Oral Language Predictors of Reading Comprehension among Elementary School Children: Does Developmental Language Impairment Make a Difference?

By

Melissa Skoczylas

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Rehabilitation Science

Faculty of Rehabilitation Medicine
University of Alberta

Abstract

Research comparing reading comprehension tests has consistently found that these tests, all designed to measure the construct of reading comprehension, tap various and frequently different component skills. A number of studies have examined the relationship of oral language skills to reading comprehension test scores and have found significant predictive ability of some language skills. The studies examining these tests have sometimes included children with language impairment in their participant group, but have not investigated whether the pattern of results might differ for children with and without language impairment. The current study extends this body of research in two ways: 1) the ability of language skills to predict reading comprehension test scores was examined in relation to language group status (typical development and language impairment); 2) predictors were examined for two reading comprehension tests, one of which had not been examined previously. The two tests included in this study were the *Kaufman Test of Educational Achievement—2nd Edition (KTEA-II)* (not previously studied) and the *Woodcock Reading Mastery Test—Revised/Norms Update (WRMTR).* The *KTEA-II* includes passages of increasing length followed by question-response tasks. The *WRMTR* includes short passages with one word missing. The task is to provide the missing word (a cloze task). Participants were 54 students in grades 4 through 6 (M age 10.08, SD .63, range 9.08-11.5). Thirty students had typical language development; 24 had language impairment. Oral language measures included receptive and expressive measures of vocabulary, morpho-syntax, and narrative language. Other variables tested included decoding, working memory and nonverbal IQ. These variables were not oral language skills and were identified as potential predictors in the literature. Through a series of hierarchical multiple regression analyses, it was found that the two reading comprehension tests compared tapped different language skills, although vocabulary and decoding were strong predictors of both tests. In addition to vocabulary and decoding, *KTEA-II* scores were predicted by narrative skills and *WRMTR* scores were predicted by syntax. The second contribution of this study was the comparison of results for children with and without developmental language impairment. A stratified analysis of the data revealed that the pattern of predictors noted above was only true for children with typically developing language. For children with LI, only decoding and vocabulary were predictive. This result leads to a practice recommendation that reading test results for children with LI be supported by detailed language assessment to support selection of therapy goals.

The methodology of this study also made a contribution to research comparing reading comprehension tests. Language sample analysis was included in the measurements of oral language skills, an approach not previously used in this type of research. The predictive value of receptive and expressive language measures was compared via multiple regression. The results showed that both modes could provide useful results, with a large proportion of shared variance between receptive and expressive measures. It is suggested that the Simple View of Reading, which views reading as the product of decoding plus oral language comprehension, be slightly extended so that the "comprehension" construct be understood as "oral language skills" to more clearly incorporate expressive language skills in the context of reading comprehension.

Preface

This thesis is an original work by Melissa Skoczylas. The research project, of which this thesis is a part, received research ethics approval from the University of Alberta Research Ethics Board, Project Name "Oral Language Predictors of Reading Comprehension among Elementary School Children: Does Developmental Oral Language Impairment Make a Difference?" No. Pro00041690, September 23, 2013.

Acknowledgements

This work was financially supported by the Social Sciences and Humanities Research Council of Canada and the University of Alberta.

I wish to thank Phyllis Schneider, my supervisor, for her guidance and support over the course of my degree. I also wish to thank Denyse Hayward and Alison McInnes, members of my supervisory committee. Thank you all for your expertise and support. Your mentorship has been invaluable.

My research project was made possible by the participation and support of children, families and school staff. I thank all those who kindly shared their time and welcomed me into their schools.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

1. *CASL: Comprehensive Assessment of Spoken Language*
2. *CELF 4: Clinical Evaluation of Language Fundamentals, 4th Edition*
3. *CELF 5: Clinical Evaluation of Language Fundamentals, 5th Edition*
4. FS: Familiar sequences (working memory) (*CELF 4* subtest)
5. *KBIT 2: Kaufman Brief Intelligence Test, 2nd Edition*
6. *KTEA-II RC*: Reading Comprehension Subtest, *Kaufman Test of Educational Achievement, 2nd Edition*
7. *KTEA-II: Kaufman Test of Educational Achievement, 2nd Edition*
8. LWR: Letter and Word Recognition subtest (*KTEA-II* subtest)
9. MLTU-w: Mean length of T-Unit in words (from language sample analysis)
10. NDW: Number of different words (from language sample analysis)
11. NLAI: Narrative language ability index (overall composite score for the *TNL*)
12. NR-T: Number repetition, total score (digit span) (*CELF 4* subtest)
13. NVIQ: nonverbal IQ score (from *KBIT 2)*
14. NwD: Nonsense Word Decoding subtest (*KTEA-II* subtest)
15. PCS: Proportion of complex sentences (from language sample analysis)
16. *TNL: Test of Narrative Language*
17. WM: working memory composite score (*CELF 4* composite score)
18. *WRMTR PC*: Passage Comprehension subtest, *Woodcock Reading Mastery Test Revised—Norms Updated version*
19. *WRMTR: Woodcock Reading Mastery Test Revised—Norms Updated version*

Chapter 1 Introduction

The story of this project begins many years ago, in a Grade 1 classroom. It was the day of our yearly progress test in writing. As a classroom speech-language pathologist, I was helping to proctor the exam, and moved about the room prompting students to give their best efforts. One student in particular was in need of encouragement. When the hour-long exam ended, I saw he had written just three letters, “IGH”. I asked him what his work said. He responded simply, “I go home.”

This child’s response to writing encapsulates the frustration of so many children I have known who struggle to acquire literacy skills, despite honest effort. Such experiences are also frustrating to the professionals seeking to help these children, when all our best efforts fail. My own frustration and the desire for a stronger understanding of literacy led me to embark on a PhD program focused on literacy in children with developmental language difficulties. My original research focus was in assistive technology (AT) for literacy. I was interested in how using tools such as text readers and speech-to-text might help children who had difficulties accessing literacy in standard ways. But in planning a study to investigate AT, I ran into a significant problem: finding the best way to measure reading comprehension in children with language impairment. The search for a solution to this problem led to the current study. Before addressing the problem and resultant research questions directly, it will be necessary to establish some background regarding reading comprehension and children with developmental language impairment.

Background

Reading is a vitally important skill in current society. In order to read effectively, one must be able to efficiently move through a series of processes from decoding individual words to building up a coherent understanding of the text. Underlying these processes are an individual's oral language skills, abilities to understand and use spoken language. Deficits in oral language skills, then, should be expected to lead to reading comprehension difficulties in at least some children. Consistent with this expectation, it is generally accepted by the research community that oral language difficulties put children at risk for long-term reading problems (see for example Nation & Norbury, 2005). Both decoding and comprehension difficulties are common, although not present in every child with oral language problems (Bishop & Snowling, 2004). Thus, many children who are evaluated using reading comprehension tests will have oral language difficulties, yet the impact of impaired oral language skills on the results of reading comprehension tests has received limited research attention.

Many terms are used to describe children with difficulties in oral language skills, such as *language learning difficulties* and *developmental language disorders* (Bishop, 2014). Subgroups are labelled as having *specific* or *nonspecific language impairment* based on level of nonverbal IQ (Catts, Fey, Tomblin & Zhang, 2002). In this study the term *developmental language impairment* (abbreviated as *LI*) will be used. This term designates a group of children with language difficulties that arise in the natural course of development without obvious cause, in the acquisition of a child's first language. It will be used to refer specifically to spoken language as opposed to written language. Less specific assumptions about IQ are made with the LI label, as opposed to SLI (specific language impairment) which typically includes a requirement for non-delayed nonverbal skills (typically above standard score 85). For LI, IQ above 70 is a general guideline, to differentiate these children from those with global developmental delays (Catts, Fey, Tomblin & Zhang, 2002). Given that specific language impairment, a subset of LI that includes a requirement of normal nonverbal IQ skills, has a prevalence of about 7% (Tomblin et al., 1997), LI affects a considerable percentage of children.

Let us turn our attention back to the problem I encountered while planning to study AT for literacy. It was necessary to know the best method of measuring reading comprehension in children with language difficulties, in order to evaluate the impact of AT interventions. As a first step, my colleagues and I completed a literature review focusing on studies that evaluated reading comprehension tests and which included measures of language skills as part of this evaluation (Skoczylas, Schneider, & Suleman, in press). We found 7 studies meeting inclusion criteria (Cutting & Scarborough, 2006; Eason, Goldberg, Young, Geist & Cutting , 2012; Francis et al., 2005; Keenan, Betjemann & Olson., 2008; Kendeou, Papadopoulos & Spanoudis, 2012; Nation & Snowling, 1997; Spear-Swerling, 2004). In each of these studies, which employed a variety of oral language constructs, results showed that reading comprehension tests were not equivalent. That is to say, reported results showed that each test appeared to sample a different set of underlying skills. The tests also varied in their level of dependence on decoding skill. None of these studies included presence/absence of language impairment as an independent variable or evaluated children with LI as a separate group. The original question (what is the best method of measuring reading comprehension for children with oral language impairment?) remained unanswered. Yet this would seem to be a necessary question to answer for the following reasons. The skill profiles of children with LI are different from the typical population, and therefore interpreting reading comprehension test results may require different approaches or additional information. Given that these tests sample different underlying language skills, use of one test could limit our attention to the constructs relevant to that test, even though there may be other areas of strength and need to consider for individual students. Children with LI frequently have varied strengths and weaknesses (Catts et al., 2002, Pennington & Bishop, 2009; Simkin & Conti-Ramsden, 2006); it may be that existing measures of reading comprehension provide only part of the picture in assessment of their skills. Since test results have significant impact on educational decisions made for these children, it is essential that the way their oral language skills impact test performance is more clearly understood. As a first step toward addressing these issues, the body of work that compares reading comprehension tests and illuminates the components of this complex construct should be extended to a group of children with LI.

Purpose

The current study was designed to address the research gap noted above, in comparing standardized reading comprehension tests for children with LI. It was hypothesized that different language variables would be predictive for students with and without LI. Two tests that differ in their reading stimuli and response tasks were selected for comparison. These tests were hypothesized to sample different underlying language skills due to test characteristics. The study was designed both to reveal differences in the skills tapped by the two tests as well as whether the presence/absence of LI made a difference to those results. The study focuses on children in Grades 4 to 6 as this group of children has been less-studied in the research base, yet are expected to learn from reading. Hence effective reading comprehension is a key academic skill expected in this age group.

Rationale

Developmental language impairment (LI) can have long-term impact on an individual. LI has been identified as a significant risk factor for literacy problems in school-age children (Catts et al., 2002, Nation et al., 2004). Up to 50% of children with reading comprehension difficulties identified during their school years can be shown to have pre-existing oral language problems (Nation et al., 2004). For children known to have specific language impairment more than 60% can be expected to develop reading problems (Catts et al., 2002). Children with LI and lower nonverbal skills are at even higher risk of reading problems (Catts et al., 2002). Even when oral language difficulties appear to resolve in preschool, these children are still at increased risk of literacy problems (Catts et al., 2002; Scarborough & Dobrich 1990; Simkin & Conti-Ramsden, 2006; Stothard et al., 1998).

Despite the research base linking early and persistent language difficulties to literacy problems, the specific relationships between oral language skills and reading development remain unclear. This is particularly the case for reading comprehension (Oakhill & Cain, 2007); the oral language correlates of decoding skills are relatively better-studied. When oral language skills are included in studies, researchers have sometimes factored out language variables, or studied a limited variable set (Oakhill & Cain, 2007). Specific issues should be considered when designing research to identify the oral language skills that determine reading comprehension, including carefully specifying the constructs under consideration (Barnes et al., 2007; Skoczylas, Schneider, & Suleman, in press), identifying environmental, task and text characteristics that impact functioning (Barnes et al., 2007, Francis et al., 2005; Skoczylas, Schneider, & Suleman, in press), and carefully selecting literacy assessments that clearly tap the constructs of interest (Barnes et al., 2007; Cain & Oakhill, 2007). This last strategy is complicated by the fact that different reading comprehension assessments do not provide equivalent results as they tap different underlying cognitive and language skills (Cutting & Scarborough, 2006; Francis, Fletcher, Catts & Tomblin, 2005; Keenan, Betjemann, & Olson, 2008; Kendeou, Papadopoulos, & Spanoudis, 2012; Nation & Snowling, 1997; Spear-Swerling, 2004).

To date, standardized reading comprehension tests have not been evaluated specifically for children with LI (Skoczylas et al., in press). Yet due to the high prevalence of reading problems associated with LI, these children may often be the ones assessed using these tests in practice. Understanding the relationship of atypical language learners' skills to assessment results both assists in understanding what is being tested (i.e., clarifying underlying constructs of the assessment) and in interpreting test results correctly.

The results of children with LI may differ quantitatively or qualitatively from children with typically developing language. For example, poor comprehenders, a group of children with poor reading comprehension in the presence of adequate decoding skills, can have a variety of skill profiles. There is no one central language or cognitive deficit that leads to poor comprehension in this group (Cain & Oakhill, 2007). Currently, it is not well understood how atypical language skills may affect the ability of children to perform adequately on reading comprehension tests. Test characteristics such as response type (cloze, multiple choice, etc.) and stimulus text characteristics (Cutting & Scarborough, 2006; Francis et al., 2005; Keenan, Betjemann & Olson, 2008) may significantly impact results differentially for children with LI relative to children with typically developing language. For example, expressive language weakness can impact the ability to answer reading comprehension questions (Ricketts, 2011; Spooner, Baddeley, & Gathercole, 2004).

The current study is designed to examine the question of how oral language skills relate to scores on standardized reading comprehension tests for children with and without LI. Two instruments were selected, the *Kaufman Test of Educational Achievement—Second Edition*

(KTEA-II) and the *Woodcock Reading Mastery Test—Revised/Norms Update (WRMTR).* The *WRMTR* was selected in order to facilitate comparisons with the extant literature, as it has been previously studied. The *KTEA-II* was selected as a measure that has a variety of question types, including those that are predicted to tap higher-level language skills known to be relevant to the development of reading comprehension in older students. It has the added advantage of being a test in common usage in local school districts, making the results of this study relevant to local stakeholders as well as to those in other areas who have adopted this test.

This study offers three unique contributions to the research on comparing reading comprehension tests. First, this appears to be the first study that analyzes the results in terms of presence/absence of LI. Whether differences between the groups are found or not, these results will be of value. If LI makes a difference in terms of what predicts reading comprehension, then this could help to inform selection and interpretation of reading comprehension test results. If LI makes no difference, then current practices in test selection are validated in this regard. Second, this study evaluates the reading comprehension subtest of the *Kaufman Test of Educational Achievement, Second Edition*. This test has not been evaluated in a study of this type. Examining the skills that are tapped by this test should assist in interpretation of test results. Third, this study has two novel approaches to methodology. Language sample analysis is a method that will be used to measure expressive language skills. This method has not previously been included in a study of this type. Since language sample measures provide an alternative to standardized tests, it will be of interest to see how they function as predictors. The use of expressive measures is also of interest. In most studies, the focus has been on receptive measures. To compare the two approaches, the relative contributions of receptive and expressive language measures will be examined.

Research Questions

1. Given receptive and expressive measures of vocabulary, syntax and narrative discourse for children in grades 4 to 6, which of these oral language scores will predict scores on two concurrently measured standardized reading comprehension tests?
2. Will there be an effect of group membership such that significant predictors for children with developmental language impairment will differ from significant predictors for children with typically developing language?

Overview of the Thesis

The current study is a correlational design with multiple regression analysis, examining oral language predictors of scores on two standardized reading comprehension tests, in children with and without LI in grades 4 to 6. The research questions focus upon naturally occurring relationships among reading, oral language and cognitive skills. In Chapter 2, the literature is reviewed to present current theory and findings regarding knowledge about relationships among language, language impairment, and reading. Possible predictor and control variables are identified and evaluated in the context of this study. Methodological issues to consider in this type of research are presented. Chapter 3 presents the methodology of the study and detailed description of participants, measures, procedures and data analysis. As a brief summary of methods, participants were recruited from a local school district as existing groups; therefore no experimental manipulations were applied. Participants completed a battery of oral language, cognitive, word reading and reading comprehension tasks. Multiple regression techniques were used to determine which oral language, word reading and cognitive scores combine to predict reading comprehension scores for the reading comprehension tests. Language status was included as a predictor to identify differences between the participant groups (children with and without LI). The rationale for selecting multiple linear regression as the statistical technique is detailed in the "Data analysis" section. Chapter 4 presents the results of the regression analyses in two parts. The first presents the original planned regressions, including problems and questions that arose during these analyses. The second part presents the analyses that were undertaken to address the problems and questions from Part 1. Finally, the results of the study are discussed and linked to the existing research base in Chapter 5.

Chapter 2 Literature Review

A significant body of literature exists regarding reading comprehension and language in school-aged children. The first section of this chapter, “Oral Language and Reading”, presents key theories and findings regarding oral language, reading and language impairment. Next, potential independent variables for the current study are identified and evaluated in the section “Relevant Variables in Language and Literacy Research”. A summary of this review presents the variables selected for the current project. The sections “Methodological Issues in Studying Reading Comprehension” and “Measuring Oral Language” present methodological issues that were considered in planning the current study. The section “Hypothesized Results” extends from this literature base to provide hypotheses for the current study’s research questions.

Oral Language and Reading

Oral language, in the simplest terms, refers to the ability to understand and use spoken words in communication with others. In the context of reading research, the related term “listening comprehension” is frequently used; it generally refers to some aspect(s) of receptive language. It has been defined by various measures, such as receptive vocabulary and the ability to respond to questions about a passage one has heard (Skoczylas, Schneider, & Suleman, in press). The importance of listening comprehension to reading is captured in the Simple View of reading (Gough and Tunmer, 1986). This model presents reading as “the product of decoding and comprehension” (Gough and Tunmer, 1986, p.7). Within this framework, listening comprehension and reading comprehension are equivalent in the context of normal development (Gough & Tunmer, 1990). The Simple View has been validated empirically, including by principle components factor analysis of reading comprehension tests, demonstrating two factors corresponding to decoding and comprehension (Bishop & Snowling, 1997; Keenan, Betjemann & Olson, 2008). Recent work evaluating the components of the Simple View has demonstrated that decoding and comprehension are not entirely separable, due to the influence of vocabulary on both components as well as the influence of comprehension on decoding ability (Tunmer & Chapman, 2012). Perfetti, Landi and Oakhill (2005) clarified the Simple View by characterizing the effects of inferencing, semantics, syntax, and working memory on the development of reading comprehension. Language skills are generally conceptualized in five domains: semantics, morphology, syntax, pragmatics and phonology. Each domain can be expected to impact reading comprehension as they are involved in the component processes of reading (Snowling and Hulme, 2005). As will be explored below, the Simple View is a useful lens to consider reading, but by itself may not encourage the assessment of all oral language skills necessary to fully understand the components of reading comprehension in children with developmental language impairment (see also Ricketts, 2011 regarding working memory and nonverbal ability).

An extensive literature base demonstrates that oral language and literacy skills are closely related in normal development. More recently, work has focused on the relationship of oral language to reading comprehension. In early reading stages, word recognition, which includes decoding and identification of the word in memory, is a strong predictor of reading comprehension. As a reader gains decoding skill, reading comprehension becomes more clearly tied to listening comprehension. In mid-to-late adolescence, reading and listening comprehension are almost perfectly correlated among children with typical language development, consistent with the Simple View of Reading (Adlof, Catts & Little, 2006; Oakhill & Cain, 2007). As children develop skill in reading, the relationship between oral language and reading changes and children begin to acquire new language from the reading material itself (see for example Catts & Kamhi, 2005). Juel (1988) found that poor readers fell behind good readers in listening comprehension by Grade 4, even when they started with equivalent listening comprehension. Similar results were obtained by Share & Silva (1987); poor readers showed less growth in language skills than good readers between 3 and 11 years of age. This is termed the *Matthew Effect*; such an effect can put a child at risk of academic delay as a result of limited language and literacy development (Stanovich, 1986). Not all studies have confirmed the Matthew effect (see for example Protopapas, Sideridis, Mouzaki, & Simos, 2011). It does appear that even though the gap between good and poor readers may not widen, neither does it close. When evaluating such studies, it is important to note the upper ages of the participants; if they are within the early school grades, then real effects of impaired oral language skills may be missed (Stothard, Snowling, Bishop, Chipchase & Kaplan, 1998). A child may have adequate language skills to cope with the concrete topics and basic language forms addressed until about Grade 3, yet struggle in later grades when material becomes more abstract and more linguistically challenging.

Given that children with LI can have various levels of difficulty with one or more of the five language domains (semantics, syntax, morphology, pragmatics and phonology) and that these domains are expected to impact reading comprehension, it is to be expected that their oral language skills will impact their ability to learn to read. A comprehensive review by Bishop and Snowling (2004) comparing specific language impairment (SLI) and dyslexia presented a variety of research results demonstrating the pervasive effects of oral language difficulties on reading development. Studies examining families at risk for reading difficulties revealed consistent evidence of preschool language difficulties within these families (Bishop & Snowling, 2004). The review also provided evidence that both decoding and comprehension difficulties are to be expected in children with SLI. However, they found that speech difficulties alone are not typically predictive of reading troubles. They can be associated with increased reading problems when severe and persistent or associated with additional language deficits (Bishop & Snowling, 2004). A particularly interesting result was reported by Stothard and colleagues (1998). They followed up a group of children who had "resolved" language difficulties at 5 ½ years of age as reported by Bishop and Adams (1990). In the re-evaluation at 15 years of age, the children were found to have made lower than expected gains in reading accuracy. The existence of "late emerging poor readers" has been reported by other authors (see for example Adlof et al., 2010). Much reading research has focused on Grades 1 to 3; however, the observation of late emerging poor readers and the increasing language demands in the upper grades suggests that research should also focus on language and literacy in older students. If older children are not considered, relevant phenomena may be missed, as noted above in the discussion of the Matthew effect. Stahl and Hiebert (2005) state, "As the process of word recognition demands fewer resources, comprehension of written language would become more like comprehension of oral language." (p.174). This statement predicts that reading comprehension impairments should mirror oral language impairments in students with LI and adequate decoding skills. A sizable body of work is accumulating relating to poor comprehenders, students who have difficulty with reading comprehension but not with decoding. Consistent with the above prediction, this group overlaps with children with LI in terms of displaying oral language problems, even though they are not typically clinically identified as having language problems (Bishop & Snowling, 2004; Nation et al., 2004). As expected from their adequate decoding skills, poor comprehenders tend to have adequate phonological skills (Bishop & Snowling, 2004; Nation et al., 2004).

Currently, it is difficult to link a given profile of oral language skills to expected reading outcomes. Some models of reading comprehension do specify oral language skills and relate them to the reading process. An example is the model presented by Perfetti, Landi and Oakhill (2005, p. 229) (see also Cain 2010, p. 151). The Construction Integration Model of reading provides a theoretical basis for the current study. The Construction Integration Model presents key aspects of reading comprehension in the fluent adult reader. The outcome of reading is conceptualized as a mental model of the text, formed of both the material inherent to the text (the "text base") as well as material from the reader's interpretation of the text (the "situation model") (Kintsch & Rawson, 2005). Thus the Construction Integration Model clearly expresses the idea that two readers approaching the same text can come away with two different understandings of the text. These differences can be due, for example, to the receptive language skills involved in forming the text base or to inferencing and background knowledge used to form the situation model. In the Construction Integration Model, the specific role of language skills is expressed primarily in terms of extracting propositions at lower levels of processing (Kintsch, 1998).

For the purposes of this study, it was important to consider the language domains typically evaluated in clinical practice (phonology, semantics, syntax, morphology and pragmatics) and how they might impact reading comprehension. The Construction Integration Model was selected as a useful platform to expand from in terms of language skills. The proposed expansions are captured as the Oral Language/ Construction Integration (OL/CI) Model. We formulated these expansions with the goal of providing a highly-specified account of the current evidence regarding the role of language skills in reading comprehension among children with and without LI (Skoczylas & Schneider, 2014). The OL/CI model represents linkages between elements of the Construction Integration Model and language skills in the five domains. In creating the OL/CI model, the literature was reviewed for evidence regarding language skills that appeared to be causally linked to reading comprehension (see literature presented in “Relevant Variables in Language and Literacy Research”). These language skills are captured in the links between various elements in the model. Other non-language elements that are internal to the reader and affect reading comprehension are represented, such as motivation and cognitive skills (Guthrie & Wigfield, 2005). Also integrated into the model are important factors external to the reader, including text characteristics and environmental conditions. Even though the Construction Integration Model is presented as a model of adult performance, it was found to be useful in a developmental context by building in a small number of assumptions. These assumptions are included in a set of extensions to the Construction Integration Model that I have proposed. First, it is assumed that the model represents functions needed for complete and successful performance. It is further assumed that developmental progress and deficits can be represented by adjusting the linkages in the model. For example, if a process is not expected at a given age, that link is assumed to be non-functioning. If a process is limited by deficits, then the link can be interpreted as weak or inconsistent, resulting in altered performance “downstream” from that link. This is a helpful approach when dealing with children with LI, as it encourages the consideration of all relevant skills, consistent with the multiple deficit approach (Pennington & Bishop, 2009). This approach incorporates the idea that performance is measured at a behavioural level, and therefore a variety of underlying skill profiles may result in a similar performance. The Construction Integration Model with the proposed extensions illustrated is included in Appendix A.

Relevant Variables in Language and Literacy Research

An interesting variety of constructs have been investigated relative to reading comprehension. These range from oral language skills to cognitive skills such as working memory, to personal and environmental variables such as motivation and socio-economic status. These constructs are presented here in categories that emerged naturally from the literature: general control variables, word recognition and related skills, oral language skills, and cognitive-linguistic skills. Each variable will be evaluated for applicability to the current study.

General controls.

Mother's educational level. Adlof and colleagues (2010) found that mother's educational level (related to socio-economic status), measured when the child was in kindergarten, predicted children's reading comprehension in Grade 2 and Grade 8. For the current study, statistical testing was used to determine if the two participant groups were equivalent on this measure.

Age. Reading comprehension appears to depend upon different skills in younger and older students. In younger students, more variance is accounted for by decoding skills than by language comprehension. As decoding skills become automatized, more reading comprehension variance is related to language comprehension than to decoding. In the case of this study, age-based standard scores are used for all dependent and independent variables. Use of age-based standard scores ensures that the students' skills are compared to an appropriate criterion of performance. For this reason, age was not used as a separate predictor.

Gender. The poor comprehender profile (readers who have relatively poorer reading comprehension in the presence of adequate decoding skills) appears to be more common in girls (Adlof et al., 2010). Girls also appear to be more commonly affected by late-emerging reading difficulties (Adlof et al., 2010). However, approximately 65% of the children with LI are expected to be boys (McGuinness, 2005). Because LI status was only determined after testing in the current study, the distribution of males and females in the participant groups could not be matched, but any significant inequality was noted for possible impact on the findings.

IQ. Cognitive skills are often controlled for in reading comprehension research as the tasks of interest tap these skills (Bowey, 2005). Catts and colleagues have argued that IQ measures tap language skills and therefore limiting participants based on IQ will reduce valid variance related to oral language skills (Catts et al., 1999). Nonverbal IQ has commonly been used as a control variable in reading studies, perhaps as an option to control variance related to general intelligence, without pulling out language-related variance associated with verbal IQ. However, differences have been found in the reading results of children with specific language impairment (SLI) and nonspecific language impairment (NLI) showing that lower nonverbal IQ was associated with poorer reading (Catts et al., 2002). These two diagnoses differ in that SLI is associated with normal nonverbal IQ, whereas NLI involves nonverbal IQ below expected levels for a child's age. Corresponding results were found by the same authors in children without language impairment; lower nonverbal IQ was associated with lower reading outcomes. Finally, nonverbal IQ measured in kindergarten showed independent prediction of reading skills (comprehension and word reading) in Grades 2 and 4 (Catts et al., 2002). These authors acknowledged that their results did not specify the nature of the relationship between their nonverbal tasks and the reading tasks, allowing for the existence of possible higher-level language covariates. In another study, kindergarten nonverbal IQ was found to predict reading impairment in Grade 8, but not Grade 2 (Adlof et al., 2010). Again the authors suggested the influence of higher-level factors, such as problem-solving skills. At least one review found that nonverbal IQ accounted for variance in reading comprehension scores among children with LI (Ricketts, 2011).

Bishop and Snowling (2004) take a different stand and suggest that nonverbal IQ is simply not an appropriate variable in this type of research as there is insufficient evidence to link it to reading development. They cite the general trend away from the use of IQ-discrepancy formulas for specifying developmental disorders such as dyslexia. This opinion is shared by Dethorne and Watkins (2006), who examined the relationship between nonverbal IQ and language abilities, and found limited evidence to support using IQ as a selection criterion for language group membership (e.g., SLI). Cutting and Scarborough (2006) found that full-scale IQ was not a significant predictor of reading comprehension.

Despite these various opinions, nonverbal IQ continues to be a typical control variable in studies relating oral language abilities to reading comprehension skills. In the context of this study, nonverbal IQ was measured so that its use as a predictor could be tested.

Word recognition and related skills

Decoding and word recognition. Word recognition is the first step in comprehension; if one cannot read the words on the page, one cannot comprehend. Accordingly, measures of decoding and word identification predict reading comprehension in the early stages of reading instruction (De Jong & van der Leij, 2002; Muter et al., 2004) and later reading comprehension for those with LI (Palikara, Dockrell, & Lindsay, 2011) . Word recognition is generally thought to be related to phonological, morpho-syntactic and semantic language skills (see for example Bishop & Snowling, 2004; Perfetti, 2007). In terms of cognitive skills, short-term memory is closely related to word recognition (Swanson, Howard, & Saez, 2007). The appropriate measure used to evaluate word recognition has not always been agreed upon. It seems that to sample relevant skills, nonword decoding, exception word reading and regular word reading are all revealing (Bishop & Snowling, 2004). Nonword decoding provides a measure of grapheme-phoneme correspondence, and is less dependent on other language skills.

Once word recognition is adequate, other variables account for the majority of variance in reading comprehension. In assessing children with language impairment, it is important to recognize that their individual skill profiles vary, and so age-related expectations of word recognition may not apply. For children with LI, who may have deficits in the complexity of their semantic networks (Paul & Norbury, 2012), one might expect vocabulary difficulties to negatively impact word recognition. Previous research with children with language impairment suggests that both vocabulary and decoding scores will remain predictive of reading comprehension after they are no longer relevant for typically-developing peers (Palikara, Dockrell, & Lindsay, 2011). Although controlling word recognition may impact variance related to semantics, word recognition also constrains reading comprehension and so must be accounted for. In the current study, both nonword reading and word recognition were measured.

Letter identification / print knowledge. Letter identification significantly predicts reading outcomes (Adlof et al., 2010; Catts et al., 2002); however, it is unclear what this measure actually samples. Training in letter names does not appear to enhance reading skills (Bowey, 2005). It may serve as an estimate of experience with literacy activities or correlate with relevant cognitive skills (Bowey, 2005; Catts et al., 2002; Paris, Carpenter, Paris & Hamilton, 2005). Measures of letter identification are of necessity subject to ceiling effects as there is a limited set to be learned, typically to mastery in a narrow window of time (Paris et al., 2005). Paris and colleagues caution that statistical analyses of letter identification and other print knowledge variables must be treated cautiously. As knowledge to be mastered, subject to a narrow timeline of development, such data are not normally distributed. For children beyond Grade 3, even in a clinical sample, the majority of students will have achieved ceiling. Therefore this variable was not measured in the current study as it is unlikely to add significant predictive value to a regression model for an older elementary school group.

