WORD AND PERSON EFFECTS ON DECODING ACCURACY

By

Jennifer K. Gilbert

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

August, 2010

Nashville, Tennessee

Approved:

Donald L. Compton

Lynn S. Fuchs

Steve Graham

David S. Cordray

Georgine M. Pion

To my parents and my other half

ACKNOWLEDGEMENTS

This research would not have been possible without the funding of Grant R305B040110 from the U. S. Department of Education, Institute of Education Sciences; and Core Grant HDR01056109 from the National Institute of Child Health and Human Development to Vanderbilt University.

In addition, many people have been irreplaceable throughout my journey in graduate school. My advisor, Don Compton, is certainly at the top of that list. He took the time to teach and guide me without ever neglecting to value me as a person. His generosity and insight have made a significant ($p < .0001$) impact on my life and career. I am grateful to Doug and Lynn Fuchs for hiring me as a research assistant as an undergraduate. Their work sparked my interest in special education research and continues to inspire me in my young career. Thanks to my superiors would be incomplete without acknowledging David Cordray and Georgine Pion for encouraging me to be a real scientist, and introducing me to the ExpERT program.

The support and love of my family and community has meant more than they know, and I will always be thankful for each of them. In particular, I am thankful for my friend and colleague Devin Kearns. I have benefitted greatly from his relentless pursuit of understanding the fundamentals of everything. I am also thankful for my closest of friends Nikki Davis who has provided the wisdom and companionship that I very much needed throughout my schooling. And, I am thankful for my husband, my biggest fan. Although he never quite figured out all my statistical mumbo jumbo, his encouragement and unconditional love gave me the freedom to be myself and try my best each day. Finally, I am thankful to the Lord Jesus for leading me down this path and being faithful all of my days.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Importance of Decoding

According to stage theories of reading acquisition, decoding skill, or the ability to apply letter-sound relationships to words, is critically important in learning to read (Chall, 1983; Ehri, 1991; Frith, 1985; Gough & Hillinger, 1980; Mason, 1980; Perfetti, 1992). Certainly this is the core assumption underlying the self-teaching hypothesis in which children's practice of translating written text into phonological units acts as a self-teaching mechanism whereby orthographic representations (e.g., knowledge structures comprised of a word's letter pattern, pronunciation, and/or meaning) are established and subsequent fluent word recognition flourishes (Share, 1995). In one study establishing the importance of decoding to later reading, Jorm, Share, Maclean, and Matthews (1984) found that two groups of first year students who were matched on sight word knowledge but differed on decoding ability had significantly different progress in reading over the next two years. Students who had superior decoding skill faired better. Although it has been noted that the utility of decoding wanes somewhat over time as individuals develop more efficient methods of word recognition (Brown & Deavers, 1999; Reitsma, 1983; Treiman, Goswami, & Bruck, 1990), experienced readers who have mostly abandoned this reading strategy for familiar words still resort to it when faced with reading unfamiliar words (Brown & Deavers; Ehri; Ehri & Wilce, 1985).

Approaches to decoding are not uniform, however. There is evidence that children make use of two kinds of decoding strategies: letter-by-letter recoding and recoding by analogy, or recognizing and applying orthographic units larger than a single letter (Ehri, 1991, 1998; Goswami, 1986, 1998; Ziegler & Goswami, 2005). Successful letter-by-letter recoding requires knowing grapheme-phoneme correspondences (GPCs; or letter-sound relationships), retrieving those relationships for each grapheme from long-term memory, blending all the GPCs, and correctly producing a word. Successful recoding by analogy requires recognizing the correspondence between the orthography of the current word and that of similar words stored in memory, having the stored words well-represented in the lexicon, retrieving the stored words, and using an analogy in which the sounds associated with the common orthographic components are applied to the current word. Both strategies suggest that decoding is affected by both person (e.g., GPC knowledge) and word (e.g., number of similar words) characteristics. In the following sections, prior research on both types of effects is discussed.

