1-1-2016

Quantile Regression Analyses of the Component Skills in Various Comprehension Tests

Anh Ngoc Hua
University of Denver

Quantile Regression Analyses of the Component Skills in Various Comprehension Tests

---

A Dissertation

Presented to

the Faculty of Social Sciences

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

Anh N. Hua

August 2016

Advisor: Janice M. Keenan

Author: Anh N. Hua
Title: Quantile Regression Analyses of the Component Skills in Various Comprehension Tests
Advisor: Janice M. Keenan
Degree Date: August 2016

Abstract

Many studies to date have examined cognitive factors that drive individual differences in reading comprehension. However, these studies often focused on typical readers, and it is not clear whether their findings apply similarly to readers performing in the extreme ends of the distribution, i.e., poor and good readers. In this dissertation, we used quantile regression on a sample of 834 children (age 8-18) to advance our understanding of the relative importance of different component processes of comprehension not just for the typical but also for poor and skilled readers. In Study 1, we examined how the relative importance of components of the Simple View of Reading, namely word recognition and listening comprehension, might vary across different skill levels of reading comprehension. Because there are large differences between tests in the component skills they assess, reading comprehension is defined by five different tests. This is to determine how generalizable our findings are across tests. In Study 2, we deconstructed listening comprehension into vocabulary and working memory to see whether their contributions to reading comprehension beyond decoding skills also vary across reading skills. In Study 3, we determined whether the contributions of vocabulary and working memory found for reading tests generalize to listening tests.

We found that, for three out of five reading tests, the contributions of the component processes vary as a function of reading performance levels. Therefore, the results previously found for typical readers are not always generalizable to poor and skilled readers. Additionally, working memory is a reliable component of listening comprehension only for some reading comprehension tests whereas vocabulary is a much more robust component of listening comprehension across all reading comprehension tests and readers. Finally, we found that reading and listening comprehension rely on the same language processes of vocabulary and working memory once differences in decoding skills are taken into account. Interestingly, we also found some evidence that working memory may be more influential in reading comprehension than in listening comprehension. We discuss the implications these findings have for diagnosis, instruction, and research.

Acknowledgements

First, I would like to express my sincere gratitude to my advisor, Jan Keenan, who has played a very influential role in helping me to become a better writer and a better thinker over the years. Her passion for research and science has been a great source of inspiration and her encouragement and devotion to my success have made graduate school a very nourishing experience for me. Next, I would like to thank Bruce Pennington for his intellectual support. In addition to being my instructor in a couple of classes, Bruce served both on my master and dissertation thesis committees. I know I can always count on him for asking thought-provoking questions and thereby helping me grow in the process. I would also like to thank Kathy Green from Morgridge College of Education who taught me Item Response Theory and who also serves as the outside chairperson. From Kathy, I have a much better understanding of how the Rasch Model can be used to inform our understanding about our own measurements. I would like to thank Omar Gudino also for his assistance with the statistical aspects of this dissertation.

My dissertation was supported by a grant P50 HD27802 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development to the Colorado Learning Disabilities Research Center. I would like to thank Dick Olson for the word and nonword reading data, Bruce Pennington for the working memory data, Erik Willcutt and John DeFries for vocabulary and the PIAT, and Jan Keenan for the reading and listening comprehension data. I would also like to thank the families for their participation and our undergraduate research assistants for transcribing and scoring our data. Finally, I would like to thank my husband, Eric Chwang, for his support and love during my graduate career.

Table of Contents

List of Tables

Chapter 1. Introduction

Overview of Dissertation

As any reading researcher or any teacher knows, large individual differences exist in how quickly children learn to decode words, integrate sentences, and construct meaning from reading a passage. Interestingly, these differences continue to exist even among the most educated adults (Jackson & McCelland, 1979). An important goal in reading research, therefore, is to understand the nature and etiology of these differences.

In response to this goal, many studies have investigated the cognitive factors that drive these differences using ordinary least squares (OLS) regression (e.g., Cain & Oakhill, 1999; Cutting & Scarborough, 2006; Keenan, Betjemann, & Olson, 2008; Nation & Snowling, 1997). Although these studies have been greatly helpful in advancing our knowledge of the nature of reading comprehension, it is important to acknowledge that the analyses they used only estimate average effects. Therefore, they may be limited in what they can tell us about readers who fall at the extreme ends of the reading comprehension distribution, i.e. low-tail poor comprehenders and high-tail good comprehenders. Because comprehension is the product of many cognitive skills operating in tandem and because comprehenders are highly heterogeneous in terms of their reading profiles, it may be that the primary determinants of performance at the low tail could be different than those at other points in the distribution.

Consider the following example. Readers who score below the 10^{th} percentile on a reading comprehension test may do so because their comprehension is highly constrained by word decoding difficulties; as a result, word decoding may be the only predictive factor in their reading comprehension scores. Average performers, on the other hand, may be much less constrained by word decoding difficulties and therefore their comprehension may be explained both by their word decoding skill and their listening comprehension skill, whereas for top performers, listening comprehension may be the only predictive factor because there is no variability in their decoding scores. What this example demonstrates is the possibility that there may be important subpopulation differences among readers, and thus, what is highly predictive of reading comprehension for average comprehenders may not necessarily be equally predictive of comprehension scores across the distribution.

The method that has been used to investigate these potential differences in the components of reading comprehension across skill levels has been to define subgroups and compare their performance on different variables. For example, one might perform a median split on the full sample and compare relative importance of different skills (decoding, working memory, etc.) for the bottom half of the sample versus the top half. Or one might use a cutoff (e.g., 20^{th} percentile) to define a subgroup of poor performers and typical performers. This subgroups approach has the benefit of specifically examining processes in the groups of interest, rather than examining the average as in the OLS regression approach; but as we review below, there are a number of measurement problems associated with the subgroup approach.

What is needed instead is a methodology that combines the power of OLS regression with the specificity of the groups approach, but without the associated measurement issues. Quantile regression, an analytical method developed by Koenker and Basset (1978), seems to be a technique that does exactly that. Because it can assess directly the relationship between the predictors and an outcome variable at various quantiles of that outcome variable while using all the data points in the full sample, it allows us to examine relationships at the extremes of the distribution and compare them to other quantiles. It is therefore more flexible than the traditional OLS regression approach and more powerful than truncating the sample into subgroups.

The current study uses quantile regression to extend our theoretical understanding of the contributions of various component skills to reading comprehension as a function of reading comprehension performance. In Study 1, we will examine how the relative importance of the components of the Simple View of Reading (SVR; Gough & Tunmer, 1986), namely word recognition and listening comprehension, might vary across reading comprehension quantiles. We are particularly interested in knowing whether these components contribute similarly across quantiles or whether it might be the case that word recognition skills account for less variance as performance increases, while listening comprehension skills account for more variance. We will define reading comprehension as five different measures commonly used to assess reading comprehension. This is because we know that tests used for comprehension assessment can vary in length, text types (expository or narrative), and format of assessment (e.g., cloze test, multiple-choice questions, open-ended questions, retelling) and these test differences have consequences for which component skills are being assessed (Cutting &

Scarborough, 2006; Keenan, Betjemann, & Olson, 2008; Nation & Snowling, 1997) and in whom they identify as poor comprehenders (Keenan et al., 2014; Keenan & Meenan, 2014). By using so many tests, we provide a rigorous assessment of whether the quantile differences are generalizable across all reading tests.

In Study 2, we will break down the SVR component of listening comprehension that was employed in Study 1 into components of vocabulary and working memory and examine whether there are differences across reading comprehension quantiles in how these components contribute to comprehension. Conceptually, listening comprehension covers a wide range of skills, some of which include vocabulary knowledge, working memory, inference generation, prior knowledge, and motivation. However, in practice, the measures that assess listening comprehension often vary in their complexity (phrases to lengthy texts) and when one's listening comprehension performance suffers, it is not clear what specifically drives the poor performance. Therefore, it is important to break down listening comprehension into its more well-defined components. The question addressed in this study is whether vocabulary and working memory contribute unique variance to reading comprehension across different comprehension levels once decoding skills are taken into account.

In Study 3, we will assess the relative contributions of vocabulary and working memory to listening comprehension across quantiles of *listening comprehension* tests. There is an ongoing debate regarding whether listening comprehension is essentially the same as reading comprehension once differences in decoding skills are taken into account. The results of this study will shed additional light on this important question. Specifically, if the patterns of results for this study approximate those found for reading comprehension in Study 2, they suggest that what is applicable to reading comprehension is also applicable to comprehension in general. If the results differ, then they will illuminate how, and perhaps, why listening comprehension may differ from reading comprehension.

In this dissertation, we use quantile regression to advance our understanding of the relative importance of different component processes of comprehension across performance levels. The hope is that this more nuanced information can be used to improve our understanding of the nature of test scores, so if component skills in comprehension change as a function of the child's skill level, then we can take this into account when interpreting a particular test score. We can also use this knowledge to help those who do intervention determine which component skills poor readers may need to improve the most. Finally, our research can increase awareness among reading researchers of the value of using quantile regression to capture information specific to readers of varying skills without having to truncate the full sample.

Analytic Methods for Studying Individual Differences in Reading Comprehension

Ordinary Least Squares Regression

Ordinary least squares (OLS) regression is one of the most commonly used techniques in research that investigates individual differences in reading comprehension. This method is highly popular among social sciences researchers partly because of its mathematical simplicity and interpretability.

OLS regression has clearly advanced the field by providing us a much more sophisticated understanding of what reading comprehension is, and how it may change as a function of a child's development (Keenan et al., 2008) and the type of test (Cutting & Scarborough, 2006; Keenan et al., 2008; Nation & Snowling, 1997). However, despite its advantages, this method along with its extensions (e.g., analysis of variance, multilevel and structural equation models) suffers a basic limitation in that it only captures the mean effects among the variables. Under ideal conditions, modeling the mean may be sufficient to provide a parsimonious description of the relationship between the predictors and the response distribution (Hao & Naiman, 2007). However, because heavy-tailed distributions commonly occur in social science research, modeling the mean can become an imprecise way of summarizing information across an entire distribution. More specifically, when floor and ceiling effects are present, regression parameters estimates can be strongly skewed and the associated *p* values will be unreliable if these normality issues are not addressed (J. Cohen, P. Cohen, West, & Aiken, 2003). Researchers have tried to address this issue by transforming the data, but when this happens, the original metric is lost and the results can be difficult to interpret. Other researchers have ignored the normality violation altogether, leading to biased regression coefficients and unreliable *p* values. Additionally, because the homoscedasticity assumption in linear regression often fails, focusing on central tendencies may not capture important subpopulation differences (Hao & Naiman, 2007). Therefore, despite linear regression's advantages, its inherent limitations make it difficult for reading researchers to assess accurately the relationships between reading components and reading comprehension for the extreme groups.

Truncating Sample into Subgroups

Truncating the full sample into groups of different comprehension skills has been the method of choice for examining individual differences in comprehension processes in extreme groups. A cutoff is used to define groups, which are then compared on the variables of interest in order to see how they differ. This method provided important insights about group differences, such as how poor comprehenders process information differently from other groups, and the nature and stability of their comprehension deficits. For example, Catts, Adlof, and Weismer (2006) selected three groups of readers: (1) specifically poor comprehenders, who scored below the 25th percentile in reading comprehension and above the 40th percentile in word recognition, (2) specifically poor decoders, who scored below the 25th percentile in word recognition and above the 40th percentile in reading comprehension, and (3) typical readers, whose word recognition and reading comprehension scores were between the 40th and 84th percentiles. They found that specifically poor comprehenders have deficits in language comprehension with normal phonological skills; specifically poor decoders, on the other hand, have deficits in phonological processing with relatively intact language comprehension skills. This double disassociation in language difficulties provided empirical support for the simple view of reading.

Truncating the full sample into control and poor reading groups at some low tail cutoff or at the median split based on their reading achievement, and then contrasting them on a number of cognitive variables, has provided extensive information on how poor comprehenders process information differently and how stable their deficits can be over time. However, there are some limitations associated with this approach. First, true dichotomies are not common in psychology (Preacher, Rucker, MacCallum, & Nicewander, 2005) and dichotomizing a continuous variable such as reading comprehension at some arbitrary cutoff loses most of the numerical information about individual differences in the original distribution. Further, dichotomizing a variable gives the illusion that there is a genuine break in the reading comprehension distribution and that individuals separated by the cutoff differ greatly from each other in their cognitive profiles. Second, assuming normal distribution, dichotomization can both attenuate and inflate population correlations, and the impact on the correlation coefficients is greater as the cutting point moves further away from the mean (MacCallum, Zhang, Preacher, & Rucker, 2002). Third, because the selection criteria used to identify poor comprehenders vary considerably across studies, the arbitrariness of these cutoffs has a substantial effect on the cognitive profiles of those who are defined as poor comprehenders (Keenan et al., 2014). This, in turn, may contribute to inconsistencies in the literature regarding group differences. Finally, just like linear regression, this method only estimates average effects. Thus, even though this method has been very popular among reading researchers, it introduces unintended measurement issues that could greatly bias the results.

