table 3). I first examined the strength of the relationship between the scores on the two reading comprehension tests, using the Pearson product-moment correlation coefficient. Preliminary examination of the data showed that there were no violations of the assumptions of normality, linearity, and homoscedasticity (Pallant, 2007). The correlation between the scores on the two measures for the entire sample was $r = .75$, $p < .001$. Thus, scores from the two tests were strongly correlated (Akoglu, 2018), with one test accounting for just over half the variance in the other (i.e., 56.3%).

The relationship between scores on the two tests of reading comprehension was next analyzed by grade. For each grade level, there was a strong, positive correlation between scores on the two reading comprehension tests, $r = .71$, .72, and .80, $< .001$, for Grades 1, 2, and 3 respectively. Thus, for each grade level, in turn, one reading comprehension test accounted for 50.4%, 51.6%, and 64.0% of the variance in the other test.

To determine if the correlations between scores on each of the two reading comprehension tests were statistically significantly different between grade levels, z values (z_{obs}) were calculated for each grade level comparison (Pallant, 2007). For the comparison between Grade 1 and 2, $z_{obs} = -.25$, for Grade 1 and 3, $z_{obs} = -1.38$, and for Grade 2 and 3 $z_{obs} = -1.09$. Z_{obs} values for each comparison were greater than -1.96 and less than 1.96, indicating that correlation coefficients at each grade were not statistically different.

Contributions of Pseudoword Decoding and Oral Language Skills to Measures of Reading Comprehension

I next examined the relative contributions of pseudoword decoding and oral language skills to each of the two reading comprehension measures. These analyses address whether decoding contributes more to one of the tests (the Passage Comprehension subtest with the cloze format) and whether oral language skills contribute more unique variance to the second test (the Comprehension subtest of the Gates-MacGinite with the multiple-choice format). Preliminary analyses were conducted to ensure no violation of the assumptions of linearity, multicollinearity, homoscedasticity, independence of observation, and multivariate normality (Pallant, 2007). For each test, I conducted two hierarchical regression analyses; the first with decoding entered as the second step after grade, and the second regression with oral language skills entered in the second step after grade (note in the regression tables, I only present the results for the oral language block, as the contribution of each individual oral language skill was not relevant to the purpose of this research).

As seen in Table 4 (Regressions 1 & 2), at the second step, decoding accounted for 41.0% of the variance in the Passage Comprehension test and 49.0% of the variance in the Gates-MacGinitie; furthermore, oral language skills entered at the third step in each regression accounted for an additional and significant 3.0% and 8.0% of the variance in the Passage Comprehension test and Gates-MacGinite test, respectively. When examining oral language entered immediately after grade at Step 2, these accounted for 23.0% of the variance in Passage Comprehension and 34.0% of the variance in Gates-MacGinite scores (see Table 4, Regressions 3 & 4). When decoding was entered at the third step in each regression, it accounted for an additional and significant 22.0% and 23.0% of the variance in the Passage Comprehension test and Gates-MacGinite test, respectively

To test whether oral language predicts more variance in each reading comprehension measure based on grade, I examined interaction terms, as the fourth step in each of Regression 1 and 2 in Table 4. I also examined whether the contribution of decoding skills decreases with grade for each test, as predicted by the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990). To create interaction terms, I first created centered variables for each of the oral language skills and decoding and multiplied this by grade.

For each of the two measures of reading comprehension, grade, decoding, and oral language were entered into the equation as the first three steps. Next, one of the interaction terms was entered in Step 4 in separate regressions. The interaction terms for grade by vocabulary, morphological awareness, and decoding were each not significant in the equations for the Passage Comprehension test (p = .23, p = .65, *&* p = .15, respectively) nor for the Gates-MacGinite test (p = .23, p = .20, p = .08). It should be noted that the interaction of Grade x Decoding approached statistical significance for the Gates-MacGinitie test. Upon further examination, this did not reflect the expected trend as predicted by the Simple View of Reading. Decoding was associated with a trend toward greater changes in reading comprehension in Grade 3 students (R^2 = .63), followed by Grade 2 students (R^2 = .53), and Grade 1 students (R^2 = .29). Although there was no statistical trend for the Passage Comprehension test, the ranking of the absolute magnitudes of the numbers was similar (R^2 = .50, .48, and .23, for Grades 3, 2, and 1, respectively).

The syntactic awareness by grade interaction term was not significant for the Gates-MacGinitie test (p = .14); however, it accounted for a statistically significant 3.0% of unique variance in reading comprehension (= .47*; p* = .002) as measured by the Passage Comprehension test. Upon further examination, syntactic awareness in Grade 3 was associated with greater changes in scores on the Passage Comprehension test (R^2 = 0.43) than was seen for Grade 2 students (R^2 = 0.13) and Grade 1 students (R^2 = 0.003).

Consistency of Identification of Students with Reading Comprehension Deficits

To further investigate the comparability of these two reading comprehension tests, I compared the tests in terms of the consistency of students that each test identified as having the lowest standardized test scores. Tests are often used this way in practice to identify who would be considered to have a deficit in reading comprehension. Students scoring at or below two pre-determined cut points were identified to determine the degree of overlap across the two measures in terms of the likelihood that a child categorized as having a comprehension deficit by one test would also be diagnosed by the other test. The two cut points were at or below the 16^{th} percentile and the 25^{th} percentile; these were chosen as they are frequent cut-points used in research and practice (e.g., Keenan et al., 2014; Elwér et al., 2013), and the 16^{th} percentile corresponds to 1 SD below the mean. Consistency of identification represents the percentage of overlap in the identification of students who are categorized below each cut point; that is, identified as the poorest performers are on each test (Keenan et al., 2014).

Students scoring at or below the 16^{th} percentile made up 20.9% of the full sample for the Passage Comprehension test and 23.8% of the full sample for the Gates-MacGinite comprehension test. There were 36 students identified on the Passage Comprehension task as scoring at or below the 16^{th} percentile. Of those, 27 (75.0%) were also identified by the Gates-MacGinite test. Conversely, there were 41 students identified on the Gates-MacGinite task with the 16^{th} percentile cut point, and 27 (65.9%) of those students were also identified by the Passage

Comprehension task. Thus, 25.0% of students identified by the Passage Comprehension test at or below the 16th percentile were not identified by the Gates-MacGinite test. Furthermore, 34.2% of students identified by the Gates-MacGinite task at this cut-point were not also identified by the Passage Comprehension task.