Prior Research

Person effects

GPC knowledge is an undisputed prerequisite to successful decoding (Ehri, 1998; Share, 1995, 1999). If children do not know grapheme-phoneme relationships, applying those relationships in order to read unfamiliar words would be impossible. Treiman et al. (1990) investigated the role of GPC knowledge in a study of pseudoword reading accuracy. Tests of pseudoword reading are the "purest" measure of decoding skill as pseudowords must be decoded by application of GPCs rather than recognized by sight (Gough & Tunmer, 1986, p. 7). Treiman et al. showed significantly high correlations between GPC knowledge and pseudoword reading for first grade children. Moreover, the authors found that the correlation was somewhat higher

(.73) among low-frequency pseudowords than high-frequency pseudowords (.61), where frequency was defined in terms of the rime unit (or the first vowel and any subsequent letters in a one-syllable word). This finding lends support to the hypothesis that accurate decoding depends partially on both person and word characteristics.

The current investigation extends Treiman et al.'s (1990) work by considering the effect of GPC knowledge as specific to each word and each student. Treiman et al. found that children who knew more GPCs were predicted to decode more words than children who knew less GPCs. The authors used the total number of known GPCs to predict the total number of correctly decoded words. In this study, we asked a finer-grained question: How does a child's knowledge (versus lack of knowledge) of the GPCs within a word affect whether that child will correctly decode that particular word? Details about how GPC knowledge was coded to be person- and word-specific in the current study can be found in the description of the research questions.

Consider four relations between word-specific GPC knowledge and decoding accuracy. First, a student knows all the GPCs within a word and decodes the word accurately. Likewise, a student does not know all the GPCs within a word and does not decode the word accurately. These first two situations are intuitive. Another possibility is that a student knows all the GPCs within a word but decodes the word inaccurately. Or, it could be the case that a student does not know all the GPCs within a word yet decodes the word accurately. These latter two cases require exploration of factors beyond GPC knowledge that affect decoding. The first model in this study simply determines the influence of word-specific GPC knowledge on decoding accuracy while allowing random variation to exist among both students and words. After establishing the influence of word-specific GPC knowledge, additional variables were entered into a model to further explain variability in decoding accuracy. It is to those other potential explanatory factors that I now turn.

In addition to GPC knowledge, phonemic awareness is necessary for decoding (Juel, Griffith, & Gough, 1986). Phonemic awareness is the ability to hear, identify, and manipulate sound units in words (National Institute for Literacy). Byrne and Fielding-Barnsley (1991) demonstrated a unique contribution of phonemic awareness to decoding in an experiment where phonemic awareness instruction was provided to preschool children. The same result emerged in the 1-year follow-up (Byrne & Fielding-Barnsley, 1993). This relationship between phonemic awareness and decoding is most clear in Kindergarten and grade 1 as the relationship generally declines over time (Byrne, Fielding-Barnsley, Ashley, & Larsen, 1997). Because phonemic awareness has been established as a co-requisite to GPC knowledge in reading acquisition by compelling prior research (Byrne & Fielding-Barnsley, 1991; Bradley & Bryant, 1983; Ehri & Sweet, 1991; Hatcher, Hulme, & Ellis, 1994), phonemic awareness will be used in the current investigation as a potential contributor of decoding skill in young children. As an extension to the phonemic awareness literature, the model in the current study allows the influence of phonemic awareness to interact with word characteristics which may provide insight into the types of words (i.e., words with a complex vowel) for which decoding is most affected by phonemic awareness skill.

Although phonemic awareness and GPC knowledge are co-requisites for decoding, these two factors are insufficient to ensure adequate decoding ability (Byrne & Fielding-Barnsley; 1991; Byrne et al., 1997). One other potential factor may be naming speed. Hogaboam and Perfetti (1978) have shown that poor readers are slower at naming nonwords than good readers even when accuracy is the same between groups. Slower nonword naming may be due to a slower cognitive process called rapid automatized naming (Perfetti, 1986). Prior research would suggest that children exhibiting the double-deficit of poor phonological awareness and slow naming speed may be most at risk of experiencing reading difficulties (Wolf & Bowers, 1999, 2000). Bowers, Sunseth, and Golden (1999) hypothesized and found evidence that slower letter naming interferes with the full processing of letters in a word, thereby negatively affecting decoding skill. The statistical contribution of this ability to quickly retrieve phonological information from memory has been shown by Compton (2003) and Manis, Doi, and Bhadha (2000). Manis et al. found that letter naming speed accounted for more variance than digit naming speed, so the effect of the former is examined in the current study.