Rapid Automatized Naming (RAN). RAN is a somewhat controversial correlate of reading skills. Although early RAN skills do predict later reading (see for example Adlof et al., 2010), interpretation of results is unclear (Bowey, 2005). As with letter identification, multiple constructs may be sampled by this measure. Typically, deficits in RAN are associated with phonological processing difficulties (Bowey, 2005; Catts et al., 2002). RAN measures may also sample letter knowledge (in the case of RAN tasks with letters) and aptitude for paired associate learning (McGuinness, 2005). This last skill is implicated in learning to link letter sounds with letter forms, an essential skill for decoding, and thus may explain the predictive value of RAN (McGuinness, 2005). It is possible that the RAN task used by Catts et al. (2002) contains a processing element that goes beyond phonological processing. The task required the children to name coloured animals (e.g., red cow) presented in random order. This was to eliminate learning effects of letter or number naming (i.e., some children might not know all letters or numbers). As the colours are atypical, the child would need to suppress the relevant automatic colour association (e.g., brown cow) in order to respond correctly. Suppression of irrelevant information has been posited as a feature of comprehension processing (Barnes, Johnston, & Dennis, 2007; McNamara & Magliano, 2009). Therefore the RAN task might sample processing skills associated with "higher-level" functions. Children who were faster on this task might be faster at suppressing irrelevant information in forming a mental representation of text. This is supported by the result reported by Catts and colleagues that RAN was a significant predictor of reading comprehension in both Grade 2 and 4. Based on the difficulty of interpreting the results gained from this variable, its inclusion was unlikely to provide explanatory power in the context of the current study.

Phonological awareness (PA). A sizable literature base has accumulated regarding the role of phonological awareness in reading acquisition. Although it is consistently found to predict reading skills (see for example Adlof et al., 2010; Catts et al., 2002), the causality of this relationship remains a matter for debate. As pointed out by Bowey (2005), some level of PA is necessary to comprehend an alphabetic writing code. However, it has also been demonstrated that PA is strongly affected by learning to read (Morais & Kolinksy, 2005). PA also becomes less predictive over time (see Oakhill & Cain, 2007), likely reflecting its relationship with word recognition. Accordingly, Oakhill and Cain (2007) note that results showing that PA predicts reading comprehension may have word recognition as an intervening variable. PA tasks may also be tapping working memory due to the need to remember and manipulate sound units (see for example De Jong & van der Leij, 2002). PA may also be a kind of "learnability measure" that taps cognitive processes that are key to reading in an alphabetic code (e.g. auditory sequential memory, coping with decontextualized information) (Bowey, 2005). Finally, it is challenging to isolate phonological and lexical semantic skills as they are intertwined in normal word recognition (Bishop & Snowling, 2004; Nation, 2005; Snowling & Hulme, 2005). Poor comprehenders are typically found to have PA skills within the normal range (Nation, 2005), but phonological processing difficulties (including PA difficulties) are often found to be central in specific language impairment (Bishop & Snowling, 2004).

Based on the above considerations, PA was not directly sampled in this study. PA is primarily of interest in how it impacts decoding skills; thus decoding itself was measured.

Phonological memory. Closely related to phonological awareness is phonological memory. Particular attention has been paid to nonword repetition as a potential marker for SLI and dyslexia; however the jury is still out on what exactly is revealed by a deficit in nonword repetition. It is generally agreed that nonword repetition measures tap phonological memory. Some authors suggest a nonword repetition deficit is a reliable indicator of language difficulties, even subtle ones (see for example Nation et al., 2004). Catts and colleagues, however, suggest that a nonword repetition deficit is indicative of dyslexia, attributing most (but not all) of the observation of poor nonword repetition in children with language impairment to comorbidity of the two disorders (Catts et al., 2005). The results of their study designed to tease out differences between SLI and dyslexia demonstrated a continuum of ability demonstrating that the most severe nonword repetition deficits can be associated with dyslexia (Catts et al., 2005). This result is consistent with the observation that poor comprehenders, as a group, do not show difficulties with nonword repetition (Nation et al., 2004). Poor comprehenders have adequate decoding skills by definition, and therefore should not be expected to show deficits associated with decoding weakness such as nonword repetition problems. Based on the preceding evidence, ability with nonword repetition is most closely associated with decoding skills. Although nonword repetition was not sampled directly in the current study, a nonword decoding task was used as previously discussed.

It must also be noted that phonological memory can be conceptualized as a component of short term memory (Swanson et al., 2007). Consistent with the above-noted research, measures designed to tap short-term memory in readers of different abilities (word recognition, reading comprehension and IQ) sorted reading groups similarly (Swanson et al., 2007). Skilled readers and poor comprehenders, who display similar word recognition skills, outperformed poor readers with decoding difficulties on STM tasks. Compare these results also with results presented in the section below on working memory.

Sentence imitation is another task heavily dependent upon phonological memory (Bowey, 2005), although it is also used as a grammar task. Note that both children with SLI and poor comprehenders have been found to have difficulties with this task (Nation et al., 2004). The sentence imitation task will be further examined in the section on syntax measures. It was not selected as an explanatory variable for this study due to the difficulty of interpreting what a deficit/strength on this task actually indicates about oral language skills. It is included in the composites used to identify language impairment.

Oral reading fluency. Oral reading fluency encompasses accuracy, rate and prosody, although prosody is not frequently measured (Stahl & Hiebert, 2005). Fluency has been used as a measure of reading comprehension as they typically correlate well (Fuchs et al., 2001). It is generally recognized that slow, effortful decoding will constrain comprehension as the demands of the task may exceed the cognitive capacities of the reader (see for example, Nation, 2005; Stahl & Hiebert, 2005). Despite this recognition, the use of oral reading fluency to measure reading comprehension has been questioned based on a number of concerns. There are observations of students who are able to read at an expected rate but do not comprehend the text adequately; these children are sometimes called “word callers” (Paris et al., 2005). Compare this with descriptions of “hyperlexia”, a term describing fluent reading combined with poor comprehension, sometimes associated with autism (see for example Snowling & Hulme, 2005). In both cases, reading fluency is not reflective of reading comprehension. The premise of the Simple View is that high word recognition accuracy is necessary for reading comprehension; necessarily then low accuracy (and consequently low fluency) is associated with low comprehension. Given this association, low reading fluency is an effect and not a cause (Paris et al., 2005). Accordingly, training in oral reading fluency does not appear to significantly improve reading comprehension (Oakhill & Cain, 2004). Considering fluency measures for children with expressive language impairment, their reading comprehension may be underestimated as they may have a reduced rate of speech that is inconsistent with their level of comprehension (Ricketts, 2011). As noted previously, poor comprehenders have reading accuracy skills that are not consistent with their level of reading comprehension.

Stahl and Hiebert (2005) suggest that rate of reading may be useful in younger students who are developing word recognition skills, but by Grade 3 most of the variability is reduced. These authors also note that increasing text complexity will increase the load on cognitive and language skills as children move up in school, shifting the relevant predictors away from word recognition and related measures (Stahl & Hiebert, 2005).

Given the above evidence, oral reading fluency was not expected to be a reliable predictor of reading comprehension, in the context of this study focusing on students in Grade 4 to 6 with variable profiles of oral language, decoding and reading comprehension.

Oral language skills

Oral language skills have been measured in a variety of ways. In the following section, measures of receptive and expressive language are considered first. Then measures of overall language skill are considered. Next, composite measures of language skills in semantics, morpho-syntax and narrative skill are reviewed. Finally, measures in the specific domains of semantics and morpho-syntax are reviewed.

Oral language comprehension /receptive language. It is widely accepted that oral language comprehension (also called receptive language) is closely tied to reading comprehension. In the Simple View of reading, they are posited as the same processes once automatic word recognition is achieved (Gough and Tunmer, 1986). Both concurrent and longitudinal studies suggest that listening comprehension is important for reading comprehension. The construct of "listening comprehension" is not defined equivalently in all studies, with measures among studies varying from receptive vocabulary to the ability to answer questions about orally presented passages (Skoczylas et al., in press). Evidence from adults (Nation, 2005), poor comprehenders (Nation, 2005) and children with specific language impairment (Kelso et al., 2007) suggests the close relationship of listening and reading comprehension. Early listening comprehension skill has been identified as one of the most important predictors of later reading comprehension (De Jong & van der Leij, 2002; Van den Broek, 2005). In studies showing that early reading comprehension is predictive of later comprehension (see for example Catts et al., 2002), it is likely that some of the variance controlled by the autoregressor is shared with listening comprehension. Since comprehension is a broad term, selection of comprehension (i.e., receptive language) variables will be explored below in the context of specific areas of oral language.

Oral language production / expressive language. The Simple View of Reading specifies oral language comprehension as a component of reading comprehension, but does not refer to expressive language skills. A number of studies have demonstrated the relationship of productive or expressive oral language skills to reading comprehension. Simkin and Conti-Ramsden (2006) conducted research with groups of children with specific language impairment (SLI) who varied in terms of receptive language and/or expressive language deficits. A key result from Simkin and

Conti-Ramsden's study was that 73% of children with only an expressive oral language deficit showed reading comprehension problems, a result not predicted by the Simple View of Reading. It has also been found that correct responses to inferential questions on narrative text were more strongly correlated to expressive than receptive language skills (McClintock, Pesco, & Martin-Chan, 2014). For the population of children with LI, inclusion of expressive language measures has the potential to reveal important relationships between oral language and reading comprehension (Skoczylas, Schneider, & Suleman, in press). Expressive language will be discussed further in specific areas of oral language.

Severity of language disability on standardized tests / overall severity. Rather than measuring receptive or expressive oral language skills, some studies have examined overall severity of language difficulties in relation to reading comprehension skills. Results consistently indicate that the more severe and persistent the oral language difficulties, the higher the risk of reading problems (Bishop & Adams, 1990; Catts et al., 2002; Simkin & Conti-Ramsden, 2006; Skebo et al., 2013). Simkin and Conti-Ramsden (2006) reported that word reading and reading comprehension were affected more in children with expressive language difficulties than those with resolved language difficulties; children with combined receptive and expressive deficits were more severely affected than those with only expressive deficits. These authors also commented on significant variation between individuals in their sample of 11 year-old children. As was noted regarding profiles of poor comprehenders, heterogeneity appears to be the rule rather than the exception when examining oral language skill profiles that lead to reading comprehension problems. This heterogeneity is potentially due to many cognitive and linguistic factors contributing to one behavioural outcome, consistent with the multiple-deficit approach (Bishop & Snowling, 2004; Pennington & Bishop, 2009). In designing a study investigating oral language predictors of reading comprehension, the above research suggests that it is important to capture both receptive and expressive language skills and a range of oral language skills, to best represent individual skill profiles.

Language composites. Some studies have examined oral language skills in a more holistic approach, reporting results that include more than one language domain in composite scores. As with other research, results suggest a close relationship between the development of good oral language skills and good reading comprehension. Language skills measured as vocabulary and narrative ability have been found to predict reading comprehension after the initial decoding-focused stage (Storch & Whitehurst, 2002). Similarly, oral language measured as a composite of vocabulary, syntax, and narrative retelling predicted reading comprehension in Grade 2 (Catts et al., 1999). Concurring results were reported by Snyder and Downey (1991) who found that narrative skills were significant predictors of reading comprehension in children with typical reading skills. The constructs of vocabulary and syntax are further explored below in the sections on semantics and morpho-syntax. Narratives are considered in the section on cognitive-linguistic skills.

Semantics. Semantics, or language content, deals with meaning. Although vocabulary is sometimes equated with this domain, semantics may be considered at the word, sentence and text/discourse level (Scott, 2011), as well as in receptive and expressive modes. In reading research, word-level semantics (vocabulary) has received a great deal of attention.

Vocabulary has consistently been found to strongly predict reading comprehension, even remaining relevant years later (see for example Bowey, 2005; Paris et al., 2005); however, the overall consensus is that vocabulary and reading comprehension develop together, each supporting the development of the other (see for example Oakhill & Cain, 2007; Perfetti et al., 2005). In fact, vocabulary has been found to be more affected by reading than other areas of semantics (Morais & Kolinsky, 2005). Consistent with these views, there are significant gaps in vocabulary between students with high and low vocabulary knowledge, and the gap increases with age (Perfetti et al., 2005). It is also notable that vocabulary may be a proxy for other skills; receptive vocabulary skills have been used to estimate verbal IQ for example (see for example Catts et al., 2005). Vocabulary also relates to background knowledge; if you have knowledge of a topic, you likely have the specialized vocabulary that goes with it (Stahl & Hiebert, 2005).

Diversity in the words a child uses suggests a broad vocabulary that is used to good effect in expressing specific ideas. Students with developmental language impairment often have vocabulary weaknesses, demonstrating restricted breadth and depth of knowledge of words, meanings, and their interrelationships (Paul & Norbury, 2012). Work with poor comprehenders has revealed vocabulary deficits in semantic relation tasks, but not corresponding phonological tasks (Nation, Clarke, Marshall, & Durand, 2004; Nation & Snowling, 1998). Deficits in both expressive and receptive vocabulary are reported for this group (Nation, 2005). According to the Lexical Quality Hypothesis (Perfetti, 2007), rich and elaborated knowledge of vocabulary is central to efficient reading and consequently to comprehension. At this time, it remains unclear what kind of mental representation of text is constructed by children with the vocabulary deficits typically seen in LI (Perfetti et al., 2005). It therefore appears that measures of vocabulary, particularly those revealing depth of knowledge such as word relationships, are important in measuring oral language skills that relate to reading comprehension, especially for children with LI.

Beyond the word level, results are more difficult to interpret. Sentence comprehension tasks, for example, involve understanding meaning that is impacted by how a sentence is constructed (syntax) (Scott, 2009). Cloze tasks (used in some reading comprehension tests) may be affected by knowledge of word meanings as well as knowledge of sentence structure (Oakhill & Cain, 2007). Therefore, both syntactic and semantic skills are involved and were included in the constructs measured in the current study. Syntax is further explored in the next section.

Morphology and Syntax/Grammar. Syntax is a rule-based system for combining words into sentences. Morphology has to do with words, meaningful word parts, and how to combine them. The two domains together are often referred to as "grammar" or "morpho-syntax". The evidence for morpho-syntax as a predictive variable for reading comprehension is inconclusive.

Some studies have found grammar scores to predict reading ability. Catts and colleagues (2002) found that a grammar composite accounted for more variance in later reading comprehension than composites in other language domains, in a group of children with language impairment that was identified in kindergarten. These researchers also found that grammar skills added additional predictive value beyond word recognition and reading comprehension when using Grade 2 data to predict Grade 4 reading comprehension. Adlof and colleagues (2010) created models to predict eighth grade reading status (impaired/not) from kindergarten language scores. Their best model included two grammatical measures, grammatical completion and sentence imitation (Adlof et al., 2010). Catts and colleagues commented on the involvement of working memory in their grammar tasks; phonological working memory in particular is thought to be involved in the observed connection between morpho-syntax and comprehension (Oakhill and Cain, 2007). Developmental effects may be at play as well. Cain and Oakhill (2004) reported that comprehension of grammar was predictive of reading comprehension in 9 year olds, but not 8 year olds.

In considering the role of syntax in reading comprehension, it is important to consider results with older students. With increasing school grades, the texts students encounter will become more complex, including increased sentence complexity. This could potentially put more strain on the language system, resulting in late-emerging reading difficulties among students with restricted syntax skills (Scott, 2009).

The diagnosis of language impairment frequently includes observed difficulties in morpho-syntax, such as difficulty marking past tense and understanding and use of complex sentences (Adlof et al., 2010; Leonard, 2000; Nation et al., 2004; Scott, 2009). Among poor comprehenders, difficulties with understanding morpho-syntax (Stothard & Hulme, 1992) and syntactic awareness (a metalinguistic ability, including ability to judge grammaticality and to repair grammatical errors) (Cain & Oakhill, 2007) have been observed. These difficulties can be expected to impact text-level reading comprehension as sentence processing will be compromised. Completion of cloze tasks may also be affected as syntactic information could assist the child in guessing the correct word to fill the blank.

As a whole, the preceding evidence on morpho-syntax skills suggests that they are likely to be relevant to reading comprehension in an older elementary population. Moreover, the evidence indicates that morpho-syntax measures, including measures of complex syntax, are likely to reveal language differences among children with and without LI that could impact their scores on reading comprehension tests.

Cognitive-linguistic skills and processes

Beyond basic language skills, a variety of higher-level processes have been identified as central to reading comprehension (see for example Loscacio, Mahone, Eason, & Cutting, 2010). Various authors have described these processes as cognitive, linguistic or both. At the level of discourse processing (either oral or written), it seems prudent to regard the component skills of cognition and language as inter-related. This was demonstrated in a key study by Oakhill, Cain and Bryant (2003b cited in Perfetti, Landi & Oakhill, 2005). Three discourse-level skills predicted reading comprehension: making inferences, comprehension monitoring, and comprehension of narrative structure. These skills were significant even with the autoregressor (earlier reading comprehension skill) controlled. Central to these processes is the concept of "standard of coherence". This term is used to refer to the reader's internal awareness that the text should make sense (see for example Perfetti et al., 2005). Reading comprehension can be viewed as the set of processes leading to a mental representation that is consistent with the reader's standard of coherence. Areas reviewed here will include narrative skills, knowledge, working memory, inferencing and comprehension monitoring.

Narrative skills. Narratives are an interesting measure to examine oral language skills. Comprehension and production of narratives are highly integrative discourse-level skills, requiring the use of all language domains, background knowledge and cognitive skills (such as working memory and inferencing) for an effective performance. These task characteristics allow the researcher to tap higher-level processes known to be related to reading whole texts. Narratives have additional benefits in being more similar than other spoken language forms to written language in terms of vocabulary and use of complex sentences (Westby, 1991 cited in Oakhill & Cain, 2007). Most of the texts school children will encounter, at least until Grade 3, are narrative. Accordingly, narrative skills have been found to predict academic success (Feagans & Applebaum, 1986). Some of the association between narrative skills and academic success is likely related directly to reading comprehension. Snyder and Downey (1991) reported that narrative retelling explained a significant amount of reading comprehension variance in typically developing 8 to 14 year olds. Comprehension of narrative structure has been found to predict reading comprehension in 7 to 10 year olds and also predicted the amount of growth in reading comprehension (Oakhill & Cain, 2007). These latter two results correspond with what is known regarding typical development of narrative skill, with production of internally consistent stories evident by 4 to 8 years of age, followed by fine-tuning from 8-10 years of age (Fayol, 2004). Associated skills such as stating a main idea and selecting important details are later developing, with growth in later elementary and beyond (Oakhill & Cain, 2004).

Narrative ability, particularly knowledge of story structure, is intertwined with the cognitive skills of inferencing and working memory. Experience with reading stories, and with everyday scripts, will bootstrap a child's understanding of story structure. Over time, this will facilitate story comprehension and allow inferencing as the child has internalized the expected series of events (Fayol, 2004; Oakhill & Cain, 2004). As will be demonstrated in the section below on working memory, solid knowledge of story schemas provides a mechanism to support and extend text processing by relieving cognitive demands. That is, if children understand how stories work, they are able to slot relevant details into an existing structure, rather than having to abstract that structure as processing proceeds (see for example Fayol, 2004; Hayward, Schneider, & Gillam, 2009). Van den Broek and colleagues (2005) identified understanding of narrative structure as significant to the development of comprehension across media (listening, viewing and reading).

As a complex cognitive-linguistic task, narratives can be expected to challenge students with oral language difficulties. Poor comprehenders have been shown to have reduced awareness of story structure; they tend to have difficulty in identifying main ideas and in production tasks require the support provided by pictures and titles (Cain & Oakhill, 2007). Research evidence points to poor narrative skills such as awareness of story structure as a causal factor in reading comprehension difficulties (Fayol, 2004; Cain & Oakhill, 2007; Perfetti et al., 2005). Accordingly, poor comprehenders are observed to have difficulties in producing written narratives that correspond with their difficulties in producing spoken narratives, both reflecting poor story structure and reduced information (Nation & Norbury, 2005). Overall, children with LI have a number of difficulties with narratives consistent with their difficulties in the language domains detailed above. Their stories contain fewer, shorter sentences and reflect poorer morpho-syntax skills than peers with typically-developing language (Fey, Catts, Proctor-Williams, Tomblin & Zhang, 2004). Their stories display less overall complexity, in terms of both story elements and number of episodes, reflecting both semantic and pragmatic weaknesses (Fey et al., 2004).

Narrative tasks appear to provide a useful vehicle to measure higher-level cognitive-linguistic skills that are likely to predict reading comprehension ability and to differentiate between students with typically developing language and those with LI. In terms of oral expressive language, narrative samples will likely elicit the most complex language from children, providing the closest approximation of oral language to literate language. In addition, as a discourse-level skill, narrative tasks include integrative skills such as working memory and the ability to make inferences.

Knowledge. Integration of personal knowledge and experiences with the information provided in the text is a central process in currently-accepted models of reading comprehension (see for example Kintsch, 1998). This integration process results in the construction of a situation model of the text; information in the situation model may be verbal, visual, emotional and experiential (Kintsch & Rawson, 2005). Nation (2005) discusses two aspects of knowledge availability that can affect comprehension: activation and retrieval. Both must be efficient in order to optimally support comprehension. Knowledge is entangled with semantics; as previously noted, high knowledge in a given area is also indicative of high vocabulary in that area. Nation (2005) reconciles these knotty observations of knowledge, vocabulary and retrieval by suggesting that what is really relevant is not disentangling these elements, but considering the efficiency with which they are integrated in text comprehension. This view suggests that the choice to sample narrative comprehension is appropriate in the current study; these tasks provide the opportunity to integrate old and new information.

Kintsch (1998) posits “retrieval structures” in his Construction-Integration theory of comprehension, which act as a work-around for limited working memory capacity. Essentially, he suggests that long-term memories (i.e., knowledge) related to a text can remain activated and available for integration without bogging down working memory. Automatic activation is associated with expertise and practice in a given knowledge domain (Kintsch, 1998; Kintsch & Rawson, 2005). This could be one reason why reading comprehension both affects and is affected by accumulated reading experience; good readers are likely to read more, providing more experience to facilitate reading processes (Oakhill & Cain 2004). Good general background knowledge has the potential to impact reading comprehension assessment as well; for example, it has been demonstrated that many questions on the *Gray Oral Reading Test* can be answered from background knowledge alone, without reading the test passages (Keenan & Betjemann, 2006). These ideas are further explored on the next section on working memory.

In the current study, knowledge is not evaluated separately as such a measure was not deemed feasible. Vocabulary, narrative skill and working memory measures address some aspects of background knowledge as implicated in reading comprehension.

Working memory. Working memory is well-established as a predictor of reading comprehension ability (Georgiou, Das & Hayward, 2008; Kendeou et al., 2012; Perfetti et al., 2005; Ricketts, 2011). Working memory can be conceptualized as a system for performing mental “work” with information. As such, it requires both storage and processing capacities, including a system for allocating attention usually labelled as the central executive (Boudreau & Costanza, 2011, citing Baddeley, 1996). Different results have been reported for working memory (WM), which taps both processing and storage and implicates the central executive, and short-term memory (STM), which taps only passive storage (see for example Swanson et al., 2007). Measures of working memory (WM) can be used to differentiate readers of different skill levels. Skilled readers had higher working memory capacity than poor comprehenders, who had higher working memory capacity than poor decoders. Poor readers (those with low verbal IQ) had the lowest working memory capacity of the groups studied (Swanson et al., 2007).

Swanson and colleagues contrasted these results with the existing literature which links differences in reading comprehension skill to oral language abilities (citing Nation et al., 1999 and Stothard and Hulme, 1992); Swanson’s group found that WM measures accounted for unique variance when STM, phonological skills and verbal cognitive skills (general information and vocabulary) were controlled (Swanson et al., 2007). Concurring results were obtained by Goff, Pratt and Ong (2005); they found that STM memory tasks were less predictive of reading comprehension than tasks involving the integration of old and new information. Cain and Oakhill (2007) presented evidence that WM relates both to listening and reading comprehension, and that poor comprehenders have more difficulty with WM tasks than do good comprehenders. In general, STM deficits are found to be more typical of dyslexic readers than of poor comprehenders (Bishop & Snowling, 2004; Catts, Hogan, Adlof, & Weismer, 2005; Swanson et al., 2007). Specific vocabulary effects are also noted; poor comprehenders have more success recalling concrete than abstract words (Bishop & Snowling, 2004).

It is important to note that results regarding working memory as a predictor of reading comprehension are not conclusive. Cutting and Scarborough (2006) found that verbal memory was not a significant predictor of reading comprehension. They used a variety of measures, including story recall, nonword repetition, digit span, and sentence span in a study that was similar in design and analysis to the current study. In a study of executive functions in reading comprehension, Kieffer, Vukovic and Berry (2013) did not find working memory to play a significant role in a path analysis that also included language and decoding.

Swanson and colleagues (2007) posit WM as a domain-general skill, not the result of reading, suggesting that its impact on reading comprehension is causal. Perfetti and colleagues (2005) discuss how working memory is somewhat fluid and subject to effects of background knowledge. In accordance with Kintsch and Kintsch (2005), these authors state that engaging in reading practice will facilitate or automatize the cognitive processes involved, effectively increasing working memory. An example of this was provided in the narratives section of this review, in that knowledge of narrative structure supports WM. Therefore, although WM may be a foundational cognitive skill for reading, reading practice has the potential to augment WM.

Note that there is controversy on how best to conceptualize the relevant memory processes; connectionist models rely on the network concept, with activation spreading to related information in place of separate working and long-term memory (McNamara & Magliano, 2009). Despite the differences in conceptualization, both working memory (information processing) and connectionist approaches acknowledge that the ability to manipulate and integrate new and old information is central to good reading comprehension.

In the current study it is necessary to include working memory as a possible predictor of reading comprehension scores. In addition to its relationship to reading comprehension, working memory is also thought to be tapped by many measures of oral language ability. Since measures of oral language ability are used as independent variables in the regression analysis for this study, it is necessary to account for the effects of working memory inherent in these tasks. "Functional working memory" (FWM) is a term that has been use to refer specifically to the working memory processes important in processing verbal information, the central executive and phonological short-term memory (Boudreau & Costanza, 2011).

The contrasts noted between WM and oral language contributions to reading comprehension skill speak to the problem of isolating constructs in a complex system. Many tasks used to sample oral language skills require significant memory skills. For example, the Formulating Sentences subtest of the *Clinical Evaluation of Language Fundamentals—4th Edition* (Semel, Wiig, & Secord, 2003) requires a child to listen to a word, view a scene, and create a grammatically correct sentence using that word and relating to the given scene. This task requires working memory capacity in combination with adequate receptive vocabulary and expressive language skills. The Recalling Sentences (sentence imitation) task of the same instrument requires a child to listen to a series of sentences of increasing complexity, repeating each one verbatim. This task requires both adequate memory and linguistic skills. In evaluating language using such tasks, it will be difficult to determine which skills are most relevant. The sentence imitation task can be viewed as a kind of language catchall that is useful for screening purposes, but is too complex to be diagnostic in itself (Adlof, 2010). An example from my clinical experience is telling. I had two students achieve similar scores on this sentence imitation task, yet for very different reasons. One kept the gist of the sentence, but because of poor morpho-syntax skills could not repeat the same structures. Another student tried to reproduce words that sounded like the stimulus sentence, producing semantically-unacceptable utterances. This suggested weak oral comprehension, as the child was relying on phonological memory. Swanson and colleagues (2007) make the important point that higher- and lower-level deficits may co-occur in an individual, as can difficulties of integration of these skills. Nation (2005), after evaluating these concerns relative to reading comprehension research, described the search for clear predictors as potentially “futile”. In general, selecting oral language measures with appropriate memory load is expected to provide the most realistic estimate of oral language skills. This is the approach taken in the proposed study. In addition, two working memory tasks are included in the study battery. In addition to a functional working memory measure, the well-established digit span task (repeating lists of numbers forward and backward) was included. Using another assessment method, language sample measures have the potential to more effectively isolate oral language skills from memory loads imposed by contrived language tasks. Three language sample measures were selected for use in this study; the use of language sampling is a novel approach in reading comprehension research.

Metacognition and inferencing. It is difficult to make clear divisions between different cognitive and linguistic constructs related to reading comprehension. It is likely that processes at different levels interact in complex ways, making differentiation a complex goal (Nation, 2005). Thus, authors differ in how they “carve up” the cognitive and linguistic skills involved. In this review, phonological awareness, although a metalinguistic skill, was discussed in proximity to related decoding skills. In this section, comprehension monitoring and inferencing will be the focus. Oakhill and Cain (2007) have identified these skills as strong predictors of reading comprehension.

Both comprehension monitoring and inferencing are closely related to the concept of standard of coherence. It is generally accepted that readers attempt to build a mental model of the text that is coherent, in other words one that makes sense to the reader (see for example Perfetti et al., 2005). Establishing coherence will typically involve making inferences, the process of "filling in the gaps" which are found in most texts. Comprehension monitoring is an ongoing process that allows a reader to detect coherence problems, then to take action to re-establish coherence. It has also been called continuity monitoring and is considered an automatic process in connectionist models of reading (McNamara & Magliano, 2009). Domain-general cognitive skills including the central executive and the individual's standard of coherence can be conceptualized as an "umbrella" under which text-processing activities occur. In this approach, comprehension monitoring and automatic inferencing are considered to be co-occurring processes that arise from neural activation that is associated with ongoing text processing. At the moment when a reader notices a coherence problem (i.e., something in the mental model of the text is not consistent with the standard of coherence), the central executive allocates attention to text processing. This switches the previously-automatic processes of comprehension monitoring and inferencing to intentional, conscious processes, such as strategy use. Once the coherence problem is resolved, the system switches back to automatic processes.

Inferencing, both automatic and conscious, depends upon the availability of topic knowledge, schemas and scripts (Barnes et al., 2007). According to connectionist models of reading comprehension, automatic inferencing proceeds through spreading activation, relying on well-established connections; note that some authors differentiate this automatic activity from true (conscious) inferencing (McNamara & Magliano, 2009). In the approach taken in the current study, conscious inferencing is triggered by the failure of automatic processes to establish coherence in the mental model of the text. It could also be triggered by a conscious effort or motivation to deeply understand a text (Cook, Chapman, & Gamino, 2007). The ability to make inferences has been suggested as a causal factor in reading comprehension, both based on comparisons of poor comprehenders and reading-age matched children, and the results of intervention studies (Oakhill & Cain, 2007). However, recent work suggests that memory for information in the text is the significant factor in inferencing accuracy among poor comprehenders rather than a problem with inferencing itself (Hua & Keenan, 2014).

Comprehension monitoring, the ability to note difficulties in comprehension, is a key process in triggering conscious text processing. In the absence of comprehension monitoring, the reader has no cause to alter the method of reading. Standard of coherence is again invoked; the reader must expect the text to make sense in order to "be on the lookout" for problems (Perfetti et al., 2005). Some evidence for comprehension monitoring problems in children with oral language impairments is again found in the research with poor comprehenders. These children appear to have more difficulty recognizing problems in texts than stronger comprehenders, particularly when the mismatched pieces of information are separated from each other in the text (Nation, 2005). As an additional example from my practice, I have found that junior high students with LI typically have difficulty flagging words that they do not know in a text, a basic comprehension monitoring task.

As discussed above, inferencing and comprehension monitoring are considered by some authors as integral to reading comprehension (Perfetti et al., 2005; Van den Broek et al., 2005). However, in the current study, both inferencing and comprehension monitoring are conceptualized as part of the cognitive architecture that supports reading comprehension. The focus in this study is on component linguistic skills of reading comprehension rather than these over-arching cognitive skills; therefore comprehension monitoring and inferencing were not directly studied.

Motivation. Students will vary not only in component skills of reading, but also in their own motivation to read. Guthrie and Wigfield (2005) presented a cognitive-motivational model of reading, which builds on the Construction-Integration model of comprehension (Kintsch, 1998) by adding a set of "motivational processes" to the set of cognitive processes (Guthrie & Wigfield, 2005, p.189). The motivational processes speak to the reader's goals, self-perceptions, level of interest and desire to read relative to a given text. Guthrie and Wigfield (2005) review a variety of studies relating motivation to reading comprehension, demonstrating that intrinsic motivation in particular predicts reading comprehension (citing Wang and Guthrie, 2004). Guthrie and Wigfield (2005) also consider the impact of motivational processes on assessment, pointing out that characteristics of the test and environment will interact with motivation to affect test outcomes (e.g., availability of choices, whether items are sequenced by difficulty). Motivational characteristics of students with reading difficulties are of special concern for children with language impairment, who often have accompanying literacy problems. They may be particularly at risk for learned helplessness and anxiety as the result of long-term problems with reading; both learned helplessness and anxiety can be linked to trouble engaging adequately with reading assessments (Guthrie & Wigfield, 2005). In assessing children with suspected language and literacy delays, then, it will be prudent to attend to factors such as time limits, interest level of the text and physical characteristics of texts (e.g., visually confusing) that may negatively impact scores (Guthrie & Wigfield, 2005). Reader factors (such as motivation) and environmental factors (such as time pressure) which may affect reading comprehension are included in the extensions I have proposed to the Construction Integration model (see Appendix

A). Since standardized tests are the stimuli of choice in this study, it is not possible or desirable to modify materials to manipulate such characteristics as listed here. Thus these constructs were not selected as predictors in this study. However, I attempted to minimize the effects of these factors in data collection as much as possible. Examiners took time to establish a rapport with students, allowed the students to take breaks as needed, made efforts to break up lengthy testing sessions, and provided encouragement to the participants as permitted by test protocols.