Considering the 25th percentile cut point, there were 43 students identified by the Passage Comprehension task. Of those, 41 (95.4%) were also identified by the Gates-MacGinite test. Conversely, there were 65 students identified on the Gates-MacGinite task, and 41 (63.1%) of those students were also identified by the Passage Comprehension task. These findings demonstrate that 4.7% of students identified by the Passage Comprehension test at or below the 25th percentile were not identified by the Gates-MacGinite test. This proportion increases to 36.9% when examining poor performers on the Gates-MacGinite comprehension test that were not identified by the Passage Comprehension test at the same cut-off. Table 5 presents this information concerning the overlap in students identified by each test for ease in making these numerical comparisons.

Contextualizing the Grade 1 Analyses and Results

For the sample of Grade 1 students, the Passage Comprehension subtest was administered at Time 1, and again at Time 2, when students were in their Grade 2 year. For this Grade 1 sample, reading comprehension data used throughout the analyses reported above are those from Grade 2, as the Gates-MacGinitie was not administered to these students in Grade 1. In order to consider any potential differences resulting from examining reading comprehension in the second year for these students, performance from the Passage Comprehension scores obtained in Grade 1 were compared to those obtained in Grade 2. Table 6 shows the bivariate correlations with other variables in this study for Passage Comprehension scores obtained at Grade 2 and Grade 1. Scores on Passage Comprehension have a correlation of moderate strength (r = .64).

I next conducted two hierarchical multiple regressions to examine the contribution of each of decoding and oral language to Passage Comprehension scores at each of the two time points. As seen in Table 7, when entered in Step 1, decoding accounted for 31.0% of the variance in Grade 1 reading comprehension and 19.0% of the variance in Grade 2 reading comprehension. Oral language skills, entered at the second step in each regression, accounted for an additional 11.0% of the variance in Grade 1 and 19.0% of the variance in Grade 2 Passage Comprehension scores. It appears that when students were tested in Grade 2, less of the variance in Passage Comprehension scores could be accounted for by decoding and more of the variance could be accounted for by oral language skills.

Clause Composition and Word Frequency Across the Two Tests

Syntactic complexity increases the difficulty of a reading task, which may be particularly detrimental to struggling readers (Scott & Balthazar, 2013). One of the ways that sentences become more syntactically complex is by the process of subordination when one or more clauses are combined into one sentence. In an effort to examine the syntactic complexity of the two tests of reading comprehension, each item was examined for the number of clauses contained in each sentence. A Chi-square test for independence indicated that there was no statistically significant difference (χ^2 (2, n = 245) = 1.22, p = .54) in the number of clauses per sentence between the two measures of reading comprehension. For the Gates-MacGinitie test (Level B and C), all items were analyzed as students were exposed to all test items during administration. In the Gates-MacGinitie Level B, 65.0% of sentences had only one clause, with 35.0% of sentences had two or more clauses. In the Gates-MacGinitie Level C, 64.5% and 35.5% had one, or two or more clauses, respectively. Items 1-25 on the Passage Comprehension subtest were examined in this analysis as these items would be inclusive of the items that second and third-grade students may have completed (25 is two standard deviations above the mean; recall Grade 1 students were in Grade 2 at the time of administration). Of the sentences examined on the Passage Comprehension subtest, 74.3% contained one clause and 25.7% contained two or more clauses.

Next, I examined word frequencies for each measure of reading comprehension, as one indication of the vocabulary difficulty for each measure. Word frequency is one of the best methods to estimate word difficulty (Breland, 1996). The assumption is that difficult words are those that are appearing least often in print. Frequencies were obtained using MCWord, an orthographic word form database (Medler & Binder, 2005). MCWord is based on the CELEX database which includes a large number of English word forms.

Words contained in the test items from the complete Gates-MacGinitie and items 1-25 on the Passage Comprehension task, were recorded and entered into the MCWord database. I excluded function words from the analyses. The database provides the raw frequencies and the values transformed by natural log. These latter values were used as raw frequency scores and did not meet normality. I conducted three *t*-tests to compare the word frequencies from each test of reading comprehension. When the Gates-MacGinitie test (Level B-for Grade 1 sample; word raw frequencies, M = 458.83, SD = 1092.20) was compared to the Passage Comprehension subtest (word raw frequencies, M = 697.63, SD = 1558.75), there was a significant difference, t (521) = -2.03, p =.04. The magnitude of the difference in means (mean difference = -238.80, 95% CI: -469.67 to -7.93) was small (η^2 = .008), with the Passage Comprehension test having a higher overall word frequency. When the Gates-MacGinitie test (Level C; M = 459.30, SD = 1048.36) was compared to the Passage Comprehension subtest, there was also a significant difference in word frequencies, t (562) = -2.13, p =.03. The magnitude of the difference in means (mean difference = -238.33, 95% CI: -458.27 to -18.38) was also small (η^2 = .008), again with the Passage Comprehension test having higher overall word frequency. When the Gates-MacGinitie (Level B) was compared to the Gates (Level C), there was not a significant difference in word frequencies, t (737) = -.006, p > .10.

Discussion

The goal of this study was to examine the measurement of reading comprehension across two common measures for early elementary school students. My findings demonstrated that pseudoword decoding accounted for similar amounts of variance in the two measures, while oral language skills contributed more to scores on the multiple-choice test than to scores on the cloze test. Furthermore, the prediction by each oral language skill was similar across grades, except for syntactic awareness which predicted more variance in cloze test scores with each increasing grade. Upon examination of students that were identified as scoring below the 16th and 25th percentile, I found that 34.2% and 36.9% of students identified by the multiple-choice test were not identified by the cloze test. This highlights the inconsistencies between these two measures. Finally, I provide an initial examination of the vocabulary and syntactic complexity contained within the passages of these two tests. Overall, this study will help to inform my ongoing conceptualization of the measurement of reading comprehension in my future role as a school psychologist.