Findings from Compton (2000) provide evidence that the skills discussed in the previous paragraphs all contribute significantly to decoding skill when considered simultaneously. The author found that phonemic awareness, letter-sound knowledge, and rapid naming speed at the beginning of grade 1 significantly predicted decoding at the end of grade 1. The author also included a measure of advanced graphophoneme knowledge (knowledge of phonemes that make up various grapheme clusters), which is encompassed in the test of GPC knowledge used in the current study. Compton's four predictors accounted for 63% of the variance in decoding skill.

Given the evidence that working memory plays a role in acquiring new phonological information (Baddeley, 1986; Baddeley, Gathercole, & Papagno, 1998; Hulme & Mackenzie, 1992), perhaps a portion of the unexplained variance in decoding reported by Compton (2000) might be accounted for by working memory. Through confirmatory factor analysis, Kail and Hall (2001) found that working memory was significantly related to decoding skill while short-term memory was not and that the effect of working memory held even when the authors included phonological awareness in the model. Furthermore, Conners, Atwell, Rosenquist, and

Sligh (2001) found evidence that phonological working memory was the most reliable difference between good and poor decoders compared to age, IQ, phonemic awareness, and a general language composite. Abu-Rabia, Share, and Mansour (2003) also found significant differences in working memory when comparing students with reading disability to age-matched and reading-matched controls, which indicates that poor working memory is not simply an artifact of poor reading skill. Because of the prior research on working memory and the fact that decoding (especially of a pseudoword) is not an automatic event but one in which individuals must make concerted effort to connect letters to sounds (Stanovich, 1991), the current study includes a measure of working memory to potentially explain individual differences in decoding accuracy. In the current investigation, phonemic awareness, rapid naming speed, and working memory are entered as predictors in a model along with the previously entered word-specific GPC knowledge (including graphophoneme knowledge) to explain variability in decoding accuracy. Word-level influences on decoding are also considered.

Word effects

The relation between word type and performance in decoding has been the topic of much research. Most studies on this topic measure naming latency, or the amount of time required to decode a word. In one such study, Fredericksen and Kroll (1976) found that words and nonwords containing secondary vowels (e.g., ai and ou) took longer to read than primary vowels (e.g., a and o). In other words, reading complex graphemes takes longer to read than simple graphemes even when the number of phonemes, or the sounds, a person must produce is the same. Marinus and de Jong (2008) also reported a difference in latency favoring primary vowels, although in their study this was only true for students with dyslexia. Shifting from latency to accuracy, Olson, Forsberg, Wise, and Rack (1994) found that children made more errors when attempting to read words with consonant clusters than single consonants. Snowling (1981) noted that the complex consonant deficit was particularly noticeable for children with dyslexia when compared to children with typical reading skills. Finally, Siegel and Faux (1989) found that children with reading disabilities had more trouble with both vowels and consonant clusters than their normally-achieving peers. Though past research has demonstrated the difficulty of complex vowels and consonants compared to their single counterparts, in the current study we compare the effect of grapheme complexity for vowels and consonants by systematically varying vowel complexity and consonant complexity in a list of pseudowords and then using a dummy variable to directly compare the difference between the two in our sample of at-risk readers. Knowing the types of words that are most difficult to decode could potentially lead to improved instructional practices.

Another word feature that influences decoding is frequency. According to the activation model of decoding (McClelland & Rumelhart, 1981), words that share features with other words are more easily decoded than words that have fewer shared features. Bowey (1990) and Booth and Perfetti (2002) have determined that a common word feature English-speaking children are sensitive to is the rime unit, or the final vowel-consonant cluster in a single-syllable word. In two studies, Calhoon and Leslie (2002) and Leslie and Calhoon (1995) investigated the effects of rime-neighborhood size and word frequency on word and nonword reading across grades 1-3. Rime-neighborhood size is the number of one-syllable words containing the same rime unit. In general, the authors found that both rime-neighborhood size and word frequency affected reading in the early elementary years. They also found an interaction of rime-neighborhood size and word frequency such that children attempting to read low-frequency words were more affected by rime-neighborhood size than when they were attempting high-frequency words. Carlisle and

Katz (2006) found similar results, although their investigation was focused on word families, or all the morphemic variations of a word (e.g., sit, sat, sitting), instead of rime neighbors. Because the above investigations did not account for both person and word variance, results may be biased in that standard errors for covariates may have been too small (Janssen, Schepers, & Peres, 2004). Thus, we examined the effects of rime-neighborhood size, average rime-neighbor frequency, the interaction of size and frequency, and grapheme complexity while accounting for all necessary sources of variance to determine the relative contribution of each to decoding accuracy. Because pseudowords have no frequency of their own, average rime-neighbor frequency was used as a proxy for word frequency because it represents the average frequency of a word that contains the same rime unit as the pseudoword itself.