Summary: Variables selected for the current study

It is evident that there are more interesting potential predictors of reading comprehension than can be considered in a single study. To evaluate all the potentials would be more than unwieldy; in terms of participant time it would simply be impossible. The goal for this study was to select predictors that can do dual duty, language variables that can be both strong predictors of reading comprehension and that can reliably differentiate between LI and typically-developing language in an older elementary-school population. Conveniently, these two goals appear to align rather nicely. Oral language measures including **vocabulary, morpho-syntax and narrative skills** were all expected to function well as predictors in a regression analysis. These predictors were complemented with a set of controls chosen to measure powerful known predictors including **word reading skills and working memory**. Three variables were included as group matching variables, rather than as independent variables carrying significant explanatory power. These **included mother's years of education, gender and nonverbal intelligence**.

With selection of strong potential predictors complete, it is necessary to consider methods for measuring these predictors in the context of a regression analysis.

Methodological Issues in Studying Reading Comprehension

There are a variety of methodological issues that have been highlighted regarding studying the underlying skills involved in reading acquisition. One of the key concerns is the definition of "reading" as a construct. In earlier development, the focus of the learner's energies is on learning to decode written text. Once word recognition is well on its way, reading comprehension becomes central (Bowey, 2005). Accordingly, Saarnio, Oka, and Paris (1990) found that decoding determined reading comprehension levels in Grade 3 students. By Grade 5, text-level processes to derive meaning were predominant in predicting reading comprehension. Early on, word reading and reading comprehension are closely related, with up to 81% shared variance (Bowey, 2000, 2005). This figure drops to 66% by fourth grade (Bowey, 2000). Therefore, studies need to be specific about whether they are examining word recognition or text-level reading comprehension; different underlying abilities are at play for these two processes (Bowey, 2005). In the case of the current study, reading comprehension is the dependent variable, and word reading skills are included in the predictors/independent variables. It was of interest to see how word reading skills predict reading comprehension in children with and without LI, given that word reading deficits are more common in children with LI.

If word reading skills are of interest, then clear explanation of the measurement method is necessary. There are a variety of ways to define word recognition. In this study, the concept is defined operationally by the tasks used: nonword decoding and ability to read real words out of context (i.e., printed words on a page, no sentences).

Statistical considerations are also key in evaluating reading skill predictors. Floor and ceiling effects are a concern when evaluating constrained skills such as letter recognition (Bowey, 2005; Paris et al., 2005). In this study, age-appropriate nonword decoding and real word reading tasks were selected. Both measures appear to have sufficient range to avoid floor/ceiling effects. The language tasks were also selected to avoid floor and ceiling effects.

Selection of control variables is of concern in reading comprehension research when oral language is also of interest. For example, some authors recommend verbal intelligence as a control for reading studies (see for example McGuinness, 2005). Yet controlling verbal intelligence will reduce the valid amount of variance associated with oral language, as these two constructs overlap. Whenever two constructs overlap, controlling for one will affect covariates (Bowey, 2005). In this study, nonverbal intelligence was measured as a possible predictor as it does appear to have relationships with language and reading skills. Accordingly, these relationships were evaluated and are reported in the Results section. Word reading variables, which involve vocabulary, are equally of concern. It was necessary to select statistical procedures that would allow relationships among the independent variables to be recognized and accounted for. These considerations are further explored in the Data Analysis section.

Drawing from the existing research base, this study requires carefully defined constructs, operationalized to provide age-based standard scores free of floor and ceiling effects. Both predictor and control variables (here “control” is used to refer to non-oral language variables that are established reading comprehension predictors in the literature) must be selected with an eye to possible inter-relationships so that these can be considered in the regression analysis. With this discussion of general methodology complete, the next section deals more specifically with issues in measuring oral language abilities.

Measuring Oral Language

The conceptualization and operationalization of oral language skills must be considered as factors in the design of the current study. Existing research on reading comprehension has included various measures of oral language, with a variety of approaches (Skoczylas, Schneider, & Suleman, in press). Relevant language constructs were identified in the literature review; operationalization of these constructs is covered in detail under "Independent/Predictor Variables" in the Measures section. Besides the selection of relevant variables, there are measurement techniques to consider. In this study, multiple regression was selected as the analysis method so interval/ratio data were preferred. Age is an important consideration in this elementary school sample; using age-based standard scores removes the necessity of including age as a separate variable. Research methods stipulate the use of measures with high validity and reliability. With all these considerations, the majority of informal methods, such as criterion-referencing, are not appropriate. Standardized assessment tools are frequently used in practice to diagnose language impairment and to reveal areas for further assessment. These tools can have strong psychometric properties and provide the required norm-referenced scores. The SLP's stock-in-trade standardized assessments include measurements of both receptive and expressive skills, with subtests sampling semantics, syntax and so on. Well-constructed instruments will target the skill of interest directly. For example, a task purporting to measure receptive semantics should not have high expressive language or short-term memory demands. Careful task analysis of instruments is necessary to reveal any demands that are additional to the intended target. Adhering to these standards, a number of standardized language measures were selected for use in the current study and are described in the Measures section.

Standardized tests are one accepted method for measuring language skills. Yet language is a social tool, and may perhaps be best measured in valid communicative contexts. As assessment tasks move farther away from natural communication tasks and artifacts creep in, it may become less clear what skills are actually being measured. Language sample analysis consists of a set of well-validated procedures that permit the use of natural language interactions to provide measurement of expressive language skills (see for example Heilmann, Nockerts, & Miller, 2010; Rice, Redmond, & Hoffman, 2006). The elicitation context itself is important. For school-aged children, narrative samples are appropriate as they may encourage more grammatical complexity than conversational samples (Westerveld, Gillon, & Miller, 2004). Much of the reading material students are expected to comprehend in the elementary years is narrative, so this sample also approximates the kind of language the children are likely to encounter in authentic reading tasks (see for example Cain, 2010). Recent research has provided norms for a narrative sampling method using materials from a standardized test of narratives (Justice et al., 2006). In the selected task, the examiner shows a picture and reads an example story to the child. Following this, the child makes up a story based on another single picture. Justice et al. (2006) reported that this relatively unconstrained format permits room for more variability in participant performance. Standard language sample measures include semantics, morphology and syntax.

To summarize, both standardized language assessments and language sampling in a narrative context are valid methods to evaluate oral language skills in school-aged children. With the selection of relevant variables and discussion of methodological issues complete, it is possible to extend the results of the literature to hypothesize results for the research questions.

Hypotheses

Research question 1

Given receptive and expressive measures of vocabulary, syntax and narrative discourse for children in grades 4 to 6, which of these oral language scores will predict scores on two concurrently measured standardized reading comprehension tests?

The first research question focuses on the oral language predictors of two reading comprehension tests, including whether or not predictors differ for the two tests. Tests with different task demands were selected, based on existing research that suggests different tasks tap different underlying skills (Skoczylas, Schneider, & Suleman, in press). These different underlying demands are hypothesized to result in differences in the significant predictors for the two tests. Hypothesized results are presented below for each reading comprehension test.

Woodcock Reading Mastery Test—Revised/Norms Update (WRMTR)

Based on previous research, the cloze format of the *WRMTR* Passage Comprehension (PC) task should be more related to decoding skills than language comprehension per se, as reported by Francis et al. (2005). The task is also hypothesized to relate to vocabulary as knowledge of single words is important in order to select the missing item (Keenan, Betjemann & Olson, 2008). Short texts are not hypothesized to tap text integration skills as measured by the narrative variables, although previous research indicates that working memory is tapped by this test (Keenan, Betjemann & Olson, 2008; Kendeou et al., 2012). Syntactic skills would be involved if knowledge of sentence structure facilitates selection of the missing word. Thus, the receptive and expressive syntax measures are hypothesized to predict the *WRMTR* PC task.

Kaufman Test Of Educational Achievement—Second Edition (KTEA-II)

The *KTEA-II* Reading Comprehension (RC) subtest has not been previously examined in a study of this kind. The subtest requires the child to complete increasingly challenging reading tasks. Beyond the most basic questions, the child reads passages and answers questions based on the passages. The final items require the child to re-order passages. More complex passages (relative to those in the *WRMTR*) in this task are hypothesized to tap complex syntax skills. Therefore the receptive and expressive syntax scores are expected to predict the *KTEA-II* RC task. As this task is expected to be language-loaded, less variance on this task should be predicted by word recognition and decoding than for the *WRMTR* PC task. Some variance is expected to be related to these lower-level reading skills; reading comprehension is necessarily limited by decoding and word recognition. Inferential questions, included in the questions on this test, are expected to tap higher-level skills and therefore narrative scores should predict performance on this task. The later questions on the *KTEA-II* RC include more complex integrative language skills. Again, this task is expected to tap higher-level text processing skills as represented by the narrative variables, in higher-achieving students who will attempt the more complex items.

Research question 2

Will there be an effect of group membership such that significant predictors for children with developmental language impairment will differ from significant predictors for children with typically developing language?

The second research question focuses on the possible differences between predictors for the two participant groups. Previous research with children with LI suggests that vocabulary and decoding scores will remain predictive of reading comprehension after they are no longer relevant for typically-developing peers (Palikara, Dockrell, & Lindsay, 2011). This effect was hypothesized to be present in results for both tests with decoding and vocabulary being stronger predictors in the group of children with LI. Nonverbal intelligence (NVIQ) has been found to be depressed among children with language impairment when standardized tests are used (Dethorne & Watkins, 2006); therefore this score may be correlated with language status. As the reasons for this relationship are unclear, it will be statistically evaluated if necessary, at the last step of regression to avoid pulling out variance that may be validly associated with language scores. The use of language sample scores may reduce the impact of NVIQ (Dethorne & Watkins, 2006).

Conclusion

The existing research base informed the selection of a set of constructs that are hypothesized to reveal language skills underlying two reading comprehension tests and differences between the two participant groups. The next task is to present measurement methods and operationalization of these constructs. Chapter 3 presents specific tools selected to measure the selected variables.

Chapter 3 Methods

In this chapter, methods specific to the current study are presented, including participants, measures, procedures and data analysis.

Participants

Fifty-five children in grades 4, 5 and 6 from local schools participated in the study. The data from one student was omitted from analysis as several key predictors were missing. Thus the sample size for analysis was 54. The sample included 27 males and 27 females. In the group with language impairment (below criterion on *CELF* Core Language), there were 11 females and 13 males. In the group with typically developing language, there were 16 females and 14 males. The age distribution was as follows: 23 nine year olds; 25 ten year olds; 6 eleven year olds. Fewer 11 year olds were available due to province wide year-end tests in Grade 6. These students could not be absent from class during two of our testing phases. Additionally, fewer 11 year olds volunteered to participate during recruitment. Students were recruited from schools with special programs that might be expected to include children with LI.

Performance on the *CELF 4* or *CELF 5* Core Language score (above criterion vs. at/below criterion) was selected for the determination of language impairment in the regression analyses. This variable had two main advantages. First, it split the sample into reasonably equal groups, helpful for statistical analysis. The second reason for selecting the Core Language criterion relates to applicability. It is useful to frame results in terms of a readily-available metric. In this case, the *CELF* is the primary diagnostic instrument used locally by speech-language pathologists. The *CELF* score was selected as a straightforward metric available for all participants. Further explanation of the approach to determining language status, and other variables considered for the analysis, is provided in Appendix B. Based on the *CELF 4/5* Core Language score criterion, 24 children were classified as having LI and 30 were classified as having typically developing language (TDL). Table 1 presents the characteristics of the participant sample in terms of age, Core Language Score from the *CELF 4/5* and group matching variables including nonverbal IQ and mother's years of education.

Table 1

Participant characteristics for all participants (N = 54), children with TDL (N = 30) and children with LI (N = 24)

	Mean			Std. Deviation			Range		
	All	TDL	LI	All	TDL	LI	All	TDL	LI
Age	10.08	9.98	10.22	0.63	0.63	0.64	9.08-11.50	9.08-11.08	9.08-11.50
Core language	90.11	100.6	77.0	15.09	11.09	6.85	60-121	86-121	60-85
NVIQ	99.31	105.37	91.75	16.16	16.83	11.72	66-131	67-131	66-115
Mother's years of education	14.92[a]	14.75	15.14	5.09	4.84	5.50	2-30	2-26	5-30

Note: NVIQ = Nonverbal intelligence on the *Kaufman Brief Intelligence Test*

[a] Five were unreported; mode was 12; N = 27 for TLD, N = 22 for LI

Correlations among the above variables were not significant, except for nonverbal IQ and Core language score ($r = .56$, $p < .001$). The relationship between nonverbal IQ on the *KBIT* and the Core Language Score appears to be in line with expected results. The population correlation between verbal and nonverbal IQ on the *KBIT* for 9, 10 and 11 year olds are 0.51, 0.40 and 0.44 respectively. The correlation between oral language (as measured on the *Kaufman Test of*

Educational Achievement—II) and nonverbal IQ on the *KBIT* is 0.45. Finally, the expected mean for nonverbal IQ on the *KBIT* for children with speech and language diagnoses (mean age of 9;9) is 88.4, SD 18.8, rather lower than the population mean of 100. Thus lower Core Language Scores are expected to be associated with lower NVIQ scores on this instrument. These data suggest that the associations noted in the current dataset are consistent with population estimates and do not represent skewed results. NVIQ was further evaluated in the regression analyses as the groups were unmatched in terms of this variable (F = 11.31, p = .001) (see Chapter 4 Results). Language status (as LI/TDL) was unrelated to age (F = 1.87, p = .177), gender (F = .291, p = .592), or mother's years of education (as an estimate of SES) (F = .070, p = .793). The mean score for Core Language was significantly different between the children with LI and the children with TDL (F = 83.10, p < .001).

Exclusion criteria. Children were required to pass a hearing screening as hearing difficulties can impact language development and could have affected test results. One child failed the hearing screening and did not participate further. Initially, children with English as a second language were excluded as second language differences can present as similar to developmental oral language difficulties and would act as a confounding variable. Once recruitment began, it became clear that this stringent criterion would limit the sample. The school division from which we recruited has a culturally diverse student population. Therefore, teachers were instructed to send recruitment packages for children with English as a second language in the home only if they judged them to be competent users of the English language. If the child displayed language difficulties and had English as a second language, the child was excluded (i.e., the teachers did not send home recruitment packages). Children with diagnoses of neurological problems such as seizure disorders or traumatic brain injury were excluded as these diagnoses may impact oral language skills differently than LI and therefore confound results. Again, teachers were instructed not to send recruitment packages to students with these diagnoses. Conditions that may co-occur with LI such as ADHD were not sufficient to exclude children as long as the child was able to adequately complete the testing.

Student eligibility was determined by classroom teachers based on the criteria provided to them. To assist teachers in understanding the criteria, I met with staff over a lunch break to provide a brief overview of the study and to explain the criteria. This gave teachers a chance to clarify any points of concern in person. Teachers were provided with contact information in case of later questions. This appeared to be effective as some teachers did contact me with questions during recruitment.

Measures

The measures for this study included two reading comprehension tests and tasks for oral language skills, word reading, working memory and nonverbal intelligence. These variables were measured using standard assessment tools and methods. They are summarized in Table 2 which lists the complete test battery for the study and subtests, measures, or composite scores used from each standardized test or assessment task.

Table 2

Oral language subtests, composites and tasks used as predictor variables

Test or Assessment Method	Subtest, composite or task	Construct measured
Comprehensive Assessment of Spoken Language (CASL)	Synonyms	Receptive vocabulary
	Antonyms	Expressive vocabulary
	Paragraph Comprehension	Receptive syntax
Test of Narrative Language (TNL)	Narrative Comprehension score	Receptive narrative skills
	Oral Narration score	Expressive narrative skills
	Narrative Language Ability Index (NLAI)	Overall narrative skills
Language sampling during narrative task from the *TNL*	Mean length of T-unit in words (MLTU-w)	Expressive syntax
	Proportion of complex sentences (PCS)	Expressive syntax
	Number of Different Words (NDW)	Expressive vocabulary
Kaufman Test of Educational Achievement, Second Edition (KTEA-II)	Nonsense Word Reading (NwR)	Decoding (phonetic reading)
	Letter and Word Recognition (LWR)	Decoding (word recognition)

	Reading Comprehension (RC)	Reading comprehension
Woodcock Reading Mastery Test—Revised/Norms Update (*WRMTR*)	Passage Comprehension (PC)	Reading comprehension
Kaufman Brief Intelligence Test (KBIT-2)	Nonverbal intelligence (NVIQ)	Nonverbal intelligence
Clinical Evaluation of Language Fundamentals—4th Edition (CELF-4)	Familiar Sequences (FS)	Working memory
	Number Repetition—Total (NR-T)	Working memory

Note: all scores were entered as age-based scaled/standard scores

Colour-coding legend

	Language tests
	Reading tests
	Cognitive tests

Dependent Variables: Reading Comprehension Scores

Two standardized reading comprehension tests were selected. Each provides one norm-referenced score for reading comprehension. The first is from the *Woodcock Reading Mastery Test, Revised/Norms Update*, and has been previously studied. The second test, from the *Kaufman Test of Educational Achievement*, *2nd Edition*, is in common usage in local schools. This test has not been previously included in a regression study of this kind. Tasks were chosen that are hypothesized to tap different underlying skills, based on published results and task analysis of task/response formats.

***Woodcock Reading Mastery Test—Revised/Norms Update* Passage Comprehension subtest** (*WRMTR* PC; Woodcock, 1998). The *WRMTR* PC task requires the student to read short (two to three sentences) texts with one missing word each and complete a cloze task. The start point for this test is determined by the participant's grade of enrolment, with testing back to easier questions as needed to establish a basal.

The reliability and validity of this test appear to be acceptable. As reported in the *WRMTR* manual, split-half reliability for form G or H is 0.92 in Grade 3 and 0.73 in Grade 5. The same figures for G and H combined are 0.96 and 0.84. In terms of content validity, item response is open-ended, to eliminate guessing. Classical item selection techniques and Rasch modelling were used in item development. In terms of concurrent validity, the manual reports correlations of 0.83 (*Iowa Test of Basic Reading*, Total Reading), 0.87 (*PIAT* Reading), 0.92 *WJ* Reading Achievement), and 0.88 (*WRAT* Reading) in Grade 3 for the Total Reading Score. Correlations in Grade 5 for the same measures were 0.78, 0.78, 0.87, and 0.86.

***Kaufman Test of Educational Achievement—Second Edition* Reading Comprehension subtest** (*KTEA-II* RC; Kaufman &Kaufman, 2004). The start point of this test is determined by the score on the Letter and Word Recognition (real word reading) subtest of the *KTEA-II,* with testing back to easier items if needed to establish a basal. The *KTEA-II* RC task consists of three types of items. Initially, students read words and identify the picture for that word, then read sentences indicating instructions they follow. Next, longer passages (approximate range is 50-225 words) are presented. Vocabulary level is controlled so that it does not exceed the target grade level of the passage. Question responses are short oral answers or multiple choice responses and include both literal and inferential responses. The final questions of the *KTEA-II* RC require the student to re-order the text and respond to inferential questions.

The reliability and validity of this test appear to be acceptable. According to the *KTEA-II* manual, split-half reliability varies from 0.89-0.95 for the grades in the proposed study. Alternate form reliability (adjusted) for the grades in the proposed study is 0.76. Inter-rater reliability (using form A) was 0.93 for grade 3 (the only grade in the sample reported). In terms of validity, confirmatory factor analysis supported the subtest and composite format of the *KTEA-II*. The reading comprehension subtest of this instrument correlated at 0.69 with the reading comprehension subtest of the *WIAT-II* and 0.85 with the reading comprehension subtest of the *PIAT-R/NU*. It correlates at 0.65 with the Academic Applications cluster of the *WJ-III* (which includes reading comprehension.

Independent/Predictor Variables

Three domains of language were sampled: vocabulary, syntax and narrative skill. The variables were expected to sample (a) oral language variance that was likely to predict reading comprehension scores and (b) oral language variance that was likely to reveal any significant differences between the two participant groups. Based on existing research, measures of both receptive and expressive skills were deemed important. Accordingly, there are six types of independent variables: receptive vocabulary, expressive vocabulary, receptive morpho-syntax, expressive morpho-syntax, receptive narrative skills and expressive narrative skills. This format is consistent with the approach from Tomblin, Records and Zhang (1996) in diagnosing specific language impairment. These measures have a broad range to avoid basal and ceiling effects. Measures were chosen to minimize memory demands that exceed those found in typical language tasks. They are all recorded as standard scores. Each measure will be described in detail in the following section.

Language sample measures

The benefits of language sample measures compared to standardized testing include avoiding excessive memory load. The narrative micro-structure analysis methods described by Justice et al. (2006) were used to collect and score narrative language samples. Children created an oral story in response to a pictured scene as part of the *Test of Narrative Language* (*TNL)*. The stories were audio recorded and transcribed into the *Systematic Analysis of Language Transcripts (SALT)* software using standard coding. Justice and colleagues provide means and standard deviations for these scores as collected during standardization of the *TNL*; these were used to calculate z scores for the complex sentence measure. For the sentence length and vocabulary diversity measures, the comparative SALT database for the *TNL* was used.

Reliability of transcriptions was evaluated for a random sample of 10% of transcripts for each of the two research assistants. In each case, 98% to 100% agreement was achieved. Melissa Skoczylas coded each sample for complex sentences. Reliability was checked by Dr. Phyllis Schneider, who independently scored 10% of the sample for coding. Agreement was from over 95% to 100%. In the few cases of disagreement, the two researchers reached a consensus on the appropriate coding for that structure, and I reviewed the remaining transcripts for occurrences of that form and made any necessary changes.

The language sample measures were as follows:

- **Mean Length of T-unit in words (MLTU-w)** (expressive syntax): average spoken sentence length; calculated by SALT. A T-unit is defined as a main clause with any connected dependent clauses (0 or more).

- **Proportion of complex sentences (PCS)** (expressive syntax)**:** number of T-units containing an independent clause and one or more dependent clauses (hand coded by researcher, frequency count completed by SALT) divided by the number of T-units used by the child (calculated by SALT program); age-based z-scores calculated from data in Justice et al. (2006).
- **Number of different words (NDW)** (expressive vocabulary): total number of different words the child used (a measure of lexical diversity); calculated by the SALT program.

Receptive syntax measure

In order to complement the expressive syntax measures noted above, a measure of the student's ability to understand syntax was selected. The **Paragraph Comprehension** subtest of the *Comprehensive Assessment of Spoken Language* (*CASL*) consists of a series of orally-presented paragraphs, containing sentences of increasing syntactic complexity. Students are asked five to seven questions per paragraph. The student responds by pointing to one of four possible pictures for each question. It is not designed to measure vocabulary or ability to make inferences. Memory is required for this task, but the test author states that the memory load required is consistent with natural language situations. To reduce memory demands, the passage is read to the student twice.

The reliability and validity of the CASL appear to be acceptable. Split-half reliability for the age range in the proposed study ranged from 0.84-0.90 for Antonyms, 0.79-0.88 for Synonyms and 0.76-0.88 for Paragraph Comprehension. Test-retest reliability (corrected for variability of the norm group) for the age range in the proposed study was 0.89 for Antonyms, 0.81 for Synonyms and 0.86 for Paragraph Comprehension. These values were computed on a sample between 8:0 and 10:11 years of age. Construct validity was demonstrated in a variety of ways. Analysis of raw scores demonstrated expected patterns of change as participant age increased. Confirmatory factor analysis for ages 7-21 using standardization data confirmed the 3 factors posited in the theoretical basis for the test: lexical/semantic, syntactic and supralinguistic. The *CASL* showed relatively high correlations to the *Oral and Written Language Scales*, another standardized language test based on the same conceptual model but with different measures. In a validity study conducted with a group of children with language impairments and a comparison group, the clinical group scored on average scored lower than the comparison group in every subtest of the *CASL*. The author notes that individual children in the clinical group had mixed abilities and did not necessarily have lower scores on every subtest.

Standardized vocabulary measures

In order to augment the NDW score from the language sample measures, two subtests from the *Comprehensive Assessment of Spoken Language (CASL)* were selected. These represent receptive and expressive tasks that tap depth of vocabulary knowledge. The **Synonyms** subtest provides a standard score for receptive vocabulary. The child is given a word and must select a synonym from a spoken list. The memory load is limited as each choice is evaluated as given. The **Antonyms** subtest provides a standard score for expressive vocabulary and word retrieval. Given a word, the child must verbalize an antonym. This task is less taxing on sentence-level expressive skills than something like providing a definition. Reliability and validity information for the *CASL* were presented in the previous section on the Paragraph Comprehension subtest from this test.

Narrative measures

The *Test of Narrative Language* (*TNL*) provides scores for receptive and expressive discourse-level skills. The **Narrative Comprehension** score measures recall and understanding of story information across 3 levels of picture support: (1) no picture support, (2) five sequenced pictures, (3) single picture. The **Oral Narration** score measures ability for the child to tell a complete story including essential story grammar elements, using correct morpho-syntax across three levels of picture support (as described above). The oral narration standard score is calculated from performance on the three story levels. For the retell condition, children are scored according to the number of key words from the story that they use in their retell. For the sequenced story condition, children are scored according to temporal and causal connections between events, morpho-syntaxand overall judgments of story quality. The single-picture condition is scored for story grammar elements, vocabulary and morpho-syntax, and overall judgments of story quality. The *TNL* also provides a total test score, the **Narrative Language Ability Index**, that serves as a useful measure of overall narrative skills. The test manual indicates that this is the most reliable of the three measures.

The reliability and validity of the *TNL* appear to be acceptable. Reliability for the Narrative Comprehension score was reported 3 ways: internal consistency (0.87), test-retest (0.85) and inter-scorer (0.94). For the Oral Narration subtest, the reliability coefficients were: internal consistency (0.76), test-retest (0.82) and inter-scorer (0.90). Note that internal consistency (as Cronbach's alpha) was also calculated specifically with a group of children with language delay and the coefficients remained acceptable (Narrative Comprehension 0.89; Oral Narration 0.92).

The authors present a description of the validation of content: "Four demonstrations of content-description validity are offered for the *TNL* subtests and composite. First, a detailed rationale for the format of each subtest is presented. Second, we describe the rationale for the way we measured narrative language. Third, the validity of the items is ultimately supported by the results of the "conventional" item analysis procedures used to choose items during the developmental stages of test construction. Fourth, the validity of the items is reinforced by the results of differential item functioning analysis, used to show the absence of bias in a test's items" (p. 49). Criterion-prediction validity was demonstrated by large correlation coefficients (corrected for restricted range and attenuation) with the *Test of Language Development—Primary: 3rd Edition*: 0.78 for Narrative Comprehension and 0.81 for Oral Narration. The Oral Narration subtest scores also showed correlations from 0.66 to 0.79 with traditional language sample analysis measures.

Construct-identification validity was demonstrated with hypothesis testing methods. The test scores were expected to correlate to chronological age, test scores were expected to differentiate groups known to differ on language skills, and the formats of the test should reflect the theoretical model of the test. Age correlated 0.50 with Narrative Comprehension and 0.57 with Oral Narration; coefficients were significant at $p < 0.0001$. A comparison of 76 children with language impairment and 76 children with typically developing language was conducted. The sensitivity index was 0.92 and specificity was 0.87, demonstrated adequate ability to identify individuals on language status. Two principal component analyses were conducted. A one factor solution was obtained that was interpreted as validating the "narrative" construct of the test. Exploratory factor analysis for a 2 factor solution was interpreted as demonstrating the narrative comprehension and oral narration constructs. Confirmatory factor analysis was used to confirm the structure of the test scores, with 2 factors representing the comprehension and narration scores.

Control and matching variables

These variables are included to control for factors that are known to predict reading comprehension but are not the main areas of interest in this study.

Nonverbal IQ, *Kaufman Brief Intelligence Test (KBIT-2*; Kaufman &Kaufman, 2004). Since nonverbal IQ does correlate to reading and language skills, this variable was included for possible statistical control rather than used as a selection criterion; to select for nonverbal IQ would result in a sample that could be truncated on reading and/or language skills that correlate to nonverbal IQ. The reliability and validity of the *KBIT-2* appear to be acceptable. Median internal reliability for the nonverbal score was 0.86 for ages 4-18 years. Test-retest reliability for the nonverbal score had a mean of 0.83. The Verbal and Nonverbal scores of this instrument show moderate correlations. The median correlation for the *KBIT*-2 Composite with the global scores of the *WASI, WISC-III, WISC-IV* and *WAIS-III* was 0.83.

Word-level Reading / Decoding Skills. Significant variability was expected in decoding skill in the group of children with language impairment. Children in upper elementary can have persistent decoding difficulties, so it is important to control for this skill. Real word decoding was selected as an important correlate of reading comprehension. As word recognition has been shown to be related to vocabulary, and vocabulary measures are included here, it is desirable to have both nonword decoding and word recognition so that these can be investigated separately. In addition, Spear-Swerling (2004) found that in poorer readers, nonword decoding was related to reading comprehension difficulties even when real-word decoding was adequate. Previous research has shown that there is significant shared variance between listening comprehension and word-level reading skills in predicting reading comprehension scores. For reasons of feasibility, lower-level reading measures were selected from one of the achievement tests included in the study. Since administration of the word reading subtest of the *KTEA-II* is required prior to administering the reading comprehension subtest, it was reasonable to select the measures from the *KTEA-II*.

***KTEA-II* Nonsense Word Decoding** (NwD) subtest**.** Students are presented with a set of printed non-words of increasing difficulty, which they read aloud. The words are composed of letter strings that conform to the rules of English pronunciation. Words include common letter patterns, suffixes and inflections. The reliability and validity of this subtest appear to be acceptable. Split-half reliability varies from 0.93-0.95 for the grades in the proposed study. Alternate form reliability (adjusted) for the grades in the proposed study is 0.90. The nonsense word subtest of this instrument correlated at $r = .86$ with the pseudoword decoding subtest of the *WIAT-II* and $r = .66$ with Reading Recognition score of the *PIAT-R/NU*. It correlates at $r = .75$ with the Basic Reading Skills cluster of the *WJ-III*.

KTEA-II Letter and Word Recognition (LWR) subtest. The task starts with knowledge of letter names and sounds, progressing to early high-frequency words and moving then to words with less predictable pronunciations. The reliability and validity of this subtest appear to be adequate. The split-half reliability coefficients for the ages in this study are 0.95 – 0.96. In terms of validity, the relationships of the *KTEA-II* subtests and composites were validated using confirmatory factor analysis. Correlations with the Word Reading subtest of the *WIAT-II* for the grades in this sample are $r = .79$ to.83. Correlations with the Reading Recognition subtest of the *PIAT-R/NU* for the grades in this sample are $r = .73$ to .85. It correlates with the Basic Reading Skills cluster of the *WJ-III* at $r = .79$ to .83.

Working memory. On the *CELF-4* **Familiar Sequences** task, students perform a series of timed working memory tasks with automatized information such as alphabet, counting and days of the week. Tasks involve re-ordering and manipulating this familiar knowledge. See Appendix C for an evaluation of this task as a measure of working memory. On the *CELF-4* **Number Repetition** tasks, students repeat increasingly long lists of numbers, first in the order presented and then in reverse. The scores on the two tasks are combined to yield the *Number-Repetition—Total* (NR-T) score. The Familiar Sequences and Number Repetition—Total scores are used to determine the overall Working Memory Composite, as per procedures presented in the *CELF 4* Test Manual. The reliability and validity of these subtests appears to be adequate. The test-retest correlations for Familiar Sequences ranged from r = .78 - .81 (correlation for 11 years not reported). For Number Repetition—Total, the test-retest correlations ranged from r = .78 - .79 (correlation for 11 years not reported). In terms of validity, the overall model of the *CELF-4* was validated via factor analysis. In addition, studies regarding special populations were conducted. These studies showed that the *CELF-4* scores, including the memory tasks, differentiated between the clinical groups (children with language learning disabilities, children with autism, children with cognitive delay, and children with hearing impairment) and the matched samples.