When the association between scores from the Gates-MacGinitie Reading Tests—Second Canadian Edition (level C, Form 3; MacGinitie & MacGinitie, 1992) and the Passage Comprehension subtest of the Woodcock Reading Mastery Tests, Third Edition (Woodcock et al., 2001) was examined, there was a strong, positive correlation. This finding contributes to an existing body of literature that has presented mixed results on the strength of the relationships between tests of reading comprehension. Previous research (Nation & Snowling, 1997) found a similarly strong correlation between a cloze test and an open-ended answer test in 7–9-year-olds. Conversely, Keenan and colleagues (2008) examined the relationship between scores on the Passage Comprehension subtest (Woodcock et al., 2001), and the GORT (Wiederholt & Bryant, 1992) and found a moderate correlation. The GORT and the Gates test both require students to answer multiple-choice questions about the text. However, in the GORT, questions are read to the student rather than the students reading the questions themselves as in the Gates test. Based on this difference in format, the two tests may draw differentially upon pseudoword decoding skills leading to scoring differences (Keenan et al., 2008). Additionally, Keenan and colleagues (2008) examined a large age range (8-18 years) which included students much older than those in the current study. These differences in task demands and age are potential explanations for why the strength of the associations differed from what was found in the current study.

In my study, the strength of the associations between tests of reading comprehension did not differ by grade. These findings converge with Nation and Snowling (1997) who found a strong correlation between two reading comprehension measures in 7–9-year-olds. Results of the current study contrast with those of Keenan and Meenan (2014) who found that there was a greater association between tests for younger children than older children. However, Keenan and Meenan (2014) examined a large age range and drew comparisons between a younger (M = 9.32 years) and an older age group (M = 13.78 years). The current sample is most like the younger group in Keenan and Meenan’s (2014) study in which there was a high degree of consistency between tests.

In this study, I also sought to examine the relative contribution of decoding and oral language to predicting differences in reading comprehension scores on each of the two reading comprehension measures. As stated by the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), these are the two main skill areas that contribute to individual differences in reading comprehension. Both decoding and oral language skills accounted for statistically significant unique variance in reading comprehension scores across both tests of reading comprehension when entered as the second or third step in the hierarchical regressions. Decoding, when entered before oral language skills, accounted for 41.0% of the variance in the cloze format test and 49.3% percent of the variance in the multiple-choice test. After decoding was taken into account, the amount of unique variance accounted for by oral language skills were 3.3% and 7.5% for each test, respectively. Therefore, in the current study, 44.3% of the variance was accounted for by components of the Simple View of Reading for the cloze test and 56.8% for the multiple-choice test. These proportions are less than what Cutting and Scarborough (2006) reported. They found that decoding and oral language contributed 49.0% to 72.0% of the variance in scores for different tests of reading comprehension. This could be due in part to Cutting and Scarborough (2006) not including a cloze measure or due to differences in the ways that oral language was measured which is discussed in more detail below.

In the hierarchical regression analyses, decoding explained significant unique variance in reading comprehension scores above and beyond what could be explained by grade and oral language skills (22.0% and 23.0% in Passage Comprehension and Gates-MacGinite scores, respectively). I predicted that decoding would account for more variance in the cloze task. What was perhaps surprising about my findings was that decoding accounted for about the same amount of variance on both tests when entered before or after oral language skills. Previous research (Keenan et al., 2008) found that when comparing tests with similar formats to those used in this study, the amount of variance accounted for by decoding was much higher in a cloze task than it was for a multiple-choice answer test. It could be the slightly younger age of the sample in the current study that would explain the similar amounts of variance in decoding for both tests. As stated by The Simple View of Reading, decoding is the primary source of individual differences for young readers (Gough & Tunmer, 1986; Hoover & Gough, 1990). Additionally, decoding and accurate word reading are largely neglected in instruction in local elementary schools. As a result, there may be greater variability in students' decoding skills in this particular sample. Furthermore, in the current study, a measure of word recognition was not included to accompany the decoding measure as was done in the other two studies.

I also expected that oral language would account for more unique variance in performance on the multiple-choice test compared to the cloze test, as found in Keenan and colleagues (2008). Indeed, results in the current study showed that oral language skills accounted for 7.5% of the variance in Gates-MacGinitie scores, while they accounted for 3.3% in Passage Comprehension scores. These results contribute to the small body of research in this area suggesting that cloze format tests rely largely on a reader's skill in decoding rather than tapping their oral language skills (e.g., Nation and Snowling, 1997). In the current study, decoding accounted for an equal amount of variance in the two tests of reading comprehension. What differed was the amount of additional variance accounted for by oral language (this is also true when oral language skills are entered first; i.e., regression 3 and 4). This is consistent with previous research showing that comprehension tests involving a multiple-choice answer format have a greater unique contribution from oral language skills than tests that employ a cloze format (e.g., Keenan et al., 2008). In the current study, the amount of unique variance in Gates scores that was accounted for by oral language appears to be less than Cutting and Scarborough (2006) who found that oral language accounted for 15.3% of unique variance in scores on the same measure. This difference may have to do with the differences in the way that oral language was measured. Cutting and Scarborough's (2006) measures encompassed additional oral language skills in lexical and sentence factors that were entered together at the same step in their regression (i.e., expressive vocabulary, knowledge of word meanings, syntactic comprehension, and sentence processing). These additional measures may contribute to the additional variance that Cutting and Scarborough (2006) were able to explain.

On examination of the interactions between grade and students' oral language skills, only syntactic awareness interacted with grade and only for the Passage Comprehension test. Stronger syntactic awareness was associated with greater changes in reading comprehension for each increasing grade. This suggests that as students move through the grades, having well-developed syntactic awareness predicts more variance in one's ability to gain meaning from text. A potential reason for why syntax may contribute to variance in Passage Comprehension scores has to do with the way that students in this province are taught to read. In early elementary reading instruction, students are taught word solving strategies based on guessing and using cues from the text, one of which is syntax (Davis et al., 2021). For example, in the sentence "*The dog played with the* ___," the reader may recognize that the missing word is likely a noun and may be able to choose the correct answer to fill in the blank. It could be that as grade-level texts become more syntactically complex, individual differences in syntactic awareness become more predictive of variance in scores on this test. Additionally, passages within the Passage Comprehension test are generally shorter than those in the Gates test; therefore, readers may have been able to get away with using sentence-level syntax cues to choose the correct answer.