Rationale for Current Investigation

The purpose of this investigation was to extend the literature on decoding by bringing together two lines of research, namely person and word factors that affect decoding, using a crossed random-effects model. There are two points of rationale for conducting the study. First, although prior research has documented factors that affect decoding, results are somewhat disjointed because not all factors have been explored together in one model. As a solution, we combined factors found to be significant in prior research into one model so that relative contributions could be established. Furthermore, a crossed random-effects model allows for the investigation of person-by-word covariates in addition to person-specific and word-specific covariates. This type of model was recently used by Kim, Petscher, Foorman, and Zhou (2010) to assess the effects of letter-name knowledge (person-by-item covariate) on the probability of knowing letter sounds while accounting for variance across letters and persons. Also, Piasta and

Wagner (2010) recently utilized a crossed random-effects model to assess the effects of person characteristics, letter characteristics, and instruction on letter-name acquisition.

The second point of rationale is that most research to the present has utilized aggregated data either across persons or across words. The problem with aggregated data is loss of information. Loss of word information occurs when person-level analyses are conducted while loss of person information occurs when word-level analyses are conducted. For instance, Manis et al. (2000) found that letter naming speed affected accuracy of decoding. However, that analysis was based on the total accuracy score, which does not provide specific information about the pseudowords that were being decoded. Information such as how the effect of letter naming speed may differ depending on which words are being decoded was lost in the aggregate analysis. Similarly, Calhoon and Leslie (2002) found that the frequency of a rime in a pseudoword affected how accurately the pseudowords were decoded. The effect of rime frequency for each word was aggregated across individuals, and further investigation of those individuals was not conducted nor was the variability amongst them taken into account. Because decoding is an activity that involves both the person doing the decoding and the word being decoded, an analysis that takes both effects into account has the potential to be more informative than focusing on either effect separately.

A crossed random-effects model allows person covariates, word covariates, and person-by-word interactions to be investigated simultaneously while also accounting for variance in the responses that is due to the persons and the words. Potential problems arise when random variance due to persons and words exist in the outcome but both sources are not included in the model. Baayen, Davidson, and Bates (2008) have demonstrated that inflated variance components and deflated standard errors result when item or person variance is not taken into account. Other model comparison papers have provided similar warnings (Luo & Kwok, 2009; Moerbeek, 2004; Rabash & Browne, 2007), although they have not dealt specifically with person-by-item data. Random effects have associated error terms that account for the unlikelihood that a set of covariates perfectly explain the variance of the outcome between units (Janssen et al., 2004). Treating people as random effects is well-established in the hierarchical linear modeling (HLM) literature for it is reasonable to assume that no combination of person covariates will be able to explain all the variance in the outcome between people. The same argument can be invoked for test items such as words; it is not likely that a set of word predictors can explain all the variance in the outcome that exists between the words. Therefore, modeling random person and item effects simultaneously provides unbiased estimates of both, whereas conducting analyses at either the person- or item-level may produce biased results due to ignored dependencies in the data. The analysis for the current study (a crossed random-effects model) allows a detailed and accurate investigation of the factors that affect decoding because the unit of analysis is the intersection of each person and each word, and inter-word variance and inter-person variance are taken into account (Janssen et al.).