Mother's years of education. As described in the literature review, this measure is related to socio-economic status. Parents self-reported this information on the study consent forms.

Procedures and Ethical Considerations

As per University guidelines, this study passed a Research Ethics Board review. All procedures for contact with participants, confidentiality and data management were conducted following procedures approved by the Ethics Board.

Signed informed consent from the children's parents and assent from the children were obtained for all participants prior to any testing. Since children with no previous language diagnosis were tested, there was the potential to identify students with possible language difficulties. There was also the potential to identify hearing problems. Parents were informed of this possibility before consent was obtained, and were given the option of allowing the researchers to share screening results with school staff for further investigation. The data were collected and scored by three qualified speech-language pathologists.

With elementary-aged children as the participants, schools were the natural context for this study. The study ran in a number of schools in one local school district, during regular school hours. Students worked individually with a researcher, in a quiet room, such as an extra classroom, office or conference room. The full test battery required about 2 hours 45 minutes to administer, with a range of one to four sessions per student. Typically two sessions were required to limit fatigue and to accommodate classroom needs.

Students were called to the office and were introduced to the researcher by school staff. The student and researcher proceeded to the testing room. For later sessions, students were either called by the office or met at the classroom by the researcher, according to school preferences. The session began with informal chat to put the student at ease. Then the researcher read an assent letter aloud. Since the study sample was meant to include students with a wide range of language and reading abilities, this procedure was intended to ensure every participant could understand the letter. Any questions were answered and the students were given the opportunity to give or refuse assent. If they assented, then the hearing screening was performed with a portable audiometer per current guidelines at 20 dB, at 1000, 2000 and 4000 Hz for both ears. After the child passed the screening, the researcher began testing with the *CELF* Core Language subtests. The order of testing at this point depended on student needs and on scheduling. For example, if only a short time remained before recess, then a short task was selected. Researchers tried to collect the most key information first (*CASL* subtests, *TNL*, and *KTEA-II* subtests). Tasks were alternated as needed to keep up student interest. Language-loaded or long tasks were interspersed with shorter and less demanding tasks. Students were given regular breaks and breaks as requested. Sometimes session length was determined by school scheduling. For example, teachers were able to identify blocks when students needed to be back in class. Testing was discontinued at any time by participant request. Once all the tasks were finished, the participant was thanked and returned to regular school activities. Because some students were undergoing assessment as part of their regular schooling, school procedures and policies were followed in this regard.

Data Analysis

Rationale for selection of multiple regression

The intent of the analysis was to identify component skills contributing to standardized reading comprehension scores among elementary school students. Of primary interest in this study is whether language status (as impaired or not impaired) predicts scores on two reading comprehension tests. Also of interest is whether significant predictors differ between the two reading comprehension tests. Existing evidence shows that the predictor variables in this analysis can be expected to show complex patterns of relationship, both with one another and with the outcome variables. Thus an analysis technique is required that is suited to revealing relationships among a number of variables. This technique must allow the modelling of variance structure to help reveal significant relationships including new and previously-studied variables. Factor analysis could be used for this purpose as it would permit the modelling of variance and covariance structure directly. However, factor analysis has a limitation in that large sample sizes are required. Multiple linear regression allows for model building to account for predicted variance patterns based on existing evidence and theory. Using hierarchical entry, the models can include a number of predetermined steps to test these patterns. Although a structure does not emerge from the data in the same way as in factor analysis, strong models can be built as long as the study has sufficient power. It is important to state that the term "predictor" is used throughout this thesis in referring to the independent variables included in the regression models. This is consistent with accepted parlance for multiple regression but should not be interpreted as indicating "prediction" in a longitudinal sense. The current analysis is based on concurrent data, not longitudinal. The term "predictor" is used to refer to the variables that combine in a regression model to best explain (or in regression terms, predict) the dependent variable, in this case reading comprehension scores. The current analysis was computed using IBM SPSS Statistics software (version 23).

Planned regression models

Hierarchical regression was employed, following the method reported by Cutting and Scarborough (2006). Sets of planned comparisons were made to contrast variance patterns in the constructs of interest, in particular language status (impaired/not) and oral language skills (vocabulary, morpho-syntax and narratives and receptive/expressive language). Each process was run twice, one for each of the two reading comprehension tests.

As an initial check, a model was calculated for each of the reading comprehension tests, with all predictors and controls entered in a single block.

The planned comparisons are presented in Table 3 below. In Model 1 for each reading comprehension test language status as impaired/not impaired was entered first. Next the language skill composites were entered. Memory and decoding skills were entered next, as basic cognitive and reading skills thought to underlie reading comprehension (consistent with Cutting & Scarborough, 2006). In the second and third iterations, these blocks were shuffled. This determined the amount of unique and shared variance for the variable blocks (see Model 1a, 1b, 1c, and 1d). Nonverbal IQ (NVIQ) was added last as the groups are unmatched on these variables. Since the significance of NVIQ to reading is poorly understood, this variable has limited explanatory power.

Next, it was of interest to evaluate the language constructs individually (see Model 2). In this iteration, variables that were not predictive in Model 1 could be eliminated to preserve power for the comparisons of interest. Thus, NVIQ and Working Memory were omitted.

Finally, receptive vs. expressive language scores were evaluated in Model 3. The models including all iterations are presented in Table 3.

Table 3

Planned comparisons for Models 1, 2, and 3

	Model 1				Model 2			Model 3	
	a	b	c	d	a	b	c	a	b
CELF 4/5 Core/ Lang Status	1	2	3	3					
Working Memory	3	1	1	2					
Decoding	3	1	1	2	4	4	4	3	3
Vocab Composite	2	3	2	1	1	3	2		
Receptive Vocab								1	2
Expressive Vocab								2	1
Syntax Composite	2	3	2	1	2	1	3		
Receptive Syntax								1	2
Expressive Syntax								2	1
Narrative Composite	2	3	2	1	3	2	1		
Rec. Narrative score								1	2
Exp. Narrative score								2	1
NVIQ	4	4	4	4					

Note. *CELF4/5 = Clinical Evaluation of Language Fundamentals, 4th or 5th Edition*; NVIQ = nonverbal IQ

The regression process laid out in Field (2013) was employed. Data were reviewed for linearity and outliers following regression analysis (with N = 54, Central Limit Theorem suggests normality is likely; Field, 2013). For each regression, residuals were checked for violations of assumptions. Any concerns are reported with the model results in Chapter 4.

Chapter 4 Results

The results of the study are presented in two main sections. The first section presents the results of the planned regressions, with any concerns that arose during these processes. In general, these concerns were the result of interesting relationships among the predictor variables. As previously noted, a key factor in the analyses was to identify and account for such inter-relationships. The second section presents analyses that were undertaken to address the concerns that arose in the original regressions.

Data Cleaning and Preparation

Prior to analysis, all standard scores were transformed to a distribution with a mean of 100, standard deviation of 15, to eliminate numerical artifacts in the regression procedures. This was necessary because although all data were recorded as standard scores, some subtests had a distribution with the mean of 10 and standard deviation of 3. The transformation procedure also assured that the regression coefficients could be directly compared to assess effect sizes for the various predictors, with the exception of language status which was a categorical variable. When it was of interest to compare language status to other variables, standardized regression coefficients were compared. A total of eight data points were missing for language predictor variables (two for one participant, six for another). In each case imputation by the mean for that predictor was used, following the advice of a statistical consultant. Digit span was unavailable for six participants. For these students, the score for Familiar Sequences was imputed for their Working Memory composite. Following data entry and calculation of new variables (described in the next section), a data integrity check was performed. Five participant data sets were selected, spread throughout the sample. For each data set, values in the database were compared with original data. This was intended to catch any systematic errors that could have occurred moving between programs (Access, Excel, and SPSS) or data files within these programs. In addition, the Core Language Score and *KTEA* Reading Comprehension Score for every participant was checked. No systematic inconsistencies were found.

Planned Regression Models

Determination of predictors to include in the regression model.

Based on power calculations using the *Power Analysis and Sample Size (PASS)* software (version 12), power of 80% can be achieved in this study with a sample size of 54, assuming R^2 of 0.2, α = 0.05 with eight predictor and three control variables. In this dataset, multi-collinearity effects were expected since language skills naturally function in an integrated fashion. Narrative skills, for example, are known to integrate both semantics and syntax. Multi-collinearity can affect the power of the planned analysis. Therefore, prior to performing the regression, the correlation matrix for all variables entered into the regression analyses (see Table 4) was analyzed to identify high correlations that might suggest multi-collinearity and to inform creation of theoretically-valid composite variables. Use of such composites can limit the effects of multi-collinearity on regression results as variance-splitting among closely related variables can be avoided. Any relevant observations of the correlation matrix are presented in the following sections on formation of composite variables.

Table 4

Correlation matrix of predictor and control variables

	Decoding	Synonyms	Antonyms	NDW	Vocab (adj.)	MLTU	PCS	Para-graph	Narr comp	Oral narr	NLAI	Fam. Seq	NR-T	WM	NwD	LWR	*WMRT*	*KTEA*	*KBIT*
Decoding	1	.342*	.496**	.353**	.448**	.130	.258	.143	.471**	.417**	.529**	.763**	.519**	.739**	.943**	.937**	.629**	.741**	.382**
Synonyms		1	.757**	.249	.936**	.134	.095	414**	.119	.385**	.286*	.285*	.378**	.339*	.205	.444**	.536**	.565**	.260
Antonyms			1	.296*	.938**	.198	.258	.519**	.338*	.385**	.418**	.427**	.470**	.483**	.352**	.587**	.656**	.646**	.286*
NDW				1	.291*	-.040	.106	.457**	.451**	.780**	.710**	.250	.105	.184	.325*	.339*	.248	.386**	.344*
Vocabulary (adj.)					1	.177	.189	.498**	.245	.411**	.376**	.380**	.451**	.439**	.298*	.551**	.636**	.646**	.291*
MLTU-w						1	.680**	.067	.268	-.020	.173	.210	.398**	.322*	.037	.212	.336*	.180	.281*
PCS							1	.208	.331*	.120	.303*	.300*	.404**	.373**	.154	.335*	.269	.278*	.371**
Paragraph								1	.288*	.567**	.505**	.045	.116	.017	.005	.271*	.240	.367**	.325*
Narrative comp									1	.515**	.858**	.543**	.325*	.511**	.415**	.472**	.276*	.475**	.414**
Oral narrative										1	.871**	.386**	.200	.314*	.353**	.433**	.361**	.456**	.538**
NLAI											1	.526**	.318*	.480**	.448**	.549**	.380**	.555**	.557**
Familiar sequences												1	.534**	.911**	.697**	.738**	.549**	.663**	.509**
NR-T													1	.856**	.424**	.552**	.566**	.529**	.488**
WM														1	.656**	.734**	.611**	.678**	.533**
Nonword Decoding															1	.768**	.529**	.562**	.281*
LWR																1	.657**	.837**	.441**
WMRTR																	1	.678**	.361**
KTEA-II																		1	.377**
KBIT																			1

- Oral language variables
- Working memory variables
- Reading variables
- Nonverbal IQ

Note: NDW = Number of different words; Vocab (adj.) = Vocabulary composites calculated without NDW; MLTU = Mean length of T-Unit in words; PCS = Percent complex sentences; Paragraph = Paragraph comprehension; Narr comp = Narrative comprehension; Oral narr = Oral narration; NLAI = Narrative Language Ability Index; Fam Seq = Familiar sequences; NR-T = Number repetition-total score; WM = Working memory composite score; NwD = Nonword decoding; LWR = Letter and word recognition; *WRMT = Woodcock Reading Mastery Test- Revised/Norms Updated,* Passage comprehension subtest*; KTEA = Kaufman Test of Educational Achievement, 2nd Edition,* Reading comprehension subtest; *KBIT= Kaufman Brief Intelligence Test, 2nd Edition*

* $p < 0.05$ level (2-tailed); ** $p < 0.01$ level (2-tailed).

Using the purposeful selection method, a linear regression was also used to evaluate predictors. Using this method, those variables with coefficients greater than 0.25 in a simple linear regression with the dependent variables would be included in the model. If measures of a single language construct (e.g., vocabulary) show strong correlations with each other ($r > 0.8$; Field, 2013) and exceed the 0.25 inclusion criterion, either the variable with the highest correlation with the dependent variable may be used, or a composite variable can be created. Composites were created only when there was sound theoretical reason to do so, that is in cases where the component variables fell under the same language construct (e.g., vocabulary). Besides the scrutiny of the correlation matrix, the SPSS collinearity diagnostics were examined following the analyses. Tolerance greater than 0.2 was considered acceptable. This criterion was met for all models.

Correlations among the predictor and control variables. The observed correlations among scores for working memory, word reading and reading comprehension were compared to reported correlations for these or similar measures. This process was undertaken as a reliability check for the current sample. Since many of these variables have been measured previously, it was of interest to determine if the current observations were broadly in line with reported values. If observations were consistent with existing reports, it would suggest that the current methods and sample were comparable to those used by other researchers. Overall, current results were consistent with expected values. Expected values are from the standardization data of the tests that were used in the study. When available, reported correlations between the tests are included in the table (superscripted as “R”). When such correlations were not available, approximations based on similar tests are included (superscripted as “A”). Sources of reported correlations and approximations are described following the table.

Table 5

Observed and expected correlations for reading and working memory variables

Variable	NwD		*KTEA-II*		*WRMTR*		Fam Seq.	NR-T	
	Obs	*Exp*	Obs	*Exp*	Obs	*Exp*	Observed	Obs	*Exp*
LWR	0.77**	0.78^{R}	0.84**	0.63^{R}	0.66**	0.59^{A}	0.74**	0.55**	$0.61 - 0.68^{A}$
NwD	--		0.56**	0.51^{R}	0.53**	0.48^{A}	0.70**	0.42**	$0.64 - 0.67^{A}$
KTEA-II			--		0.68**		0.66**	0.53**	$0.48 - 0.60^{A}$
WRMTR					--		0.55**	0.57**	
Fam Seq.							--	0.53**	0.5^{R}

Note. Observed (Obs) correlations are as measured in the current study. Expected (Exp) correlations are superscripted "R" for correlations reported in the test manuals or "A" for approximations based on similar tests or values reported in the test manuals. Correlations among LWR, NwD and *KTEA* scores are as reported in the *KTEA-II* test manual; NR-T and FS are as reported in *CELF 4* test manual. *WRMTR* correlations for LWR and NwD are approximations based on values reported for *Word Identification* (similar to LWR*)* and *Word Attack* (similar to NwD) in the *WRMTR* test manual (Grade 5 values). Correlations noted between LWR and NwD with NR-T are approximations using reported values for correlations between these reading scores and the Working Memory subtest on the *Woodcock-Johnson Tests of Cognitive Ability—Third Edition* as reported in the *KTEA-II* manual. A correlation between *WRMTR* and *KTEA-II* reading comprehension was not available. NwD = Nonword decoding; *KTEA-II = Kaufman Test of Educational Achievement, 2nd Edition,* Reading comprehension subtest; *WRMTR = Woodcock Reading Mastery Test-Revised/Norms Updated,* Passage Comprehension subtest; Fam Seq = Familiar sequences working memory subtest from the *Clinical Evaluation of Language Fundamentals, 4th Edition;* NR-T = Number repetition—total from the *Clinical Evaluation of Language Fundamentals, 4th Edition.*

** significant at $p < 0.01$; * significant at $p < 0.05$.

Letter and Word Recognition (LWR) and the *KTEA* reading comprehension score are correlated consistently with expectations, but at a higher magnitude than the expected value (r = .63 for 9 to 12 year olds). Nonsense word decoding was correlated similarly to the two reading comprehension tests.

The working memory measures had a variety of significant correlations with reading measures. The Familiar Sequences score was significantly correlated with both nonsense word reading and real word reading. Digit span (NR-T) was also significantly correlated to these measures, but at lower levels (r = .55 and r = .42). In fact it is interesting to note that digit span appears to be similarly correlated to all the reading and working memory measures at around the r = .4 to .5 level.

Since working memory is expected to be tapped by language tests, it is also of interest to review these relationships. Based on the number of significant correlations among working memory and language scores seen in Table 4, most of the oral language variables are also tapping working memory skills. It also appears that the two working memory subtests are capturing related but somewhat different aspects of working memory as represented in these tasks. Note that no significant correlations were observed with Paragraph Comprehension or with Number of Different Words. In terms of regression, it appears to be appropriate to combine the two memory variables into a composite, available from the *CELF 4* scoring.

Word reading measures and language variables. Letter and Word Recognition was more highly correlated to *every* language score than was Nonsense Word Decoding. Word recognition is related to vocabulary knowledge and as expected LWR was more highly correlated to Synonyms and Antonyms, both measures of depth of vocabulary, than to other language measures. It is interesting to note that LWR was not as highly correlated to Number of Different Words. It was interesting and somewhat unexpected to note that LWR was strongly correlated with the narrative composite, NLAI.

Evaluation of predictors with purposeful selection method. The purposeful selection method validated the selection of these variables as potentially predictive in a multiple regression analysis. Only MLTU-w failed the selection process relative to the *KTEA-II*, but passed relative to the *WRMTR*. Since the analyses were intended to compare results across the two tests, MLTU-w was retained as a possible predictor rather than risk losing its value to the *WRMTR*.

Formation of language composites. As planned, the language variables were combined into theoretically-based composites following the intentional selection process. Composites were used when necessary to preserve power, as predictors in the regression analyses. These did not include variables from the *CELF 4* or *CELF 5* tests (used in the calculation of the participants' Core Language scores). Since these Core Language scores were used to determine participant group status, it was more statistically valid not to include these subtest scores as predictors. The distributions of the variables were graphed. All appeared to be reasonable approximations of normal with the exception of Number of Different Words. Narrative Comprehension and Narrative Language Ability Index (NLAI) were somewhat borderline approximations of normal distributions.

For use in Model 3, composites of receptive and expressive skills were created as necessary in the areas of vocabulary and syntax. Please see details in the following sections.

Vocabulary composite (VC). Synonyms and Antonyms were significantly correlated to one another (r = .76, p < .001). These variables were more highly correlated with each other than with any other language score. Both Synonyms and Antonyms were significantly correlated to Word Classes (WC) (r = .43, p = .001; r = .42, p = 0.002). Antonyms was significantly correlated to Number of Different Words (NDW) (r = .30, p = .030). NDW was also significantly correlated to WC (r = .30, p = .025). The correlation between Synonyms and NDW was similar (r = .25, p = .069). NDW showed statistically significant correlations to variables that tapped discourse-level skills (Paragraph Comprehension, Narrative Comprehension, and Oral Narration). NDW appeared to have a different distribution than the other scores, which may call into question the mathematical validity of including it in a composite with the other vocabulary scores. Accordingly, the Vocabulary Composite was calculated with and without NDW, and the collinearity statistics following the regressions were examined for any impact. The vocabulary composites were calculated as unit-weighted means, as follows:

- Vocabulary Composite = PVC = (Synonyms+Antonyms+NDW)/3
- Receptive Vocabulary Score=Synonyms
- Expressive Vocabulary Composite= Antonyms + NDW/2

Morpho-Syntax Composite (SC). Formulated Sentences (FS) was significantly correlated with a number of language measures. Its highest correlation was with Recalling Sentences (RS) (r = .68, p < .001). It was also significantly correlated with Paragraph Comprehension (PC), our receptive syntax measure (r = .41, p = .002). Note that FS was also significantly correlated to Synonyms and Antonyms (r = .56, p < .001; r = .64, p < .001). This is not a concern as the ability to make semantically-plausible sentences reasonably subsumes knowledge of word meanings. The correlation between RS and PC was more moderate (r = .29, p = .032). MLTU-w and Proportion of Complex Sentences (PCS) were correlated at the highest level of magnitude (r = .68, p < .001). This was consistent with the factor analysis reported by Justice et al. (2006) in terms of narrative microstructure analysis. Both MTLU-w and PCS were moderately correlated to the Concepts and Directions score (CD) (r = .33, p = .015; r = .30, r = .026), a receptive syntax measure. CD had higher correlations to the syntax measures FS and RS (r = .62, p < .001; r = .58, p < .001). The correlation with PC was statistically significant (r = .42, p = .002). Note that CD was also significantly correlated to the Vocabulary measures, but this is expected given that they are all measures of semantics in some ways. The syntax composites were calculated as unit-weighted mean scores, as follows:

- Syntax Composite=PSC= (PC+ PCS + MLTU-w)/3
- Receptive Syntax Score: PC
- Expressive Syntax Composite: (PCS+MLTU-w)/2

Narrative composite (NLAI). For this composite, the score for Narrative Language Ability Index from the *Test of Narrative Language* was used. This score is derived from the receptive and expressive narrative subtests of that instrument, Narrative Comprehension (NC) and Oral Narration (ON). Based on the correlations from the sample data, this appears to be a valid approach. The two were significantly correlated (r = .52. p < .001) and had few strong correlations with other variables. Oral Narration showed two other relatively high correlations, Number of Different Words (r = .78, p < .001) and Paragraph Comprehension (r = .57, p < .001). Number of Different Words was calculated from a language sample from among the Oral Narration tasks. As another discourse task, it is sensible that Paragraph Comprehension should be related to narrative skills. Since scores from the *Test of Narrative Language* were used, no additional calculations for narrative composites were done.

Summary of variables selected for regression models.

The preceding sections detailed how the predictor variables for the regression models were selected. Descriptive statistics for the variables included in the models are provided in Table 6, reported for all participants and divided by participant group (children with and without LI).

Table 6

Descriptive statistics for dependent and independent variables for all participants (N = 54), children with language impairment (LI) (N = 24) and children with typically developing language (TDL) (N = 30)

Variables	Range			Mean			Std. Deviation		
	All	LI	TDL	All	LI	TDL	All	LI	TDL
Synonyms	65.00-150.00	65.00-119.00	79.00-150.00	101.44	94.75	106.80	14.16	10.96	14.29
Antonyms	71.00-135.00	71.00-114	82.00-135.00	99.70	91.88	105.97	14.38	9.91	14.44
NDW	68.50-172.45	68.50-172.45	79.00-170.20	101.37	94.49	106.88	22.65	21.97	22.00
Paragraph comprehension	70.00-132.00	70.00-125.00	92.00-132.00	104.81	98.83	109.60	12.81	13.97	9.59
MLTU-w	73.45-140.65	73.45-140.65	77.50-135.85	100.25	99.09	101.18	14.71	17.82	11.90
PCS	58.75-133.75	58.75-133.5	73.75-115.00	94.12	89.87	97.52	14.98	18.45	10.63
Narrative Comprehension	75.00-135.00	75.00-125.00	85.00-135.00	104.07	98.96	108.17	14.47	14.37	13.42
Oral Narration	70.00-130.00	70.00-120.00	86.00-130.00	94.82	87.92	100.33	14.60	15.32	11.52
KBIT	66.00-131.00	66.00-115.00	67.00-131.00	99.31	91.75	105.37	16.16	11.72	16.83
WMRTR[1]	71.00-120.00	71.00-99.00	75.00-120.00	95.02	90.42	98.83	11.21	7.32	12.50
KTEA-II[1]	67.00-132.00	67.00-103.00	68.00-132.00	94.54	85.83	101.50	15.70	10.02	16.06
Composite variables									
Vocabulary	78.15-141.32	78.15-113.82	86.77-141.32	100.84	93.70	106.55	13.25	9.54	13.14
Vocabulary (adjusted)	73.00-141.50	73.00-116.50	84.50-141.50	100.57	93.31	106.38	13.37	9.27	13.42
Expressive Vocabulary	76.88-148.60	76.88-130.23	86.23-148.60	100.54	93.18	106.42	15.10	12.30	14.71
Syntax	76.73-118.73	76.73-118.57	87.23-118.73	99.73	95.93	102.77	10.59	12.50	7.71

Expressive Syntax	66.10-130.53	66.10-130.53	81.85-117.55	97.18	94.48	99.35	13.60	17.12	9.74
NLAI (narrative)	73.00-136.00	73.00-124.00	85.00-136.00	99.56	92.75	105.00	14.82	14.04	13.26
Decoding	68.5-126.50	74.00-103.00	68.50-126.50	94.16	89.04	98.25	13.63	10.05	14.85
Working Memory	55.0-123.00	63.00-115.00	55.00-123.00	88.33	80.33	94.73	16.91	14.51	16.14

Note. NDW = Number of different words; MLTU-w = Mean length of T-Unit in words; PCS = Proportion of complex sentences; *WRMTR = Woodcock Reading Mastery Test- Revised/Norms Updated,* Passage comprehension subtest*; KTEA-II = Kaufman Test of Educational Achievement, 2nd Edition,* Reading comprehension subtest; *KBIT= Kaufman Brief Intelligence Test, 2nd Edition*; NLAI = Narrative Language Ability Index

[1] Dependent variables

Results of planned regressions.

The results of the planned regressions are presented in three models, following the order introduced in Table 3.

Model 1 Language status vs. Control factors vs. Language skill factors. Model 1 was designed to examine the main constructs of interest in this study, language status (as impaired/not impaired) and language skills, and to account for any significant impact of the control variables. The first regression analysis presented, Model 1a, places language status on the first step of the analysis as a key predictor of interest. Next, specific language skills are entered as a block, represented as composites for vocabulary, syntax and narrative skills. Then controls are entered, starting with the variables that are expected to have more explanatory value, decoding and working memory, and finally nonverbal IQ (NVIQ). Models 1b through 1d simply shuffle the order of entry of these variables to highlight patterns of shared and unique variance. Table 7 presents these models with the results for each step, including the predictors entered.

Regression coefficients for each predictor are as calculated for the final model step unless otherwise noted. Note that at each step the predictors are entered as a block.

Table 7

Regression results for Model 1

	KTEA-II					*WRMTR*				
Model 1a	*R*	R^2	R^2_Δ	β	Std β	*R*	R^2	R^2_Δ	β	Std β
Step1	.501	.251	.251*			.377	.142	.142*		
Language Status				*2.128*	.136				*-.195*	-.017
Step 2	.703	.495	.244*			.604	.365	.223*		
Syntax				.114	.077				.175	.167
Narrative				.006	.006				-.193	-.257
Vocab				.370	.313*				.347	.413*
Step 3	.837	.701	.206*			.754	.568	.203*		
Decoding				.431	.374*				.252	.306
Working Memory				.234	.252				.196	.296
Step 4	.842	.708	.007			.754	.568	.000		
NVIQ				-.114	-.117				.000	.000
Model 1b	*R*	R^2	R^2_Δ			*R*	R^2	R^2_Δ		
Step1	.766	.587	.587*			.666	.444	.444*		
Decoding										
Working Memory										
Step 2	.797	.635	.048*			.674	.454	.010		
Language Status					(.243*)					
Step 3	.837	.701	.066*			.754	.568	.114*		
Syntax										
Narrative										
Vocab										
Step 4	.842	.708	.007			.754	.568	.000		
NVIQ										
Model 1c	*R*	R^2	R^2_Δ			*R*	R^2	R^2_Δ		
Step1	.766	.587	.587*			.666	.444	.444*		
Working Memory										
Decoding										
Step 2	.831	.691	.104*			.753	.568	.123*		
Syntax										
Vocab										
Narrative										

Step 3 Language Status	.837	.701	.010	.754	.568	.000
Step 4 NVIQ	.842	.708	.007	.754	.568	.000
Model 1d	R	R^2	R^2_Δ	R	R^2	R^2_Δ
Step1 Syntax Narrative Vocab	.681	.464	.464*	.598	.358	.358*
Step 2 Language Status	.703	.495	.031	.604	.365	.007
Step 3 Decoding Working Memory	.837	.701	.206*	.754	.568	.203*
Step 4 NVIQ	.842	.708	.007	.754	.568	.000

Note. β in parentheses is for step 2; all others are for final step of the model. The non-standardized β for Language Status is in italics as a reminder that it is on a different metric than the other variables and therefore cannot be directly compared.

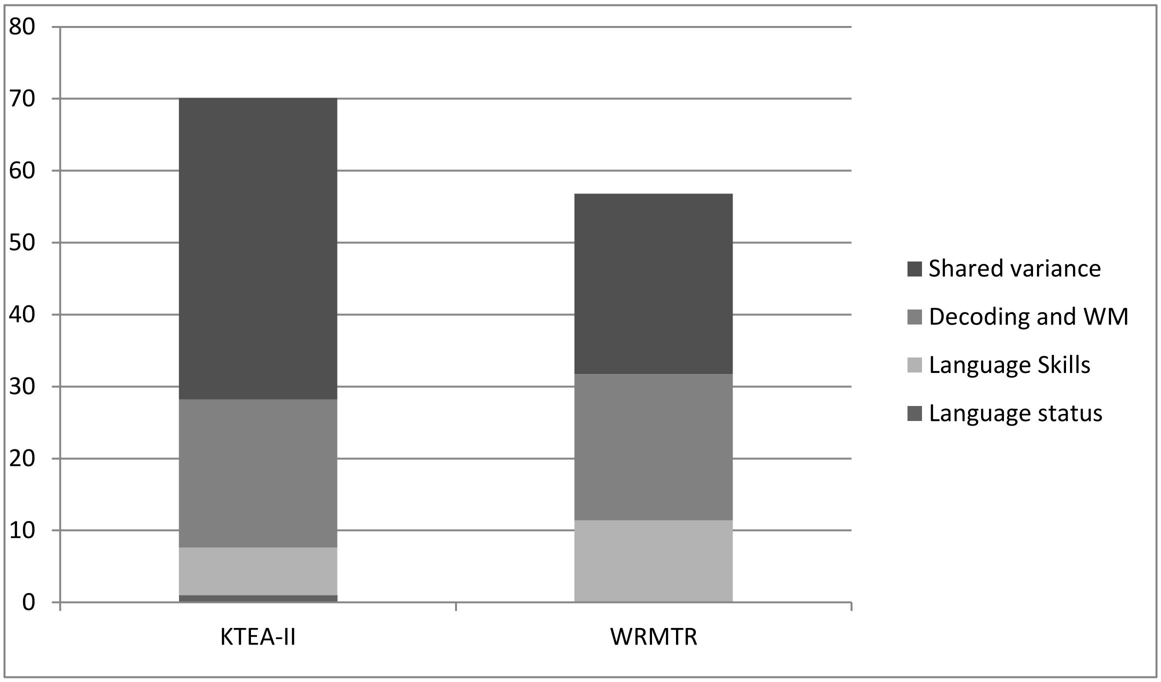
*significant at $p < .05$.

Regression diagnostics. For both models (*KTEA-II* and *WRMTR*), assumptions of parametric analyses were checked. All conditions were acceptable, although some potential outliers were identified. Outliers are addressed in detail in the section on Follow-Up Analyses later in this chapter. There were indications of multicollinearity effects as predicted in the analysis of the correlation matrix. In particular, decoding, vocabulary and working memory appeared to be related (these variables all had their highest variance loadings on the same eigenvalues; method per Field, 2013, p. 342). The regression coefficient for NLAI in the *WRMTR* models was negative, an unexpected effect possibly due to shared variance or outlier(s) affecting that variable. The same effect, an unexpected negative β value, was noted for NVIQ in the *KTEA-II* models. Note that these negative β values were not statistically significant. Note also that NLAI and NVIQ were significantly correlated (r = .557, p = .000). Full regression diagnostics are reported in Appendix D for Model 1a only to demonstrate the methods used to evaluate the analyses.

Significant results for Model 1. Notice in Model 1a that Language Status is a useful predictor of reading comprehension in step 1 for both tests and in step 2 for the *KTEA-II*. For the *KTEA-II*, Language status alone explained 25% of variance in reading comprehension scores. For the *WRMTR*, it explained 14% of variance. However, by the fourth step and final step in each model two other predictors appeared to be explaining most of the variance: decoding and vocabulary. NVIQ did not appear to contribute significant unique variance in the final step (step 4), for either reading comprehension test. Looking at the final steps, these models account for 71% of variance in *KTEA-II* scores and 57% of *WRMTR* scores. Following the method used by Cutting and Scarborough (2006), the unique and shared variance of the blocks are represented graphically in Figure 1. Note that the values from Step 3 of each model were used to determine the unique contribution of the other predictors since NVIQ had minimal unique variance to report. The variance proportions represented then include the influence of NVIQ.