An examination of which students were identified as having reading comprehension difficulties or deficits by each test was undertaken and has important implications for children. In general, there was a relatively high likelihood that a student who was identified by the cloze test would also be identified by the multiple-choice test. 75.0% of students who were identified by the cloze task as scoring at or below the 16th percentile were also identified by the multiple-choice test. This proportion increased to 95.4% when the cut point was set at or below the 25th percentile. In contrast, 65.9 % and 63.1% of students who were identified by the multiple-choice test were also identified by the cloze test, for the 16th and 25th percentile cut points, respectively. The tests are thus somewhat inconsistent in who is identified as having difficulties with reading comprehension.

These findings show that many students in this sample would be considered to have a difficulty in reading comprehension by the Gates-MacGinitie reading comprehension test who would be missed by the Passage Comprehension subtesm000047.71ETQq0.00000912 0 6Th72.0tBT/F1 8.04 Tf format test, students with weak oral language skills would be more likely to be identified. As proposed by Keenan et al. (2008), reading comprehension tasks that use single sentences or very short texts, such as Passage Comprehension, may offer minimal context for identifying words. This forces students to rely largely on correctly decoding words in the passage (sometimes only a single word needs to be decoded) to achieve the correct response. For the Gates-MacGinitie test, passages were generally longer than in the Passage Comprehension test. It is possible that for the longer passages, in addition to successful decoding of the words, students must rely more on language understanding and may need to build a mental model of the passage to achieve the correct answer. In any case, the fact that students with reading comprehension difficulties on this test with generally longer passages and different question answering formats would be completely missed by another measure is disconcerting.

I undertook an examination of the vocabulary and syntax characteristics of the test items to compare the two measures. Vocabulary, as well as syntax knowledge and skills, have each been shown to contribute to reading comprehension (e.g., Ouellette, 2006; Metsala et al., 2021). As proposed by the Simple View of Reading (Gough & Tunmer, 1986; Hoover & Gough, 1990), these become essential with the development of skilled reading comprehension (Lonigan et al., 2018; Gough & Tunmer, 1986; Hoover & Gough, 1990). Students that have a strong vocabulary are more likely to be successful at comprehending written texts (Ouellette, 2006). Ouellette (2006) found that both vocabulary breadth and vocabulary depth contribute uniquely to reading comprehension. Word frequency counts are one way to estimate vocabulary difficulty, giving a sense of the vocabulary demands a text may be placing on a young reader (Breland, 1996). The basic assumption of word frequency is that words that appear less frequently in print and spoken language represent more difficult and less well-known words. It is important that words being used in the assessment of reading-related skills, such as reading comprehension, show a balance of including words that are appearing frequently enough in a language to be reasonable material for young readers as well as words that ard in the aoo fung n ty e a asset4(re)7(ast m(d ir)-4y-3(heETQq0.00000

ree

I also investigated the syntactic complexity of the sentences found in the test items for each reading comprehension measure. Increasing syntactic complexity makes comprehension more difficult, particularly for struggling readers (Scott & Balthazar, 2013). If a reader does not understand how a sentences' structure determines its meaning, comprehending the text becomes more challenging (Scott & Balthazar, 2013). As syntactic awareness has been a predictor of reading comprehension in previous research (e.g., Guo et al., 2011; Tong & McBride, 2017), the syntactic complexity in standardized tests is important. Results showed that syntactic complexity, as measured by clausal composition, did not differ by the test of reading comprehension. This suggests that both tests may place similar amounts of syntactic demands on young readers at this age. Syntactic complexity as measured by clausal composition does not appear to be implicated in score and identification differences between tests.

Implications and Limitations of the Current Study

This study, along with similar past research, has implications for the way we choose, use, and interpret measures in the assessment of reading comprehension. It is important that psychologists and others administering such measures realize that measures of reading comprehension, particularly with young children, may mostly tap decoding skills rather than comprehension of the language in the text. That being said, certain test formats may provide a more balanced assessment of the two major constructs contributing to reading comprehension, even in younger students. Tests that have students read passages and answer questions (in this study, through multiple-choice format), appear to capture more variance due to oral language skills in addition to decoding. Tasks with single sentences or short passages that rely on a cloze format appear to rely on oral language to a lesser extent. Assessors may need to interpret reading comprehension measures as just one source of information toward determining a student's reading comprehension ability. Other sources of information may need to include performance on major contributors to difficult comprehension tasks in the classroom. As implicated in theoretical models of reading comprehension, these might include vocabulary, syntactic awareness, and morphological awareness, as well as general knowledge. Typically, a psychoeducational assessment may be limited to measures of listening comprehension, vocabulary, word reading, and decoding; however, other aspects of oral language are often not assessed. By viewing reading comprehension as a complex process that requires the contribution of many skills or sources of knowledge, the limitations of current measures will become apparent.

The current study, like all research, has limitations to be considered. One potential limitation is the relatively small sample size in each grade level. This study was unique in that it examined reading comprehension in early elementary school readers, many of whom would be in the relatively early stages of learning to read. With smaller sample sizes, there are concerns about generalizability to the larger population. The high degree of convergence with some past research (Nation & Snowling, 1997) does lend confidence in the findings of this study.

A more limiting aspect of the current study was that the reading comprehension measure for the Grade 1 sample was completed when the students were in Grade 2. Therefore, I was not able to examine reading comprehension in Grades 1-3. This is a limitation because my examination covers a smaller grade range for reading comprehension measures meaning that results are not generalizable to Grade 1 reading comprehension. Future studies that include measures of Grade 1 reading comprehension may find significant interactions with grade across the oral language or decoding measures. Additional analyses completed to contextualize the differences between students' Passage Comprehension scores in Grade 1 and Grade 2 show that Passage Comprehension standard scores were quite similar across the two grades, as shown by a moderate correlation. There appeared to be, however, differences in the relative contributions of decoding and oral language in Grade 1 and Grade 2 Passage Comprehension when compared to the full sample of Grade 2 and Grade 3 students. It is possible that had Grade 1 reading comprehension been included, there would have been differences in the relative contributions of these component skills, and the grade interaction.