Four research questions are asked in accordance with the aim of the proposed study: the first involves a person-by-word covariate, the second involves a person-by-word covariate plus person covariates, the third involves a person-by-word covariate plus word covariates, and the fourth involves a person-by-word covariate along with person and word covariates plus interaction terms between persons and words. The following questions were answered using crossed random-effects models:

1. What is the effect of word-specific GPC knowledge on decoding accuracy? Students attempted 20 pseudowords on the Targeted Nonword Test (TNT). This dependent

variable was assigned a value of 1 for a correct answer and 0 for an incorrect answer. Students were also given a separate test of GPC knowledge. The GPCK variable was specific to each student attempting each word on the TNT. If the student knew all the GPCs within a pseudoword, GPCK equaled 1. Otherwise, GPCK equaled 0 to signify that a student did not know all the GPCs within the pseudoword. Our hypothesis was that there would be a significant difference in the predicted probability of correct decoding between knowing and not knowing word-specific GPCs, with the former case having the higher probability.

2. What is the relative importance of naming speed, phonemic awareness, and working memory in accounting for person variance in decoding accuracy after controlling for word-specific GPC knowledge? How much variance do these factors account for? In accordance with the prior research discussed in the introduction, we hypothesized that all student factors entered in the model would have significant influence on decoding accuracy.
3. What is the relative importance of average rime-neighborhood frequency, rime-neighborhood size, an interaction of neighborhood frequency and size, and grapheme (vowel versus consonant) complexity in accounting for word variance in decoding accuracy after controlling for word-specific GPC knowledge? How much variance do these factors account for? Each pseudoword contained either a complex consonant cluster or a complex vowel cluster, so one dummy variable was created in which pseudowords with a complex vowel cluster were assigned a value of 1 and a value of 0 was assigned to pseudowords with a complex consonant cluster. We expected to find a significant interaction between rime size and rime frequency, which has also been found in previous

total-score analyses. Furthermore, we expected that pseudowords with complex vowels would be significantly more difficult to decode than pseudowords with complex consonants, as suggested by trends in past research.

4. Are there significant person-by-word interactions? First, we consider an interaction term between working memory and rime frequency. Our hypothesis is that students with poorer working memory might perform worse on pseudowords with infrequent rimes than frequent rimes, while students with better working memory might perform similarly in both cases. Poorer performance on words with the infrequent rimes could be due to the more likely letter-by-letter decoding which we think may be more taxing on working memory than decoding by analogy (Siegel & Faux, 1989). Second, we consider an interaction between phonemic awareness and grapheme (vowel versus consonant) complexity. Prior research has suggested that children with poor phonological skills have more difficulty with complex vowel clusters than complex consonant clusters (Marinus & de Jong, 2008; Siegel & Faux). Knowing which types of words struggling readers have the most difficulty with may provide information about how word-reading instruction could be targeted more effectively to benefit struggling readers.

CHAPTER II

METHOD

Participants and Procedures

Participants in the study were part of a larger study investigating the efficacy of supplemental reading tutoring for first grade students who were at risk for later reading difficulties. In the larger study, 287 students from 56 classrooms and 15 schools were consented and screened. The screening battery included measures of accuracy of decoding and word reading, efficiency of decoding and word reading, and rapid naming speed. A factor score was created from the screening battery and the lowest 250 students based on factor score were chosen for further testing. Of the given measures, only the Backward Digit Recall subtest of the Working Memory Test Battery (Pickering & Gathercole, 2001) was relevant to the current study. Students who completed all pretests (N = 215) were randomly assigned to one of four groups. Three of the groups were intervention groups in which students were provided supplemental reading instruction, while the control group did not receive any instruction beyond their classroom reading instruction. We did not expect that working memory at the end of the year would be differentially affected by group membership, but we did test group equivalence of working memory at the beginning of the year to ensure the groups began the year with equivalent scores. Indeed, group equivalence was established with an ANOVA, $F(3, 219) = 0.33$, $p = .80$.

Because decoding accuracy is the outcome and GPC knowledge is included as a covariate, it is worth mentioning that the supplemental reading instruction incorporated practice of GPCs and decoding. It is likely that the regular classroom reading instruction also incorporated these activities as the district's academic standards state that children in first grade should be using GPCs to decode words (Tennessee/Metro Nashville Public Schools, n.d.). One additional note is that treatment was not considered in the current set of analyses because we are concerned with how children's skills affect decoding regardless of how they acquired those skills.