Matching Groups

When groups of specifically poor comprehenders are selected and compared with typical comprehenders, they are often matched on variables such as chronological age and decoding skills, either by equating samples on mean performance levels or by using case-by-case matching. By matching the groups on age and decoding skills, researchers can then isolate certain cognitive processes of interest from potentially confounding effects of these variables in order to see whether the cognitive processes of interest are directly related to differences in comprehension.

The matching method is quite appealing because of its conceptual simplicity in allowing researchers to examine the specificity of certain cognitive processes in reading comprehension failures by removing the effects of decoding. However, a major limitation of this approach is that the internal and external validity of the results may be threatened due to regression to the mean. That is, because researchers likely have selected extreme cases from each population of interest, regression to the mean likely occurs. When this happens, the results can be biased as the scores of the samples may regress toward their own population mean, thereby threatening the internal and external validity of any conclusions drawn regarding group differences. Another limitation, especially for studies that employ case-by-case matching, is that by forcing the two samples to be equal in variability as well as in the mean performance levels, important information about the respective populations of interest may be obscured (Jackson & Butterfield, 1989). Thus, despite its conceptual simplicity, it has major limitations that may make it difficult for the reader to trust the results.

How Quantile Regression Can Address Previously Discussed Limitations

Quantile regression is an extension of OLS regression in that it also examines the relationships between X and Y. However, instead of generating a single coefficient that characterizes the relationship, it predicts Y based on the score of X at specific points in the distribution of Y. As discussed in more detail later, it uses all the data points to estimate relations between two variables at various locations across the distribution of scores through asymmetric weighting of the values across the distribution using bootstrapping, data resampling, and statistical inference. In doing so, it generates a unique intercept and a slope value for each point in the distribution of the outcome variable. Due to this method of estimation, quantile regression makes no assumptions of the variance in the residual error terms and is robust to non-normal errors and outliers (Koenker, 2005). As such, this method is superior to OLS regression because it can capture the potentially nonlinear relationships among the variables while making no normality assumptions.

Because quantile regression allows us to investigate the differential relations between the component skills of reading comprehension as a function of a child's comprehension skill level while using all the data points, we avoid the issue of truncating our sample into multiple comprehension subgroups based on some selection criteria that may or may not be consistent with those used in other studies. Moreover, by using all the data points, we maintain the statistical power to detect meaningful relationships among our variables. In other words, instead of basing our conclusions about the sources of individual differences on studies that use only a small number of poor comprehenders and controls selected from a larger sample, quantile regression uses all the data points from the entire sample. Another advantage of quantile regression is that multiple predictors can be used, thus allowing us to examine the specificity of certain cognitive processes in reading comprehension across the reading comprehension distribution while controlling for the effects of other confounding factors, such as word recognition and vocabulary, without having to match children on those variables.

Despite the clear advantages of quantile regression over the traditional analytic approaches, only a handful of reading studies have employed this method. Specifically, Catts, Petscher, Schatschneider, Bridges, and Mendoza (2009) used quantile regression to examine how children's oral reading fluency in first and second grade was related to reading comprehension in third grade, and found that the predictive validity was stronger for those with higher oral reading fluency scores than those with lower scores, especially in first grade. Petscher and Kim (2011) used quantile regression to examine the predictive validity of oral reading fluency for reading comprehension as a function of students' oral reading fluency scores, and found that fluency and reading comprehension were poorly correlated, $r = .10$, at the 10^{th} percentile of oral reading fluency scores, whereas they were highly correlated, $r = .95$, at the 90^{th} percentile. These results demonstrate that for those who score high on fluency, they also score high on reading comprehension whereas for those who score low in fluency, they could either score high or low on the comprehension task. Petrill, Logan, Sawyer, and Justice (2014) used quantile regression to examine the association between frequency of storybook reading and emergent literacy in children who were at risk for language impairment both as a function of levels of emergent literacy and frequency of storybook reading. They found that the correlations between emergent literacy skills and frequency of storybook reading are higher when storybook reading was infrequent and much lower when storybook reading was high. Additionally, they found that the association between frequency of storybook reading and emergent literacy was highest at higher quantiles of emergent literacy for print knowledge. As such, these results suggest that for children with language impairments, the relationship between the home environment and emergent literacy skills is conditional on the quality of the home environment as well as the child's literacy skills. Another study, conducted by Logan and colleagues (2012), used quantile regression to determine if the genetic and environmental contributions remain stable across the reading distribution. They found higher heritability at the lower quantiles for word and nonword decoding measures and lower heritability at the lower and higher ends of the vocabulary measure. Additionally, they found no meaningful differences in heritability estimates across the phonological awareness distribution. Finally, Tighe and Schatschneider (2014) used quantile regression to assess the predictive utility of morphological awareness and vocabulary knowledge at multiple points along the reading comprehension distribution in adult basic education students. They found that morphological awareness had the greatest unique predictive ability at lower quantiles of reading comprehension whereas vocabulary showed the greatest unique predictive ability at higher quantiles. Clearly, these studies demonstrated the utility of quantile regression in revealing nonlinear relationships that may have been missed entirely by the traditional statistical methods.

Chapter 2

Research Questions & Method

Study 1: How Does the Simple View of Reading Apply across Reading Comprehension Quantiles and Tests?

The current study takes advantage of the benefits of the quantile regression approach to extend our theoretical understanding of the contributions of the Simple View of Reading components, namely word recognition and listening comprehension, to reading comprehension as a function of a child's comprehension skill. We know that both of these components account for all the reliable variance in average reading comprehension performance (e.g., Catts, Adlof, & Weismer, 2006; Hoover & Tunmer, 1993; Joshi, Williams, & Wood, 1998; Vellutino, Tunmer, Jaccard, & Chen, 2007), but we do not yet know whether the relative contributions of these skills are consistent for children whose reading comprehension scores lie at the extreme ends of the distribution. As previously mentioned, studies using the groups approach suggest that the relationships may vary across the reading comprehension distribution (Curtis, 1980; Gough, Hoover, & Peterson, 1996; Keenan et al., 2008), such that word recognition plays a dominant role in average reading comprehension during the primary grades or when a child is a less skilled word reader, whereas listening comprehension plays a dominant role in average reading comprehension beyond the primary grades or when a child is a more skilled word reader. According to Paris (2005), these findings are consistent with the notion that skills like decoding are considered to be more constrained relative to other reading skills in that they are learned and mastered within a brief developmental span and therefore, do not explain enduring individual differences, whereas vocabulary and comprehension are the least constrained skills in that they continue to develop over time and explain enduring and meaningful individual differences. However, the empirical evidence of the varying relationships between reading components and reading comprehension across different reading comprehension skill levels in previous research is indirect because the variables of interest in these studies were either word recognition or chronological age and not reading comprehension itself.

Given the imperfect correlations among word recognition, chronological age, and reading comprehension, the current study examines the relative importance of word reading and listening skills directly as a function of children's reading comprehension scores. We will do so using five different commonly used reading comprehension tests: Woodcock Johnson Passage Comprehension (WJPC-3, Woodcock, McGrew & Mather, 2001), Peabody Individual Achievement Test (PIAT, Dunn & Markwardt, 1970), Gray Oral Reading Test (GORT-3, Wiederholt & Bryant, 1992), and both Retellings and Comprehension Questions from the Qualitative Reading Inventory (QRI-3, Leslie & Caldwell, 2001). These tests are the same ones used by Keenan et al. (2008) and shown via OLS regression to be very different from each other in the skills they assess. Namely, word decoding accounted for far more variance than listening comprehension on the WJPC and the PIAT, but listening comprehension accounted for the majority of the variance in the other tests.

Because vocabulary and comprehension are the least constrained type of skills (Paris, 2005), we expect listening comprehension's role to increase as reading comprehension increases. However, since word recognition is influenced by both the most and the least constrained skills of decoding and vocabulary (e.g., Betjemann & Keenan, 2008; Ouellette, 2006; Plaut, McClelland, Seidenberg, & Patterson, 1996; Ricketts, Nation, & Bishop, 2007), it is not clear whether word recognition's influence in reading comprehension will necessarily decline among the better readers. Our hypothesis about quantile differences is also tempered by what we know about test characteristics. That is, in addition to the previously mentioned characteristics that discriminate one test from another, i.e., length and format of assessments (e.g., cloze test, multiple choice comprehension questions, open-ended questions or retellings), tests may also differ in how items change in complexity as the difficulty level increases. For some tests, harder items may simply involve passages with more infrequent and phonologically complex vocabulary and this increase in the variance of vocabulary knowledge could lead to a greater correlation between word recognition and reading comprehension at the higher quantiles. For other tests, harder items may involve things such as more integration of information across longer passages and this increase in the variance of oral language skills could lead to a greater correlation between listening comprehension and reading comprehension at the higher quantiles. Alternatively, there may be no increase in the variance associated with word knowledge or higher-level language skills among items that are considered to be the most difficult by the tests' manuals. In such a case, no quantile differences may be observed for these tests. Thus, by using many tests that are sufficiently different from each other, we provide a rigorous assessment of how test differences could contribute to whether quantile differences are obtained or not.

Study 2: Do Vocabulary and Working Memory Contribute Consistently to Reading Comprehension across Quantiles?

The second goal of the current study is to assess some of the main components of listening comprehension, i.e., vocabulary and working memory, to determine the relative contributions of these component skills across quantiles of reading comprehension. We selected vocabulary and working memory because each of these variables alone has received extensive empirical support in their contribution to reading comprehension (e.g., Carpenter & Just, 1988; Daneman & Carpenter, 1980; Landauer, 2007; Perfetti, 2007). However, while vocabulary has received extensive empirical support in its unique contribution to reading comprehension (e.g., Braze, Tabor, Shankweiler, & Mencl, 2007; Cain, Oakhill, & Bryant, 2004; Cornoldi, De Beni, & Pazzaglia, 1996; Landauer, 2007; Oakhill & Cain, 2012; Oakhill, Cain, & Bryant, 2003; Ouellette, 2006; Perfetti, 2007; Seigneuric & Ehrlich, 2005; Swanson & Berninger, 1995; Yuill, Oakhill, & Parkin, 1989), working memory has only received mixed support. For example, Nation, Adams, Bowyer-Crane, and Snowling (1999) and Stothard & Hulme (1992) found that the relationship between working memory and reading comprehension is mediated by vocabulary. Similarly, Tighe, Wagner, and Schatschneider (2015) found that working memory is not a unique predictor of reading comprehension across third, eighth, and tenth graders after controlling for decoding, verbal skills, and nonverbal reasoning. Other researchers found that working memory contributes unique variance to reading comprehension when decoding and vocabulary are controlled (Cain, Oakhill, & Bryant, 2004; Cornoldi, De Beni, & Pazzaglia, 1996; Oakhill, Cain, & Bryant 2003; Sesma, Mahone, Levine, Eason, & Cutting, 2009; Swanson & Berninger, 1995; Yuill, Oakhill, & Parkin, 1989). Longitudinally, Seigneuric and Ehrlich (2005) found that working memory does not contribute unique variance to reading comprehension during first and second grade. However, it accounts for an additional 6% of the variance in third grade reading comprehension scores. These researchers concluded that once decoding skills are mastered, higher-order skills such as working memory start to emerge as important determinants in reading comprehension.

It is clear that the research on working memory's role in average reading comprehension remains inconclusive. The fact that working memory sometimes uniquely predicts reading comprehension and sometimes does not motivates us to think deeper on the similarities and differences across studies that may have led to the inconsistent outcomes. A notable difference across previous studies may include differences in the reading assessments. In particular, we learned from Keenan and colleagues' research that, poor comprehenders, when assessed by the PIAT and WJPC, have significantly worse working memory skills compared to poor comprehenders who are assessed by the QRIs and the GORT (Keenan et al., 2014; Keenan & Meenan, 2014). Keenan and colleagues note that format features of the tests determine working memory demands of different tests. In particular, they note that the PIAT requires that each single-sentence text be held in memory as children look at four pictures to determine which picture best represents the meaning of the sentence. Similarly, they note that the cloze format of the WJPC requires each passage blank be held in memory while reading the rest of the passage and considering various word choices to fill in the blank (Keenan et al., 2014; Keenan & Meenan, 2014). It is therefore possible that the inconsistent findings observed in the extant literature regarding working memory's unique predictive validity may have been partly due to differences in the reading comprehension tests across studies.