Finally, my examination into the oral language features of these reading comprehension measures was limited in the depth of analysis about the language aspects of each measure. I analyzed word frequency and the clause composition of sentences to give an initial description of these measures. There are, however, additional ways that vocabulary and syntax can be analyzed. For example, age of acquisition (the age at which a word was learned) is another variable that is commonly measured to determine whether words used in stimuli for research and assessment are developmentally appropriate for the age of the child (Kuperman et al., 2012). Additionally, canonical word order is a straightforward way that syntactic complexity could be analyzed in a reading measure (Scott & Balthazar, 2013). Thus, I provide an initial description of factors that should be considered when examining reading comprehension measures.

The research on reading comprehension, especially in young readers who are early in their developing reading skills, has been inconsistent with regards to comparing and contrasting measures of reading comprehension. In the present study, two measures of reading comprehension with differing formats were highly correlated and the strength of the association did not change with age. Furthermore, decoding contributed similar amounts of significant unique variance to the two tests for this group of early elementary school readers. In contrast, the contribution of variance by oral language differed by test. This difference may have influenced the students in this study that scored below cut-off points typically used in practice to identify those with reading comprehension deficits.

In summary, this study contributes evidence that current measures of reading comprehension fall short in taking into consideration the many skill areas that contribute to this complex cognitive activity. Examining measures of reading comprehension provides insight that test developers rarely provide, which is what the most popular standardized measures are precisely measuring, and how this differs between tests. As evidenced by my findings, different tests identify or fail to identify different students as scoring below the average range. School psychologists and others administering measures of reading comprehension must begin to conceptualize reading comprehension differently and avoid reducing it to a single score from an individual measure.

References

Afflerbach, P., Pearson, P. D., & Paris, S. (2017). Skills and strategies: Their differences, their relationships, and why they matter. In K. Mokhtari (Ed.), Improving reading comprehension through meta cognitive reading strategies instruction (pp. 33–49). Rowman & Littlefield.

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

Alexander, P., & Kulikowich, J. (1991). Domain knowledge and analogic reasoning ability as predictors of expository text comprehension. *Journal of Literacy Research*, 23(2), 165–190. https://doi.org/10.1080/10862969109547735

American Speech-Language-Hearing Association. (n.d.). What is speech? What is language? https://www.asha.org/public/speech/development/speech-and-language/

Apel, K. (2011). What is orthographic knowledge? *Language, Speech & Hearing Services in Schools*, *42*(4), 592–603. https://doi.org/10.1044/0161-1461(2011/10-0085)

Apel, K. (2014). A comprehensive definition of morphological awareness: Implications for assessment. *Topics in Language Disorders*, *34*(3), 197–209. https://doi.org/10.1097/TLD.0000000000000019

Breland, H. M. (1996). Word frequency and word difficulty: A comparison of counts in four corpora. *Psychological Science*, *7*(2), 96–99. https://doi.org/10.1111/j.1467-9280.1996.tb00336.x

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ*, *314*(7080), 572–572. https://doi.org/10.1136/bmj.314.7080.572

Barnes, M. A., Dennis, M., & Haefele-Kalvaitis, 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*(3), 216–241. https://doi.org/10.1006/jecp.1996.0015

Boulineau, T., Fore, C., Hagan-Burke, S., & Burke, M. (2004). Use of story-mapping to increase the story-grammar text comprehension of elementary students with learning disabilities. *Learning Disability Quarterly*, 27, 105–121.

Cabell, S. Q., & Hwang, H. (2020). Building content knowledge to boost comprehension in the primary grades. *Reading Research Quarterly*, *55*(S1), S99–S107. https://doi.org/10.1002/rrq.338

Cain, K. (2016). Reading comprehension development and difficulties: An overview. Perspectives on Language and Literacy, 42(2), 9–16.

Cain, K., & Bignell, S. (2014). Reading and listening comprehension and their relation to inattention and hyperactivity. *The British Journal of Educational Psychology*, *84*, 108–124. https://doi.org/10.1111/bjep.12009

Cain, K. & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, 76(Pt 4), 697–708. https://doi.org/10.1348/000709905X69807.

Carlisle, J. F. (1995). Morphological awareness and early reading achievement. In *Morphological aspects of language processing* (pp. 189–209). Lawrence Erlbaum Associates, Inc.

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

Catts, H.W., Hogan, T. P., & Adolf, S. M. (2005). Developmental changes in reading and reading disabilities. In H.W. Catts & A. G. Kamhi (Eds.), *The connections between language and reading disabilities*. Mahwah, NJ: Erlbaum.

Catts, H. W., Adlof, S. M., & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language & Hearing Research*, *49*(2), 278–293. https://doi.org/10.1044/1092-4388(2006/023)

Catts, H. W., Bridges, M., Little, T., & Tomblin, J. B. (2008). Reading achievement growth in children with language impairments. *Journal of Speech, Language, and Hearing Research*, 51, 1569–1579.

Catts, H. W., & Kamhi, A. G. (2014). Prologue: Reading comprehension is not a single ability. *Language, Speech, and Hearing Services in Schools*, *45*(3), 73–76. https://doi.org/10.1044/2017_LSHSS-16-0033

Cervetti, G., Wright, T., & Hwang, H. (2016). Conceptual coherence, comprehension, and vocabulary acquisition: A knowledge effect? *Reading & Writing*, *29*(4), 761–779. https://doi.org/10.1007/s11145-016-9628-x

Chen, C., & Liu, Y. (2020). *The role of vocabulary breadth and depth in IELTS academic reading tests*. https://doi.org/10.17863/CAM.51399

Clark, M. K., & Kamhi, A. G. (2014). Influence of prior knowledge and interest on fourth- and fifth-grade passage comprehension on the Qualitative Reading Inventory–4. *Language, Speech, and Hearing Services in Schools*, 45, 291–301.