After tutoring ceased in the spring, students were given several tests including the Assessment of Grapheme-Phoneme Correspondence Knowledge (AGPCK; Kearns, 2006), the Rapid Letter Naming subtest of the Comprehensive Test of Phonological Processing (Wagner, Torgesen, & Rashotte, 1999), the Sound Matching subtest of the Comprehensive Test of Phonological Processing (Wagner et al.), and the experimenter-created TNT. Data analysis was conducted on the data of 196 students who were not missing scores on any relevant variables. Of the 196 students, the gender composition was nearly equal between females (47.96%) and males (52.04%). The majority (65.59%) of the sample, with available information (n = 186), was receiving a free or reduced price lunch. Although the racial make-up of the sample was varied (40.82% African American, 38.78% Caucasian, 9.18% Hispanic, 7.65% Biracial, 1.53% Kurdish, 1.02% Asian, and 1.02% Other), none of the students were classified as English-language learners. Finally, there were 20 students (10.75%) with a special education label and 15 (8.06%) who were repeating first grade.

Students were enrolled in 15 schools, 13 of which were eligible for Title I funds. School Title I eligibility is based upon the criteria that at least 35% of students in that school reside in low-income homes and are qualified for a free or reduced price lunch. Further, although 9 of the 15 schools met Adequate Yearly Progress (ATP) for the 2009-2010 academic year according to the standards of the No Child Left Behind Act (2002), 4 schools were given the Target label (failed to meet at least one benchmark) and 2 were given the School Improvement label (failed to meet the same benchmark for at least two consecutive years). The majority of teachers of the participating students were Caucasian (83.93%) females (89.83%) who had been teaching an average of 15.69 years. Most teachers reported that their age was between 40-49 years old and reported taking more than 12 credit hours of college coursework in reading.

Measures

Decoding

Accurate decoding was assessed with the Targeted Nonword Test (TNT). The TNT is a 20-item experimenter-created measure designed to explore the relation between decoding accuracy and other factors, not to measure individual differences in decoding skill. In this way, the pseudowords on the TNT were *targeted* because they were created to answer specific research questions. See the Appendix for the pseudowords that comprise the TNT. Pseudowords were used instead of real words because students' unfamiliarity with pseudowords ensures that children attempt to read the words by using decoding skills rather than by using memory, or reading by sight (Gough & Tunmer, 1986). All pseudowords were comprised of four letters and three phonemes. Frederickson and Kroll (1976) showed that word length affects reading accuracy, and we wanted to control for that effect. In addition, we eliminated potential pattern effects by using only the CVC (consonant-cluster vowel-cluster consonant-cluster) grapheme pattern; Calhoon and Leslie (2002) also followed this format.

Words on the TNT varied systematically in two ways. The first source of planned variability was rime frequency, defined in terms of the rime-neighborhood size (Treiman et al., 1990; Ziegler, Stone, & Jacobs, 1997). Half the pseudowords contained frequent rimes (more than 2 rime neighbors) while half contained infrequent rimes (2 or less rime neighbors). Rime size as opposed to rime frequency was used to create the test items because the former has been shown to be more influential when examined simultaneously (Carlisle & Katz, 2006). Furthermore, we decided to only consider the grade K-3 word corpus because it is likely grade 1 children are exposed to those words in texts more than any other grade-level text. The average rime-neighborhood size for the frequent pseudowords was 8.90 (range 3-16) while it was 0.90 (range 0-2) for the infrequent pseudowords. The average rime-neighbor frequency for frequent pseudowords was 117.54 and was 20.90 for infrequent ones. (More information on rime size and rime complexity can be found in the Rime Frequency subsection.)

The other source of systematic variation was grapheme cluster complexity. Complexity was defined as two letters composing one phoneme (e.g., digraphs). Half the pseudowords contained a complex vowel grapheme cluster (e.g., ea) and a simple consonant grapheme (e.g., n) and half contained a complex consonant grapheme cluster (e.g., th) and a simple vowel grapheme (e.g., e). Marinus and de Jong (2008) showed a tendency for dyslexic children to make more errors and take longer in pronouncing vowel digraphs than consonant clusters. One limitation of the Marinus and de Jong study was that the position of the vowel cluster always followed the consonant cluster, which confounds cluster type with cluster position. In the measure, the cluster order was reversed to shed more light on the complexity versus position issue. Fredericksen and Kroll (1976) also showed a grapheme complexity effect although their outcome of interest was naming latency as opposed to accuracy.