Figure 1

Variance proportions of major constructs in Model 1



In models for both reading comprehension tests, the largest β values (regression coefficients) are for decoding and vocabulary. The next largest β is working memory, although it does not reach statistical significance. It is also important to note that there is the possibility of collinearity between decoding and working memory. Although tolerances are above criterion, decoding and working memory have the lowest tolerances; the two variables also both load on the same small eigenvalue, as does vocabulary. An odd effect is present for narratives in the *WRMTR* models—the β is a rather large negative value. Narrative ability also had a negative β value in the *KTEA-II* models in step 3, before NVIQ was taken into account. These observations again suggest that something unexpected is going on. Since the simple correlations between the *KTEA-II* and NLAI (the narrative composite measure) and the *WRMTR* and NLAI are positive, this negative β will need further explanation. Although these negative β values are not statistically significant, obtaining negative values rather than positive values as expected indicates that the overall model should be scrutinized for unanticipated relationships among the variables. There was a concern with Number of Different Words noted in the formation of the language composite variables, as it appeared to correlate more consistently with discourse level measures rather than vocabulary. This could be causing the problem. Since the vocabulary composite is holding so much variance, the NDW term could be pulling out variance that is related to narrative skill. Another curiosity to note regarding NLAI: it correlates rather highly to decoding (r = .529, $p > .001$). One final observation of interest: for the *WRMTR*, the β for Syntax is relatively large, although not statistically significant. Clearly more digging will be required to uncover these relationships satisfactorily. For now, the results of Models 2 and 3 await.

Model 2 Discrete language constructs (Vocabulary, Syntax and Narratives). Model 2 was designed to more specifically examine the language constructs for this study: vocabulary, syntax and narratives. Accordingly, this set of models shuffles the order of entry among the following variables: syntax, narratives and vocabulary, followed by decoding. It was of interest to contrast the variance proportions of the three language composites, as well as note the impact of decoding. Language status was not entered, as based on the results of Model 1 it would hold little unique variance once language skills had been accounted for. The results of Model 2 are presented in Table 8.

Table 8

Regression results for Model 2

	KTEA-II					*WRMTR*				
Model 2a	R	R^2	R^2_Δ	β	Std β	R	R^2	R^2_Δ	β	Std β
Step1	.655	.429	.429*			.569	.324	.324*		
Vocab				.409	.345*				.327	.390*
Step 2	.667	.445	.016			.597	.356	.032		
Syntax				.156	.106				.220	.210
Step 3	.681	.464	.019			.598	.358	.002		
Narrative				-.006	-.005				-.173	-.231
Step 4	.815	.665	.201*			.729	.531	.173*		
Decoding				.628	.545*				.415	.504*
Model 2b	R	R^2	R^2_Δ			R	R^2	R^2_Δ		
Step1	.363	.132	.132*			.379	.144	.144*		
Syntax										
Step 2	.572	.327	.196*			.449	.201	.058		
Narrative										
Step 3	.681	.464	.136*			.598	.358	.156*		
Vocab				(.199)	(.188)					
Step 4	.815	.665	.201*			.729	.531	.173*		
Decoding										
Model 2c	R	R^2	R^2_Δ			R	R^2	R^2_Δ		
Step1	.555	.308	.308*			.380	.144	.144*		
Narrative				(.588)	(.555*)				(.285)	(.380*)
Step 2	.675	.456	.148*			.569	.324	.180*		
Vocab										
Step 3	.681	.464	.008			.598	.358	.034		
Syntax										
Step 4	.815	.665	.201*			.729	.531	.173*		
Decoding										

Note. β in parentheses are for step at which they appear; all others are for step 4 of the model. See Appendix D for regression diagnostics for Model 2a

*significant at $p < .05$

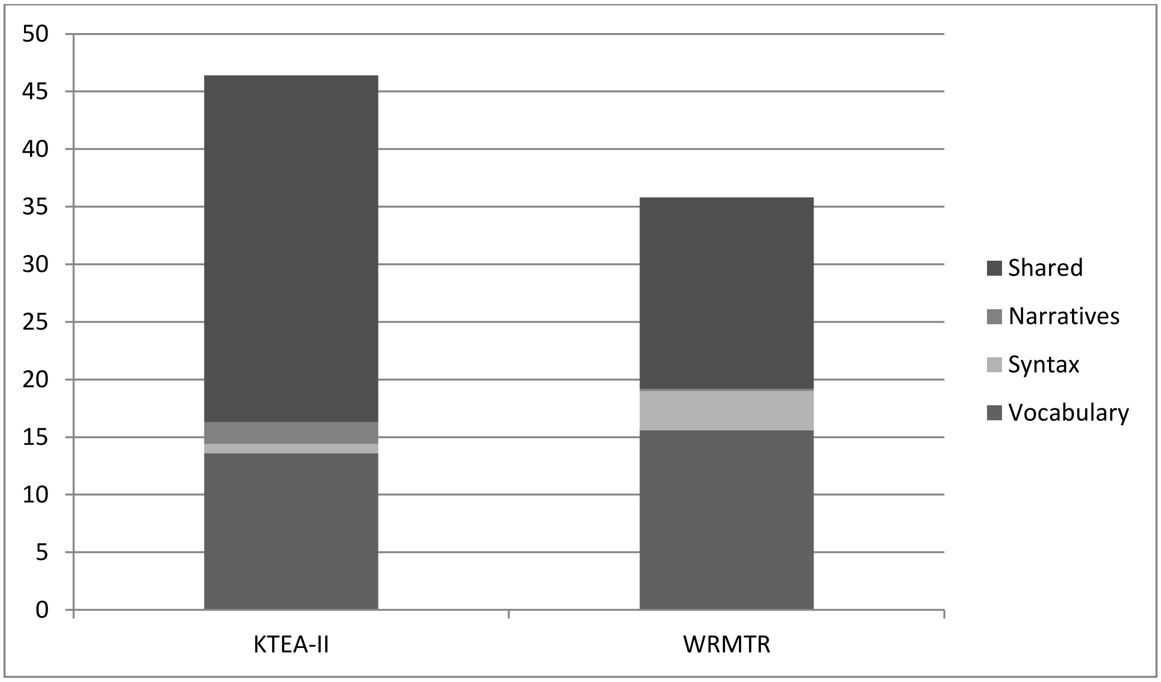
Significant results for Model 2. The language constructs account for 46% of the variance in *KTEA-II* reading comprehension scores before taking into account decoding and language status. For the *WRMTR*, the language constructs account for only 36% of variance. Note that for both tests, the strongest predictors overall were vocabulary and decoding as observed in Model 1. It is interesting to note that each of the language constructs was a significant predictor on its own for both tests, entered in the first step. Both vocabulary and narratives accounted for more variance on the *KTEA-II* than on the *WRMTR*. Syntax, however, accounted for similar amounts of variance on the two tests.

These models also reveal another important clue about the narrative variable. Note that narratives alone accounted for 31% of variance on the *KTEA-II* and 14% of variance on the *WRMTR*. Note also that the βs for these terms were positive (step 1, Model 2c). The regression coefficient for narratives only becomes negative when decoding is included in the model. Thus collinearity may be one effect at play here.

The unique and shared variance of the language constructs of vocabulary, syntax and narratives can be segmented and represented graphically (see Figure 2), following Cutting and Scarborough (2006) with one major difference. Cutting and Scarborough parcelled out decoding variance prior to examining specific language variance. Due to the observations noted above regarding the apparent collinearity effects in the current dataset, decoding variance was not removed prior to analyzing the language data. It should be understood that different results would be reported if decoding were to be taken into account first. The variance proportions for Model 2 are shown in Figure 2.

Figure 2

Variance proportions for language constructs in Model 2



Model 3: Receptive and Expressive Skills. Model 3 is designed to contrast the contributions of receptive versus expressive language skills. Accordingly, Model 3a enters receptive skills in vocabulary, syntax and narrative skills in the first block. Each of these areas was measured as a single variable: the Synonyms subtest score, Paragraph Comprehension subtest score and Narrative Comprehension composite score. The second block enters expressive skills in the three language domains. Oral Narration was the narrative task. Expressive syntax was entered as a composite of mean length of T-unit in words and proportion of complex sentences. Expressive vocabulary was entered as a composite of the Antonyms subtest score and Number of Different Words. Step 4 entered decoding. Model 3b repeats the analysis with expressive skills entered first. The results of Model 3 are presented in Table 9.

Table 9

Regression results for Model 3

	KTEA-II					*WRMTR*				
Model 3a	R	R^2	R^2_Δ	β	Std β	R	R^2	R^2_Δ	β	Std β
Step1	.700	.489	.489*			.577	.333	.333*		
Receptive skills										
Synonyms				.329	.296*				.230	.293*
Narrative Comprehension				.160	.148				-.083	-.107
Paragraph Comprehension				.154	.126				-.039	-.045
Step 2	.719	.517	.028			.651	.424	.091		
Expressive skills										
Expr. Syntax				.027	.023				.175	.215
Expr. Vocab				.094	.091				.115	.156
Oral Narratives				-.112	-.104				.003	.004
Step 3	.837	.700	.182*			.745	.555	.131*		
Decoding				.628	.545*				.377	.459*
Model 3b	R	R^2	R^2_Δ			R	R^2	R^2_Δ		
Step1	.619	.384	.384*			.563	.317	.317*		
Expressive skills										
Expr. Syntax										
Expr. Vocab										
Oral Narratives										
Step 2	.719	.517	.134*			.651	.424	.107*		
Receptive skills										
Synonyms										
Narrative Comprehension										
Paragraph Comprehension										
Step 3	.837	.700	.182*			.745	.555	.131*		
Decoding										

Notes. For Expressive syntax in Step 2 of Model 3a for the *WRMTR*, $p = 0.55$.

*significant at $p < 0.05$

Significant Results for Model 3. As before, vocabulary and decoding seem to maintain their status as top predictors. Model 3 adds clarity to this picture as the Synonyms measures has the largest β of all the language scores, for both reading comprehension tests. Synonyms is the receptive vocabulary score. A new observation can be made as well: for the *WRMTR*, the expressive syntax composite has the next highest β (.175, $p = .055$), suggesting syntax may be playing a role for reading comprehension on that test. Note that the narrative variables still show negative β values. In this model, the effect was also noted with Paragraph Comprehension for the *WRMTR*.

As with the previous models, we can parse out the variance for each model iteration. For this model, the contrast of interest was receptive vs. expressive language skills. As for Model 2, decoding is entered last in order to preserve language variance for the comparison. The following graph (Figure 3) is based on the results of steps 1 and 2 and therefore do not consider decoding. As before, pulling out decoding variance first would yield different results.

Figure 3

Variance proportions for receptive and expressive language skills in Model 3

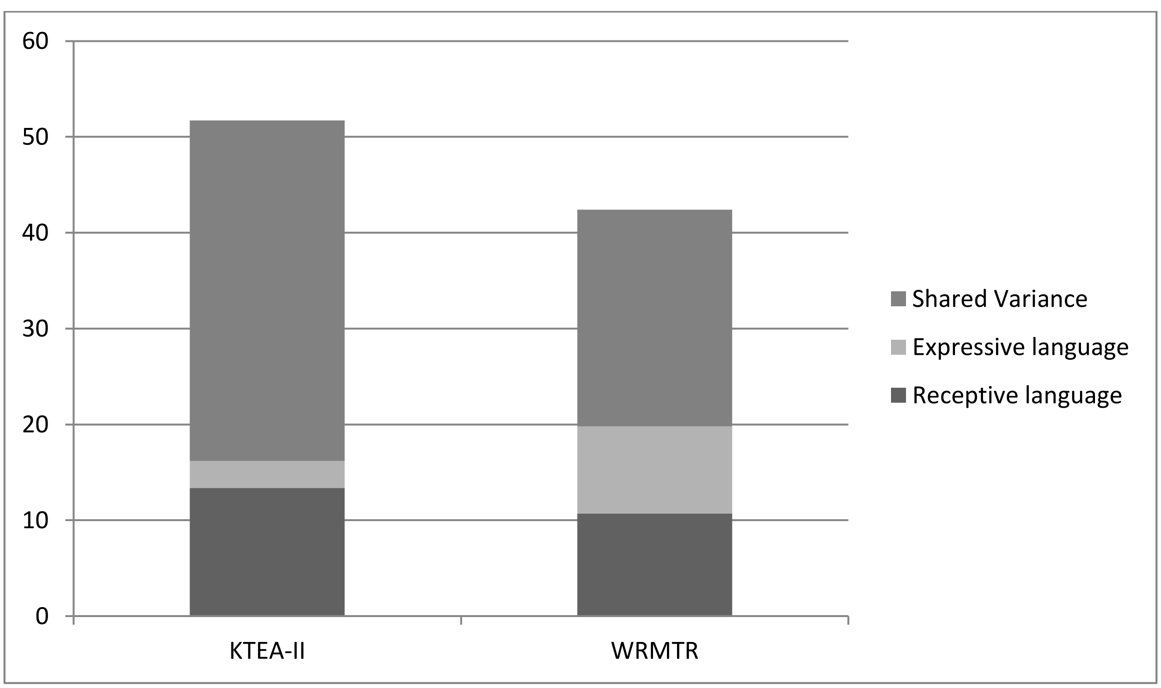


Figure 7 reveals that the *WRMTR* loads more heavily on expressive language skills, as measured here, than does the *KTEA-II*. As with all the models so far, more variance overall is explained for the *KTEA-II* than for the *WRMTR*. The amount of shared variance is considerable for both tests.

Summary of planned regressions.

The short story of models 1, 2 and 3 appears to be "it's all about decoding and vocabulary." The extended version, however, is more complex. Language status turned out to be significant when taken as a single predictor, a key result of the study. This was true for the *KTEA-II* even when decoding was entered first. All three language constructs-- vocabulary, syntax, and narratives-- were shown to hold unique variance in predicting reading comprehension, with differences evident between the two reading comprehension tests. Overall, the models accounted for more variance in *KTEA-II* scores than in WRTMR scores. Language status alone accounted for more variance on the *KTEA-II* (25%) than on the *WRMTR* (14%). Narrative skills accounted for more variance in *KTEA-II* scores whereas syntax accounted for more variance in WRMTR scores. Expressive language skills were shown to hold unique variance in predicting scores for both reading comprehension tests, with this effect being stronger for the *WRMTR* than for the *KTEA-II*.

To fully understand the impact of Language Status on these results, a stratified analysis is required to identify how the predictors for the two participant groups may differ. Because this analysis will drastically reduce sample size, it will be necessary to preserve power by selecting only the most predictive variables to include in these analyses. In the planned regressions, decoding and vocabulary were clearly strong predictors. In the models above, there were indications of specific predictors that should be considered, even though some did not reach statistical significance in these models. Synonyms seemed to be a strong component of the vocabulary measures. Expressive syntax was relevant to the *WRMTR*. Resolving the collinearity effects that seemed to obscure the results of the planned regressions should clarify which predictors to select for the stratified analysis. There appears to be a problem with the vocabulary composite, a problem that was initially identified from analyzing the correlation matrix. Number of different words (NDW) seems to part company from the vocabulary construct, and would rather mingle with the discourse variables such as Paragraph Comprehension and NLAI. As planned (as a result the scrutiny of the correlation matrix during the formation of language composites), the vocabulary composite was recalculated to include only Synonyms and Antonyms and the regression models re-run. If NDW is pulling out variance that reasonably belongs to NLAI, this process should clarify the results and perhaps solve the mystery of the negative β for narratives. Finally the impact of outliers on the models should be examined.

Follow-Up Analyses

This section includes descriptions of three additional data analyses. First, Models 1, 2, and 3 were recalculated with a modified vocabulary composite to attempt to resolve the collinearity problem reported in the previous section. Next, the results of the planned stratified analysis are reported, with results for children with LI and with typically developing language. Finally, the outlier analysis is reported.

Adjusting the Vocabulary Composite.

As discussed above, the vocabulary composite appeared to be causing collinearity problems that could be obscuring the impact of other language variables. From the correlation matrix review, the most likely culprit for this variance-splitting appeared to be Number of

Different Words (NDW). Model 1 was recalculated exactly as before, but with a new vocabulary composite created by averaging the Synonyms and Antonyms subtest scores. The results are displayed in Table 10.

Table 10

Regression results for Model 1 with Vocabulary composite adjusted

	KTEA-II					*WRMTR*				
Model 1a	*R*	R^2	R^2_Δ	β	Std β	*R*	R^2	R^2_Δ	β	Std β
Step1	.501	.251	.251*			.377	.142	.142*		
Language Status				*1.816*	.116 (.501*)				-*.526* *(4.205)*	-.047 (.377*)
Step 2	.739	.545	.295*			.663	.439	.297*		
Syntax				.068	.046				.134	.127
Narrative				.147 (.340)	.139 (.321*)				-.056 (.088)	-.074 (.117)
Vocab				.346	.295*				.335	.404*
Step 3	.843	.716	.165*			.767	.589	.149*		
Decoding				.445	.386*				.262	.318*
Working Memory				.173	.121				.137	.207
Step 4	.846	.716	.006			.767	.589	.000		
NVIQ				-.099	-.102				.015	.022
Model 1b	*R*	R^2	R^2_Δ			*R*	R^2	R^2_Δ		
Step1	.766	.587	.587*			.666	.444	.444*		
Decoding Working Memory										
Step 2	.797	.635	.048*			.674	.454	.010		
Language Status										
Step 3	.843	.711	.076*			.767	.589	.134*		
Syntax Narrative Vocab										
Step 4	.846	.713	.005			.767	.589	.000		
NVIQ										

Model 1c	R	R^2	R^2_Δ	R	R^2	R^2_Δ
Step1	.766	.587	.587*	.666	.444	.444*
Working Memory						
Decoding						
Step 2	.839	.704	.117*	.766	.587	.143*
Syntax						
Vocab						
Narrative						
Step 3	.843	.711	.007	.767	.589	.001
Language Status						
Step 4	.846	.716	.006	.767	.589	.000
NVIQ						

Model 1d	R	R^2	R^2_Δ	R	R^2	R^2_Δ
Step1	.729	.531	.531*	.663	.439	.439*
Syntax						
Narrative						
Vocab						
Step 2	.739	.545	.014	.663	.439	.000
Language Status						
Step 3	.843	.711	.165*	.767	.589	.149*
Decoding						
Working Memory						
Step 4	.846	.716	.006	.767	.589	.000
NVIQ						

Note. Values in parentheses are for the first step the variable was entered. The β value for Language Status is in italics as a reminder that it is on a different metric than the other β values.

*significant at $p > 0.05$

As previously discussed, it was also possible that collinearity between LWR and NLAI was splitting variance. Although the decoding composite was calculated with both LWR and nonword decoding, it was of interest to check what the impact of using just nonword decoding would be. Table 11 presents Model 1a exactly as in the previous table, except that nonword decoding is used instead of the decoding composite.

Table 11

Regression results for Model 1a with nonword decoding

	KTEA-II					*WRMTR*				
Model 1a	R	R^2	R^2_Δ	β	Std β	R	R^2	R^2_Δ	β	Std β
Step1	.501	.251	.251*			.377	.142	.142*		
Language Status				*1.656*	.106				*-.468*	-.042
Step 2	.739	.545	.295*			.663	.439	.297*		
Syntax				.062	.042				.154	.147
Narrative				.211	.199				-.057	-.076
Vocab				.406	.346*				.366	.440*
Step 3	.814	.663	.118*			.765	.585	.146*		
NwD				.165	.156				.203	.271[a]
Working Memory				.308	.332*				.153	.231
Step 4	.819	.671	.008			.765	.585	.000		
NVIQ				-.115	-.118				.015	.022

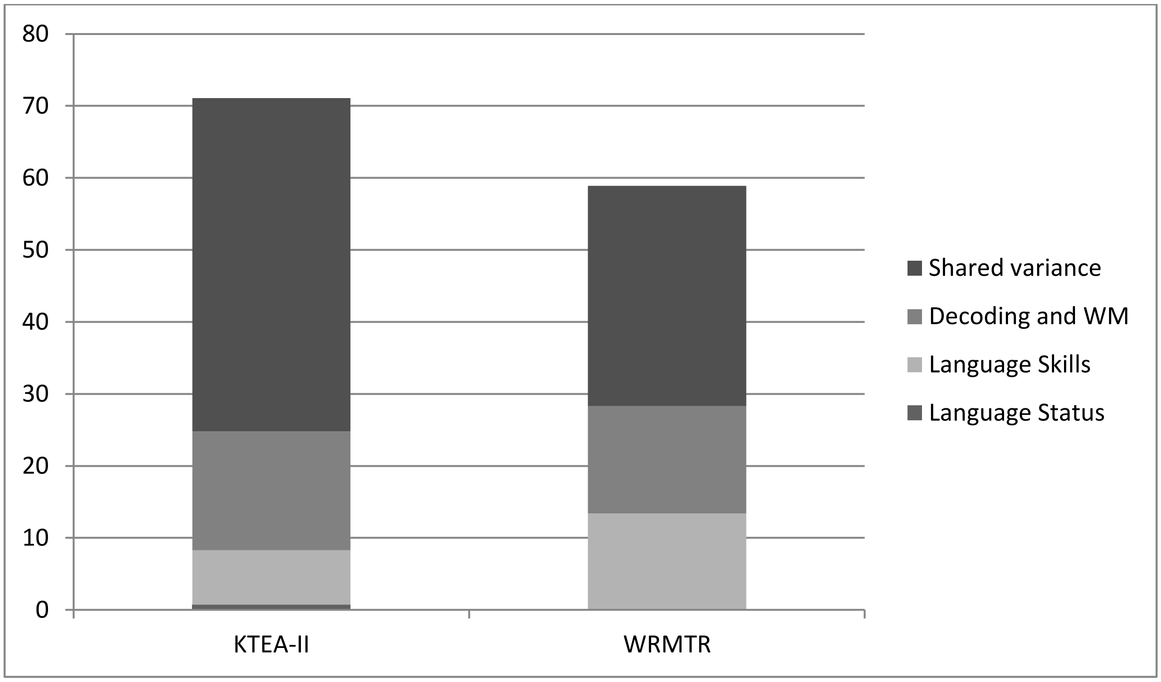
Notes. Non-standardized coefficients for Language status are italicized as a reminder that they cannot be directly compared with coefficients of the other variables. Std. β = Standardized β; NwD = Nonword decoding; Vocab = Vocabulary composites; NVIQ = Nonverbal intelligence.
[a] $p = .053$.
*$p < .05$

Significant results for Model 1 with Vocabulary adjusted. The first result of interest is that removing NDW from the vocabulary composite appears to have resolved the problem of narratives having a negative regression coefficient in the *KTEA-II* model. This suggests that NDW was pulling variance associated with discourse-level skills to the vocabulary composite. It is also quite interesting to note that the narrative variable has gained influence in step 2 of Model 1a. That is, before decoding is entered, narratives are a significant predictor of reading comprehension. Note for the *WRMTR*, the narrative coefficient is positive (although not statistically significant) until decoding variance is removed. At that point the coefficient is negative, but also close to 0, suggesting that is it a very poor predictor of reading comprehension on the *WRMTR*; the confidence interval for the narrative coefficient includes 0. The picture presented in the adjusted model is more consistent with the observations of the correlation matrix. In addition, the adjusted R^2 values are closer to the R^2 values in the model with vocabulary adjusted, which suggests that this model provides a closer estimate of population values. Overall, this model appears to provide a better fit to the observed data. Using the nonword decoding score had minimal impact on the model overall. As expected from the correlation matrix, the models with nonword decoding are slightly less predictive, the regression coefficients of the language terms increase and the impact of the change is less on the *WRMTR* model. Note that the use of nonword decoding does change the apparent importance of decoding, but the change is absorbed largely by changes in the β value for WM. This supports the observation that the decoding and working memory composites were sharing variance.

The graph of variance proportions was corrected using the model in Table 11 which includes the decoding composite. The resulting graph is presented in Figure 4.

Figure 4

Variance proportions for Model 1 with adjusted vocabulary composite



Based on the results of Model 1, it appears that the adjusted vocabulary composite improves the model. Accordingly, Models 2 and 3 were also recalculated with the adjusted vocabulary composite. The results for Model 2 adjusted are presented in Table 12.

Table 12

Regression results for Model 2 with vocabulary adjusted

	KTEA-II					*WRMTR*				
Model 2a	*R*	R^2	R^2_Δ	β	Std β	*R*	R^2	R^2_Δ	β	Std β
Step1	.646	.418	.418*			.636	.404	.404*		
Vocab				.413	.352*				.337	.405*
Step 2	.660	.435	.017			.654	.427	.024		
Syntax				.081	.055				.158	.150
Step 3	.729	.531	.096*			.663	.439	.011		
Narrative				.144	.136				-.054	-.072
Step 4	.832	.692	.160*			.755	.569	.131*		
Decoding				.574	.498*				.370	.449*
Model 2b	*R*	R^2	R^2_Δ			*R*	R^2	R^2_Δ		
Step1	.363	.132	.132*			.379	.144	.144*		
Syntax				(.538)	(.363*)				(.397)	(.379*)
Step 2 Narrative	.572	.327	.196*			.449	.201	.170		
Step 3 Vocab	.729	.531	.204*			.663	.439	.237*		
Step 4 Decoding	.832	.692	.160*			.755	.569	.131*		
Model 2c	*R*	R^2	R^2_Δ			*R*	R^2	R^2_Δ		
Step1	.555	.308	.308*			.380	.144	.144*		
Narrative				(.588)	(.555*)				(.285)	(.380*)
Step 2 Vocab	.729	.531	.223*			.654	.427	.283*		
Step 3 Syntax	.729	.531	.000			.663	.439	.012		
Step 4 Decoding	.832	.692	.160*			.755	.569	.131*		

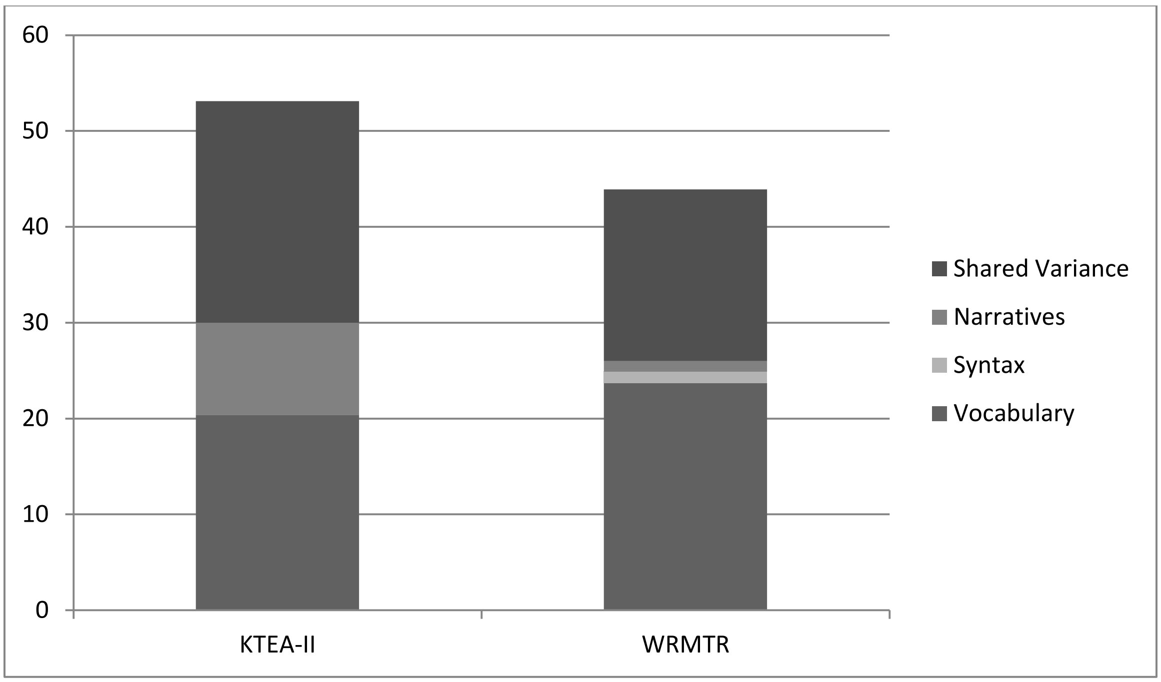
Notes. β in parentheses are for step at which they appear; all others are for step 4 of the model
*significant at $p < .05$

As with Model 1a, the effect of using nonword decoding rather than the decoding composite was checked. Results were only moderately affected. The regression coefficients for language became slightly larger, as expected. As previously noted, the impact was more significant in the *KTEA-II* model than in the *WRMTR* model. More importantly, the relative importance of the language terms remained unchanged. Decoding and vocabulary were still significant, with NLAI next for the *KTEA-II* and syntax for the *WRMTR*.

The variance proportion graph for Model 2 resulting from the adjusted vocabulary composite is presented below in Figure 5.

Figure 5

Variance proportions for Model 2 with adjusted vocabulary composite



As before, *KTEA-II* scores are more predictable with the measures used in this study. From this graph it can be seen that *WRMTR* scores appear to load more heavily on vocabulary and syntax, whereas as *KTEA-II* scores are more related to vocabulary and narrative skills. As before, a large proportion of the variance is shared.

The results for Model 3 with the adjusted vocabulary score are presented in Table 13. Note that the regression diagnostics for Model 3 suggested some collinearity between Synonyms and Antonyms (tolerance is still above 0.2), and somewhat non-normally distributed residuals for the *KTEA-II*. The variance proportions for the adjusted Model 3 are presented in Figure 6.

Table 13

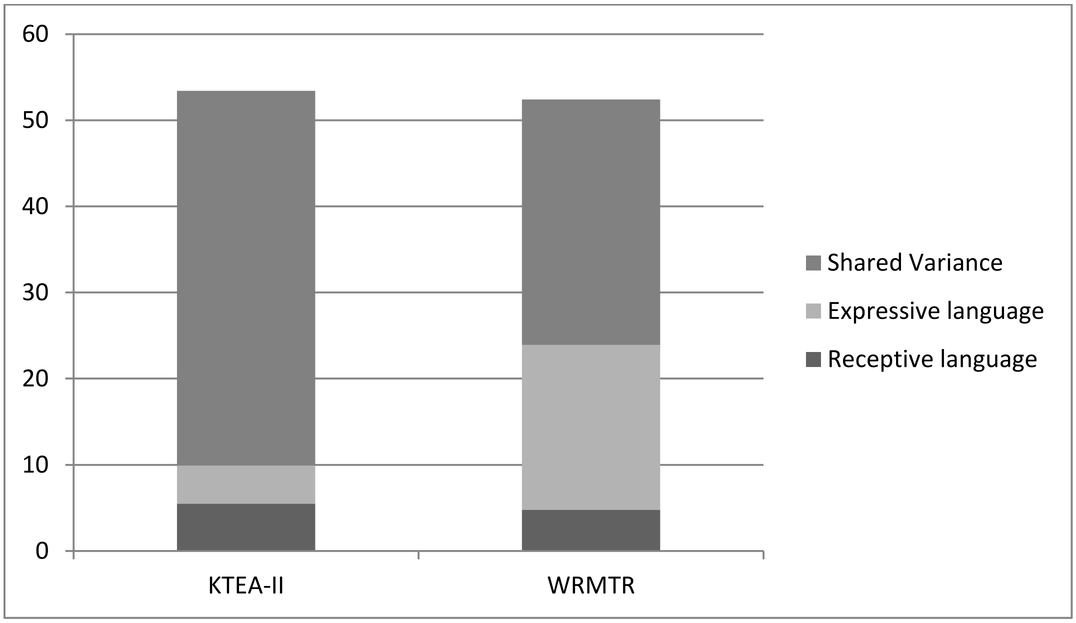
Regression results for Model 3 with vocabulary adjusted

	KTEA-II					*WRMTR*				
Model 3a adjusted	R	R^2	R^2_Δ	β	Std β	R	R^2	R^2_Δ	β	Std β
Step1	.700	.489	.489*			.577	.333	.333*		
Receptive skills										
Synonyms				.304	.274*				.047	.060
Narrative Comprehension				.160	.147				-.121	-.156
Paragraph Comprehension				.149	.121				-.129	-.149
Step 2	.730	.534	.044			.724	.524	.191*		
Expressive skills										
Expr. Syntax				.023	.020				.161	.197
Antonyms				.078	.071				.345	.445*
Oral Narratives				-.053	-.049				.124	.163
Step 3	.836	.699	.165*			.777	.604	.079*		
Decoding				.623	.541*				.307	.373*
Model 3b	R	R^2	R^2_Δ			R	R^2	R^2_Δ		
Step1	.692	.478	.478*			.690	.476	.46*		
Expressive skills										
Step 2	.730	.534	.055			.724	.524	.048		
Receptive skills										
Step 3	.836	.699	.165*			.777	.604	.079*		
Decoding										

* $p < .05$

Figure 6

Variance proportions for Model 3 with adjusted vocabulary composite



This model demonstrates that both receptive and expressive language skills account for unique variance in predicting reading comprehension scores on both tests. *WRMTR* scores in particular appear to load heavily on expressive language as measured here. The amount of shared variance is considerable. For the *KTEA-II*, about 81% of variance is shared and for the *WRMTR*, about 54% of variance is shared.