Differences across working memory measures may also contribute to inconsistent findings across studies. In particular, working memory measures range from simple span tasks where participants are only required to reproduce a sequence of items without any manipulation to complex span tasks where participants are required to store and manipulate the information. In a meta-analysis of 18 studies conducted by Carretti, Borella, Cornoldi, and De Beni (2009), they found that verbal complex span tasks or tasks involving executive functions are generally better than the simple span tasks at discriminating between good and poor reading comprehenders. It is thought that these tasks reflect the processes of reading where readers need to simultaneously decode words, retrieve their semantic content, maintain previously read text in memory while integrating information and anticipating where the story is going. Other differences across studies that may contribute to inconsistent results include group selection criteria and chronological ages.

In this study, we will use the same measures across the same sample of children so that we can determine whether vocabulary and working memory contribute unique variance to reading comprehension across comprehension skill levels after controlling for each other and decoding skills. Decoding skills are defined as nonword decoding rather than word recognition. From previous research, we know that vocabulary knowledge plays a significant role in decoding when decoding is defined as word recognition (e.g., Betjemann & Keenan, 2008; Ouellette, 2006; Plaut, McClelland, Seidenberg, & Patterson, 1996; Ricketts, Nation, & Bishop, 2007). In order to quantify vocabulary's unique contributions to reading comprehension, it is important to define decoding as nonword decoding skills. Thus, when decoding skills are defined this way, we expect them to play a smaller role in reading comprehension than when decoding skills are defined as word recognition. Additionally, based on the work of Tighe and Schatschneider (2014), we expect that vocabulary will increase its influence in reading comprehension as reading comprehension skills increase. However, it remains to be seen whether it is still the case after nonword decoding and working memory are entered in the quantile regression model. Finally, since our tests differ widely in how comprehension is assessed, i.e., cloze format, multiple-choice comprehension questions, and retells, it is likely that working memory's predictive validity may not be uniform across all tests and quantiles. If this is true, we hope that by using many different tests, we can identify factors that explain when working memory directly influences reading comprehension and when it does not.

Study 3: Do Vocabulary and Working Memory Contribute Consistently to Listening Comprehension across Listening Comprehension Quantiles?

There is an ongoing debate in the field regarding whether listening comprehension is essentially the same as reading comprehension once differences in word decoding are taken into account. Some researchers argue that reading and listening comprehension involve the same cognitive processes (e.g., Joshi, Williams, & Wood, 1998; Kintsch & Kozminksy, 1977; Townsend, Carrithers, & Bever, 1987). Others argue that the differences in the stimuli drive differences between reading and listening comprehension. The latter argument is consistent with the fact that although the correlations between listening and reading comprehension are usually quite high, i.e., from r = .71 - .90, (Joshi, Williams, & Wood, 1998; Keenan et al., 2014), they are still far from unity. Those who think that they are different argue that speech contains much richer prosodic cues to sentence structure and semantic content than is provided by punctuation in written texts (e.g., Braze, Tabor, Shankweiler, & Mencl, 2007; Goldman-Eisler, 1972) and these prosodic cues provide contextual support that becomes especially important in interpreting complex or ambiguous grammatical structures (Mann, Cowin, & Schoenheimer, 1989). Others think the nature of the input is more transient during listening than reading due to the fact that listeners cannot control the speech rate and cannot re-access the original information (e.g., Keenan et al., 2014).

To explore the nature of the correlation between reading and listening comprehension, we will use three tests where there are equivalent reading and listening versions: the Woodcock Johnson, QRI-Retells, and QRI-Questions so that the reading and the listening measures parallel each other as much as possible. Our main hypothesis is that while vocabulary's contributions to listening may be similar to reading, working memory's contributions to comprehension may differ between the two modalities. Specifically, if the print signal is indeed much more impoverished than the speech signal, then it is reasonable to expect working memory to play a particularly important role in reading comprehension. This is because readers are trying to decode words and retrieve the semantic content as quickly as they can while also trying to establish meaning and coherence across sentences and paragraphs without the sufficiently rich context. Speech, on the other hand, may contain rich prosodic cues such as pitch, stress and juncture that may make accessing lexical and semantic information relatively less burdensome (Goldman-Eisler, 1972). These advantages may in turn reduce working memory demands during listening comprehension. Alternatively, working memory may play a greater role in listening than in reading comprehension. As previously mentioned, the input in listening comprehension is much more transient than that during reading, and therefore, comprehenders may rely more on attention and relatedly, working memory, during listening than during reading (Keenan et al., 2014; Miller et al., 2013).

In summary, the current study asks the following specific questions: (1) do the relative contributions of word recognition and listening comprehension to reading comprehension vary across readers of different comprehension ability and across different comprehension tests; (2) do the components of listening comprehension, i.e., vocabulary and working memory, contribute uniquely to reading comprehension after controlling for each other and decoding skills and do they vary in their relative importance across reading comprehension quantiles; and (3) do these same components contribute similarly to listening comprehension across the quantiles as they contribute to reading comprehension.

Method

The current study uses quantile regression to investigate the relationships among reading component skills and reading comprehension across the comprehension quantiles and to see how they vary across different tests. We will use data from the Colorado Learning Disabilities Research Center (DeFries et al., 1997; Olson, 2006).

Across all studies, we will determine whether the relationships between the reading components (e.g., word recognition and listening comprehension) and comprehension differ across comprehension quantiles. We will report standardized regression coefficients, or correlations, that characterize these relations across quantiles. If the correlations between each component and reading comprehension are significantly greater than zero across all quantiles, then this means that both word recognition and listening comprehension are important for reading comprehension no matter what the child's reading achievement is. If the correlations are significantly greater than zero at some quantiles and not at others, and the difference in the correlation estimates is statistically significant, then this suggests that there are quantile differences in what constitutes reading comprehension. Finally, if the correlations between reading components and reading comprehension are significantly different between those at the 50^{th} quantile and those at the others, then this suggests that what is important for an average comprehender cannot be generalized to other readers. Once we have established in Study 1 that quantile regression is an improved method compared to OLS regression in detecting different relationships for different comprehension skill levels, we will move forward with presenting only quantile regression estimates in Study 2 and Study 3 for brevity purposes.

Participants

834 children ranging in age from 8 to 18 (*M* = 11.51, *SD* = 2.54) were selected from a sample of 1850 twins and their sibs tested on the language comprehension battery (Keenan et al., 2008; Keenan et al., 2006) of the Colorado Learning Disabilities Research Center (Olson, 2006). Because of the potential non-independence of the data when related individuals, such as twins and their siblings, constitute the sample, only one twin

from each pair was selected at random to be included in the present study's analyses. All were native English speakers.

Materials

The following include all the tests with the constructs that they are intended to measure. Their descriptive statistics, such as the range of scores and the skewness, are included in Table 1.

Table 1. *Range, Minimum and Maximum Scores, and Skewness for All Measures*

	N	Range	Minimum	Maximum	Skewness
PIAT	824	5.35	-3.07	2.28	-.44
WJPC	832	7.62	-4.12	3.51	-.09
GORT	827	5.40	-2.61	2.79	.25
QRI-R-Retells	818	6.32	-2.82	3.50	.27
QRI-R-Questions	804	5.86	-4.45	1.42	-.94
WJOC	833	8.74	-4.29	4.46	-.32
QRI-L-Retells	824	5.93	-2.59	3.34	.34
QRI-L-Questions	833	5.05	-3.49	1.57	-.83
Word Recognition	834	5.94	-3.58	2.36	-.41
Listening Comprehension	834	7.57	-4.58	2.99	-.84
Vocabulary	834	5.88	-3.35	2.53	-.15
Working Memory	834	5.96	-2.38	3.58	.47

Reading Comprehension Tests. The reading comprehension tests were the Woodcock Johnson Passage Comprehension (WJPC), in which children read silently short passages and provide a missing word to demonstrate their comprehension (Woodcock, McGrew, & Mather, 2001); the PIAT of Reading Comprehension (Dunn & Markwardt, 1970), which has participants read silently a sentence and choose which of four pictures expresses the meaning of the sentence; the QRI-R-Retells and QRI-R-Questions (Leslie & Caldwell, 2001), in which grade-level expository and narrative stories (250–785 words) are read aloud, and comprehension is assessed by the number of ideas recalled from an idea checklist provided for each passage in a passage retelling and short-answer comprehension questions, respectively; the GORT (Wiederholt & Bryant, 1992), in which expository and narrative passages (80–150 words) are read aloud, and multiple-choice comprehension questions for each passage are read to the child by the examiner. The PIAT and WJPC are known to load more highly on decoding than on listening comprehension whereas the GORT, QRI-R-Questions, and QRI-R-Retells are known to load more highly on listening comprehension than on decoding (Betjemann, et al., 2011; Keenan et al., 2014).

Listening Comprehension. A composite measure of listening comprehension was based on the following tests: Woodcock-Johnson Oral Comprehension (WJOC) subtest (Woodcock et al., 2001), in which children listen to short passages and provide a missing word to demonstrate their comprehension; QRI-L-Retells and QRI-L-Questions (Leslie & Caldwell, 2001), in which participants listen to narrative and expository passages, and then retell details from the passage and answer six comprehension questions from each passage; and the KNOW-IT Test (Barnes & Dennis, 1996; Barnes, Dennis, & Haefele-Kalvaitis, 1996), in which children are first taught a knowledge base relevant to the story that they will listen, then answer a series of literal and inferential comprehension questions about the story.

Word Decoding. A composite measure of word decoding was computed from the Timed Oral Reading of Single Words (Olson, Forsberg, Wise, & Rack, 1994) and the Peabody Individual Achievement Test Word Recognition subtest (Dunn & Markwardt, 1970). The Timed Oral Reading of Single Words assessed word recognition accuracy for a series of words that increased in difficulty on the computer screen. For a response to be scored as correct, it had to be initiated with two seconds. The Peabody Individual Achievement Test Word Recognition subtest assessed word recognition by having children read across rows of increasingly difficult unrelated words until they reach an error criterion. There is no time constraint on this task.

Nonword Decoding. Nonword decoding skill was a composite of two tests of nonword reading developed by Olson et al. (1994). One assessed reading 45 one-syllable nonwords (e.g., *ter, strale*). The other assessed reading of 40 two-syllable nonwords (e.g., *vogger, strempick*).

Vocabulary. Vocabulary knowledge was assessed by both the WISC Vocabulary subtest and Peabody Picture Vocabulary Test-III (PPVT-III). The WISC Vocabulary subtest assesses children's expressive vocabulary by having participants define orally presented words (Wechsler, 1974). The PPVT assesses children's receptive vocabulary by requiring participants to choose among four pictures to represent the meaning of spoken target words (Dunn & Dunn, 1997).

Working Memory. Working memory was measured using a composite of sentence span (Daneman & Carpenter, 1980), counting span (Case, Kurland, & Goldberg, 1982), and digit span using forward and backward digit span from the WISC-R or the WAIS-R. In the sentence span task, children generate a word at the end of a simple sentence that was presented orally (e.g., "I throw the ball up and then it comes …". Then, they had to repeat their generated words in blocks ranging from two to six sentence sets. In the counting span task, children count the number of yellow dots presented on a set of cards and then repeat, in order, the number of dots that appeared on each card. The sets vary in size from two cards to six cards per set. In the digit span task, children repeat multiple series of numbers either forward or backwards. The series begin with two numbers and continue to increase in length.

Reliabilities

A factor that could influence the correlations between measures is their reliability. Unfortunately, the procedures that publishers use to calculate reliability vary considerably and therefore, it is difficult to know how comparable these values are. To have more comparable reliabilities across the tests, we followed the example of Keenan and Meenan (2014) and exploited the twin feature of our sample and computed correlations between the monozygotic (MZ) twins as a proxy for test-retest reliability. The correlations between MZ twins can be considered an estimate of test-retest reliability because MZ twins share both their genes and family environment. They are a conservative estimate of reliability because even though MZ twins share genes and family environment, nonshared environmental influences may reduce the correlation in addition to the unreliability of the test; but because nonshared environmental influences on comprehension tend to be small (Keenan, et al., 2006), the MZ correlations are generally good proxies of test-retest reliability. They are shown in Table 2.