Clemens, N. H., & Fuchs, D. (2021). Commercially developed tests of reading comprehension:

Gold standard or fool's gold? *Reading Ressrs Quarotelyt*

/dois

oliss

2)a6(,? -9(232f)]TJETQq0.00000912 0 612 792 reW* nBT/F4

Deacon, S. H., Kieffer, M. J., & Laroche, A. (2014). The relation between morphological awareness and reading comprehension: Evidence from mediation and longitudinal models. *Scientific Studies of Reading*, *18*(6), 432–451. https://doi.org/10.1080/10888438.2014.926907

Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test Third edition*. American Guidance Service.

Elosúa, M.R., García-Madruga, J.A., Gómez-Veiga, I., López-Escribano, C., Pérez, E. & Orjales, I. (2012). Habilidades lectoras y rendimiento académico en 3° y 6° de Primaria: aspectos evolutivos y educativos. Estudios de Psicología, 33(2), 207–218. https://doi.org/10.1174/021093912800676411

Elwér, Å., Keenan, J. M., Olson, R. K., Byrne, B., & Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology*, *115*(3), 497–516. https://doi.org/10.1016/j.jecp.2012.12.001

Francis, D. J., Fletcher, J. M., Catts, H. W., & Tomblin, J. B. (2005). *Dimensions Affecting the Assessment of Reading Comprehension.* In S. G. Paris & S. A. Stahl (Eds.), *Center for improvement of early reading achievement (CIERA). Children's reading comprehension and assessment* (pp. 369–394). Lawrence Erlbaum Associates Publishers.

Guo, Y., Roehrig, A. D., & Williams, R. S. (2011). The relation of morphological awareness and syntactic awareness to adults' reading comprehension: Is vocabulary knowledge a mediating variable? *Journal of Literacy Research*, *43*(2), 159–183. https://doi.org/10.1177/1086296X11403086

García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*. https://doi.org/10.3102/0034654313499616

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Graham, L., & Bellert, A. (2004). Difficulties in reading comprehension for students with learning disabilities. In B. Wong (Ed.), *Learning about learning disabilities* (pp. 251–279). Elsevier.

Graves, M.F., Ryder, R.R., Slater, W,H., & Calfee. R.C. (1987). The relationship between word frequency and reading vocabulary using six metrics of frequency. *Journal of Educational Research*, SI, 81-90.

Guthrie, J. T., Wigfield, A., Barbosa, P., Perencevich, K. C., Taboada, A., Davis, M. H., Scafiddi, N. T., & Tonks, S. (2004). Increasing reading comprehension and engagement through concept-oriented reading instruction. *Journal of Educational Psychology*, *96*(3), 403–423. https://doi.org/10.1037/0022-0663.96.3.403

Hagaman, J. L., Casey, K. J., & Reid, R. (2012). The effects of the paraphrasing strategy on the reading comprehension of young students. *Remedial and Special Education*, *33*(2), 110–123. https://doi.org/10.1177/0741932510364548

Hagley, F. (1987). Suffolk Reading Scale. NFER-Nelson.

Harris, K. R., & Pressley, M. (1991). The nature of cognitive strategy instruction: Interactive strategy construction. *Exceptional Children*, 57(5), 392–405.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799

Hoover, W. A., & Tunmer, W. E. (2018). The simple view of reading: Three assessments of its adequacy. *Remedial and Special Education*, *39*(5), 304–312. https://doi.org/10.1177/0741932518773154

Hoover, W. A., & Tunmer, W. E. (1993). The components of reading. In *Reading acquisition processes* (pp. 1–19). Multilingual Matters.

Hwang, H. (2019). The role of science domain knowledge and reading motivation in predicting informational and narrative reading comprehension in L1 and L2: An international study. *Learning and Individual Differences*, *76*, 101782. https://doi.org/10.1016/j.lindif.2019.101782

Keenan, J. M., Betjemann, R. S., & Olson, R. K. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*(3), 281–300. https://doi.org/10.1080/10888430802132279

Keenan, J. M., & Brown, P. (1984). Children's reading rate and retention as a function of the number of propositions in a text. *Child Development*, *55*(4), 1556. https://doi.org/10.2307/1130026

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*, *47*(2), 125–135. https://doi.org/10.1177/0022219412439326

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95, 163-182.

Kintsch, W., & Rawson, K. A. (2005). Comprehension. In M. J. Snowling & C. Hulme (Eds.), The science of reading: A handbook (pp. 209–226). Blackwell., https://doi.org/10.1002/9780470757642.ch12

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Kirby, J., Deacon, S., Bowers, P., Izenberg, L., Wade-Woolley, L., & Parrila, R. (2012). Children's morphological awareness and reading ability. *Reading & Writing*, *25*(2), 389–410. https://doi.org/10.1007/s11145-010-9276-5

Kozminsky, E., & Kozminsky, L. (2001). How do general knowledge and reading strategies ability relate to reading comprehension of high school students at different educational levels? *Journal of Research in Reading*, *24*(2), 187–204. https://doi.org/10.1111/1467-9817.00141

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Layton, A., Robinson, J., & Lawson, M. (1998). The relationship between syntactic awareness and reading performance. *Journal of Research in Reading*, 21(1), 5-23.

Lervåg, A., Hulme, C., & Melby-Lervåg, M. (2018). Unpicking the developmental relationship between oral language skills and reading comprehension: It's simple, but complex. *Child Development*, *89*(5), 1821–1838. https://doi.org/10.1111/cdev.12861

Leslie, L., & Caldwell, J. (2001). Qualitative reading inventory-4. New York: Longman.

Lonigan, C. J., Burgess, S. R., & Schatschneider, C. (2018). Examining the simple view of reading with elementary school children: Still simple after all these years. *Remedial and Special Education*, *39*(5), 260–273. https://doi.org/10.1177/0741932518764833

MacGinitie, W. H., & MacGinitie, R. K. (1992). Gates-MacGinitie reading tests (2nd Canadian ed.). Nelson Canada.

MacGinitie, W.H., MacGinitie, R.K., Maria, K. & Dreyer, L.G. (2000). Gates-MacGinitie Reading Tests –Directions for Administration (levels 7/9 & 10/12) (4th Ed.). Riverside.