The frequency and complexity dimensions create four cells into which the pseudowords were categorized: frequent rimes with complex vowel clusters, frequent rimes with complex consonant clusters, infrequent rimes with complex vowel clusters, and infrequent rimes with complex consonant clusters. The Appendix contains the pseudowords in their respective cells. All words began with a single consonant, and stop and continuous consonants were counterbalanced across the four cells. Liquid sounds such as /l/, /r/, /w/, and /y/ were not used as initial phonemes because these sounds are more difficult for young children to produce than other phoneme types (Preisser, Hodson, & Paden, 1988). All words were consistent (or regular) in that none of the rime units had real words with illegal pronunciations.

Interrater reliability was established by asking someone other than the original test administrator to rescore each item on the TNT by listening to tape recordings of the tests. The interrater reliability, based on 20% of the tests, was 92.28% for total test score and ranged from 77.50-100.00% for each individual item. All items except one had a reliability exceeding 87.00%. The Kruder-Richardson 20 internal consistency coefficient for the measure was 79.65%. Though the items on the TNT were all created to assess decoding accuracy, we did not expect an especially high internal consistency because each pseudoword required potentially different knowledge (e.g., word-specific GPCs) for giving a correct response. Validity of this measure has not been established. Students were given 1 of 4 randomly ordered stimuli sheets.

Phonemic awareness

Phonemic awareness was assessed with the Sound Matching subtest of the Comprehensive Test of Phonological Processing (Wagner et al., 1999). In this assessment, children are provided four pictures, one of the stimulus and one each of the 3 response options. The test administrator reads aloud all stimuli and response options, while simultaneously pointing to the corresponding pictures of each. The administrator asks the child to indicate which of three response options begins or ends with the same phoneme as the stimulus. Test developers reported a Cronbach's alpha of .93 for this 20-item subtest for 5- and 6-year olds in the normative sample. Using item response theory, test developers determined that the Sound Matching subtest had content validity. They also reported that the subtest's correlations with word and nonword reading were .58 and .49, respectively. These moderate correlations would be expected as phonemic awareness is somewhat but not perfectly related to word reading and decoding. In sum, the validity of the subtest was deemed sufficient (Wagner et al.).

Rapid naming

Rapid naming was assessed with the Rapid Letter Naming subtest of the Comprehensive Test of Phonological Processing (Wagner et al., 1999). For this assessment, children are presented with a sheet of paper containing randomly ordered exposures of the letters *a*, *c*, *k*, *n*, *s*, and *t*. Letters are presented in 4 rows of 9 letters each for a total of 36 letters. Students are asked to say the names of the letters as quickly as possible. The test administrator records the time it takes the student to name all 36 letters. Students complete the process twice on two randomly ordered stimulus sheets. The final score is the number of seconds taken to read both lists. Cronbach's alpha in the normative sample was .89 and .89 for 5- and 6-year-olds, respectively. Evidence supporting concurrent validity of this subtest consists of significant correlations with measures of word reading accuracy (.47-.52) and rate (.44) as well as overall reading (.59). The predictive validity of the subtest was established by high, significant correlations with later word reading (.53-.61), decoding (.65), and overall reading (.49) (Wagner et al.).

Rime frequency

There were two indices of rime frequency, rime-neighborhood size and average rime-neighborhood frequency. The rime-neighborhood size of each pseudoword was determined using the Educator's Word Frequency Guide (Zeno, Ivens, Millard, & Duvvuri, 1995). This database of words was created from texts used in schools including textbooks, literary works, and popular fiction and nonfiction writings. It contains frequency count information for words by grade. Rime-neighborhood size in the current study was the number of monosyllabic, monomorphemic words that appeared at least once in 1,000,000 words in texts from grades K-3. Rimes were chosen such that all neighbors (if any) shared a common pronunciation.

The average rime-neighbor frequency was also calculated from numbers contained in the Educator's Word Frequency Guide (Zeno et al., 1995). As defined by Zeno et al., the frequency of a word represented how often a word appeared in 1,000,000 words of text and was weighted with the word's dispersion across different subject areas. The average rime-neighbor frequency was calculated by summing the frequency of each neighbor (as defined by rime-neighborhood size) across grade K-3 texts and dividing the total by the rime-neighborhood size. This measure provides information about the average frequency of a rime neighbor.