Table 2. *Monozygotic Correlations as Estimates of Test-Retest Reliability of Measures in Descending Order (N=268)*

Measures	MZ Correlations
Word Recognition Composite	.88
Nonword Composite	.85
Vocabulary Composite	.84
Listening Comprehension Composite	.79
WJPC Reading Comprehension Test	.75
PIAT Reading Comprehension Test	.73
WJOC Listening Comprehension Test	.69
Working Memory Composite	.66
QRI-L-Retells	.63
QRI-L-Questions	.62
QRI-R-Retells	.60
QRI-R-Questions	.55
GORT Reading Comprehension Test	.52

Data Analyses: Quantile Regression Analytic Approach[1]

Quantile regression is an extension of OLS regression. In OLS regression, the data are used to find a single regression line that minimizes the sum of squared errors. The best fitting line in OLS regression is the one that passes through the expected mean of the outcome distribution that is conditioned at every value of the predictor variable. Thus, the equation that represents a simple OLS regression model can be written as:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad (1)$$

where β_1 is the slope that represents the strength of the relationship between *x* and *y* and β_0 represents the intercept, i.e., the expected value of *y* when *x* is zero, and ε_i represents the error term. When the assumptions of linearity, homoscedasticity, and independence of *x* values are met, OLS regression provides the most parsimonious and unbiased

[1] This section was taken from Hua, A. N. & Keenan, J. M. (under review).

estimates of the population parameters. In quantile regression, the model can be written as:

$$y_i = \beta^{(\tau)} + \beta^{(\tau)} x + \varepsilon^{(\tau)} \tag{2}$$

where the only notational difference is the superscript "τ", which represents the quantile where the estimation occurs. Here, estimates of the intercept and slope coefficients that correspond to the specific quantiles are the ones that minimize the weighted sum of absolute errors. For example, in estimating the unique parameters for the relationships between x and y at the 50^{th} quantile of y, the resulting median-regression line must pass through a pair of data points with half of the data lying above the regression line and the other half lying below. That is, half of the errors are positive, and half are negative (Hao & Naiman, 2007). There are typically multiple lines with this property, and among these lines, the one that minimizes the absolute residuals between the observed and the predicted y value is the solution for that quantile. Similarly, for the solution at the 10^{th} quantile, 10% of the data points fall below the regression line and 90% fall above and the solution is the one that minimizes the absolute residuals between the observed and the predicted y value for that quantile. It is important to emphasize that in estimating the solutions, quantile regression uses all the data points in the sample by employing a weighting procedure where the weight is $1-\tau$ for points below the fitted line and τ for points above the line (Hao & Naiman, 2007). For example, in estimating the relationships between x and y for the 10^{th} quantile, the observations below the line are given a weight of .90 and the ones above the line receive a smaller weight of .10. As a result, 90% of the data points lie above the fitted line leading to positive residuals, and

10% lie below the line leading to negative residuals. To estimate the relationships for the 90th quantile regression, points below the line are given a weight of .10, and the rest have a much greater weight of .90 leading to 90% of the observations have negative residuals and the remaining 10% have positive residuals (Hao & Naiman, 2007). Lastly, in quantile regression, no assumptions are made about the distributional form of the error term. This critical feature allows quantile regression equations to be fitted to data with non-normal distributions. For a more thorough explanation of the estimation theory behind quantile regression, readers are encouraged to read seminal sources (e.g., Koenker, 2005; Hao & Naiman, 2007).

Quantile regression analyses are to be conducted using the R statistical software package (R Development Core Team, 2011). In all analyses, we chose to estimate the relationships among the variables at the 10th, 50th, and 90th quantiles. This decision is arbitrary, but it allows us to relate our findings to other studies that defined comprehension skills for poor, average and skilled readers as such. We will provide confidence intervals to characterize not only the statistical significance but also the precision of the estimates. Additionally, we formally test whether the correlations among our reading variables differ significantly across the quantiles by conducting Wald tests, which provide a χ^2 statistic with degrees of freedom, and p values for the test of quantile differences (Koenker & Bassett, 1982). Finally, because between-quantile coefficient testing may be considered an instance of multiple hypothesis testing (Petscher & Logan, 2013), we conducted Bonferroni corrections to minimize type I errors. Results are significant, therefore, if the p value is less than alpha (.05) divided by the number of quantiles tested (3), which is $p = .017$ (.05/3).

Preliminary Data Analyses: Age Regressions

Prior to all analyses, possible linear and nonlinear effects of age will be removed by regressing the outcome variables on age and age squared against the full sample. This allows us to assess the relations between reading components and comprehension regardless of the child's age.

Sample Size and Power

As with any inference-based method, issues with sample size and power must be considered when planning a study. Unfortunately, although a considerable body of research is available on power for traditional methods that focus on the means, much less is known about sample planning for quantile regression (Petscher et al., 2013). Thus, the general recommendations are not different from those provided for conditional means model: larger sample sizes are better than smaller. The current study's sample size of 834 is small in comparison to other studies that employed this method (Catts et al., 2009; Logan et al., 2012; Petscher & Kim, 2011), but larger in comparison to others (Petrill et al., 2014; Tighe & Schatschneider, 2014). However, the fact that quantile regression utilizes all the data points to generate quantile estimates suggests that the current sample size may be adequate.

Missing data was minimal for most of the individual variables, i.e., less than 2% of the full sample had missing data. The one exception was for the PPVT-III. This measure was not part of the testing battery until 2006, so roughly 40% of the full sample had missing data on this variable. However, because the current study uses composite vocabulary scores, and given that PPVT-III and WISC Vocabulary are highly correlated, $r = .76$, we decided to keep all of these cases as they all at least had data for WISC Vocabulary. Therefore, less than 1% of the full sample had missing data on the composite scores. This small amount of missing data is unlikely to bias our results.

Tests of Normality and Homoscedasticity

When assumptions of normality and homoscedasticity in multiple linear regression do not hold, quantile regression provides more accurate statistical estimates than OLS regression does (Hao & Naiman, 2007). Such was the case for some of our comprehension measures and regression models. For the reading comprehension tests, skewness statistics, as shown in Table 1, indicated more skew for the PIAT (-.44) and QRI-R-Questions (-.94) compared to the WJPC (-.09), GORT (.25), and QRI-R-Retells (.27). For the listening comprehension tests, skewness statistics indicated more skew for the QRI-L-Questions (-.82) than the QRI-L-Retells (.34) and WJOC (-.32). Results from Breusch-Pagan tests (Breusch & Pagan, 1979), which regressed the squared residuals on predictor variables, suggested heteroscedasticity for three regression models in Study 1(GORT: $BP = 23.60$, $p < .001$, $df = 2$; QRI-R-Retells: $BP = 10.50$, $p = .005$, $df = 2$; QRI-R-Questions: $BP = 36.42$, $p < .001$, $df = 2$). Similarly, in Study 2, Breusch-Pagan tests suggested heteroscedasticity for four regression models (PIAT: $BP = 7.96$, $p < .05$, $df = 3$; GORT: $BP = 12.53$, $p < .01$, $df = 3$; QRI-R-Retells: $BP = 15.87$, $p = .001$, $df = 3$; QRI-R-Questions: $BP = 65.73$, $p < .001$, $df = 3$). Finally, in Study 3, Breusch-Pagan tests suggested heteroscedasticity for two regression models (QRI-L-Questions: $BP = 59.97$, $p < .001$, $df = 2$; WJOC: $BP = 6.07$, $p < .05$, $df = 2$). These findings show that quantile regression analyses may be especially helpful in allowing us to capture nonlinear relationships among our variables of interest that the traditional OLS regression may miss entirely.

Chapter 3. Study 1:

How Does the Simple View of Reading Apply across Reading Comprehension Quantiles and Tests?[2]

OLS and Quantile Regressions of the PIAT Reading Comprehension Test

The first column of Table 3 shows the results from OLS multiple regression analyses. Similar to Keenan et al. (2008), we found that individual differences in word recognition explain considerably more of the variance in the PIAT Reading Comprehension test than individual differences in listening comprehension do; the correlation of PIAT comprehension with word recognition is r = .61 and with listening comprehension is r = .25. Results of quantile regression are displayed in columns 2 – 4 of Table 3. What quantile regression shows is that the OLS regression finding that word recognition explains significantly more variance than listening comprehension in the PIAT holds regardless of where the child is in the distribution. This can be seen from the fact that confidence intervals of the reading components' estimates do not overlap with each other. Word recognition and listening comprehension are both significantly predictive of PIAT comprehension across all comprehension quantiles, as evidenced by the size of the estimates (*rs* = .62 - .54 and *rs* = .35 - .22, respectively) and the confidence intervals not including zeros. There are no statistically significant differences in the

[2] This section was taken from Hua, A. N. & Keenan, J. M. (under review).

correlations between reading components and reading comprehension across quantiles (*p*s > .017), although there is a small decline in the correlation sizes as reading comprehension increases. Additionally, the confidence intervals of quantile regression largely overlap with those of OLS regression. In sum, what we learn from quantile regressions is that the OLS regression estimates of the influence of word recognition and listening skills in comprehension apply similarly across the PIAT distribution.

OLS and Quantile Regressions of the WJPC Reading Test

The results for the WJPC are quite similar to those for the PIAT. Examining the first column of Table 3 with the OLS multiple regression results, we see that as in Keenan et al. (2008), word recognition explains more variance (r = .56) in the WJPC test than listening comprehension does (r = .33). Furthermore, quantile regression shows that word recognition explains significantly more variance than listening comprehension does in the WJPC across all the quantiles, as evidenced by the confidence intervals of the reading components' estimates not overlapping with each other. Word recognition and listening comprehension are highly predictive of WJPC scores across all quantiles, as evidenced by the estimates (*rs* = .53 - .61 and *rs* = .32 - .33, respectively) and the confidence intervals not including zeros. Moreover, there are no quantile differences in the correlations between the reading components and reading comprehension (*ps* > .017). Thus, for both the PIAT and the WJPC, the similarity between the quantile and OLS regression estimates, as well as their largely overlapping confidence intervals, show that the quantile regression estimates do not differ from those of OLS regression.

OLS and Quantile Regressions of the GORT Reading Test

In contrast to what was found for the PIAT and WJPC, OLS multiple regression analysis for the GORT shows that word recognition explains *less* variance ($r = .23$) than listening comprehension does ($r = .41$), similar again to what Keenan et al. (2008) found. But as can be seen from comparing columns 2 – 4, quantile regression analyses show that there are differences across quantiles in how much variance is explained by word recognition. The correlation estimate between word recognition and GORT performance at the 10th quantile is significantly smaller than that at the 50th quantile ($F(1, 1653) = 18.13$, $p < .001$) and at the 90th quantile ($F(1, 1653) = 9.04$, $p = .003$). No significant quantile differences were observed between the 50th and 90th quantiles, $F(1, 1653) < 1$. As we will discuss later, this finding that word recognition contributes little to performance for those in the low-tail, but instead is largely explained by listening skills, is consistent with the finding that some of the multiple-choice questions of the GORT can be correctly guessed without even reading by using background knowledge (Keenan & Betjemann, 2006). Listening comprehension explains significantly more variance in the GORT than word recognition does across all quantiles, as evidenced by the size of the correlation estimates as well as their largely non-overlapping confidence intervals, but it explains more variance at the 90th quantile than that at the 10th quantile, $F(1, 1653) = 6.64$, $p = .01$. In short, quantile regression shows that for the GORT, the factors that explain reading comprehension depend upon the reader's performance level, and that OLS regression analysis greatly overestimates the relationship between word recognition and comprehension for poor performers.

OLS and Quantile Regressions of the QRI-R-Retells Test

OLS regression shows that, like Keenan et al. (2008), listening comprehension explains significantly more variance (r = .46) than word recognition does (r = .19) on QRI-R-Retells. Quantile regression shows that word recognition and listening comprehension account for significant variance across all quantiles (*rs* = .13 - .28 and *rs* = .37 - .48, respectively). However, whereas the correlations between word recognition and QRI-R-Retells do not differ significantly across quantiles (*ps* > .017), the correlation between listening comprehension and QRI-R-Retells at the 10^{th} quantile is significantly smaller than that at the 50^{th} quantile, $F(1, 1635) = 6.92$, $p = .01$. No other quantile differences were observed, *ps* > .017. In short, these results show again that what explains reading comprehension can depend on the reader's level. Although OLS regression estimates are very similar to those of quantile regression for the role of word recognition in QRI-R-Retells, they overestimate the relationship between listening comprehension skills and reading comprehension performance at the 10^{th} quantile.