Magnusson, C. G., Roe, A., & Blikstad-Balas, M. (2019). To what extent and how are reading comprehension strategies part of language arts instruction? A study of lower secondary classrooms. *Reading Research Quarterly*, *54*(2), 187–212. https://doi.org/10.1002/rrq.231

Medler, D.A., & Binder, J.R. (2005). MCWord: An On-Line Orthographic Database of the English Language. http://www.neuro.mcw.edu/mcword/

Metsala, J. L., Sparks, E., David, M., Conrad, N., & Deacon, S. H. (2021). What is the best way to characterise the contributions of oral language to reading comprehension: Listening comprehension or individual oral language skills? *Journal of Research in Reading*, *44*(2). https://doi.org/10.1111/1467-9817.12362

Moravcsik, J. E., & Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. *Canadian Journal of Experimental Psychology*, *47*(2), 360. https://doi.org/10.1037/h0078823

Muhid, A., Amalia, E. R., Hilaliyah, H., Budiana, N., & Wajdi, M. B. N. (2020). The effect of metacognitive strategies implementation on students' reading comprehension achievement. *International Journal of Instruction*, *13*(2), 847–862. https://doi.org/10.29333/iji.2020.13257a

Nation, K., & Snowling, M. J. (2000). Factors influencing syntactic awareness skills in normal readers and poor comprehenders. *Applied Psycholinguistics*, *21*(2), 229–241. https://doi.org/10.1017/S0142716400002046

Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, *67*(3), 359–370. https://doi.org/10.1111/j.2044-8279.1997.tb01250.x

Nation, K., Snowling, M. J., & Clarke, P. (2005). Production of the English past tense by children with language comprehension impairments. *Journal of Child Language*, *32*(1), 117–137. https://doi.org/10.1017/S0305000904006555

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read : an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

Neale, M, D, (1989), The Neale Analysis of Reading Ability-Revised. NFER.

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

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

Pallant, J. (2007). *SPSS survival manual. A step-by-step guide to data analysis using SPSS version 15* (Third Edition). Open University Press.

Perfetti, C., & Stafura, J. (2014). Word knowledge in a theory of reading comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Radvansky, G.A., Zwaan, R.A., Curiel, J.M., & Copeland, D.E. (2001). Situation models and aging. *Psychology and Aging*, *16*, 145–160.

Raudszus, H., Segers, E., & Verhoeven, L. (2019). Situation model building ability uniquely predicts first and second language reading comprehension. *Journal of Neurolinguistics*, *50*, 106–119. https://doi.org/10.1016/j.jneuroling.2018.11.003

Recht, D. R., & Leslie, L. (1988). Effect of prior knowledge on good and poor readers' memory of text. *Journal of Educational Psychology*, *80*(1), 16–20. https://doi.org/10.1037/0022-0663.80.1.16

Reutzel, D. R., Smith, J. A., & Fawson, P. C. (2005). An evaluation of two approaches for teaching reading comprehension strategies in the primary years using science information texts. *Early Childhood Research Quarterly*, *20*(3), 276–305. https://doi.org/10.1016/j.ecresq.2005.07.002

Robertson, E.K. & Deacon, S.H. (2019). Morphological awareness and word-level reading in early and middle elementary school years. *Applied Psycholinguistics*, 40(4), 1051–1071. https://doi.org/10.1017/S0142716419000134

Scott, C. M., & Balthazar, C. (2013). The Role of Complex Sentence Knowledge in Children with Reading and Writing Difficulties. *Perspectives on Language and Literacy*, *39*(3), 18–30.

Shanahan, T., Callison, K., Carriere, C., Duke, N. K., Pearson, P. D., Schatschneider, C., & Torgesen, J. (2010). Improving reading comprehension in kindergarten through 3rd grade: A practice guide (NCEE 2010-4038). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://whatworks.ed.gov/publications/practiceguides

Sparks, E., & Deacon, S. H. (2015). Morphological awareness and vocabulary acquisition: A longitudinal examination of their relationship in English-speaking children. *Applied Psycholinguistics*, *36*(2), 299–321. https://doi.org/10.1017/S0142716413000246

Sparks, E. & Metsala, J.L. (2021) Morphological Awareness Predicts Reading Comprehension in First Grade Students. Manuscript submitted for publication.

Spencer, M., Wagner, R. K., & Petscher, Y. (2019). The reading comprehension and vocabulary knowledge of children with poor reading comprehension despite adequate decoding: Evidence from a regression-based matching approach. *Journal of Educational Psychology*, *111*(1), 1 van der Schoot, M., Horsley, T. M., & Lieshout, E. C. D. M. van. (2010). The effects of instruction on situation model construction: An eye fixation study on text comprehension in primary school children. *Educational Psychology*, *30*(7), 817–835. https://doi.org/10.1080/01443410.2010.510600

van Dijk, T., & Kintsch, Walter. (1983). *Strategies of discourse comprehension*. Academic Press.

Wechsler, D. L. (1992). *Wechsler Individual Achievement Test*. Psychological Corporation.

Wiederholt, J. L., & Bryant, B. R. (1992). *Gray oral reading tests: GORT-3*. Pro-ed.

Wiig, E. H., & Secord, W. (1992). *Test of Word Knowledge (TOWK)*. APA PsycTests.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock Johnson III tests of achievement*. Riverside.

Zhang, L., & Seepho, S. (2013). Metacognitive strategy use and academic reading achievement: Insights from a Chinese context. *Electronic Journal of Foreign Language Teaching*, 10(1), 54-69.

Zwaan, R.A., Langston, M.C., & Graesser, A.C. (1995a). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science*, *6*, 292–297.

Zwaan, R.A., Magliano, J.P., & Graesser, A.C. (1995b). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning Memory and Cognition*, *21*, 386–397.

Zwaan, R. A., & Radvansky, G. A. (1998). Situatiion models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162. https://doi.org/10.1037/0033-2909.123.2.162

Table 1

Means and Standard Deviations for Study Measures for Grades 1, 2, and 3 Students

	Full Sample (*n* = 181)		Grade 1 (*n* = 65)		Grade 2 (*n* = 57)		Grade 3 (*n* = 59)	
Measure	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Reading Comprehension (PC; SS)	96.59	14.98	98.53	13.62	92.19	14.78	98.71	15.59
Reading Comprehension (G-M; SS)	47.27	9.53	47.79	8.55	45.58	9.63	48.32	10.37
Decoding (RS)	10.73	7.02	8.05	5.91	9.38	5.96	14.98	7.36
Vocabulary (RS)	45.22	6.69	41.67	5.86	45.12	5.95	49.22	6.04
Morphological Awareness (RS)	11.18	3.02	10.03	3.17	11.10	2.51	12.53	2.79
Syntactic Awareness (RS)	8.25	3.04	6.92	3.01	8.35	2.87	9.61	2.52

Note: SS = standard score. RS = raw score. PC = Passage Comprehension. G-M = Gates-MacGinite.