Stratified Analysis: Looking for Effects of Language Impairment.

The next follow-up analysis is to split the sample into two groups, children with LI and children with typically developing language (TDL), to address Research Question 2. This necessarily reduced the sample size, so only the most predictive measures for each test were included. Power analyses were completed for these regressions, revealing that even with the reduced sample size (N = 24 for LI), there was over 80% power to detect effects of at least R^2 =

0.4 at α = 0.05 with three independent variables tested. With N = 30 (the TDL group), power is increased to detect effects of $R^2 = 0.3$ (that is there is sufficient power to detect smaller effects). From the planned analyses and informed by the power analysis, a set of three predictors was selected for each reading comprehension test. This set consisted of two language variables and the decoding composite.

For both the *KTEA-II* and the *WRMTR*, it was clear that the adjusted vocabulary composite (average of Synonyms and Antonyms scores) should be included. For the *KTEA-II*, the other variable of interest was narrative skill (measured as the NLAI composite). This was based on the variance proportion analyses, which suggested that narrative skill was more relevant to this test than was syntax. This is not to say that syntax is not involved, but rather it was better to focus the power of the analysis on the most important predictors. For the *WRMTR*, the variance proportions suggested that syntax was important. From the syntax measures, Mean Length of Utterance in Words (MLTU-w) appeared to be a valuable predictor, based on its regression coefficient in the first regression analysis including all the language variables. Finally, the decoding composite was included for both tests. The model was calculated separately for the two participant groups, children with LI and children with TDL. The results are presented in Table 14.

Table 14

Regression analysis of KTEA-II and WRMTR scores for LI and TDL groups

	KTEA-II					*WRMTR*					
TDL n=30	R	R^2	R^2_Δ	β	Std β	TDL n=29	R	R^2	R^2_Δ	β	Std β
Step 1	.763	.583	.583*			Step 1	.690	.476	.476*		
Vocab				.484	.404*	Vocab				.345	.374*
NLAI				.292	.241	MLTU-w				.317	.305*
				.526	.435*						
Step 2	.808	.653	.070*			Step 2	.761	.579	.103*		
Decoding				.397	.367*	Decoding				.306	.360*
LI n=24	R	R^2	R^2_Δ	β	Std β	LI n=24	R	R^2	R^2_Δ	β	Std β
Step 1	.300	.090	.090			Step 1	.441	.195	.195		
Vocab				.221	.205	Vocab				.290	.367*
NLAI				.045	.063	MLTU-w				.076	.185
Step 2	.732	.536	.446*			Step 2	.692	.479	.284*		
Decoding				.679	.682*	Decoding				.388	.534*

Notes. β values in faded print show NLAI coefficients before Decoding is entered into the model. *KTEA-II= Kaufman Test of Educational Achievement, 2nd Edition; WRMTR = Woodcock Reading Mastery Test-Revised/Norms Updated;* TDL = Typically-developing language skills; LI = Language impairment; Vocab = Vocabulary composite; NLAI = Narrative Language Ability Index from the *Test of Narrative Language;* MLTU-w = Mean language of T-unit in words.
* $p < 0.05$

The results of these models demonstrate that there were differences between the two reading comprehension tests and between the participant groups. Both tests included vocabulary and decoding as strong predictors, consistent with the initial regression results.

The results for children with TDL can be compared between the two reading comprehension tests. It was interesting to see that the β values of the vocabulary terms were higher than those for decoding on both tests, suggesting that vocabulary was a stronger predictor of reading comprehension than was decoding in this group. One could argue that the TDL models indicate that narrative language skills were somewhat predictive of reading comprehension scores on the *KTEA-II*. Before decoding is entered, the β value for narratives in the *KTEA-II* model is actually the largest β obtained for that test. Note that there continues to be evidence of collinearity between narratives and decoding, as indicated by the large reduction in the β value for narratives from step 1 to step 2 of this model. This could be obscuring the impact of narrative skills. For the *WRMTR*, the model suggests that syntax as measured by MLTU-w was an important predictor for this reading comprehension scores in this participant group. In fact, the β values for vocabulary, MLTU-w and decoding were quite similar in magnitude, suggesting that they all contributed similarly to the model. It was also intriguing to note that the models accounted for 65% of the variance on the *KTEA-II* and 58% of the variance on the *WRMTR*, with just three predictors.

The LI models returned interesting results, but must be interpreted with caution due to the limited sample size. The residuals for this model were not normally distributed, so the significance tests for the model may be questionable. The estimation of model parameters is not affected by this observation. For both tests, decoding was the strongest predictor, with this effect being stronger on the *KTEA-II* than on the *WRMTR*. Vocabulary remained the second strongest predictor in terms of β values. Vocabulary was more important (and reached statistical significance) in the *WRMTR* model. For both tests narratives and syntax proved to be poor predictors in this group. In short, the results suggest that decoding was the primary determinant of reading comprehension in this group and vocabulary was the next most predictive, consistent with the findings reported by Palikara, Dockrell, and Lindsay (2011).

Checking for impact of decoding skills

Given the results presented here, decoding skill appeared to merit further investigation. It was possible that decoding skill was actually the factor behind the differences between the results for LI and TDL, so an additional stratification was performed. The sample was split on decoding ability at -1 SD based on the expected population distribution (M = 100, SD = 15), rather than on language status as in previous models. This resulted in two groups, one with decoding ability within or above the expected range by age (N = 38) and one group with decoding ability below expectations (N = 16). As with the previous stratification, only the most promising predictors were included to preserve power. In this case, it was of interest to compare the tests on the same variable set and there was adequate power to do so. Since the low decoding group was too small to have adequate power for a regression analysis, only the results for the stronger decoders are reported in Table 15. Trends for the lower decoders are reported below.

Table 15

Regression results for higher decoders (n = 38)

	KTEA-II						*WRMTR*				
	R	R^2	R^2_Δ	β	Std β		R	R^2	R^2_Δ	β	Std β
	.737	.542	.542*				.674	.454	.454*		
Vocab				.516	.567*	Vocab				.414	.573*
MLTU-w				-.001	-.001	MLTU-w				.186	.268*
NLAI				.270	.322*	NLAI				.002	.004

Notes. *KTEA-II= Kaufman Test of Educational Achievement, 2nd Edition; WRMTR = Woodcock Reading Mastery Test-Revised/Norms Updated;* Vocab = Vocabulary composite; MLTU-w = Mean language of T-unit in words; NLAI = Narrative Language Ability Index from the *Test of Narrative Language*.
* $p < 0.05$

The results of this analysis are quite similar to the results of the group of children with TDL, with the relative magnitudes of the regression coefficients following the same pattern as for the TDL group. Vocabulary is the strongest predictor for both groups, followed by NLAI for the *KTEA-II* and by MLTU-w for the *WRMTR*. For the low decoders, only NLAI (for the *KTEA-II*), syntax as MLTU-w (for the *WRMTR*) and vocabulary were entered. Only vocabulary was significant for both reading comprehension tests. In terms of effect size, however, the β for NLAI on the *KTEA-II* was close to the magnitude of the β for vocabulary (β = .409 and .562 respectively). Since it is likely that the groups of children with TDL and the high decoders likely include many of the same participants, these results seem sensible. The relationship is not perfect, as the high decoder group had 38 participants, while the group of children with TDL included 30 children. The overall mean of the sample for decoding was 94 (SD = 13.6), lower than the expected mean of 100. A one way ANOVA confirmed that the difference between the means for children with and without LI was significant (F = 6.74, p = .012). The distribution of decoding skill between the language skill groups is shown in Table 16.

Table 16

Decoding skill by language status for participant groups

Decoding Skill	Language Status		Decoding Score
	LI	TDL	
Low	9	7	77. 7 (4.3)
High	15	23	101.1 (9.6)
Decoding Score	89.0 (10.1)	98.3 (14.9)	

Notes. Decoding scores are recorded as group means, with standard deviations in parentheses. LI = language impairment; TDL = typically developing language.

Note that there were more high decoders in the group with TDL than in the LI group, but that there were still more high decoders than low decoders in the LI group. Also there were low decoders in the TDL group. Therefore it is not the case that decoding status alone is responsible for the differences between children with LI and children with TDL in this study. It is important to keep in mind that this simple approach collapses interval data into categorical data. The true picture includes distributions of both language and decoding skill. To more fully understand how decoding and language skill are interacting to determine reading comprehension, it will be helpful to look in more detail at outliers—those cases whose results are poorly predicted by the models reviewed here.

Outlier Analysis.

Outliers are cases that may affect the estimation of model parameters because they fall outside the expected boundaries of the data (Field, 2013). As such, it is important to consider their effects upon the regression models and their interpretation. Outlier analysis can proceed in more than one way. If the Cook's distance (see Appendix D for explanation of this term) for a case is greater than one, you can remove the case from the data set and refit the model. The other method is to leave the case in, but to evaluate where the goodness-of-fit falls apart (Field, 2013). In the current dataset, the Cook's distances were within criteria, so the cases were examined individually. An initial step was to create boxplots for the reading comprehension variables as the original standard scores. When this was done, no outliers were obvious among the *KTEA-II* scores. On the other hand, there were 4 scores outside the expected range on the *WRMTR* scores. However, when the boxplots were generated for the participant groups separately, no outliers were observed. Thus scores that were considered extreme in the overall distribution were not extreme considering groups by LI and TDL. The low score was within the expected range for children with LI and the three high scores were within the expected range for the children with TDL. At this point there was no reason to consider deleting cases.

Following this initial step, outliers identified during the regression diagnostics for each model were reviewed. The default criterion for identifying outliers in SPSS is a score more than 3 SD from predicted values. Because cases that were poorly predicted by the models were of interest, a criterion of 2 SD was used (Field, 2013). For each case, the dataset was examined to look for patterns that could be related to the poor prediction of reading comprehension for that case. A total of seven unusual cases were identified. For six of those seven cases, the models had overestimated the reading comprehension scores; the predicted score was more than 2 SD higher than the observed value. These cases were reviewed to look for patterns. The results of this case-by-case review are summarized in Table 17.

Table 17

Summary of unusual cases (predicted scores were > 2SD above/below observed)

Regression Model	Observed Score	Expected Score	Explanation
LI *KTEA-II*	77	93	Poor language skills, high decoding
LI *WRMTR*	71	86	Good language, lower decoding and WM
TDL *KTEA-II*	84 (same case as one from model 1a)	105	Language scattered, decoding and vocabulary high
TDL *WRMTR*	115 (same case as one from model 1a)	96	Language and WM average; only WRMT high
Model 1a / no NDW	Observed Score	Expected Score	Explanation
KTEA-II	84	105	Somewhat scattered language skills, high vocabulary and decoding
	88	107	High skills overall but lower comprehension
WRMTR	115	96	All close to average except WRMT; possible artifact
	75	94	WM low, some scattering in language skills
	77	93	Scattered language; higher vocabulary; lower WM [a]
Model 1b/no NDW	No new cases		
Model 1c/ no NDW	No new cases		
Model 1d/ no NDW	No new cases		
Model 2 and 3	No new cases		

Notes.LI = Language impairment; TDL = Typically developing language; *KTEA-II = Kaufman Test of Educational Achievement, 2nd Edition; WRMTR = Woodcock Reading Mastery Test—Revised/Norms Updated;* NDW = Number of different words; WM = Working memory.
[a] NDW was lower so fit was worse than initial model

The results of this analysis indicate that scattered skills were poorly taken into account by the models. This makes sense; the results showed clearly that vocabulary and decoding were the strongest predictors. It is important to note, however, that other skills are significantly impacting reading comprehension for at least some children.

Cases with high Mahalanobis distances in Model 1 were also reviewed (this metric reflects how far the predicted value is from the mean of the dependent variable (Field, 2013, p. 307). These were all valid cases; they tended to be children who scored quite high overall. There was no strong evidence in this analysis that any of these cases should be removed from the dataset. It is possible that doing so might have improved the model statistics, but it would make the models less relevant to the actual sample, thereby reducing external validity.

Summary

The results of this study were quite consistent. Vocabulary and decoding were robust predictors of reading comprehension for both reading comprehension tests and for both participant groups. The follow-up analyses revealed that results for children with LI contrasted with results for children with TDL. Scores for children with LI appeared to be mostly related to decoding and to vocabulary on both reading comprehension tests. For children with TDL, scores were also predicted by vocabulary and decoding on both tests. However, for children with TDL, scores were also predicted by narrative skill on the *KTEA-II*, and by MLTU-w on the *WRMTR*. Results obtained for children with higher decoding skills were similar to results for children with TDL. The outlier analysis added another useful piece of information. When cases were poorly predicted by the models, the models tended to overestimate how high the children would score on reading comprehension tests. In general, these cases occurred among children with more scattered language profiles. Even though vocabulary and decoding tended to be high in this group, other language skills were lower.

Overall, the analysis showed that the predictor variables selected for this study have complex relationships among them. It was necessary to adjust the vocabulary composite to deal with collinearity issues and improve model fit. Among the language variables, it was clear that much variance was shared rather than being unique to one variable. This was also observed when receptive and expressive measures were compared—a large proportion of the variance was shared.

In Chapter 5, these results will be discussed relative to the research questions and related to the existing research base. From these discussions, ideas for future research will be presented.

Chapter 5 Discussion

The initial results of the study were unequivocal and almost startling in their apparent clarity. For both reading comprehension tests, decoding and vocabulary were key predictors. Both held unique variance, regardless of how the predictors were entered. The message appeared to be *it's all about decoding and vocabulary.* The story became more complex as the planned regressions unfolded. Then the interplay among the predictors moved to the forefront. It slowly became clear that although decoding and vocabulary were key, other predictors were involved, and it was in fact the interrelationships among these variables that had to be understood. It will be helpful to begin the discussion by focusing on the research questions one at a time.

Research Question 1

Given receptive and expressive measures of vocabulary, syntax and narrative discourse for children in grades 4 to 6, which of these oral language scores will predict scores on two concurrently measured standardized reading comprehension tests?

The two reading comprehension tests selected for this study had quite different tasks. The *Kaufman Test of Educational Achievement, 2nd Edition (KTEA-II)* involves reading passages of increasing difficulty, followed by answering written questions. The *Woodcock Reading Mastery Test—Revised/Norms Updated (WRMTR)* is a cloze task, wherein students read passages of increasing length, each with a missing word; the response is to provide the missing word. My working hypothesis was that these tasks should tap different skills, consistent with previous research on similar tests. Models 1, 2 and 3 were designed to contrast different constructs of interest in this study. Early in the analysis, the vocabulary composite, particularly the Number of

Different Words (NDW) variable, was identified as a potential source of problems in the regression analysis. Problems were in fact observed when the models were calculated so the models were recalculated with an adjusted vocabulary composite. The results of the follow-up regression models with the vocabulary composite adjusted to remove NDW will be the focus of the discussion, as these results appeared to provide a better fit to the data.

In Model 1 the focus was on whether language status (impaired/not impaired), language skills, and decoding plus working memory were predictive. Nonverbal intelligence was also included, but carried essentially no unique variance consistent with results reported by Cutting and Scarborough (2006). All three major constructs were significant, for both tests.

Language status, as impaired/not impaired was a significant predictor for both tests when considered alone, but held unique variance only in predicting *KTEA-II* results. This appeared to be consistent with the hypothesis that the *KTEA-II* would be more likely to reveal differences between children with and without language impairments (these ideas are further explored below in the discussion of Research Question 2). However as soon as more detailed language scores were included, this predictive value disappeared. Such a result was quite reasonable in the sense that the predictor variables were all selected to differentiate between the two participant groups. Thus, they were at least as adequate in predicting language-related variance as the dichotomous language-status variable. Language status is therefore relevant to reading comprehension, but a categorical approach is less fruitful than a detailed set of predictors. Partly, this is a result of the analysis method itself. Regression depends upon variance patterns to yield results. When interval-level data is reduced to categorical, variance is lost. For this reason, language status was omitted as a predictor in follow-up regressions. However, the fact that language status was significant as a single predictor led to a stratified analysis of the data, viewing results for children with and without LI separately. These analyses represent a key contribution of this study and are discussed below in the section regarding Research Question 2.

The observation that language status predicted reading comprehension scores was consistent with expectations; however, two observations were inconsistent with hypothesized results. First, once decoding and working memory were taken into account, language skills accounted for more variance on the *WRMTR* than on the *KTEA-II*. I will consider a possible explanation for this unexpected result in the discussion of Model 2 below. Second, essentially the same amount of variance was accounted for by decoding and working memory on both the tests (about 20%). Based on existing research on the *WRMTR* and research on question-response tests similar to the *KTEA-II*, it was expected that the cloze task of the *WRMTR* would depend more upon both working memory and decoding than would the task on the *KTEA-II*. A possible explanation for this result will arise from considering the relative contributions of language and decoding variance in the existing research on reading comprehension tests.

The relative contributions of language and decoding to reading comprehension scores have been of interest in previous research of this kind, largely as a result of the application of the Simple View of Reading. In the current study, the *KTEA-II* scores were more predictable overall, with up to 72% of the variance in scores predicted by the regression models, as opposed to 59% of variance on the *WRMTR*. Although the *KTEA-II* has not been previously studied, the format of the test is similar to that of the Gates-McGinitie (G-M) as reported by Cutting and Scarborough (2006). The models reported by these authors, using oral language, decoding and memory, predicted 68% of variance on the G-M, quite a comparable result to that for the *KTEA-II* obtained in the current study. Keenan, Betjemann and Olson (2008) reported regression models using listening comprehension and decoding, which predicted 61% of variance on the

Woodcock–Johnson Passage Comprehension subtest of the *Woodcock–Johnson Tests of Achievement–III*, a task very similar to that on the *WRMTR* used in the current study. Again, the results are quite comparable between the two studies. Francis, Fletcher, Catts and Tomblin (2005) reported correlations for the *WRMTR* to language (r = .65, p < .001) and decoding (as a composite of nonword decoding and word recognition) (r = .84, p < .001), measured in Grade 4. The language measures used in that study were receptive measures of "listening comprehension" and vocabulary. In the current study, this may correspond best to Paragraph Comprehension (r = .24) and Synonyms/Antonyms (r = .64, p < .001) measures. Thus the two studies are in accord, at least relative to vocabulary. The correlation obtained by Francis et al. for decoding appears to be much higher than those obtained in the current study (for Letter and Word Recognition, r = .66, p < .01; for Nonword Decoding, r = .53, p < .01). This is a cautious observation given that different word reading measures were used. It is unclear why this marked difference should be observed. It may simply be the effect of using different measures across the studies, but one possible explanation may relate to norming. It was previously noted that the correlation between decoding and the *KTEA-II* scores was higher than expected. In the case of this study, the decoding measures were co-normed with the *KTEA-II* reading comprehension task as they came from the same instrument. Similarly, the decoding measures used by Francis and colleagues were from the *WRMTR*, and so were co-normed with the *WRMTR* reading comprehension task. Thus the higher correlations may possibly be due to the fact that decoding and reading comprehension skills should be more highly correlated when compared relative to the same group of children for each task than when compared to different groups of children for each task. This in turn suggests a possible explanation for the relatively higher contribution of decoding in predicting scores on the *KTEA-II* given that the word reading task was taken from this test. The higher correlation between word reading and reading comprehension on the *KTEA-II* results in word reading being more predictive for the *KTEA-II* than for the *WRMTR*. It may therefore be partly an artifact of task selection. Nevertheless, the results of this study are in general consistent with other research of this kind in that both language and decoding are important predictors of reading comprehension scores.

In the preceding discussion of Model 1, “language” was considered primarily as a single construct. To expand on those results, Model 2 focused on the component language skills of vocabulary, syntax and narratives and their contributions to predicting reading comprehension scores. Both reading comprehension tests depended heavily on vocabulary, with *WRMTR* having relatively more weight on this construct. The value of vocabulary in predicting reading comprehension is consistent with the Construction Integration Model with the proposed extensions, which predict that semantic knowledge will affect all levels of reading from decoding to text base to situation model. In the current study, measures of depth of vocabulary knowledge proved useful, while the measure of vocabulary breadth, Number of Different Words, caused difficulties in the regression models. With the vocabulary composite adjusted to reduce variance-splitting, it became clear that narrative skills were predictive of *KTEA-II* scores, while scores on the *WRMTR* depended more on syntax skills. These results are generally consistent with hypothesized results. Since the *KTEA-II* has longer passages, it was expected that discourse-level language skills as sampled in the narrative tasks would be more relevant to that test. I did not find evidence to support the hypothesis that skill with complex sentences would predict scores on the *KTEA-II*. This observation will be further considered in the discussion of Research Question 2, when results for children with and without LI are interpreted. As hypothesized, the cloze task of the *WRMTR* is supported by knowledge of syntax (i.e., what word *could* go in this sentence?)

and by depth of vocabulary (i.e., what contextual cues give me hints about words to fill in?). Note that this role of vocabulary may partly explain why *WRMTR* scores appeared to load more heavily on language skills per se. If vocabulary is by far and away the largest language predictor, and the *WRMTR* loads heavily on vocabulary, then the *WRMTR* will appear to load more heavily on language overall. The measures to sample vocabulary, syntax and narratives in Model 2 included both receptive and expressive tasks. The importance of the receptive/expressive dichotomy was explored directly in Model 3, which will be discussed later in this chapter, under the Methodology Contributions heading.

It is important to acknowledge that working memory was intentionally omitted in both Model 2 and Model 3. This decision is based upon the premise that working memory forms part of a cognitive architecture within which the component skills of interest function. Therefore removing working memory variance prior to examining language skills will necessarily reduce variance that should be relevant to these skills. Working memory and its role in these models is of such interest that it will be discussed later in this chapter as part of the consideration of shared variance in regression modelling.

Research Question 2

Will there be an effect of group membership such that significant predictors for children with developmental language impairment will differ from significant predictors for children with typically developing language?

The second research question in this study asks whether the predictors of reading comprehension differ for children with and without LI. The answer provided by these results appears to be “yes”. Language status as impaired/not impaired had predictive value in the regression analyses for both reading comprehension tests, although it held unique variance only for the *KTEA-II*. A stratified analysis of the data set demonstrated that decoding was the most predictive variable for children with LI followed by vocabulary for both reading comprehension tests (recall that vocabulary only reached statistical significance in the *WRMTR* model, but that the significance tests of these models were suspect due to possible violation of regression assumptions). This result was consistent with previous work that showed decoding and vocabulary to remain predictive for children with LI, after the time at which these abilities are no longer predictive for children with TDL (Palikara, Dockrell, & Lindsay, 2011). For children with stronger language skills, narrative skills and syntax skills came into play in predicting *KTEA-II* and *WRMTR* scores respectively. Note that among children with TDL, vocabulary was more predictive than decoding for the *KTEA-II* and both vocabulary and syntax were more predictive than decoding for the *WRMTR*. It was somewhat unexpected that both vocabulary and decoding remained among the strongest predictors in the group of children with TDL.

These results lead to an interesting question: are children with LI employing a different set of skills during reading comprehension tests? I will suggest that no, they are no not. Rather the results of this study could be interpreted to suggest that reading comprehension tests are less revealing for children with LI partly due to their (typically) reduced decoding abilities. This reduced sensitivity to language skills has important implications about standard practices for reading comprehension assessment.

The impact of decoding ability on reading comprehension assessment.

As alluded to above, there is an interaction between ability and test performance that should be considered in interpreting the results of the current study. One factor, demonstrated by numerous studies, is that different tests can be expected to tap different skills, for readers with sufficient decoding ability. In this study for example, vocabulary, morpho-syntax and narrative skills have been considered. The other factor is that some readers do not have sufficient decoding ability. It stands to reason that if a student is not very good at decoding, reading comprehension tests primarily test decoding—because poor decoding limits comprehension (see for example Keenan et al., 2008). So if one is evaluating a reading comprehension test that loads heavily on decoding, one is not so likely to find differences related to other language abilities—it tests primarily decoding for all. But if a test loads heavily on comprehension, one could expect to see more language-based differences. That is, a test can tap comprehension for children with good enough decoding for those skills to come into play. In this study, the *KTEA-II* was hypothesized to tap higher-level skills, and if so should be more likely to show differences for students with high vs. low language skills. The results of this study support that hypothesis. Recall that for children with TDL, language skills alone predicted 58% of the variance in reading comprehension scores on the *KTEA-II*. For children with LI, language skills alone predicted only 9% of the variance on that test. On the *WRMTR*, language skills alone predicted 48% of the variance for children with TDL. For children with LI, language skills predicted 19.5% of the variance (note that three times as much of this variance was accounted for by vocabulary than by syntax). On both tests, vocabulary was the most significant predictor for children with TDL; however for children with LI, decoding was the strongest predictor. Since children with LI as a group are expected to have lower decoding skills (Palikara, Dockrell, & Lindsay, 2011), decoding skills themselves could contribute to group differences between the two tests in this study. This possible effect was examined and was shown to be inadequate to completely explain the findings. That being said, a significant difference in decoding abilities between the groups could limit the ability to detect real differences in comprehension based on oral language skills;

lower decoders do not have the opportunity to engage language skills if they cannot adequately access the text. Consider that in the current sample about half the children with LI showed low decoding skills as well.

Although there were more low decoders than high decoders in the LI group, there were still children who had LI and decoding skills within the normal range. The outlier analysis showed that most of the children whose results were poorly predicted by the models had uneven profiles of language skills. Given these results then, the implication for practice is that language profiles should be of interest in assessing and remediating reading comprehension difficulties.

Implications for practice: Assessment of reading comprehension in children with LI.

The results of this study strongly suggest that a profile of oral language skills is an important component of a reading assessment. One can score poorly on a reading comprehension test, yet this result will not illuminate what caused the poor score. Was it primarily decoding? Even if this is so language skills may still be heavily impaired but without more detailed information it will be difficult to select goals to clearly target areas of weakness. For example, a child may have strong vocabulary and yet score low on measures of syntax or narrative ability. A reading comprehension test alone will not reveal these characteristics, although they are accepted precursors of effective reading comprehension. Since evidence strongly favours targeting comprehension in concert with decoding (see for example Cain, 2010; Rapp et al., 2007), additional assessment is necessary to choose effective targets.

A group of interest that was discussed in the literature review is identified as “poor comprehenders”. These were students with word recognition within the expected range, who had difficulties with reading comprehension. With the criteria of decoding above -1 SD and *KTEA-II* RC score less than 86, five children were identified who fit this profile. These children had varied language profiles, but the mean Core Language Score for this group was 76, and only two of the children scored above the cut-off for this measure. The same criteria were applied again, using *WRMTR* PC scores. Only two children were identified as poor comprehenders. Among this smaller group, Core Language Scores were 87 for both children. Both of these children had rather scattered language profiles. These results are reasonably consistent with previous reports that the poor comprehender group overlaps with the LI group. A more surprising observation, however, was that no children were identified as poor comprehenders by both reading comprehension tests.

On some academic assessments such as the *KTEA II*, the need for additional information about language skills is somewhat addressed by tests of listening comprehension. However, these measures may not provide detailed results in terms of the major domains of language (semantics, morphology, syntax, phonology, and pragmatics). Consider once again that the focus on "comprehension" may be limiting, particularly in relation to children with LI. The results of Model 3 indicate that language skills identified as "expressive" on a standardized language test can predict reading comprehension. One must be cautious in extending from correlational research to intervention practice; as always, correlation does not imply causation. That being said, all the independent variables in this study were selected based on research evidence suggesting that they were causally related to reading comprehension. A standard speech-language assessment would provide detailed information on receptive and expressive language skills in all the language domains. This in turn would be a valuable addition to reading comprehension assessment to target specific areas of need.

A specific example presents another case that illustrates the idea that reading comprehension performance, based on a set of underlying skills resulting in a behavioural outcome, is best explained in a profile approach. During the study, a teacher asked me to help her explain classroom observations of a student. This student appeared to have much higher comprehension than was expected based on her decoding skills. She would often make substitution errors, but substituted a word that made sense in the text yet did not orthographically resemble the target. Following my test battery, I was able to explain her results in terms of the Triangle Model of reading (Plaut, 2005). This student had strong semantic and discourse skills, but her phonological and orthographic representations were weak. As she attacked unknown words, her ability to use context would "win out" over her ability to decode, resulting in those meaningful substitution errors. This was a clear example of how detailed language assessment can support the interpretation of reading assessments and inform selection of meaningful intervention targets.

The Role of Shared Variance in Multiple Regression Analyses

It is clear from the preceding sections that the use of multiple regression techniques required the researcher to carefully attend to the relationships among the variables in the study. A number of variables were observed to have complex relationships that merit further discussion. They included vocabulary and decoding, the overall role of working memory, and narrative skills. The section will conclude with a discussion of measuring integrated skills in a complex system.

Vocabulary and Decoding: Considering the Lexical Quality Hypothesis.

In some previous studies of this type, decoding variance has been removed prior to considering language variance, or language variance has only been considered in terms of unique contributions. In the current study, Model 1 parsed out the unique contribution of language skills consistent with previous research. However, decoding was not removed prior to analyzing language variance in Model 2. Since Letter and Word Recognition (LWR) correlated to all language variables, it is impossible to know what of that variance "belongs" to decoding and what "belongs" to language. It would be possible to remove non-word decoding (NwD) first, but phonological abilities (which are tapped by this measure) are also language skills. In this study, a composite of non-word decoding and real word decoding was used as a way to cover "decoding" in a broad sense. Based on existing research, it is well accepted that decoding and oral language abilities share variance, but not how they share it. That is, it is not known where to carve them apart. If we accept that oral language abilities are foundational to reading, even though they progress recursively with reading development, it makes sense to look at the impact of language prior to removing the decoding variance-- which based on the premise of foundational language will necessarily incorporate true language variance. Evidence from this study supports this assertion; recall that variance was shared between decoding and language skills such as vocabulary and narrative skill.

This interdependence of language and decoding can be framed in terms of the Lexical Quality Hypothesis (Perfetti, 2007). This hypothesis suggests that successful reading is supported by high quality word representations that include accurate and readily-retrievable orthographic, phonological and semantic information. Consistent with this hypothesis, the strongest predictors of reading comprehension scores across analyses were vocabulary and decoding. Strong semantic and orthographic skills support high quality lexical representations underlying automatic decoding which in turn supports reading comprehension (Perfetti, 2007). Consider also the cases identified in the outlier analysis. These could be interpreted as demonstrating the impact of decoding sustained by effort and supported by poor underlying skills. In most cases, the outliers were cases where decoding and/or vocabulary were high, thus causing the models to predict high levels of comprehension. However, the predicted scores were at least 2 standard deviations higher than the scores these students actually achieved. It is possible that the performance of these students was limited by the splintered nature of their skills. Although decoding and vocabulary may be intact, weaknesses in other skills may cause students to incompletely understand the text. It may also be true that their apparent high levels of decoding are sustained by more effort than is apparent from their scores, leaving fewer cognitive resources free to apply to the integrative tasks of comprehension. Their decoding skills may be incompletely automatized although their skills were sufficient in a single-word reading task. This would be consistent with the Lexical Quality Hypothesis, which states that high quality lexical representations include quick access to grammatical representations and deep semantic representations of words. Children who have less well-developed language skills are less likely to have these automatic connections available. The need for them to do extra cognitive "work" at the decoding level could leave fewer resources available for comprehension, which would also be drawing upon splintered resources.

In addition to weak language skills, many children with LI also have poor working memory (Archibald & Joanisse, 2009). In fact, poor working memory may underlie some of the reading and language deficits observed in children with LI. This makes sense in the context of cognition as a "platform" for language—weak underlying skills will affect the behavioural outcome. Working memory is explored in more detail in the next section.

Working memory.