OLS and Quantile Regressions of the QRI-R-Questions Reading Test

OLS multiple regression analysis shows, like Keenan et al. (2008), that word recognition explains less variance (r = .17) in performance on the QRI-R-Questions than listening comprehension does (r = .52). Quantile regressions show, however, that what holds on average does not hold at the extremes. Word recognition is important for reading comprehension, but only for readers at the 10^{th} and 50^{th} quantile (*rs* = .31 - .12). At the 90^{th} quantile, its correlation with reading comprehension does not differ significantly from zero (see Table 3). Listening comprehension, on the other hand, is highly predictive of children's QRI-R-Questions performance regardless of the child's level, and explains significantly more variance than word recognition does. The comparison across quantiles shows a number of quantile differences. For word recognition, the correlation with QRI-R-Questions is significantly higher at the lower quantiles than at the higher quantiles: between the10th and 50th quantiles, $F(1, 1653) = 8.47$, $p = .003$; between the 50th and the 90th quantiles, $F(1, 1653) = 8.33$, $p = .004$; between the 10th and the 90th quantiles, $F(1, 1653) = 18.29$, $p < .001$. For listening comprehension, the correlation with QRI-R-Questions is significantly smaller at the 90th quantile than at the 50th quantile, $F(1, 1653) = 12.87$, $p < .001$. In sum, OLS regression appears to underestimate the contribution of word recognition to comprehension at the 10th quantile and overestimates both word recognition and listening comprehension's contributions to comprehension at the 90th quantile.

Table 3. *OLS and Quantile Regression Estimates with 95% Confidence Intervals (CIs) for Each Reading Test*

	OLS (95% CI)	Quantile Regression (95% CI)		
		10th	50th	90th
PIAT				
Word Rec	**.61** (.56, .66)	**.62** (.55, .68)	**.63** (.59, .69)	**.54** (.46, .62)
Listen Comp	**.25** (.20, .29)	**.35** (.23, .38)	**.22** (.19, .26)	**.22** (.17, .28)
WJPC				
Word Rec	**.56** (.51, .61)	**.53** (.49, .61)	**.53** (.47, .57)	**.61** (.55, .66)
Listen Comp	**.33** (.29, .38)	**.32** (.25, .37)	**.31** (.29, .38)	**.33** (.27, .37)
GORT				
Word Rec	**.23** (.16, .29)	**.07** (.01, .12)	**.30** (.24, .37)	**.29** (.17, .34)
Listen Comp	**.41** (.35, .47)	.**28** (.23, .44)	**.38** (.33, .44)	**.45** (.36, .51)
QRI-R-Retells				
Word Rec	**.19** (.12, .26)	**.13** (.05, .22)	**.15** (.10, .24)	**.28** (.15, .40)
Listen Comp	**.46** (.40, .52)	**.37** (.29, .49)	**.49** (.41, .55)	**.48** (.40, .52)
QRI-R-Questions				
Word Rec	**.17** (.10, .23)	**.31** (.20, .40)	**.12** (.05, .19)	**-.00** (-.03, .06)
Listen Comp	**.52** (.46, .58)	**.53** (.49, .63)	**.56** (.47, .62)	**.41** (.36, .45)

Measurement Issues and Their Impact on Current Findings

Because some of the correlations in the present results are extremely low, we checked to see if they might reflect lack of variance due to floor and ceiling effects, as opposed to reflecting a true lack of relationship. To determine whether the correlation of $r = .07$ between word recognition and GORT at the 10th quantile is partly due to the lack of variance at the 10th quantile, we examined the scatter plot between word recognition and GORT comprehension scores for those scoring at the 10thquantile. The scatter plot revealed that word recognition scores ranged roughly from a *z* score of -3 to a *z* score of +2; there was similar spread of GORT comprehension scores. Therefore, the low correlation between word recognition and GORT comprehension scores at the 10th quantile cannot be due to the lack of variance in these variables.

We followed the same procedure to determine whether the correlation of $r = .00$ between QRI-R-Questions and word recognition at the 90th quantile could also be due to lack of variance. Inspection of the distribution of the scores of the 90th quantile readers showed that while word recognition scores range widely for these 90th quantile readers from -2 to +2, the majority of the reading comprehension scores on the QRI-R-Questions cluster tightly around the ceiling. This information, together with the skewness value of -.93, demonstrated that the reason that there is no relationship found between word recognition and QRI-R-Questions performance at the 90th quantile is due to a ceiling effect.

Between Test Differences

An interesting question is whether previously reported differences between tests (Keenan, et al., 2008) are apparent at the extreme ends of the reading comprehension distribution. We examined confidence intervals of the correlation estimates provided in Table 3 to assess statistical significance between the tests at the 10th quantile and at the 90th quantile. We found that previously reported OLS regression findings, namely that comprehension is explained more by word recognition in the PIAT and in the WJPC than in the other tests, applies similarly to poor and skilled readers. This is shown by greater correlations between word recognition and comprehension for the PIAT and WJPC when compared to the remaining tests and their non-overlapping confidence intervals. In addition, at the 90th quantile, the GORT, QRI-R-Retells, and QRI-R-Questions assess listening comprehension to a much greater extent than the PIAT and WJPC do. Interestingly, however, at the 10th quantile, the only clear difference is that QRI-R-Questions assesses listening comprehension significantly more than the PIAT and WJPC. Altogether, these findings provide further support for between-test differences.

Discussion

The OLS regression results of the present study replicated test differences that have previously been reported for these reading comprehension tests (Keenan et al., 2008). However, the quantile regression findings from the present study show that a second factor that needs to be considered when interpreting test scores is the child's level of performance on the test. We found differences across performance levels in what explains comprehension for three of the five tests that we examined. Most of the quantile differences occurred for poor readers relative to the others. What that means for examiners is that using OLS regression studies to interpret why poor comprehenders score poorly can sometimes be misleading.

One example of this is with the GORT. By directly comparing OLS and quantile regressions, we learned that for poor readers, OLS regression (.23) overestimates how much word recognition explains comprehension for those performing in the 10th quantile (.07). This is a particularly important finding because it may seem counterintuitive for those who administer the test. Because the GORT requires the child to read aloud, and because it is readily apparent to an examiner when a child is struggling to decode, one is likely to think that word decoding explains comprehension scores on the GORT. However, the present study shows that is not the case for the poorest performers. Even though we found that there is considerable variance in word recognition skills among those who score at the 10th quantile, it accounts for very little of their performance.

Another test in which there is a discrepancy between OLS and quantile regressions for poor readers is QRI-R-Questions; word recognition is more important in explaining poor comprehenders' scores than typical or high performers' scores. OLS regression also overestimates the contribution of poor readers' listening comprehension skills to performance when comprehension is assessed by QRI-R-Retells. In short, the insight from this work is that researchers and clinicians should not assume that whatever has been previously found for a typical reader with a particular test will generally apply to all readers and to all tests.

Quantile differences were found for some tests (GORT, QRI-R-Questions, QRI-R-Retells) but not others (PIAT, WJPC). Why is that? The tests showing quantile differences are the tests that have more of their variance explained by listening comprehension skills than by word recognition skills, whereas the tests that do not show quantile differences are the tests that have more of their variance explained by word recognition. Is that a coincidence, or might there be reasons underlying this pattern of quantile differences? To answer this question, we will next deconstruct word recognition and listening comprehension into their more basic components of nonword decoding skills, vocabulary and working memory, to see if they can explain the quantile differences observed in this study.

Chapter 4. Study 2:

Do Vocabulary and Working Memory Contribute Consistently to Reading Comprehension across Quantiles?

In Study 1, we demonstrated that, depending on the test, quantile regression can be superior to OLS regression in capturing the relationships between component skills and reading comprehension. In this study, we deconstruct listening comprehension into vocabulary and working memory to assess the extent to which these component skills contribute unique variance to reading comprehension beyond decoding skills. Here, decoding skills are defined as nonword decoding rather than word recognition as in Study 1. As previously mentioned, this is so that we could remove any variance decoding skills may share with vocabulary. The question we are interested in here is whether each of the subskills (vocabulary, working memory) shows the same pattern across quantiles as the other. We will enter vocabulary and working memory along with nonword decoding as predictors and the five individual reading measures as the dependent variables in our quantile regression models.

Results

PIAT Reading Comprehension Test

As expected, when decoding is defined as nonword decoding rather than word recognition, its unique contributions to PIAT reading comprehension decline substantially, *rs* = .29 - .40. As can be seen in Table 4, the contributions of vocabulary (*rs* = .35 - .48), and working memory (*rs* = .10 - .22) to PIAT comprehension scores are each statistically significant and substantial across all quantiles. Interestingly, vocabulary is significantly more predictive than working memory across all quantiles. Moreover, its correlation with PIAT scores is significantly weaker at the 90th quantile than at the 50th quantile, $F(1, 1647) = 9.76$, $p = .002$. For working memory, no quantile differences were observed between working memory and PIAT reading comprehension, *ps* > .017.

WJPC Reading Comprehension Test

As was found for the PIAT, the unique contributions of nonword decoding (*rs* = .28 - .40) to WJPC Reading Comprehension shown in Table 4 also decline substantially compared to those of word recognition found in Study 1. Moreover, the contributions of vocabulary (*rs* = .39 - .46) and working memory (*rs* = .15 - .26) to WJPC comprehension are each statistically significant across all quantiles. Vocabulary is significantly more predictive than working memory across all quantiles. Finally, no quantile differences were observed between vocabulary and working memory and WJPC reading scores, *ps* > .017.

GORT Reading Comprehension Test

The most striking finding in the analyses of the GORT was that when we defined decoding by nonword decoding, its contributions to GORT reading comprehension scores were not significantly different from zero across all quantiles (Table 4). Working memory's patterns of correlations with reading comprehension were also minimal and statistically significant only for the 50th quantile readers, $r = .09$. Multiple quantile regression revealed that vocabulary is the only reliable predictor of reading comprehension across all comprehension levels, *rs* = .36 - .60, after controlling for working memory and nonword decoding. Additionally, vocabulary is significantly less correlated with GORT reading comprehension at the lower quantiles than at the higher quantiles; between the 10^{th} and 50^{th} quantiles: $F(1, 1653) = 22.62$, $p < .001$ and between the 10^{th} and 90^{th} quantiles: $F(1, 1653) = 13.11$, $p < .001$, but between the 50^{th} and 90^{th} quantiles: $p > .017$.

QRI-R-Retells Comprehension Test

As shown in Table 4, the unique contributions of nonword decoding (*rs* = .08 - .23) to QRI-R-Retells are statistically significant for the 10^{th} and 50^{th} quantiles, and are generally smaller than those of word recognition. Vocabulary, as found for the other three tests, also contributes substantially to reading comprehension, *rs* = .18 - .48. Like the GORT, working memory's contribution to reading comprehension is statistically significant only at the 50^{th} quantile, $r = .07$. Additionally, vocabulary is less predictive with QRI-R-Retells at the lower quantiles than at the higher quantiles: between 10^{th} and 50^{th} quantiles: $F(1, 1653) = 12.26$, $p < .001$ and between the 10^{th} and 90^{th} quantiles: $F(1, 1653) = 12.36$, $p < .001$. No significant differences were found between the 50^{th} and 90^{th} quantiles: $p > .017$. No quantile differences were observed between working memory and QRI-R-Retells reading comprehension scores, $ps > .017$.

QRI-R-Questions Comprehension Test

The unique contributions of nonword to QRI-R-Questions are statistically significant only at the 10^{th} and 50^{th} quantiles, *rs* = .15 and .06, respectively. These contributions are generally smaller in size compared to those of word recognition in

Study 1, although the confidence intervals of these estimates overlap to some extent. Vocabulary, like the other tests, is a reliable and robust predictor a

cross reading comprehension quantiles, *rs* = .30 - .62. Working memory's contribution to reading comprehension is not significantly different from zero across quantiles. Additionally, the patterns of quantile differences between vocabulary and QRI-R-Questions are the opposite of the GORT and QRI-R-Retells in that vocabulary and QRI-Questions are less correlated at the higher quantiles than at the lower quantiles: between the 90th and 50th quantiles: $F(1, 1653) = 8.86$, $p = .003$ and between the 90th and 10th quantiles: $F(1, 1653) = 11.84$, $p < .001$. This is due to a ceiling effect of the QRI-R-Questions, as previously demonstrated in Study 1. No significant quantile differences were found between the 10th and 50th quantiles, $p > .017$.