Table 2

Zero-order Correlations Among Raw Scores for Study Variables

Measures	1.	2.	3.	4.	5.	6.
Full Sample						
1. Decoding	-					
2. Vocabulary	.47**	-				
3. Syntactic Awareness	.42**	.41**	-			
4. Morphological Awareness	.53**	.46**	.52**	-		
5. Reading Comprehension, PC (SS)	.59**	.28**	.30**	.40**	-	
6. Reading Comprehension, G-M (SS)	.65**	.40**	.36**	.47**	.75**	-

Note. SS = standard scores. PC = Passage Comprehension. G-M = Gates-MacGinite.

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

Table 3

Zero-order Correlations for Grade 1 (top panel), Grade 2 (middle panel), and Grade 3 (bottom panel)

Measures	1.	2.	3.	4.	5.	6.
1. Decoding	-					
2. Vocabulary	.28*	-				
3. Syntactic Awareness	.12	.28*	-			
4. Morphological Awareness	.47**	.50**	.43**	-		
5. Reading Comprehension, PC (SS)	.48**	.33**	.05	.49**	-	
6. Reading Comprehension, G-M (SS)	.54**	.40**	.19	.55**	.71**	-
1. Decoding	-					
2. Vocabulary	.36**	-				
3. Syntactic Awareness	.29*	.24	-			
4. Morphological	.38**	.23	.34**	-		
5. Reading Comprehension, PC (SS)	.71**	.20	.37**	.29*	-	
6. Reading Comprehension, G-M (SS)	.73**	.41**	.42**	.29*	.72**	-
1. Decoding	-					
2. Vocabulary	.40**	-				
3. Syntactic Awareness	.60**	.39**	-			
4. Morphological Awareness	.51**	.31*	.59**	-		
5. Reading Comprehension, PC (SS)	.69**	.43**	.65**	.48**	-	
6. Reading Comprehension, G-M (SS)	.79**	.52**	.59**	.61**	.80**	-

Note. SS=standard scores. PC = Passage Comprehension. G-M = Gates-MacGinitie.

$^{**}p < .01$. $^{*}p < .05$.

Table 4

Hierarchical Regression Analyses Predicting Reading Comprehension (n = 181)

Step	Outcome Measure	Predictor	ΔR^2	β	Final β
Regression 1	Passage Comprehension				
1		Grade	.00	.00	-.35**
2		Decoding	.41**	.70**	.59**
3		Language Comprehension	.03*		
		Vocabulary		.08	.08
		Syntactic Awareness		.09	.09
		Morphological Awareness		.13	.13
Regression 2	Gates-MacGinite				
1		Grade	.00	.20	-.40**
2		Decoding	.49**	.77**	.60**
3		Language Comprehension	.08**		
		Vocabulary		.20**	.20**
		Syntactic Awareness		.10	.10
		Morphological Awareness		.15*	.15*
Regression 3	Passage Comprehension				
1		Grade	.00	.00	-.35**
2		Language Comprehension	.23**		
		Vocabulary		.19*	.08
		Syntactic Awareness		.15	.09
		Morphological Awareness		.32**	.13
3		Decoding	.22**	.59**	.59**

Regression 4	Gates MacGinite				
1		Grade	.00	.02	-.40**
2		Language Comprehension	.34**		
		Vocabulary		.32**	.20**
		Syntactic Awareness		.17*	.10
		Morphological Awareness		.34**	.15*
3		Decoding	.23**	.60**	.60**

Note. Data reported from Imputation 1.

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

Table 5

Consistency of Identification (percentages) of Poor Comprehenders Across Two Tests of Reading Comprehension

Test	16^{th} percentile cut off	25^{th} percentile cut off
PC (% of sample identified)	20.93	25.00
G-M (% of sample identified)	23.84	37.79
PC | G-M (% of sample identified by both tests)	15.70	23.84
PC (% also identified by G-M)	75.00	95.35
G-M (% also identified by PC)	65.85	63.08

Note. PC = Passage Comprehension. G-M = Gates-MacGinite.

Table 6

Comparing C 0 Scores at Grade 1 (bottom diagonal) vs. Grade 2 (top diagonal)

Measures	1.	2.	3.	4.	5.	6.
1. Decoding	-	.47**	.42**	.53**	.59**	.65**
2. Vocabulary	.28*	-	.41**	.46**	.28**	.40**
3. Syntactic Awareness	.12	.28*	-	.52**	.30**	.36**
4. Morphological Awareness	.47**	.50**	.43**	-	.40**	.47**
5. T2 Rdg Comp, PC	.48**	.33**	.05	.49**	-	.75**
6. T2 Rdg Comp, G-M	.54**	.40**	.19	.55**	.71**	-
7. T1 Rdg Comp, PC	.56**	.22	.26*	.54**	.64**	.75**

Note. Rdg Comp = Reading Comprehension. PC = Passage Comprehension. G-M = Gates-MacGinitie.

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

Table 7

Hierarchical Regression Analyses Comparing Predictions for Grade 1 Students First and Second Grade Reading Comprehension

Step	Outcome Measure	Predictor	ΔR^2	β	Final β
Regression 1	Grade 1 PC				
1		Decoding	.31**	.56**	.40**
2		Language Comprehension	.11*		
		Vocab		-.07	-.07
		Syntactic Awareness		.08	.08
		Morphological Awareness		.35*	.35*
Regression 2	Grade 2 PC				
1		Decoding	.19**	.44**	.25*
2		Language Comprehension	.13*		
		Vocab		.13	.13
		Syntactic Awareness		-.19	.19
		Morphological Awareness		.38**	.38**

Note. PC = Passage Comprehension.

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