In the current results, working memory (WM) tended not to reach statistical significance as a predictor of reading comprehension. This result is somewhat consistent with the work of

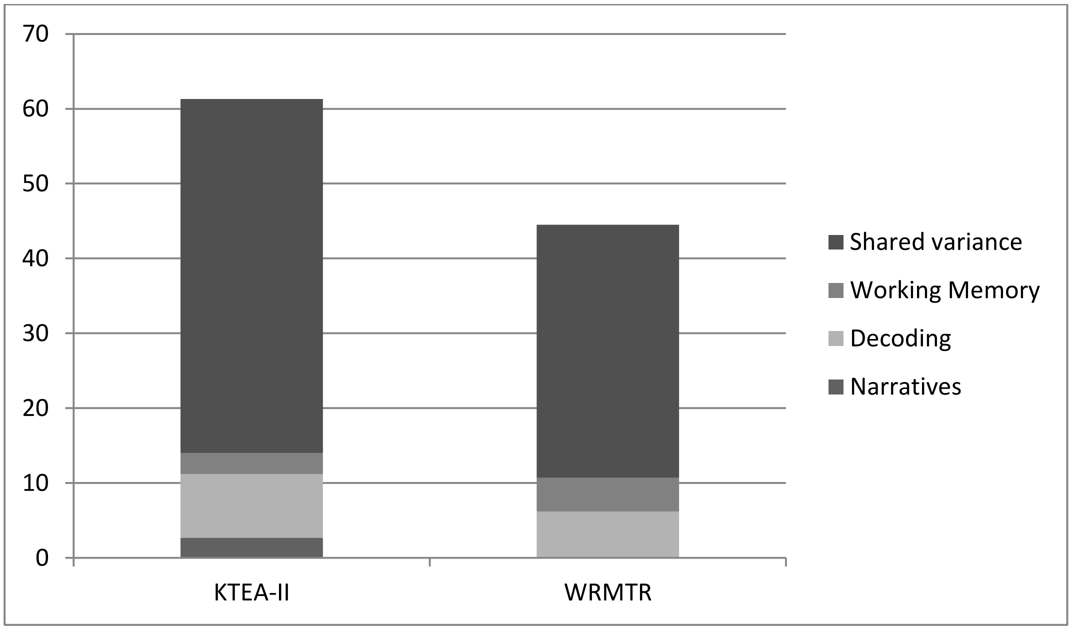
Cutting and Scarborough (2006) who found WM not to be a significant predictor of reading comprehension. They suggested that their component reading and language tasks incorporated memory variance, thus resulting in no additional contributions from specific memory measures. Despite the lack of statistical significance in the current findings, it is important to recognize that in terms of effect size working memory typically had one of the larger β values among the predictors (behind decoding and vocabulary). In addition, the correlations between working memory and the other predictors were most interesting. Working memory had a rather stable level of correlation around r = 0.5 with the reading measures, and around r = 0.3 - 0.5 with the language measures. This might be taken as support for the cognitive architecture approach presented in this paper. It was proposed that cognitive skills, including executive functions such as working memory and controlled attention, form a support network within which reading comprehension occurs. The component tasks (oral language and reading) measured here relate to working memory rather consistently which is in accordance with the concept that working memory is indeed a domain-general aptitude that supports those tasks. Therefore it could be argued that parcelling out working memory prior to language and word reading variables may remove variance that is rightly associated with these tasks. Note that based on this premise, some of the inter-correlations among the reading and language measures might be mediated by working memory. It is interesting to note that if all the variance associated with the established predictors is removed first, working memory still has a hefty β value, behind vocabulary and decoding on both tests and behind MLTU-w on the *WRMTR* (but a higher β value than NLAI on the *KTEA-II*, interestingly). Thus it appears that beyond the component skills measured here, WM contributed to reading comprehension. It is quite likely that some of that unique WM variance could be associated with other component skills not measured in this study. That is rather than working memory per se, it reflects other component tasks that engage working memory, such as making inferences or strategy use. It may also be relevant to consider that WM had the lowest mean of any of the predictors used in this study (M = 88.3, SD = 16.9). The closest mean was for the *CELF* Core Language score (M = 90.1, SD = 15.1). This score incorporates several tasks with high working memory load. Given that working memory is frequently associated with language impairment (Archibald & Joanisse, 2009) and that this sample included more children with LI than expected by population rates, the impact of working memory in this sample may overestimate the impact that would be observed in the general population. Thus the cautious interpretation of the current finding is that working memory is a moderate predictor of reading comprehension, in a group that includes a higher-than expected number of children with LI.

Narrative skills.

Narratives ended up being a problematic predictor, having relatively high simple correlations with the dependent variables and predicting significant variance in a simple regression, yet ending up with low or negative β coefficients in many of the models, in particular in *WRMTR* models. Given that much research considers narrative skills to be critical for the development of reading comprehension (see for example Botting, Simkin, & Conti-Ramsden, 2006), this finding requires further explanation. There are several possible reasons for the modest results for narrative skills as a predictor of reading comprehension. On the mathematical level, narratives correlated more highly with *KTEA-II* scores than with *WRMTR* scores. Thus, narratives would be expected to have a more robust effect as a predictor of *KTEA-II* than of *WRMTR* scores, if they are predictive at all. This is consistent with results and also with hypotheses regarding the skills tapped by these two tests. What was more surprising in the results was that NLAI lost its predictive value when certain other terms were included, particularly decoding but also the other language measures. This has two different but complementary explanations, one statistical and one theoretical. The statistical explanation is an extension of a point mentioned earlier. The measures in this study were inter-correlated in complex ways. The terms that will be most significant in the regressions will tend to be those that correlate most strongly with the dependent variables while also correlating most modestly to the other independent variables (or at least correlating more strongly to the dependent than the other independents do). These conditions do not describe NLAI very well. It correlated rather highly to both decoding and working memory, more highly in fact than to *WRMTR* scores and at a similar level to *KTEA-II* scores. Decoding and working memory both correlated more strongly to both reading comprehension tests than did NLAI. Thus, part of the result for low coefficients for NLAI was that any variance shared with these two predictors (decoding and working memory) tended to be pulled away from NLAI and associated with those predictors. This is best visualized by a set of regression models that allow the building of variance proportion charts for decoding, working memory and narratives (see Figure 7). Note that for purposes of these charts, no other predictors were used in the regression models.

Figure 7

Variance proportions for reading comprehension tests for NLAI, decoding and working memory



Although it is interesting to note the increased role of working memory on the *WRMTR* and the increased role of NLAI on the *KTEA-II*, the most important feature of this figure is the amount of shared variance, for both tests. From a theoretical standpoint, it was unexpected that narratives and decoding should be closely related. On the other hand it makes sense that WM should share variance with many variables, as many of the tasks involve this cognitive ability. The narrative test used, the *TNL*, has a fairly high working memory load on some of the component tasks (in terms of retelling and recalling specific details). Therefore regression analyses that include working memory will pull related variance from this variable. So why are decoding and narratives related? Partly, this is because both include a working memory component. Another explanation is that both abilities are related to overall level of literacy achievement in students of this age group. Children with high narrative skills are likely to have good language abilities overall with skill in applying and integrating language. Such children are more likely to be good readers as demonstrated in this and other studies. These children are therefore likely to do more reading and are more likely to be skilled automatic decoders. Since decoding is strongly correlated to the reading comprehension scores, it will pull out variance that it shares with narratives. These points form the statistical explanation of narratives as a problematic predictor. Another part of the explanation will come from a theoretical approach to the "vanishing variance" problem associated with NLAI.

From a theoretical approach, NLAI can be thought of as a composite of other language variables. Accordingly, NLAI was correlated to a number of other language variables in the current dataset and this was both reasonable and expected. This is partly due to the nature of narrative tasks: they require the integration of a number of cognitive and linguistic skills. It is partly due to the method of scoring in the *TNL*. On this instrument, different aspects of narrative skills such as use of story grammar elements are collapsed along with other abilities, such as use of grammatical sentences, which overlap with variables typically used to measure language subskills. So in this sense, putting NLAI in a regression with variables that measure those subskills is somewhat illogical. It sets up a situation where shared variance is inevitable. If the variables representing the subskills are well-chosen, they will correlate well with the predicted value (i.e., dependent variable), and in this case variance will be loaded onto any stronger predictors, which may be these subskills rather than narratives per se. This is the case in the current analysis with respect to the vocabulary composite. One might realistically ask, "What is the fix?" Based on what was observed in this dataset, it would be necessary to decide first what the real question of interest is—whether narratives per se or the set of subskills is the target. If it is subskills, then more fine-grained measures would be required, measuring the component skills of narrative ability. Perhaps a better approach, however, to the whole issue of shared variance is to take a broader look at the measurement of component skills in an integrated system. Before I turn my attention to that discussion, it is important to consider some possible limitations of narrative skills as a predictor in the age group of the current study.

Narrative skill appears to be a good predictor of reading comprehension in the elementary-school population in the existing literature. However, much of the existing research focuses on children in Grade 3 and below. When the norms for the narrative test in this study are reviewed, it appears that progress in narrative skills tapers off towards the end of the age range (up to 11 years, 12 months). On the *TNL*, the score for the 50th percentile essentially stays the same over the ages in this study. Although this is accounted for by the use of age-based standard scores, it is possible that the predictive value of the measures was reduced by a ceiling effect. Although it was possible for children to continue to develop skill in this area, the older and more skilled children may actually have plateaued on these skills, thus contributing to a reduced or negative relationship between narratives and the dependent variables. The effect might be exacerbated by the fact that the higher-achieving children with stronger language skills were likely to attempt test items that would incorporate discourse-level skills, and these children will be most subject to the plateau effect. Given this observation, the use of expository discourse might provide a broader range of performance in this age group (Coppman & Griffith, 1994). Future research in this area comparing the performance of older elementary-aged students in narrative and expository discourse could be helpful in determining when and if this ceiling effect is observed, and when a switch to expository discourse might be advisable. As noted previously, it is also important to consider the narrative test that was used in this study relative to other narrative tests used in previous research. Measures of Story Grammar usage or other measures of narrative information skill (e.g., the Information Score on the *Renfrew Bus Story*) have been shown to be predictive in past studies. The *TNL* does not provided such a score, so results obtained with this instrument may underestimate the predictive value of narrative ability when information-based scores are considered.

The discussion to this point has presented a number of interesting situations that arose as a result of inter-related variables in the analysis of the results of the current study. The approach taken to dealing with collinearity in the current dataset is considered in the following section, in the context of a theoretical approach to the integrated nature of the skills of interest.

Measuring component skills in an integrated system.

As discussed in the Literature Review, reading comprehension is not considered to be a single construct. Numerous authors have commented on the complexity of reading comprehension as a construct, and the difficulties incumbent on the researcher who attempts to isolate component skills (see for example Nation, 2005). In this study, language skills in particular were of interest. Language, like reading comprehension, cannot be considered unitary. Rather, language skills form a complex system with component skills that are partially separable, but always integrated in real-life situations. Carving up oral language, therefore, is just as problematic as carving up reading comprehension. During initial attempts to form meaningful composites as predictors, it became clear that things were going to get messy. The clearest example of this situation was the vocabulary composite. It turned out that although Number of Different Words (NDW) had been selected as a vocabulary measure, it functioned more as a discourse measure, similar to narrative abilities. On one hand, that made sense as it is measured from samples of actual discourse. The difficulty arose in terms of the regression analysis, where correlations are key. Since NDW correlated more highly with the discourse variables, but the other vocabulary variables correlated more highly with the reading comprehension tests than did the discourse variables, a split variance problem arose. This artificially created the picture of discourse-level skills as unimportant predictors. Once the variance splitting was reduced by the removal of the NDW component of the vocabulary composite, a more consistent picture was revealed, with discourse skills (as narrative abilities) gaining some modest prominence.

If all the composites are put to the same test (putting theoretical expectations aside and observing how these variables cluster) some interesting patterns emerge. The adjusted vocabulary composite is solid. Both on theory and observations, these two variables (Synonyms and Antonyms) hold together. When they were split in Model 3, the regression diagnostics indicated collinearity problems. This result supports the idea that Synonyms and Antonyms really measure aspects of one construct. As soon as one moves to the syntax variables, one encounters problems. Paragraph Comprehension was most highly correlated to Antonyms and to Oral Narratives, not to other variables in the Syntax composite. MLTU-w and Proportion of Complex Sentences were more straightforward, as they correlated strongly to one another and rather poorly with most other language variables. Finally, the narrative variables, Narrative Comprehension and Oral Narration, correlated well with each other and also with NDW. What is to be made of this picture? One solution is simply to eliminate the variables that are causing difficulty, namely NDW and Paragraph Comprehension. This approach was used in the current study to clarify results (the syntax composite was replaced by MLTU-w in the stratified analysis). Another approach would be to let the data inform the composites within the confines of an alternative theoretical framework. If one conceptualizes the constructs not as Vocabulary, Syntax and Narratives but rather as Word, Sentence and Discourse (Cain & Oakhill, 2007; Scott, 2009) then the variable associations appear to be more consistent. Word is represented by

Synonyms and Antonyms, Sentence is represented by MLTU-w and Proportion of Complex Sentences, and Discourse includes NDW, Narrative Comprehension, Oral Narration and perhaps Paragraph Comprehension. There should be a further assumption that these constructs are partially nested: the Discourse level subsumes the Sentence and Word levels and the Sentence level subsumes Word, but some independence is also observed. With this assumption, if all the constructs are used in one model, Discourse will suffer from split variance the most—some of its relevant skills are already accounted for by Word and Sentence skills. This can be related to the behaviour of the narrative variable in the current study. Regression models using the alternative constructs (calculated as averages of the component scores) predicted 49.8% of variance in *KTEA-II* scores and 45% of variance in *WRMTR* scores. Consistent with the results obtained in the reported models, word-level skills accounted for the most variance on both tests, followed by sentence skills for the *WRMTR* and discourse skills for the *KTEA-II*. When only Discourse is used, the models account for 25.4% of variance in *KTEA-II* scores and only 11.9% of variance in *WRMTR* scores. This is consistent with the hypothesis that the *KTEA-II* would load more heavily on discourse skills than would the *WRMTR*. It appears that this alternative conceptualization of Word, Sentence and Discourse can be considered complementary to the established approach of Vocabulary, Syntax and Narratives. Results obtained with the two schemes appear to be similar, thus adding credibility to the interpretation of results presented here.

The summary of this rather extended discussion of shared variance takes the form of a caution. When trying to carve up an integrated system such as language into component skills, there will inevitably be difficulties, particularly in the context of multiple regression analyses. Changes in the operationalization of the variables will make a considerable difference. Two methods of capturing expressive syntax, for example, will return different results. Within the confines of a single study, the order of entry into a hierarchical regression analysis (or the use of a different method of regression) will alter the findings. Consider for example the approach taken in this study regarding decoding. Because the correlation matrix revealed that decoding was (a) very strongly correlated to the dependent variables, and (b) quite strongly correlated to some of the language variables, some variance-splitting was anticipated. Since language constructs were of central interest in this study, the impact of decoding was sometimes intentionally disregarded. Such a decision was inconsistent with much of the existing research, which is largely framed in terms of the Simple View of Reading. However, the current study was framed in terms of the Construction Integration model of reading comprehension and the proposed extensions regarding oral language, which focus on specificity in identifying component language skills. Within the model, these component language skills contribute to and interact with decoding ability. Thus decoding was sometimes disregarded in the current analysis based the assumption that parcelling out decoding would parcel out shared variance that would be relevant to the language variables. This indeed appeared to be the case—shared variance was the rule, not the exception. The implication of all this regarding reading comprehension research is just this: exploring the construct of reading comprehension in various ways is likely to be fruitful, as long as researchers clearly specify their constructs of interest and the operationalization of these constructs (see also Skoczylas, Schneider, & Suleman, in press). With this practice in place, each alternative approach can provide additional perspective on the whole. The "true picture", if such a thing exists, is most likely to require viewing from many angles. In support of this goal, clear reporting of results will support cross-study comparison and possible meta-analysis (Skoczylas, Schneider, & Suleman, in press).

Methodology Contributions

This study makes two contributions to methodology in research on language and reading comprehension. First, results in this study were presented to compare the use of receptive and expressive tasks as predictors. Second, the use of language sample analysis was a novel approach to variable selection in research of this kind.

Receptive and expressive language contributions.

The results of Model 3 address a question of methodology in reading comprehension research: the use of expressive language measures. Much of the reading comprehension literature to date is framed in terms of the Simple View of Reading. This parsimonious model states that reading comprehension is essentially the product of listening comprehension and word recognition skills. The popularity of this model is well-placed. It has been empirically validated numerous times and provides a succinct and intuitive description of key processes in reading comprehension. Some expansions to the model have been suggested, such as the inclusion of reading speed (Cutting & Scarborough, 2006) and increased specification regarding the term "listening comprehension" (Ricketts, 2011; Skoczylas, Schneider, & Suleman, in press). What has received less focused research attention is the focus on "comprehension". This wording suggests that the focus should be on measures that evaluate language understanding and not language use (Skoczylas, Schneider, &Suleman, in press). This study is the first that I am aware of that explicitly contrasts receptive versus expressive language measures in predicting reading comprehension. If it is assumed that language functions as an integrated system, then it is quite possible that the mode of a task (i.e., receptive or expressive) may be less important than the domain (i.e., vocabulary, syntax, and so on). This point may be particularly relevant to children with LI. Children with expressive LI only are still at significant risk of reading problems (Simkin & Conti-Ramsden, 2006). The results of Model 3 strongly suggest that both receptive and expressive measures can capture language skills relevant to reading comprehension. Note that the *WRMTR* appeared to load more heavily on expressive measures; however two out of three syntax measures were expressive and syntax was an important predictor of *WRMTR* scores. So the result that *WRMTR* appears to depend more heavily on expressive language is partially a reflection of the result that the *WRMTR* tapped syntax, and that my syntax measures were largely expressive (the syntax measure most related to scores on the *WRMTR*, Mean Length of T-Unit in words, is expressive). What is really more relevant is the large amount of shared variance. This suggests that there was a great deal of overlap in terms of explanatory value between the receptive and expressive language tasks. Such a result could be interpreted to support the idea stated previously: mode is perhaps less important than domain. For example, in this study the Synonyms task (receptive) and the Antonyms task (expressive) were both selected to capture skills in the construct "depth of vocabulary". Some researchers may not be surprised by this result as they are already approaching language measures from this standpoint. Consider the measures selected by Cutting and Scarborough (2006). Although the work is framed in the Simple View, their language measures included both receptive and expressive tasks (Skoczylas, Schneider, & Suleman, in press). The results of Model 3 would appear to support such an approach.

The use of both receptive and expressive tasks appears to be appropriate to capture constructs of interest and provide predictive power in a regression approach. On a broader level, it may be useful to extend the concept of "listening comprehension" as stated in the Simple View of Reading to something like "oral language skills". Such an approach may improve the application of the Simple View to children with LI as it would more explicitly encourage the consideration of expressive language abilities.

Use of language sampling measures.

The use of language sample analysis was a novel approach in the research comparing reading comprehension tests. Expressive language skills were of interest in this study as discussed in the previous section. Language sample analysis provides measures of expressive language from samples of real language use, rather than from standardized tests. This has the potential to avoid some confounds that are of concern in this type of research, including working memory. Three language sample measures were used in this this study, including Number of Different Words, Mean Length of T-unit in words, and Proportion of Complex Sentences. These measures were selected from the Narrative Microstructure Index presented by Justice and colleagues (2006).

Number of Different Words (NDW) provides an index of the diversity of words used by a child, and was selected as a vocabulary measure. In the results, NDW came forth as a different type of vocabulary measure from Synonyms and Antonyms. NDW can be seen as a measure of vocabulary breadth, whereas the other vocabulary measures tap into depth of word knowledge. NDW had its strongest relationships to measures at the sentence and discourse level (Paragraph Comprehension, Narrative Comprehension, and especially Oral Narration). As a measure derived from language samples, it is appropriate that NDW should capture aspects of “vocabulary in action”. The above-noted language tasks represent both receptive and expressive tasks at both the sentence and discourse levels. NDW then may be capturing the productive or active aspects of a child’s vocabulary. Note that this suggestion is supported by the effect of removing NDW from the vocabulary composite. The relevance of NLAI increased, showing that some variance associated with this variable had been pulled out by NDW. The use of NDW as originally planned was not successful; however, the results suggest that NDW captures relevant language variance at the discourse level. From a statistical standpoint, the distribution of NDW was non-normal. Although predictors in a regression analysis do not need to be normally distributed (residuals should be normally distributed), it did appear that NDW was capturing student skills in a different way than did many of the other measures. This should not be viewed as a deficit however, as the point of selecting language sample measures was to look at language in a different way. Rather, it appears that further research with language sample data will be helpful in clarifying what type of information NDW provides relative to reading skills.

The remaining two language sample scores are syntax measures, Mean Length of T-Unit in words (MLTU-w) and Proportion of Complex Sentences (PCS). MLTU-w proved useful in predicting reading comprehension on the *WRMTR,* consistent with hypotheses. PCS did not prove to be predictive on its own. This was surprising as age-appropriate texts for this group should be at a level of linguistic complexity that will engage complex sentence skills (Scott, 2009). The non-predictive nature of this task has several possible explanations. It may be that expressive ability with complex sentences is not closely related to the receptive syntax skills employed in reading comprehension. It may be that this skill is relevant, but did not reach significance in this sample, which included a much higher than typical rate of children with poor language skills who did not attempt texts that would require complex sentence skills. It may also be due to a ceiling effect in this skill set similar to the effect noted in narrative skills. When the language sample norms from Justice et al. (2006) are reviewed, it can be seen that raw scores for the narrative microstructure measures (NDW, PCS and MLTU-w) actually drop off over the ages in the study. Although age-based standard scores were used, it is possible that this inherent ceiling effect limits the usefulness of these scores as predictors. It is important to note that these scores were all collected in a narrative context. As previously discussed, the age group in this study may be at the end of the useful range of narrative sampling, at least with the tasks from the *TNL*. That is to say that the narrative task itself may limit performance. It may be more revealing then to collect these language sample measures in the context of expository discourse. Further research comparing narrative and expository discourse samples could be used to test this idea (see for example Scott & Windsor, 2000).

The current study has made contributions to the methodology in research comparing reading comprehension tests. The results of Model 3 demonstrate that expressive language skills can be useful predictors of reading comprehension in a regression approach. This result provides validation for the existing research using expressive measures, and suggests that the "comprehension" construct of the Simple View of Reading as employed in the current research context could be expanded to "oral language skills" to better represent the full range of measures available. In the current study, expressive skills were measured using standardized tests and language sample analysis. The usefulness of language sample measures in this type of research was less clear. Although MLTU-w was a predictor of reading comprehension on the *WRMTR*, closer analysis of the language sample measures revealed a possible weakness, a ceiling effect present in the age range of this study. Further research will be useful to examine the use of language sample analysis in this context, with a recommendation that these measures be collected in an expository discourse context to eliminate the ceiling effect.

Limitations/Delimitations of the Study

As a doctoral project, this study was necessarily modest in scope. Accordingly, power calculations were completed to determine how many predictors could be entered into the regression and still have at least 80% power given the sample size. Composites were used to preserve statistical power when necessary. A larger sample size would have permitted the use of individual variables, which could produce more specific results. It should be noted, however, that when collinearity was observed in the current analysis, composites were actually preferable to individual variables.

Language status as impaired/unimpaired was of key interest here. One of the measures of language impairment considered for this study was previous S-LP referral and/or current diagnosis. This information turned out to be somewhat unreliable, as most children had no current S-LP service and older reports were not consistently accessible. In this district, speech and language services are provided by the health authority, rather than the school district. Procedures were in place in this study to access some speech-language results through school staff, but there was no direct contact with S-LPs. Parents were given the opportunity to report on previous service, but did not always choose to do so. Accordingly, this measure had limited usefulness. Since we collected specific language measures, we were able to use these scores to determine language status. These measures are not really diagnostic however, as they represent only performance in a limited context and fail to incorporate competence in real, daily communication tasks.

Recruitment in this study was as open as possible; within the participating schools, any child in Grades 4 to 6 who met criteria was accepted. This meant that the numbers of children with/without LI were not pre-determined. Similarly, there was no way to ensure that the groups would be matched on criteria of interest such as females/males, mother's years of education and so on. However, between-groups comparisons of these variables suggested the groups were roughly equivalent on these criteria. Efforts were made to recruit from schools that varied in geographical location in the city to encourage diversity in the sample. Typically, clinically selected populations are considered to be affected by an ascertainment bias (Catts et al., 2002; Nation et al., 2004) such that the language difficulties represented are relatively severe compared to population prevalence. By allowing the identification of language impairment among children who fail the *CELF 4* or *5* Core Language tasks, the ascertainment bias is reduced. This sample should be relatively free of other diagnoses which could potentially confound results, including frank neurological disorders, significant hearing loss, significant behavioural disorders and so on.

The results of this study must be generalized with caution. Differences between children with and without LI were of key interest, so it was desirable to have similar numbers of children in the two groups for statistical reasons. This meant that children with LI were oversampled relative to the population prevalence rate. Thus the sample recruited for this study, although appropriate for the research questions, is not representative of the population.

Conclusion and Future Directions

The current study had two main research questions. The first asked what the oral language predictors of two standardized reading comprehension subtests were. It was of interest whether those predictors differed between the tests. The second asked whether there were differences between predictors for children with language impairment and children with typically developing language. In fact, the answers to the two research questions were related, as it was observed that the significant predictors for the two language tests were different for children with typically developing language. On the *KTEA-II*, vocabulary, decoding, and narrative skill were predictive. On the *WRMTR*, vocabulary, decoding and syntax were predictive. For children with language impairment, only decoding and in one case vocabulary (for the *WRMTR)* appeared to be significant in predicting results of the two reading comprehension tests.

The finding of the first research question, contrasting predictors between the *KTEA-II* and *WRMTR* reading comprehension subtests, complements and extends the existing body of work in this area. In general, findings were in agreement with previous work. Vocabulary was a strong predictor for both tests, with narratives and syntax being less robust predictors of the *KTEA-II* and *WRMTR* scores respectively. The *KTEA-II* has not been previously studied, but existing work suggested that question-response tasks like the *KTEA-II* should be less related to decoding skills than cloze tasks, like the *WRMTR*. Such a pattern was not observed in the current dataset. Rather, decoding seemed to be equally relevant to both tests. Further research on the *KTEA-II* will be helpful in clarifying this result. It is possibly due to the current sample, which was weighted heavily toward children with LI. Exploring the results for children with LI provided more information on how language impairment affects performance on standardized reading comprehension tests.

The answer to the second research question represents a unique contribution of this study to the literature. For both reading comprehension tests, language status as impaired/not impaired was a useful predictor of reading comprehension scores although language status held unique variance only for the *KTEA-II*. A stratified analysis of the data revealed that for children with LI, vocabulary and decoding were the only significant predictors among those tested. Other language skills, including narratives and syntax, were important for children with typically developing language. Knowledge of this result should affect the interpretation of the answer to question number 1. Since the sample was weighted toward children with LI relative to the population distribution, and predictors in this group are limited to decoding and vocabulary, the overall result that decoding was equally important to both tests must be viewed with caution. Results for the full sample of this study will necessarily be skewed toward results for children with LI. Although the sample was desirable in the context of this study, further research on language predictors and on the *KTEA-II* with more population-based samples may yield different results. Another approach would be to look for differences between tests in a language with a transparent orthography. Such an approach may reduce the influence of decoding on the results overall, as phonological decoding deficits are less prevalent in languages with transparent orthographies (see for example Kendeou et al., 2012).

Given that the reading comprehension tests appear to sample a limited range of relevant language skills in children with LI, it appears that a broader view will be helpful in assessing reading comprehension among these children. It is generally accepted that language skills underlie reading comprehension and also that comprehension should be targeted alongside decoding. Since the language profile of a child with LI is unlikely to be clearly represented in the results of a reading comprehension test, standard speech-language assessments could form a useful adjunct to reading comprehension tests. Such specific language data could be used to select appropriate targets in areas such as vocabulary, syntax and narrative skills that appear to support reading comprehension. At the same time, decoding intervention appears to be necessary for many of these children. The observation of decoding skill within the expected range appears also to be suspect for some children with LI. For some children in this sample, lower levels of reading comprehension appeared to occur along with splintered language profiles but within-limits decoding skill. This was possibly due to poorly automatized word-reading skills or to the effects of an unstable base of language skills. Such effects could be the result of an overloaded cognitive-linguistic system. Rather than one or a few skills as culprits, a set of weak skills overall can be overloaded by the task demands, consistent with a multiple-deficit approach to language and reading difficulties (Pennington & Bishop, 2009). Thus the practice recommendation from the current study is that an overall profile of language and literacy skills should be compiled to interpret reading comprehension results for children with LI (see also Skoczylas, Schneider, & Suleman, in press).

The idea that a child's overall profile of skills is important in understanding reading comprehension performance is supported by the methodology contributions of this study. Results of regression modelling showed that expressive language skills, rather under-represented in the current literature, can be included among useful independent variables in reading comprehension research. It has been suggested here that it may be more important to consider the level of language (as vocabulary, syntax and narrative or alternatively by word, sentence and discourse) than to necessarily separate receptive vs. expressive skills. It was also suggested that the prevailing Simple View of Reading be slightly extended to more clearly reflect this result. Specifically, the term "comprehension" could be re-stated as "oral language skills". This would more clearly articulate the importance of considering expressive language skills. This is most important relative to children with LI who may have expressive-only deficits (Simkin & Conti-Ramsden, 2006). It is important that our theoretical approaches to reading transparently include such children so that their risk of reading difficulties is not overlooked.

The Construction Integration model of reading comprehension, with a set of proposed extensions regarding the role of oral language skills, was selected as the theoretical basis of this study. Accordingly, a number of oral language predictor variables were selected here. Although inferencing was acknowledged as part of the cognitive architecture supporting reading comprehension, no specific measure of inferencing skill was examined in this study. This was a matter of feasibility; however, it is possible that such a measure would add further important information in a study of this type. Previous research has shown inferencing to be key in reading comprehension and has revealed inferencing deficits in children with LI. Thus specific research regarding inferencing in reading comprehension among children LI will extend the results presented here.

When this study was first conceptualized, it was surprising to me that no one had specifically asked the question, "What is the best method of measuring reading comprehension for children with LI?" Given that many children who are assessed with standardized reading comprehension tests will have language difficulties (whether diagnosed or not), it appears to be quite necessary to specifically examine how our current tests represent the skills of this group. Although there is not currently an answer to my question, this study and many others looking at reading skills among children with language impairment are contributing to our understanding of literacy among children with LI. The main contribution of the current study is that users of reading comprehension tests should be aware of overall language skills among the children they test, to better interpret the results of these tests and to select measures as necessary to augment results. In short, developmental language impairment *does* make a difference in the interpretation of standardized reading comprehension test scores.

References

Adlof, S.M., Catts, H.W., & Lee, J. (2010). Kindergarten predictors of second vs. eighth grade reading comprehension impairments. *Journal of Learning Disabilities, 43*(4), 334-345.

Barnes, M.A., Johnston, A.M., & Dennis, M. (2007). Comprehension in a neurodevelopmental disorder, Spina Bifida Myelomeningocele. In K. Cain & J. Oakhill (Eds.), *Children's Comprehension Problems in Oral and Written Language* (pp. 193-217). New York, NY: Guilford Press.

Berko-Gleason, J. & Bernstein Ratner, N. (2013). *The development of language.* Upper Saddle River, NJ: Pearson Education, Inc.

Bishop, D.V.M. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language and Communication Disorders, 49*(4), 381-415.

Bishop, D.V.M. & Snowling, M.J. (2004). Developmental dyslexia and specific language impairment: Same or different? *Psychological Bulletin, 130*(8)*,* 858-886.

Botting, N. (2007). Comprehension difficulties in children with specific language impairment and pragmatic language impairment. In K. Cain & J. Oakhill (Eds.), *Children's Comprehension Problems in Oral and Written Language* (pp. 81-103). New York, NY: Guilford Press.

Botting, N., Simkin, Z., & Conti-Ramsden, G. (2006). Associated reading skills in children with a history of specific language impairment (SLI). *Reading and Writing, 19*(1), 77-98. doi:10.1007/s11145-005-4322-4

Bowey, J.A. (2005). Predicting individual differences in learning to read. In M. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 155-172). Oxford, UK: Blackwell Publishing Ltd.

Cain, K. (2010). *Reading development and difficulties.* West Sussex, UK: Blackwell Publishers

Cain, K. & Oakhill, J. (2007). Reading comprehension difficulties: Correlates, causes and consequences.. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language* (pp. 41-75). New York, NY: Guilford Press.