Table 4. *Nonword Decoding, Vocabulary, and Working Memory's Contributions to Reading Comprehension Tests*

	Quantile Regression (95% CI)		
	10th	50th	90th
PIAT			
Decoding	**.39** (.25, .51)	**.40** (.33, .45)	**.29** (.22, .35)
Vocabulary	**.40** (.32, .53)	**.48** (.43, .52)	**.35** (.31, .42)
Working Memory	**.22** (.05, .25)	**.10** (.05, .17)	**.14** (.11, .17)
WJPC			
Decoding	**.40** (.27, .51)	**.28** (.23, .33)	**.34** (.21, .37)
Vocabulary	**.39** (.34, .45)	**.46** (.40, .50)	**.42** (.35, .53)
Working Memory	**.16** (.10, .20)	**.15** (.11, .23)	**.26** (.14, .35)
GORT			
Decoding	**.04** (-.03, .09)	**.00** (-.06, .09)	**.00** (-.08, .06)
Vocabulary	**.36** (.28, .46)	**.56** (.49, .63)	**.60** (.53, .73)
Working Memory	**.04** (-.01, .14)	**.09** (.04, .17)	**.07** (-.03, .11)
QRI-R-Retells			
Decoding	**.23** (.11, .35)	**.16** (.05, .26)	**.08** (.00, .29)
Vocabulary	**.18** (.10, .28)	**.40** (.27, .48)	**.48** (.34, .59)
Working Memory	**.11** (-.04, .19)	**.07** (.03, .15)	**.10** (-.07, .23)
QRI-R-Questions			
Decoding	**.15** (.08, .38)	**.06** (.01, .15)	**-.06** (-.08, .05)
Vocabulary	**.62** (.50, .71)	**.53** (.46, .58)	**.30** (.14, .40)
Working Memory	**-.02** (-.11, .06)	**.01** (-.06, .07)	**.04** (-.05, .10)

Between Test Differences

Vocabulary is strongly predictive of reading comprehension across all tests and comprehension levels. This finding adds to the extant literature on the prominent role vocabulary plays in reading comprehension. Nonword decoding skills, on the other hand, are highly predictive of all reading comprehension levels for the PIAT and WJPC, but much less so for the remaining tests. Similarly, working memory is predictive of all reading comprehension levels only for the PIAT and WJPC, but its predictive validity in reading comprehension is not as prominent as that of nonword decoding and of vocabulary skills.

Discussion

The purpose of Study 2 was to deconstruct listening comprehension into its more basic components, i.e., vocabulary and working memory, to see whether they contribute unique variance to reading comprehension after controlling for each other and nonword decoding skills, and to see if the contributions vary between poor and skilled readers. What we found is that vocabulary is a highly influential predictor in reading comprehension, across all reading comprehension skills and all reading comprehension tests. Working memory, on the other hand, was found to be a reliable predictor for all readers only on the PIAT and WJPC.

Based on the work of Tighe and Schatschneider (2014), we had expected that vocabulary would play a significant role in reading comprehension but that its role would increase as reading comprehension performance increases. We found this to be the case only for the GORT and QRI-R-Retells. For the WJPC, its contributions remain relatively constant across readers. For the PIAT, we found the opposite pattern where its contributions are significantly less predictive at the 90th quantile than at the 50th quantile. The PIAT is designed so that as one proceeds through to more difficult texts, the vocabulary items become more multisyllabic and less familiar. Therefore, depending on the magnitude of the increase in the variance explained by vocabulary complexity, it is likely that its influence would either stay constant or increase. However, a small ceiling effect on this measure, suggested by the skewness of -.44, and the restricted range of scores at the higher end of the PIAT distribution, as shown in Table 1, may have resulted in a decrease in the correlation between vocabulary and the PIAT at the 90th quantile. This interpretation is corroborated by the fact that declining correlations at the 90th

quantile occurred for nonword decoding and working memory as well. A ceiling effect on the QRI-R-Questions may have also contributed to the decreasing role vocabulary plays in reading comprehension as reading comprehension performance increases.

We also found that working memory does not always play a direct role in reading comprehension. It is a reliable predictor for the PIAT and WJPC, but even then, its contributions to reading comprehension tend to be significantly smaller than vocabulary's. It is not a reliable predictor for the GORT and QRI-R-Retells. It significantly predicts reading comprehension beyond the influence of vocabulary and nonword decoding only for the 50^{th} quantile readers; but as previously reported, its effect sizes are extremely small, $r = .09$ and $r = .07$, respectively. For the QRI-R-Questions, working memory does not contribute any unique variance once vocabulary and nonword decoding skills are statistically controlled. Altogether, it is clear that vocabulary is a very important component of listening comprehension influencing reading comprehension for all readers and tests, whereas working memory is an important component of listening comprehension only for some tests. Next, we will determine whether what we found for reading comprehension tests in this study are also applicable to listening comprehension tests.

Chapter 5. Study 3:

Do Vocabulary and Working Memory Contribute Consistently to Listening Comprehension across Listening Comprehension Quantiles?

In this study, we will determine whether vocabulary and working memory contribute unique variance to listening comprehension tests, i.e., the WJOC, QRI-L-Retells, and QRI-L-Questions, across listening comprehension quantiles. We are particularly interested in knowing whether the patterns of correlations for the listening tests replicate those found for the reading version of the same tests. In these analyses, we will enter decoding skills (defined as nonword reading as in Study 2), vocabulary, and working memory as predictors and the listening comprehension tests as dependent variables in the quantile regression models. Even though the act of mapping letter strings to their corresponding sounds in nonword decoding is generally considered specific to reading comprehension and not to listening comprehension, we know from our own data that the zero-order correlations between our nonword measures and listening comprehension tests are substantially greater than zero. In particular, their correlation with WJOC is $r = .33$; with QRI-L-RC is $r = .26$; and with QRI-L-Q is $r = .16$. Thus, to ensure that our results are not confounded with decoding skills' effects and to ensure that our analyses are comparable to those in Study 2, we included nonword decoding along with vocabulary and working memory in our quantile regression models in order to assess the unique contributions of each skill to listening comprehension and compare them to the results as found in Study 2.

Correlations between the Reading and the Listening Comprehension Tests

It is important to determine the correlations between reading and listening tests before investigating the nature of their relationships. The WJPC and WJOC are moderately correlated at $r = .61$. Similarly, the QRI-R-Retells and QRI-L-Retells are correlated at $r = .60$. Finally, the QRI-R-Questions and QRI-L-Questions are correlated at $r = .48$. The far from perfect correlations between the reading and listening measures partly reflect the moderate reliability of the measures (MZ correlations = .55 - .75). What is equally likely is that these tests may assess different skills. For example, reading comprehension tests require children to map orthography with phonology in order for them to sound out the words, whereas listening comprehension tests do not. Moreover, as previously discussed, memory may be differentially involved across the two modalities. In order to understand these correlations better, we will present the quantile regression results for all listening comprehension tests and compare them to their reading counterparts.

WJOC Listening Comprehension Test

When nonword decoding skills are entered along with vocabulary and working memory, their unique predictive validity of WJOC scores is substantially reduced to zeros across quantiles (see Table 5). Vocabulary is significantly predictive of WJOC across all quantiles, *rs* = .58 - .66. It is interesting to note that vocabulary plays a significantly more influential role in the listening test than in the reading test, *rs* = .39 - .46. Moreover, vocabulary's contributions to the listening test are significantly greater compared to working memory's contributions. While working memory was found in Study 2 to be predictive of the WJPC across all quantiles, working memory is significantly predictive of WJOC scores only for the 10th and 50th quantiles, $r = .11$ and $r = .08$, respectively (see Table 5). Finally, just as we observed for the WJPC, there are no quantile differences between the listening comprehension components and WJOC scores, $ps > .017$.

QRI-L-Retells Comprehension Test

As shown in Table 5, nonword decoding skills are only significantly predictive of QRI-L-Retells at the 10th quantile, but the size of its predictive validity is small, $r = .08$. Vocabulary, on the other hand, is quite influential in predicting comprehension for all readers, $rs = .32 - .46$. Working memory, on the other hand, is not significantly predictive of QRI-L-Retells once vocabulary and nonword decoding skills were statistically controlled (Table 5). This contrasts slightly to what was found for the QRI-R-Retells, where working memory is significantly predictive of comprehension for readers at the 50th quantile. No quantile differences were observed between the listening comprehension components and QRI-L-Retells, $ps > .017$, which again is slightly different than QRI-R-Retells, where the correlation between vocabulary and reading comprehension at the10th quantile is significantly smaller than those at the 50th and 90th quantiles.

QRI-L-Questions Comprehension Test

As found for the WJOC and QRI-L-Retells Comprehension Tests, nonword decoding skills are not uniquely predictive of QRI-L-Questions across quantiles. Vocabulary, on the other hand, is significantly predictive for all readers, $rs = .09 - .69$

whereas working memory is not, once vocabulary and decoding skills were statistically controlled (Table 5). This was also the case for the QRI-R-Questions. Additionally, the unique correlation between vocabulary and QRI-L-Questions is significantly higher at the lower quantiles than at the higher quantiles: between 10th vs. 50 quantiles, $F(1, 1665) = 18.30$, $p < .001$; between 50th vs. 90th quantiles, $F(1, 1665) = 52.34$, $p < .001$; between 10th vs. 90th quantiles, $F(1, 1665) = 88.02$, $p < .001$. This pattern of quantile difference was also found for the QRI-R-Questions, and is likely due to a ceiling effect of the listening measure, as suggested by the skewness value of -.82 shown in Table 1.

Table 5. *Nonword Decoding, Vocabulary, and Working Memory's Contributions to Listening Comprehension Tests*

	Quantile Regression (95% CI)		
	10th	50th	90th
WJOC			
Nonword	**-.03** (-.09, .04)	**-.00** (-.04, .11)	**-.01** (-.11, .06)
Vocabulary	**.66** (.59, .70)	**.66** (.60, .70)	**.58** (.49, .73)
Working Memory	**.11** (.06, .19)	**.08** (.01, .14)	**.08** (-.05, .15)
QRI-L-Retells			
Nonword	**.08** (.01, .11)	**.04** (-.00, .11)	**.11** (.00, .23)
Vocabulary	**.32** (.23, .40)	**.38** (.28, .47)	**.46** (.34, .61)
Working Memory	**.03** (-.02, .14)	**.09** (-.02, .14)	**-.04** (-.20, .10)
QRI-L-Questions			
Nonword	**-.13** (-.24, .03)	**-.07** (-.16, .08)	**-.00** (-.04, .02)
Vocabulary	**.69** (.60, .81)	**.43** (.35, .54)	**.09** (.05, .13)
Working Memory	**.01** (-.13, .11)	**-.01** (-.11, .04)	**-.05** (-.06, .01)

Between Listening Comprehension Test Differences

Across all listening comprehension tests, we learned that vocabulary plays a prominent role for readers of varying performance skills. It is interesting to note, however, that vocabulary plays a much more influential role for all readers in the WJOC than in the remaining tests. Working memory and decoding skills, on the other hand, generally play a trivial role across all listening comprehension tests. Working memory is only significantly predictive of the WJOC, but only at the 10th and 50th quantiles. Nonword decoding skills are generally not uniquely predictive of any listening comprehension test, with the exception of a statistically significant but small correlation with QRI-L-Retells at the 10th quantile, $r = .08$. Thus, decoding skills' trivial contributions suggest that including or excluding them in the regression models would not make a difference in the over all results of the study.

Discussion

The purpose of Study 3 was to assess the relative contributions of vocabulary and working memory to listening comprehension tests in order to determine whether the results found for listening comprehension tests replicate those for reading comprehension tests. We found that there are similarities as well as differences between reading and listening tests. On the one hand, we learned that, regardless of modality, vocabulary plays a very influential role in comprehension for all readers whereas working memory generally plays a much smaller role than vocabulary does. This finding demonstrates that these processes are not specific to reading comprehension but rather applicable to general comprehension. However, what is also interesting is that working memory tends to be more influential in reading comprehension than in listening comprehension and this is particularly true for the 90th quantile readers on the Woodcock Johnson Test.

Our finding that working memory is generally more influential in reading comprehension than it is in listening comprehension suggests there may be nuanced differences between the two modalities of comprehension. When reading a passage, either silently to oneself as during the WJPC Reading Comprehension Test or aloud to the testers as during the QRI, all children need to decode, sound out the words, and retrieve their semantic content as quickly as they can while trying to establish textual coherence. The ability to coordinate all of these mental activities likely taps into working memory resources to a greater extent when compared to listening comprehension, where the demands for decoding and sounding out the words are removed. Another interpretation of our finding is that because the print signal lacks the prosodic cues inherent in speech that are thought to provide contextual support in comprehension, individual differences in working memory matter more for reading comprehension. However, because we did not directly investigate the quality of the print versus the speech signal, our data cannot directly support this interpretation. Finally, even though we used the listening and the reading versions of the same tests in order to match them as much as possible, one cannot entirely rule out the possibility that content differences may have driven these results. However, we think the possibility is unlikely since working memory is generally more predictive of reading comprehension than of listening comprehension across tests.

Our results did not clearly indicate one way or the other whether vocabulary plays a differential role in listening than in reading. This is because vocabulary was found to be significantly more predictive of WJOC than of WJPC, but it was found to be significantly more predictive of QRI-R-Questions than of QRI-L-Questions at the 90^{th} quantile. However, what is clear across all of our results is that vocabulary plays a very prominent role in comprehension. Thus, a poor or a skilled performer on these tests most likely has poor or skilled vocabulary knowledge, respectively.