Calfee, R. & Greitz Miller, R. (2005). Comprehending through composing: Reflections on reading assessment strategies. In S. Paris & S. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 215-233). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language.* Circle Pines, MN: American Guidance Service, Inc.

Catts, H.W., Fey, M.E., Tomblin, J.B., & Zhang, X. (2002). A longitudinal investigation of reading outcomes in children with language impairments. *Journal of Speech, Language and Hearing Research, 45,* 1142-1157.

Catts, H.W., Fey, M.E., Zhang, X., & Tomblin, B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*(4), 331, 361.

Catts, H.W, Adlof, S.M., Hogan, T.P., & Ellis Weismer, S. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language and Hearing Research, 48*, 1378-1396.

Cook, L.G., Chapman, S.B., & Gamino, J.F. (2007). Chapter 8 Impaired discourse gist in Pediatric Brain Injury: Missing the forest for the trees. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language* (pp. 218-243). New York, NY: Guilford Press.

Copmann, K., & Griffith, P. (1994). Event and story structure recall by children with specific learning-disabilities, language impairments, and normally achieving children. *Journal of Psycholinguistic Research, 23*(3), 231-248

Cutting, L. E., & Scarborough, H. S. (2006). Prediction of reading comprehension: Relative contributions of word recognition, language proficiency, and other cognitive skills can depend on how comprehension is measured. *Scientific Studies of Reading, 10*, 277-299.

Dethorne, L. S., & Watkins, R. V. (2006). Language abilities and nonverbal IQ in children with language impairment: Inconsistency across measures. *Clinical Linguistics & Phonetics, 20*(9), 641-658.

Eason, S.H., Goldberg, L.F., Young, K.M., Geist M.C., & Cutting, L.E. (2012). Reader-text interactions: How differential text and question types influence cognitive skills needed for reading comprehension. *Journal of Educational Psychology, 104*(3), 515-528.

Fey, M. E., Catts, H. W., Proctor-Williams, K., Tomblin, J. B., & Zhang, X. (2004). Oral and written story composition skills of children with language impairment. *Journal of Speech, Language & Hearing Research, 47*(6), 1301-1318.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics.* London: Sage.

Francis, D., Fletcher, J., Catts, H., and Tomblin, J. (2005). Dimensions affecting the assessment of reading comprehension. In S. Parks & S. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369-394). Mahwah, NJ: Lawrence Erlbaum Associates.

Fuchs, L., Fuchs, D., & Hosp, M. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical and historical analysis. *Scientific Studies of Reading, 5*(3), 239-256.

Georgiou, G. K., Das, J. P., & Hayward, D. V. (2008). Comparing the contribution of two tests of working memory to reading in relation to phonological awareness and rapid naming speed. *Journal of Research in Reading, 31*(3), pp. 302-318.

Gillam, R.B. & Peason, N.A. (2004). *Test of Narrative Language.* Austin, TX: PRO-ED, Inc.

Goff, D., Pratt, C. & Ong, B. (2005). The relations between children's reading comprehension, working memory, language skills and components of reading decoding in a normal sample. *Reading and Writing, 18* (7-9), 583-616.

Gough, P. B. & Tunmer, W.E. (1986). Decoding, reading and reading disability. *RASE*, *7*(1), pp. 6-10.

Guthrie, J.T. & Wigfield, A. (2005). Roles of motivation and engagement in reading comprehension assessment. In S. Paris & S. Stahl (Eds.),*Children's reading comprehension and assessment* (pp. 187-213). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Hayward, D.V., Schneider, P. & Gillam, R.B. (2009). Age and task-related effects on young children's understanding of a complex picture story. *The Alberta Journal of Educational Research, 55*(1), 54-72.

Heilmann, J., Nockerts, A., & Miller, J. F. (2010). Language sampling: Does the length of the transcript matter? *Language, Speech, and Hearing Services in Schools, 41*(4), 393-404.

Hua, A.N. & Keenan, J.M. (2014). The role of text memory in inferencing and in comprehension deficits. *Scientific Studies of Reading, 18*(6), 415-431.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437-447.

Justice, L.M., Bowles, R.P., Kadaverek, J.N., Ukrainetz, ,T.A., Eisenberg, S.L., & Gillam, R.B. (2006). The index of narrative microstructure: A clinical tool for analyzing school-age children's narrative performances. *American Journal of Speech-Language Pathology, 15*, 177–191.

Kaufman, A.S. & Kaufman, N.L. (2004). *Kaufman Brief Intelligence Test, 2nd Edition.* Circle Pines, MN: AGS Publishing.

Kaufman, A.S. & Kaufman, N. L. (2004). *Kaufman Test of Educational Achievement, Second Edition.* Bloomington, MN: Pearson, Inc.

Keenan, J.M. & Betjemann, R.S. (2006). Comprehending the *Gray Oral Reading Test* without reading it: why comprehension tests should not include passage-independent items. *Scientific Studies of Reading, 10*, 363-380.

Kieffer, M.J., Vukovic, R.K., & Berry, D. (2013). Roles of attention shifting and inhibitory control in fourth-grade reading comprehension. *Reading Research Quarterly*, *48*(4), pp. 333–348.

Kelso, K., Fletcher, J. & Lee, P. (2007). Reading comprehension in children with specific language impairment: An examination of two subgroups. *International Journal of Language and Communication Disorders, 42*(1), 39-57.

Kendeou, P., Papadopoulos, T. C., & Spanoudis, G. (2012). Processing demands of reading comprehension tests in young readers. *Learning and Instruction, 22*(5), 354-367.

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

Kintsch, W. & Kintsch, E. (2005). Comprehension. In S. Paris & S. Stahl (Eds.),*Children's reading comprehension and assessment* (pp. 71-92). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Kintsch, W. & Rawson, K.A. (2005). Comprehension. In M. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 209-226). Oxford, UK: Blackwell Publishing Ltd.

Leonard, L.B. (2000). *Children with specific language impairment.* Cambridge, MA: MIT Press

Leslie, L., & Caldwell, J. (2001). *Qualitative Reading Inventory–3*. New York: Addison Wesley Longman.

Loscacio, G. Mahone, M, Eason, S.H., & Cutting, L.E. (2010). Executive dysfunction among children with reading comprehension deficits. *Journal of Learning Disabilities, 43*(5), 441-454.

McClintock, B., Pesco, D., & Martin-Chang, S. (2014). Thinking aloud: effects on text comprehension by children with specific language impairment and their peers. *International Journal of Language and Communication Disorders, 49*(6), 637-648.

McGuinness, D. (2005). *Language development and learning to read.* Cambridge, Massachusetts: The MIT Press.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. In Ross B.H. (Ed.), *The Psychology of Learning and Motivation, 51* (pp. 297-384). San Diego, CA: Elsevier, Inc.

Muter, V., Hulme, C., Snowling, M.J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary and grammar skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665-681.

Nation, K. (2005). Children's reading comprehension difficulties. In M. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 248-265). Oxford, UK: Blackwell Publishing Ltd.

Nation, K., Clarke, P., Marshall, C.M., & Durand, M. (2004) Hidden language impairments in children: Parallels between poor reading comprehension and specific language impairment? *Journal of Speech, Language, and Hearing Research. 47,* 199-211.

Nation, K., & Norbury, C. F. (2005). Why reading comprehension fails: Insights from developmental disorders. *Topics in Language Disorders, 25*(1), 21-32.

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.

Nation, K & Snowling, M.J. (1998). Semantic processing and the development of word-recognition skills: evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39*(1), 85-101.

Oakhill, J.V. & Cain, K. (2004). The development of comprehension skills. In Nunes, T. & Bryant, P. (Eds.) *Handbook of Children's Literacy* (pp. 155-180). Dordrecht, Netherlands: Kluwer Academic Publishers.

Oakhill, J.V. & Cain, K. (2007). Introduction to comprehension development. In K. Cain & J. Oakhill (Eds.), *Children's Comprehension Problems in Oral and Written Language* (pp. 3-40). New York, NY: Guilford Press.

Palikara, O, Dockrell, J.E., & Lindsay, G. (2011). Patterns of change in the reading, decoding, and comprehension performance of adolescents with specific language impairment (SLI). *Learning Disabilities: A Contemporary Journal, 9*(2), 89-105.

Paris, S.G., Carpenter, R.D., Paris, A.H., & Hamilton, E.E. (2005). Spurious and genuine correlates of children's reading comprehension. In S. Paris & S. Stahl (Eds.),*Children's reading comprehension and assessment* (pp. 131-160). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Paul, R. & Norbury, C.F. (2012). *Language disorders from infancy through adolescence: Listening, speaking, reading, writing and communicating.* St. Louis, Missouri: Elsevier Mosby

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading 11*(4), 357-383.

Perfetti,C., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skill. In M. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 227-247). Oxford, UK: Blackwell Publishing Ltd.

Plaut, D.C. (2005). Connectionist approaches to reading. In M.J. Snowling & C. Hulme (Eds.), *The science of reading.* Malden, MA: Blackwell Publishing, Ltd.

Protopapas, A., Sideridis, G. D., Mouzaki, A., & Simos, P. G. (2011). Matthew effects in reading comprehension: Myth or reality? *Journal of Learning Disabilities, 44*(5), 402-420.

Rapp, D.N.,van den Broek , P., McMaster, K.L., Kendeou, P., & Espin, C.A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention, *Scientific Studies of Reading*, *11*(, 289-312.

Rice, M. L., Redmond, S. M., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language & Hearing Research, 49*(4), 793-808.

Scarborough, H.S. & Dobrich, W. (1990). Development of children with early delay. *Journal of Speech and Hearing Research, 33*, 70-83.

Scott. C. (2009). A case for the sentence in reading comprehension. *Language, Speech and Hearing Services in the Schools, 40*, 184-191.

Scott, C. M. (2011). Assessment of language and literacy: A process of hypothesis testing for individual differences. *Topics in Language Disorders, 31*(1), 24-39.

Scott, C.M. & Windsor, J. (2000). General language performance measures in spoken and written narrative and expository discourse of school-age children with language learning disabilities. *Journal of Speech, Language, and Hearing Research, 43*(2), 324-339.

Semel, E., Wiig, E.H. & Secord, W.A. (2003). *Clinical Evaluation of Language Fundamentals, 4th Edition.* San Antonio, TX: Pearson, Inc.

Semel, E., Wiig, E.H. & Secord, W.A. (2013). *Clinical Evaluation of Language Fundamentals, 5th Edition.* Bloomington, MN: Pearson, Inc.

Share, D.L. & Silva, P.A. (1987). Language deficits and specific reading retardation: Cause or effect? *British Journal of Disorders of Communication, 22*(3), 219-226

Simkin, Z. & Conti-Ramsden, G. (2006). Evidence of reading difficulty in subgroups of children with specific language impairment. *Child Language Teaching and Therapy, 22*(3), 315-331.

Skebo, C.M., Lewis, B.A., Freebairn, L.A., Tag, J., Ciesla, A.A., & Stein, C.M. (2013). Reading skills of student with speech sound disorders at three stages of literacy development. *Language, Speech, and Hearing Services in Schools, 44*, 360–373.

Skoczylas & Schneider (2014). The oral language/construction integration model of reading comprehension. Poster session presented at the Symposium on Research in Child Language Disorder, Madison, WI.

Skoczylas, M., Schneider, P., & Suleman, S. (in press). Language matters: Measuring reading comprehension in children with oral language impairment.

Snowling, M.J & Hulme, C. (2005). Learning to read with a language impairment. In M. Snowling & C. Hulme (Eds.), *The science of reading* (pp. 397-412). Oxford, UK: Blackwell Publishing Ltd.

Snyder, L. S., & Downey, D. M. (1991). The language-reading relationship in normal and reading-disabled children. *Journal of Speech and Hearing Research, 34*(1), 129-140.

Spear-Swerling, L. (2004). Fourth graders' performance on a state-mandated assessment involving two different measures of reading comprehension. *Reading Psychology, 25*, 121-148.

Stahl, S.A. & Hiebert, E.H. (2005). The “word factors”: A problem for reading comprehension assessment. In S. Paris & S. Stahl (Eds.), *Children’s reading comprehension and assessment* (pp. 161-186). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Storch, S.A. & Whitehurst, G.J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934-947.

Stothard, S.E & Hulme, C. (1992). Reading comprehension difficulties in children: The role of language comprehension and working memory skills. *Reading and Writing, 4,* 245-256.

Stothard, S.E., Snowling, M.J., Bishop, D.V.M., Chipchase, B., & Kaplan, C. (1998). Language-impaired preschoolers: A follow-up in adolescence. *Journal of Speech, Language, and Hearing Research, 41*, 407-418.

Swanson, H.L., Howard, C.B., & Saez, L. (2007). Reading comprehension and working memory in children with learning disabilities in reading. In K. Cain & J. Oakhill (Eds.), *Children's comprehension problems in oral and written language* (pp. 157-189). New York, NY: Guilford Press.

Tomblin, J. B., Records, N., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40,* 1245-1260.

Tomblin, J.B., Records, N.L., & Zhang, X. (1996). A system for the diagnosis of specific language impairment in kindergarten children. *Journal of Speech and Hearing Research, 39*, 1284-1294.

Tunmer, W.E. & Chapman, J.W. (2012). The simple view of reading redux: Vocabulary knowledge and the independent components hypothesis. *Journal of Learning Disabilities, 45*(5), pp. 453-466.

Van den Broek, P., Kendeou, P., Kremer, K., Lynch, J., Butler, J., White, M.J., & Lorch, E.P. (2005). Assessment of comprehension abilities in young children. In S. Paris & S. Stahl (Eds.),*Children's reading comprehension and assessment* (pp. 107-130). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Westerveld, M. F., Gillon, G. T., & Miller, J. F. (2004). Spoken language samples of new zealand children in conversation and narration. *Advances in Speech-Language Pathology, 6*(4), 195-208.

Woodcock, R.W. (1998). Woodcock Reading Mastery Test-Revised/NU. Allen, TX: American Guidance Service, Inc.

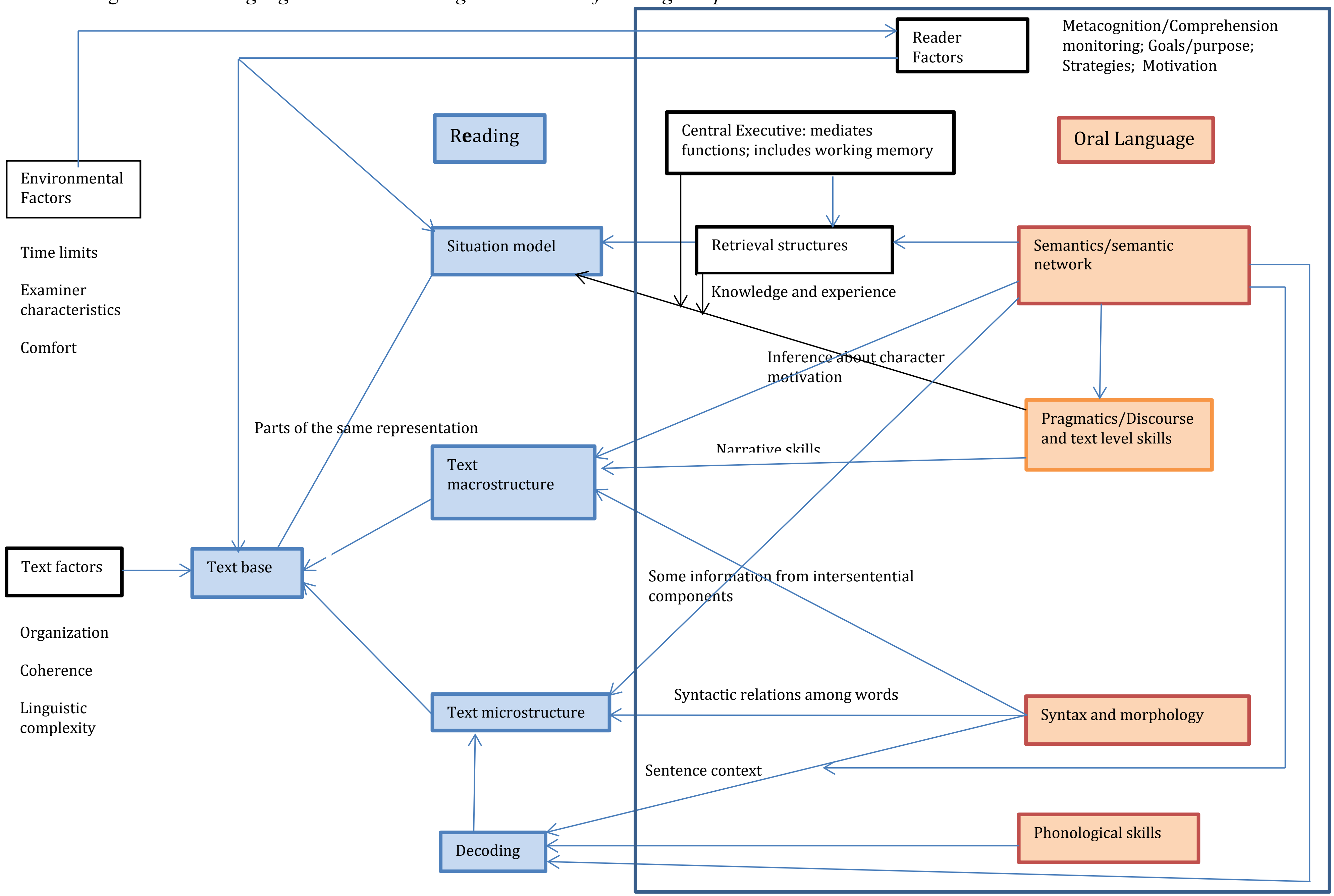
Appendices

APPENDIX A Construction Integration Model of Reading Comprehension with Proposed Extensions: Construction Integration/Oral Language Model

The Construction-Integration (CI) Model of reading comprehension (Kintsch, 1998) was adopted as the theoretical basis for this study. In order to apply this model of fluent adult performance to children, including children with language impairment, it was necessary to propose a set of extensions to the CI model (Skoczylas & Schneider, 2014). The main purpose of these extensions was to more clearly illustrate the role of oral language skills in the processes of reading comprehension. The CI model, along with the proposed extensions, is shown in Figure 8 as the Construction Integration/Oral Language Model of reading comprehension. Note that in the figure the CI model is presented on the left (under the heading "Reading"), and oral language skills are presented along the right, as the domains of semantics, syntax, morphology, phonology and pragmatics. Note that language skills are considered as part of the "Reader Factors", which are characteristics and skills that the individual brings to the reading task. The lines between the elements indicate how the components interact. For example, semantics is expected to be involved in every level of reading, from decoding to forming the text base and situation model. In some cases, there are labels on the lines which are intended to clarify the interaction or role being highlighted, or to provide an example of the interaction or role. Examples of the reader factors include working memory and motivation.

Beyond oral language skills, the extensions also capture factors such as time limits, which can affect reading performance. These factors are labelled as "Environmental Factors." Characteristics of the reading material which can affect comprehension are captured under the label "Text Factors". Examples include organization of the text and linguistic complexity.

Figure 8 *Oral Language/Construction Integration Model of reading comprehension*



APPENDIX B Determination of Language Impairment

Two main language status (i.e., typically developing language vs. language impairment) variables were selected for consideration in this study. One variable took into account all the available information: evidence of previous clinical service, score on composites of vocabulary, syntax, narrative, receptive and expressive skills (the Diagnostic Composites presented below), and performance on the *CELF 4* or *CELF 5* Core Language Score (above criterion vs. at/below criterion). The other variable was based only on the results of the *CELF 4* or *CELF 5* Core Language score (above criterion vs. at/ below criterion). As discussed in Chapter 3, the latter variable was selected for two reasons. First, it split the sample into reasonably equal groups, helpful for statistical analysis. The second reason for selecting the Core Language criterion relates to applicability. It is useful to frame results in terms of a readily-available metric. In this case, the *CELF* is the primary diagnostic instrument used locally by speech-language pathologists. In practice, results of standardized tests are complemented by functional data. In this study, the functional data available (i.e., reports of previous service, which would presumably be based upon complete assessments) was limited and reporting was possibly somewhat unreliable. Test scores used to form the diagnostic composites were subject to many of the same limitations as the *CELF* Core Language Scores (as single day observations with an unfamiliar adult in an unfamiliar location with artificial stimuli). For students who had disagreement among the measures, there was no clearly preferable method of resolving the inconsistencies. Therefore the *CELF* score was selected as a straightforward metric available for all participants. As a follow-up, Model 1 was re-run with the alternative language status variable to check the impact of this variable selection. Results were similar to those reported here.

Calculation of diagnostic composites. The language composites were constructed according to the identification method used by Tomblin et al. (1996). In short, this method involved the creation of composites for vocabulary, morpho-syntax, narratives, receptive and expressive language. The composites included all language variables available, including variables from the *CELF* language tests.

Table 18

Variables included in language composites used identifying language impairment in the non-selected method

Variable	Receptive task	Expressive task
Vocabulary	Synonyms subtest (*CASL*)	Antonyms subtest (*CASL*)
	Word classes[1] (*CELF4*)	NDW (language sample)
Grammar	Paragraph comprehension subtest (PCS; *CASL*)	MLTU-w (language sample)
		Proportion of complex T-units (language sample)
	Concepts and following directions (*CELF* 4)[2]/Semantic Relations (*CELF* 5)	Formulated sentences[2] (FM; *CELF* 4/5) Recalling sentences [2] (RM; *CELF* 4/5)
Narratives	Narrative comprehension subtest (*TNL*)	Oral Narration subtest (*TNL*)

Notes. *CASL = Comprehensive Assessment of Spoken Language; CELF4 = Clinical Evaluation of Language Fundamentals, 4th Ed.; CELF* 5 *= Clinical Evaluation of Language Fundamentals,5th Ed.;* NDW = Number of different words; MLTU-w = Mean length of T-Unit in words; *TNL = Test of Narrative Language.*
[1] For *CELF 4*, Word Classes Receptive subtest score was used; *CELF* 5 Word Classes was used—these scores were equivalent. [2] following Scott, 2009.

The standard scores of all the above variables were converted to a standard distribution with a mean of 100 and standard deviation of 15 (originally, the *CELF* subtests had a mean of 10 and standard deviation of 3. The language sample measures were reported as z-scores). The component scores were combined as unit-weighted means to calculate the value for each composite. For the Narrative Composite only, the Narrative Language Ability Index (NLAI) on the *Test of Narrative Language* was used. The NLAI is a composite of Narrative Comprehension and Oral Narration. It has a mean of 100 and standard deviation of 15, consistent with the other scores. For evaluation of each composite the criterion value (1.25 SD below mean) was 100-(1.25*15), or 81.25. The composites were calculated as follows:

- Diagnostic Vocabulary Composite =DVC = (Synonyms+Antonyms+WC+NDW)/4
- Diagnostic Syntax Composite=DSC= (FS+ RS + PC+ PCS + MLTU-w+ CD)/6
- Diagnostic Narrative Composite = NLAI as calculated in scoring *TNL*
- Diagnostic Receptive Language Composite=DRLC=(WC+CD+NC+Synonyms+PC)/5
- Diagnostic Expressive Language Composite=DELC =(FS+RS+ON+Antonyms+MLTU-w+PCS+NDW)/7

As discussed above, one criterion for language impairment from the literature was calculated as two (or more) out of five composites at -1.25 SD (Tomblin et al., 1996). Since professional speech-language pathologists collected this data, it was possible to consider clinical opinion when reviewing the application of this criterion. The use of clinical opinion was suggested when one of the clinicians noted that at least two students who had displayed marked communication difficulties during testing were not identified as language-impaired on the composite evaluation. A criterion of -1 SD on at least two out of five composites was applied as a comparison. With this criterion, three students above the original five (for a total of eight) were identified as language impaired. The three identified with this criterion included the two noted previously plus another student who had also been noted to have difficulties during testing. Thus it appears that modifying the criterion to -1 SD may enhance sensitivity without impacting specificity, based on clinical observations. Using the criterion of -1 SD increased the correlation between *CELF* Core Language Scores and the children's status on the composite evaluation.

Using -1.25 SD, the correlation was $r = .417$ ($p = .0017$), whereas using -1 SD, the correlation was increased to $r = .532$ ($p < .001$). The -1 SD criterion is more consistent with the standardized tests used in the creation of the composites. Each of the tests suggests the -1SD criterion as a cut-off for performance below age expectations (however note that the *CELF 5* suggests -1.3 SD provides the best balance between sensitivity and specificity).

APPENDIX C A Brief Evaluation of the *Familiar Sequences 1* Subtest of the *CELF- 4* As a Measure Of Working Memory

Working memory is well-established as a predictor of reading comprehension ability (Georgiou et al., 2008; Kendeou et al., 2012). In the current study on reading comprehension assessment, it was deemed necessary to include working memory as a possible predictor of reading comprehension scores. In addition to its relationship to reading comprehension, working memory is also thought to be tapped by many measures of oral language ability. Since measures of oral language ability are used as dependent variables in the regression analysis for this study, it is necessary to account for the effects of working memory inherent in these tasks.

The *Familiar Sequences* subtest is a set of tasks requiring both storage and processing of rote-learned information such as numbers to 20, the alphabet, days of the week and months. Additionally, computational tasks such as counting by 4's and 6's are included. The task is timed.

Based on current conceptualizations of working memory, storage, processing and processing speed should all be sampled (Montgomery & Windsor, 2007). The *Familiar Sequences* subtest meets these criteria. It has been recommended as a "functional working memory" (FWM) task (Boudreau & Costanza-Smith 2011). The FWM construct captures the idea that the capacity of working memory is limited and includes the necessity to allocate cognitive abilities when performing tasks (Boudreau & Costanza-Smith 2011).

Although sentence-span tasks have been shown to be predictive of reading comprehension, it is possible that these tasks may cause testing artifacts in the current study. Sentence span tasks require linguistic skills in syntax and semantics (Georgiou et al., 2008; Kendeou et al., 2012) Since we will be analysing for any differences between children with LI and children with typically developing language, a language-loaded memory task could indicate group differences that are caused by linguistic differences rather than memory differences per se.

Mental calculations are commonly included in the description of valid working memory tasks. There is a potential concern with a memory task that depends solely upon numeric information in the current study. Since a portion of the study sample will include children with learning disabilities, it is possible that memory results could be skewed lower for children with LD in math. The Familiar Sequences subtest does include mental math, but also tasks with rote-learned information such as months of the year. The variety of tasks may make this measure more inclusive of children with learning difficulties in specific areas.

Thus the *Familiar Sequences I* task appears to meet the requirements of a test of verbal working memory that is consistent with current working memory descriptions. It has an additional feasibility benefit as it is included in one of the instruments used for other measures, the *CELF 4* test.

APPENDIX D Regression Diagnostics

Procedures are as recommended by Field (2013). For each diagnostic, the label and a brief description are presented followed by the relevant data or statistic, with its criterion and the evaluation of the diagnostic. The SPSS statistical software package was used for all procedures.

Model 1a *KTEA-II*

Statistic	Criterion	Observed value	Evaluation	Comment
Durbin-Watson statistic	close to 2 is ideal; < 1 or > 3 are problems)	1.883	☑	tests that the residual terms are uncorrelated
Tolerance Values	> 0.2	all > 0.2	☑	check for collinearity effects
Residual statistics: checks for influential cases and outliers				
Cook's Distance	should not be > 1	maximum value = 0.145	☑	"measure of the overall influence of a case on the model" (Field, 2013; p. 306)
Leverage	cases should be close to average (k+1)/n where k=#predictors, n=sample size -investigate cases > (2(k+1)/n).	Observed: (2(7+1)/54) = 0.3; no cases > 0.3	☑	"influence of the observed value of the outcome variable over the predicted values" (Field, 2013; p. 307)
Mahalanobis distance	examine cases above 12	Maximum observed=14.06	potential problem	In all, 6 potential influential cases identified, with 2 being more extreme.

Covariance ratio	check values greater than 1+(3(k+1)/n=1.44; and values less than 1-3(k+1)/n = 0.56	Values observed outside the specified range; Same potentials as Mahalanobis distance identified. Data for influential cases were rechecked against original to rule out data entry errors (none found).

The graph of standardized residuals appeared to fit a normal curve and normal P-P plot of regression standardized residuals revealed no concerns with normality assumption. Scatterplots of regression standardized predicted values by regression standardized residuals revealed no concerns, nor did partial plots of individual predictors. Assumption of homogeneity of variance appears to be valid.

Model 1a *WRMTR*

	Criterion	**Observed value**	**Evaluation**	**Comment**
Durbin-Watson statistic	close to 2 is ideal; < 1 or > 3 are problems)	2.27	☑	tests that the residual terms are uncorrelated
Tolerance Values	> 0.2	All > 0.2	☑	check for collinearity effects
Residual statistics: checks for influential cases and outliers				
Cook's Distance	should not be > 1	maximum value = 0.152	☑	"measure of the overall influence of a case on the model" (Field, 2013; p. 306)
Leverage	cases should be close to average (k+1)/n where k =	Observed: (2(7+1)/53) = 0.3; no cases >	☑	"influence of the observed value of the outcome variable over the predicted

	#predictors, n = sample size -investigate cases > (2(k+1)/n).	0.3		values" (Field, 2013; p. 307)
Mahalanobis distance	examine cases above 12	Maximum observed = 14.03	potential problem	In all, 4 potential influential cases identified, with 2 being more extreme.
Covariance ratio	check values greater than 1+(3(k+1)/n=1.45; and values less than 1-3(k+1)/n = 0.55	Values observed outside the specified range; Same potentials as Mahalanobis distance identified. Data for influential cases were rechecked against original to rule out data entry errors (none found).		

The graph of standardized residuals appeared to fit a normal curve and normal P-P plot of regression standardized residuals revealed no concerns with normality assumption. Scatterplots of regression standardized predicted values by regression standardized residuals revealed no concerns, nor did partial plots of individual predictors. Assumption of homogeneity of variance appears to be valid.

Model 2a *KTEA-II*

	Criterion	Observed value	Evaluation	Comment
Durbin-Watson statistic	close to 2 is ideal; < 1 or > 3 are problems)	1.652	☑	tests that the residual terms are uncorrelated
Tolerance	> 0.2	all > 0.2	☑	check for collinearity

Values				effects
Residual statistics: checks for influential cases and outliers				
Cook's Distance	should not be > 1	maximum value = 0.135	☑	"measure of the overall influence of a case on the model" (Field, 2013; p. 306)
Leverage	cases should be close to average (k+1)/n where k=#predictors, n=sample size -investigate cases > (2(k+1)/n).	Observed: (2(5+1)/54) = 0.22; no cases > 0.22	☑	"influence of the observed value of the outcome variable over the predicted values" (Field, 2013; p. 307)
Mahalano-bis distance	examine cases above 12	Maximum observed=11.43	☑	

:

The graph of standardized residuals approximates a normal curve with some possible skewing towards the upper end. The normal P-P plot of regression standardized residuals revealed no concerns with normality assumption. Scatterplots of regression standardized predicted values by regression standardized residuals revealed no concerns, nor did partial plots of individual predictors. Assumption of homogeneity of variance appears to be valid.

Model 2a *WRMTR*

	Criterion	**Observed value**	**Evaluation**	**Comment**
Durbin-Watson	close to 2 is ideal; < 1 or > 3 are	2.053	☑	tests that the residual terms are uncorrelated

statistic	problems)			
Tolerance Values	> 0.2	all > 0.2	☑	check for collinearity effects
Residual statistics: checks for influential cases and outliers				
Cook's Distance	should not be > 1	maximum value = 0.176	☑	"measure of the overall influence of a case on the model" (Field, 2013; p. 306)
Leverage	cases should be close to average (k+1)/n where k=#predictors, n = sample size -investigate cases > (2(k+1)/n).	Observed: (2(5+1)/53) = 0.23; no cases > 0.23	☑	"influence of the observed value of the outcome variable over the predicted values" (Field, 2013; p. 307)
Mahalanobis distance	examine cases above 12	Maximum observed =11.31	☑	

:

The graph of standardized residuals approximates a normal curve with dip creating a bimodal appearance. The normal P-P plot of regression standardized residuals revealed no strong concerns with normality assumption. Scatterplots of regression standardized predicted values by regression standardized residuals revealed no concerns, nor did partial plots of individual predictors. Assumption of homogeneity of variance appears to be valid.