Finally, it is interesting to see that the zero-order correlations between nonword decoding skills and the listening comprehension tests are not insubstantial—they range from .16 to .33, and these correlations go away when vocabulary and working memory are also included as predictors in the models, demonstrating the meaningful variance decoding skills share with vocabulary and working memory. Relatedly, it suggests that vocabulary and working memory, albeit to a much smaller extent, mediate the relationship between phonological skills and listening comprehension. This finding is consistent with experimental studies that found that children who are poor at phonological skills are also poor at learning new words, a critical process in acquiring comprehension skills (Gathercole & Baddely, 1990; Gathercole, Hitch, Service & Martin, 1997) and those who have language impairments are poor at phonological skills (Graf Estes, Evans & Else-Quest, 2007).

In sum, we have learned that listening and reading comprehension largely depend on similar language processes, such as vocabulary, but working memory seems to play a greater role in reading than in listening. Next, we will discuss what all of our findings mean and the implications they have for research, diagnosis, and instruction.

Chapter 6

General Discussion

The goal of this dissertation was to use quantile regression to advance our theoretical understanding of the relative importance of different component processes of comprehension across comprehension performance levels. In order to accomplish this goal, we conducted three studies. In Study 1, we examined how the relative importance of the components of the Simple View of Reading, word recognition and listening comprehension, might vary across reading comprehension quantiles and tests. In Study 2, we broke down the component of listening comprehension into vocabulary and working memory to determine whether there are differences across reading comprehension quantiles in how these components contribute to reading comprehension. Finally, in Study 3, we examined the relative contributions of vocabulary and working memory to listening comprehension across listening comprehension quantiles.

One important finding of this dissertation is that quantile regression can be superior to OLS regression because we found that for some tests, the relationships between reading components and reading comprehension are different for readers at the extreme ends of the distribution. Across all three studies, there were instances in which OLS regression was found to either underestimate or overestimate the relationships between the reading components and comprehension. Thus, quantile regression appears to be a powerful method that allows us to ascertain information specific to the poor, average, and skilled readers without truncating the full sample into specific skill groups and thereby avoiding many methodological issues.

Another important finding of this dissertation is that tests differences have consequences for whether quantile differences are observed and the nature of these quantile differences. We have learned in recent years that interpreting the findings from these tests can be challenging because tests can differ dramatically in the extent to which they assess the component skills of reading comprehension (Cutting & Scarborough, 2006; Keenan, Betjemann, & Olson, 2008; Nation & Snowling, 1997). Some tests have most of their variance explained by word recognition skills and very little explained by oral language or listening comprehension skills. As a result, a child with weak word reading skills, but good listening comprehension skills, will appear to be a poor comprehender on tests like the PIAT and the WJPC (where the variance is largely explained by word reading), but look like a good comprehender when examined with a test like the QRI or the GORT (where longer passages provide context to facilitate word recognition). We now have learned that tests can also make a difference in whether quantile differences are observed or not. In Study 1, quantile differences were observed for the GORT, QRI-R-Retells, and QRI-R-Questions but not for the PIAT and WJPC. In order to understand the reasons underlying this pattern of quantile differences, we will first consider why quantile differences in the role of word recognition in comprehension tests occur; then we'll consider why quantile differences in listening comprehension occur.

Word recognition is a skill that is influenced by both decoding skills and vocabulary knowledge (Betjemann & Keenan, 2008; Ouellette, 2006; Plaut, McClelland, Seidenberg, & Patterson, 1996; Ricketts, Nation, & Bishop, 2007). Since decoding skills are generally considered to be a set of constrained skills that get automated within a brief developmental span, their predictive validity would likely decrease as reading comprehension performance increases. Vocabulary, on the other hand, is a skill that continues to develop across an entire lifespan and its predictive validity would likely increase as reading comprehension increases. Therefore, it is not clear whether word recognition's contributions to reading comprehension would vary at higher comprehension levels.

We think the reason the PIAT and WJPC may show no statistically significant differences in how much word recognition contributes to comprehension across quantiles is that the tests are designed so that as one proceeds through to more difficult texts, the words become more morphologically complex, multisyllabic, and less familiar. Thus, even though there is a small decline in the correlations at the 90th quantile, variability in phonological decoding and vocabulary knowledge is maintained relative to the other tests such as the QRI-R-Questions where word recognition plateaus at higher performance levels.

The same explanation applies to the GORT, which also uses less frequent and more complex vocabulary in higher-level passages. The GORT shows near equivalent correlations of word recognition and comprehension for the 50th (r = .30) and 90th (r = .29) quantiles. However, there is a deviation from this pattern of no quantile differences at the 10th quantile on the GORT, where word recognition accounts for almost no variance in comprehension. We think that finding likely reflects an uneven distribution of passage-independent items (items that can be guessed correctly without reading the passages). More passage-independent items were among the lower level passages than among the higher passage levels (Keenan & Betjemann, 2006), allowing answers based on guesses from prior knowledge (listening comprehension skills) rather than from decoding the text. In sum, we think the contribution of individual differences in word recognition to comprehension is unlikely to show the expected decline across quantiles if the test design involves increasing unfamiliar, complex vocabulary for higher-level passages.

Listening comprehension encompasses a broad range of skills, and it was not obvious whether its contributions would be constant or not across quantiles. What we found was that in the tests where more of the variance is explained by word recognition, the PIAT and the WJPC, there are no quantile differences in the contributions of listening comprehension skills to reading comprehension. To the extent that vocabulary underlies listening comprehension, the arguments made above in relation to the consistency of word recognition would account for the consistency here as well. But what can explain the trend for listening comprehension to become more predictive of reading comprehension across quantiles, as was found for the GORT and QRI-Retells? We speculate that increasing complexity of the passages, reflected in part by length, may underlie this difference because text complexity requires additional comprehension skills, such as integration of information across many sentences, sentences that may occur far apart in the text. For example, the lowest-level passage on the GORT has only 52 words, whereas the highest-level passage has 163 words (Wiederholt & Bryant, 1992). In contrast, the higher-level test items on the PIAT and WJPC do not vary much from the lower-level items in length. Being mostly single sentences, these tests do not require the mental model building, coherence processing and inferencing that are needed to comprehend longer texts. Thus, there may be a lack of quantile differences in listening comprehension skills on these tests because there is not an increase in the range of skills required by higher-level passages.

When we deconstructed listening comprehension into its basic components in Study 2, we gathered two additional important findings. The first finding is that vocabulary is a very robust component influencing reading comprehension. Interestingly, we found that vocabulary is significantly more predictive of reading comprehension among the more skilled comprehenders for the GORT and QRI-R-Retells. This is consistent with what Tighe and Schatschneider (2014) had found, even though they did not use the same reading comprehension tests as ours. However, we also found that vocabulary is consistently predictive across quantiles for the WJPC, and significantly less predictive at the 90th quantile than at the 50th quantile for the PIAT and QRI-R-Questions.

We think the inconsistent findings regarding vocabulary's role among the best readers reflect the different range of items' difficulty across tests. In particular, the TABE-Reading subtest, used by Tighe and Schatschneider (2014), is a nationally used measure in Adult Basic Education programs and like the majority of our reading comprehension tests, the questions in this test increase in difficulty across reading levels. However, as described by Tighe and Schatschneider (2014), there seems to be quite a range in item difficulty for this test. Specifically, among the easiest items, participants are simply required to recognize letters and sounds, identify simple vocabulary words, and match letters. Among the more difficult items, participants are asked to interpret graphic information, recall information as well as constructing meaning and making inferences from the passages. Therefore, due to the nature of the variability among the difficult items on this test, vocabulary's influence in reading comprehension continues to increase as reading comprehension performance increases. The PIAT and the QRI-Questions, on the other hand, have the most restricted range of scores compared to our other three tests (see Table 1) and are the most negatively skewed measures, suggesting that variability among the most difficult items is not maintained to the same extent as our other tests. This may have contributed to the diminishing correlations at the 90th quantile, with more attenuation for the QRI-R-Questions than for the PIAT.

The second finding that we have gathered when deconstructing listening comprehension into its basic components is that working memory plays an important role in reading comprehension, but only for the PIAT and WJPC. Keenan and colleagues' research (2014) found that the poor comprehenders on these tests have significantly worse working memory skills compared to the poor comprehenders on the remaining tests, and now we have learned that these differences also exist for the average and skilled comprehenders. It is consistent with previous research (Cain, Oakhill, & Bryant, 2004; Cornoldi, De Beni, & Pazzaglia, 1996; Oakhill, Cain, & Bryant 2003; Sesma, Mahone, Levine, Eason, & Cutting, 2009; Swanson & Berninger, 1995; Yuill, Oakhill, & Parkin, 1989) that found evidence for working memory's direct influence in reading comprehension.

We think there are two factors that can help explain when working memory may play a direct role in reading comprehension. The first factor, as previously introduced by Keenan and colleagues (Keenan et al., 2014; Keenan & Meenan, 2014) and previously mentioned in Chapter 2, has to do with the format features (e.g., cloze format) of reading comprehension tests. That is, if the assessment requires that children hold the content of the passages in memory while quickly trying to find a one-word answer that best represents the meaning of the passages, as in the case of the WJPC, or finding the best picture that expresses the meaning of the sentence, as in the case of the PIAT, then working memory likely plays an important role in comprehension.

A second factor may have to do with the child's vocabulary knowledge. In the current study, working memory explains unique variance in reading comprehension for the GORT and QRI-R-Retells but only for readers at the 50^{th} quantile. These results suggest that for much longer tests, working memory's influence likely emerges for readers whose vocabulary knowledge is not too impoverished or too high. That is, the correlations of $r = .36$ and $r = .18$ at the 10^{th} quantile for the GORT and QRI-R-Retells, respectively, suggest that when vocabulary is too poor, comprehension invariably suffers and there is little room for any other cognitive processes, e.g., working memory, to exert their influence in reading comprehension. On the other hand, being extremely skilled at vocabulary (i.e., correlation of $r = .60$ and $r = .48$ at the 90^{th} quantile for the GORT and QRI-R-Retells, respectively) makes it easier for readers to build a coherent representation of the text without necessarily needing to recruit working memory, especially when the tests are as long as the GORT and the QRI. If samples in previous research differed considerably in vocabulary knowledge, that may have led to working memory sometimes playing a direct role in reading comprehension and sometimes not. Future research is needed to test this hypothesis more directly and rigorously; but at the very least, our findings provide an interesting rationale for this research.

We also found that reading and listening comprehension largely depend on similar language processes of vocabulary and working memory. This is consistent with a large body of extant literature that found reading and listening comprehension are mediated by the same cognitive mechanisms (e.g., Joshi, Williams, & Wood, 1998; Kintsch & Kozminksy, 1977; Townsend, Carrithers, & Bever, 1987). However, what we also found is that while vocabulary plays a very influential role in comprehension for all readers regardless of modality, working memory is generally more predictive of reading comprehension than it is of listening comprehension.

Limitations and Future Directions

There are several limitations to this dissertation. First, some of our interpretations would have been strengthened if our measures had also been assessed using Item Response Theory (IRT). In many of our standardized measures, children are required to read passages until basal and ceiling are established and the harder items are described in the manuals as increasing in complexity in terms of length, vocabulary, and syntax, etc. Our knowledge of these test items confirms that this is true. However, the manuals did not provide any information that validates and quantifies this increase in complexity. As a result, we rely on the manuals as well as our observations of the test items in order to explain why we think quantile differences occur for some tests and not for others. IRT would have allowed us to obtain specific knowledge of item characteristics such as item's difficulty and to use this information to corroborate our explanation in a more precise and systematic manner.

Second, we did not conduct a uniform sampling of children across the distribution of comprehension ability, resulting in some estimates being less precise than others. This is evident from that fact that some of our correlation estimates have wider confidence bands than others.

Finally, when we used the reading and listening versions of the same test in order to investigate the question of whether reading comprehension applies to general comprehension once differences in word decoding are taken into account, we assumed that these tests are parallel to each other in all aspects except for the modality. However, it is unclear whether differences in content or difficulty between the reading and listening tests may have played a role in the findings. Future studies should consider investigating these differences directly and adjusting them, if necessary, in order to clarify further the nature of the relationship between reading and listening comprehension.

Conclusion

In the current work, we used a large selection of different reading and listening comprehension tests to evaluate how the predictive validity of cognitive components might change with test performance level. But of course, there are many tests that we did not evaluate. One might wonder how the present findings generalize beyond the specific tests that we used. We were fortunate that the tests we examined were sufficiently different that we could decipher factors relevant to controlling when differences might occur across quantiles so that this information can be used to understand performance on other tests. Specifically, test characteristics – such as whether the test has most of its variance explained by word recognition, whether it increases complexity of word recognition by using less frequent words, and whether the length of the passages requires integration and coherence skills – seem to control the dynamic nature between the cognitive components and comprehension across readers of varying skills.

Our study adds to the small but growing number of reading studies (e.g., Logan et al., 2012; Tighe & Schatschneider, 2014) that have employed quantile regression in uncovering important relationships between component skills that may be specific to extreme readers without truncating the full sample. By identifying the types of tests that do and do not show quantile differences, we have shown that sometimes one needs to take into account the student's performance level to understand their performance. Such information allows us to go beyond using reading comprehension tests simply to identify who are struggling readers and gives insights into why they struggle. For example, if a child performs poorly on the GORT and the QRIs, we know that it is specifically the lack of vocabulary skills that drives his/her poor comprehension. Therefore, it may be wise for teachers and parents to focus training these children on vocabulary knowledge. If, however, a child is diagnosed with poor comprehension on the PIAT and the WJPC, then it is likely that both vocabulary and working memory drive their poor performance. Another implication of the current work is that depending on the modality of comprehension poor readers are selected from, they may show different cognitive profiles. Currently, some researchers select poor comprehenders based on listening comprehension rather than on reading comprehension with the goal to select readers whose comprehension skills are not also influenced by word decoding skills, which is especially important in studying beginning readers (e.g., Elwér, Keenan, Olson, Byrne, & Samuelsson, 2013; Hua & Keenan, 2014). While our data support this decision, they also show that poor comprehenders may show different severity in their vocabulary and working memory deficits depending on whether it is a reading or listening test. Finally, future research on poor comprehenders should consider using quantile regression instead of traditional methodologies that only estimate mean effects. As previously mentioned, methods that estimate average effects are not necessarily inadequate; often they are sufficient in providing a parsimonious description of the relationships among variables (Hao & Naiman, 2007). However, because heavy-tailed distributions commonly occur in social science research, quantile regression likely gives researchers more precise estimates and nuanced insights about how relationships may differ as a function of how poor or skilled the reader is. This methodology holds great promise in allowing us to make more progress in understanding why poor comprehenders struggle with comprehension the way they do.

References

Barnes, M. A., & Dennis, M. (1996). Reading comprehension deficits arise from diverse sources: Evidence from readers with and without developmental brain pathology. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 251–278). Mahwah, NJ: Erlbaum.

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

Braze, D., Tabor, W., Shankweiler, D. P., & Mencl, W. E. (2007). Speaking up for vocabulary reading skill differences in young adults. *Journal of learning disabilities*, *40*(3), 226-243.

Breusch, T. S., Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. *Econometrica*, 47, 1287-1294.

Cain, K., & Oakhill, J. (1999). Inference making ability and its relation to comprehension failure in young children. *Reading and Writing: An Interdisciplinary Journal*, *11*, 489–503.

Cain, K. & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, *76*, 683–696. doi:10.1348/000709905X67610

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: Guilford Press.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of educational psychology*, 96, 31- 42.

Carretti, B., Borella, E., Cornoldi, C., & De Beni, R. (2009). Role of working memory in explaining the performance of individuals with specific reading comprehension difficulties: A meta-analysis. *Learning and Individual Differences*, *19*(2), 246-251.

Carpenter, P. A., & Just, M. A. (1988). The role of working memory in language comprehension. In D. Klahr & K. Kotovasky (Eds.), Complex information processing: The impact of Herbert A. Simon (pp. 31–68). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*, 386-404.

Catts, H. W., Adlof, S. M. & Weismer, S. E. (2006). Language deficits in poor comprehenders: A case for the simple view of reading. *Journal of Speech, Language and Hearing Research, 49,* 278-293.

Catts, H. W., Petscher, Y., Schatschneider, C., Bridges, M. S., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, *42,*163−176.

Cohen, J., Cohen, P., West, S.G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: Erlbaum.

Cornoldi, C., De Beni, R., & Pazzaglia, F. (1996). Reading comprehension profiles. In C.Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention*. Mahwah, NJ, US: Erlbaum.

Curtis, M. (1980). Development of components of reading skill. *Journal of Educational Psychology, 72*, 656-669.

Cutting, L., & Scarborough, H. (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.

Daneman, M. & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19, 450–466.

DeFries, J. C., Filipek, P. A., Fulker, D. W., Olson, R. K., Pennington, B. F., Smith, S. D. (1997). Colorado Learning Disabilities Research Center. *Learning Disabilities, 8*, 7–19.

Dunn, L. M., & Dunn, M. (1997). Peabody Picture Vocabulary Test–III. Circle Pines, MN: American Guidance Service.

Dunn, L.M., & Markwardt, F. C. (1970). *Examiner's manual: Peabody individual achievement test.* Circle Pines, MN: American Guidance Service.

Elwér, Å., Keenan, J.M., Olson, R.K., Byrne, B., Samuelsson, S. (2013). Longitudinal stability and predictors of poor oral comprehenders and poor decoders. *Journal of Experimental Child Psychology*, *115,* 497–516.

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

Gathercole, S.E., & Baddeley, A.D. (1990). Phonological memory deficits in language disordered children: is there a causal connection? *Journal of Memory and Language*, 29, 336–360.

Gathercole, S.E., Hitch, G.J., Service, E., & Martin, A.J. (1997). Short-term memory and long-term learning in children. *Developmental Psychology*, 33, 966–979.

Goldman-Eisler, F. (1972). Pauses, clauses, and sentences. *Language and Speech*, *15,*103-113.

Gough, P. B., Hoover, W. A., & Peterson, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds), *Reading comprehension difficulties: Processes and intervention* (pp. 1-13). Mahwah, NJ: Lawrence Erlbaum Associates.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7,* 6–10.

Graf Estes, K., Evans, J.L., & Else-Quest, N.M. (2007). Differences in nonword repetition performance of children with and without specific language impairment: a meta-analysis. *Journal of Speech, Language and Hearing Research,* 50, 177– 195.

Hao, L., & Naiman, D. Q. (2007). *Quantile Regression*. CA: Sage Publications, Inc.

Hoover, W., & Tunmer, W. E. (1993): The components of reading. In G. B. Thompson, W. E. Tunmer & T. Nicholson (Eds.), *Reading acquisition processes* (pp. 1-19). Adelaide, Australia: Multilingual Matters, Ltd.

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.

Jackson, N. E., & Butterfield, E. C. (1989). Reading-level-match designs: Myths and realities. *Journal of Literacy Research*, *21*, 387-412.

Jackson, M. D., & McCelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology, 108*, 151-181.

Joshi, R.M., Williams, K.A., & Wood, J.R. (1998). Predicting reading comprehension from listening comprehension: Is this the answer to the IQ debate? In C. Hulme & R. M. Joshi, (Eds.), *Reading and spelling: Development and disorders* (pp.319-327). Mahwah, NJ: Lawrence Erlbaum Associates.

Keenan, J. M. (2014). Assessment of reading comprehension. In C. A. Stone, E. R. Silliman, B. J. Ehren, & G. P. Wallach (Eds.), Handbook of Language and Literacy: Development and Disorders (2nd Edition). New York: Guilford Press.

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.

Keenan, J., & Betjemann, R. (2008). Comprehension of single words: The role of semantics in word identification and reading disability. Single-word reading: Behavioral and biological perspectives (pp. 191-209). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Keenan, J. M., Betjemann, R. S., Wadsworth, S. J., DeFries, J. C., & Olson, R. K. (2006). Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading, 29,* 79–91.

Keenan, J., Betjemann, R., & Olson, R. (2008). Reading comprehension tests vary in the skills they assess: Differential dependence on decoding and oral comprehension. *Scientific Studies of Reading*, *12*, 281–300.

Keenan, J. M., Hua, A. N., Meenan, C. E., Pennington, B. F., Willcutt, E., & Olson, R. K. (2014). Issues in identifying poor comprehenders. *L'Année psychologique*, *114*, 753-777.

Keenan, J. M., & Meenan, C. E. (2014). Test differences in diagnosing reading comprehension deficits. *Journal of Learning Disabilities*, *47*, 125–135.

Kintsch, W., & Kozminksky, E. (1977). Summarizing stories after reading and listening. *Journal of Educational Psychology*, *69*, 491-499.

Koenker, R. (2005). *Quantile regression*. New York: Cambridge University Press.

Koenker, R., & Basset, G. (1978). Regression quantiles. *Econometrica*, *46*, 33-50.

Koenker, R., & Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica*, *50*, 43–61.

Landauer, T. K. (2007). LSA as a theory of meaning. *Handbook of latent semantic analysis*, 3-34.

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

Logan, J. A. R., Petrill, S. A., Hart, S. A., Schatschneider, C., Thompson, L. A., Deater-Deckard, K., ... Bartlett, C. (2012). Heritability across the distribution: An application of quantile regression. *Behavioral Genetics*, *42,* 256– 267.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods*, *7,* 19-40.

Mann, V. A., Cowin, E., & Schoenheimer, J. (1989). Phonological processing, language comprehension, and reading ability. *Journal of Learning Disabilities*, *22*, 76-89.

Miller, A. C., Keenan, J. M., Betjemann, R. S., Willcutt, E. G., Pennington, B. F., & Olson, R. K. (2013). Reading comprehension in children with ADHD: Cognitive underpinnings of the centrality deficit. *Journal of Abnormal Child Psychology*, *41*, 473-483.

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

Nation, K., & Snowling, M. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, *67*, 359–370.

Oakhill, J., & Cain, K. (2012). The precursors of reading comprehension and word reading in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, 16(2), 91-121.

Olson, R.K. (2006). Genes, environment, and dyslexia: The 2005 Norman Geschwind memorial lecture. *Annals of Dyslexia, 56*, 205–238.

Olson, R., Forsberg, H., Wise, B. & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In G.R. Lyon (Ed.), *Frames of reference for the assessment of learning disabilities: New views on measurement issues.* (pp. 243–277).

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

Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly, 40,* 184-202.

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

Petscher, Y. & Kim, Y. (2011). The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *Journal of School Psychology*, *49,* 107-129.

Petscher, Y., & Logan, J. A. (2013). Quantile regression in the study of developmental sciences. *Child Development*, *85*, 861-881.

Petrill, S. A., Logan, J. A., Sawyer, B. E., & Justice, L. M. (2014). It Depends: Conditional Correlation Between Frequency of Storybook Reading and Emergent Literacy Skills in Children With Language Impairments. *Journal of Learning Disabilities, 47,* 491-502.

Petscher, Y., Logan, J. A. R., & Zhou, C. (2013). Extending conditional means modeling: An introduction to quantile regression. In Y. Petscher, C. Schatschneider, & D. L. Compton (Eds.). Applied quantitative analysis in education and social sciences (pp. 3–33). New York, NY: Routledge.

Plaut, McClelland, Seidenberg, and Patterson (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103-56-115.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: a critical reexamination and new recommendations. *Psychological Methods*, *10*(2), 178.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http:// www.R-project.org.

Ricketts, J., Nation, K., & Bishop, D. V. (2007). Vocabulary is important for some, but not all reading skills. *Scientific Studies of Reading*, *11*(3), 235-257.

Seigneuric, A., & Ehrlich, M. (2005). Contribution of working memory capacity to children's reading comprehension: A longitudinal investigation. *Reading and Writing, 18*, 617-656.

Sesma, H. W., Mahone, E. M., Levine, T., Eason, S. H., & Cutting, L. E. (2009). The contribution of executive skills to reading comprehension. *Child Neuropsychology*, 15(3), 232-246.

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.

Swanson, H. L., & Berninger, V. (1995). The role of working memory in skilled and less skilled readers' comprehension. *Intelligence, 21,* 83– 108.

Tighe, E. L., & Schatschneider, C. (2014). A quantile regression approach to understanding the relations among morphological awareness, vocabulary, and reading comprehension in adult basic education students. *Journal of Learning Disabilities,* 1-13.

Tighe, E. L., Wagner, R. K., & Schatschneider, C. (2015). Applying a multiple group causal indicator modeling framework to the reading comprehension skills of third, seventh, and tenth grade students. *Reading and Writing*, 28(4), 439-466.

Townsend, Carrithers, & Bever (1987). Listening and reading processes in college- and middle-school age readers. In R. Horowitz & S. J. Samuels (Eds.), *Comprehending oral and written language* (pp. 217-242). New York: Academic.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific studies of reading*, *11*, 3-32.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children, Revised.* New York, NY: The Psychological Corporation.

Wiederholt, L., & Bryant, B. (1992). *Examiner's manual: Gray Oral Reading Test–3.* Austin, TX: Pro-Ed.

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

Achievement.

Yuill, N. M., Oakhill, J. V., & Parkin, A. J. (1989). Working memory, comprehension ability and the resolution of text anomaly. *British Journal of Psychology, 80,* 351-361.