Word-specific GPC knowledge

Word-specific GPC knowledge (GPCK) was calculated from the Assessment of Grapheme-Phoneme Correspondence Knowledge (Kearns, 2006), which is a 48-item test designed to assess knowledge of the correspondence between graphemes and phonemes. For this assessment, students are presented with a sheet of paper comprised of 48 printed graphemes. A test administrator asks, "What sound does this letter make?" If a grapheme has more than one possible sound associated with it (e.g., a), the test administrator says, "Ok. This letter can make another sound too. What else can it say?" One point is recorded for each correct answer. This test is administered individually, and the maximum point total is 62. Reliability and validity of this measure have not been established.

GPCK scores are specific to each student and each pseudoword given on the TNT. Scores of 1 were given when a student knew all the GPCs associated with a pseudoword, or 0 otherwise. For example, the pseudoword sheb contains three phonemes: /sh/, /e/, and /b/. If the student knew the GPC for *sh* and *b* but not for *e*, GPCK was assigned a value of 0 for that student on that pseudoword. GPCK equaled 1 only when the student knew all the sounds in the pseudoword. Pseudowords on the TNT were comprised of 26 different graphemes and their 27 corresponding phonemes. The grapheme *oo* had two corresponding phonemes, /oo/ and /uu/, that were legal in the present sample of pseudowords.

Working memory

Working memory was assessed with the Backward Digit Recall subtest of the Working Memory Test Battery (Pickering & Gathercole, 2001). In this assessment, children are orally presented a list of digits. After the test administrator dictates the list of digits, children are asked to repeat the list of digits in reverse order. Children are given practice items (with feedback as necessary) to ensuring their understanding of the task. Initially, the lists contain only two digits. As the test progresses, the number of digits increases by one up to a maximum span of seven. Each span has six trials. The test is terminated if children make three or more errors in any one span. Correct answers are awarded one point and the total score is the number of correct answers. The testing manual reports a test-retest reliability of .53 for a normative sample of 5- and 6-year olds. Validity was established with correlations > .60 for other measures of central executive measures. A confirmatory factor analysis further confirmed the validity of the subtest (Pickering & Gathercole, 2001).

Data Analysis

Van den Noortgate, De Boeck, and Meulders (2003) proposed a method of assessing item-by-student data using a crossed random-effects model. In item response theory (IRT) terminology, this model is equivalent to a Rasch model with a random item parameter. In this model both items and persons are treated as random parameters, which permits the partition and explanation of variance in the outcome in terms of person and item factors. This means that the probability of a person answering an item correctly is assumed to be concomitantly dependent upon item and person characteristics. Research cited in the introduction suggests this is a sensible assumption for decoding data, and we formally test that assumption by calculating the proportion of variance in the outcome that is due to persons and words in the Base model. A likelihood ratio test comparing models with and without person and word random effects serves as further evidence for their inclusion/exclusion.

Although using IRT within a multilevel framework is not novel, treating items as random is somewhat unique. Most multilevel IRT models consider items as fixed effects (Adams, Wilson, & Wu, 1997; Fox, 2005; Maier, 2001; Van den Noortgate & Paek, 2004). This is primarily because IRT is most often used for scale construction in which the characteristics of specific items is of great interest. These types of models are classified as measurement models because the aim of the model is to describe individuals' performance on specific test items. Alternatively, explanatory models are used when researchers are less interested in the specific items (and/or specific persons) and more interested in item (or person) features and how those features can explain variability among item responses (De Boeck & Wilson, 2004). In explanatory models, person and items can be considered random effects simultaneously such that the model includes both a person- and item-specific residual to account for dependencies in the responses.

Treating words as random factors allows the variability around the decoding of words to be explained in terms of word characteristics (De Boeck, 2008; Janssen et al., 2004). Because of the nature of the research questions addressed in this paper (i.e., word covariates and person-by-word covariates), such a model is necessary. Furthermore, Baayen et al. (2008) advocate for random person and random item effects specifically when dealing with linguistic material. Their argument is based on the assumption that words and sentences used in assessment situations are only samples of the total population of words and sentences. If results of linguistic tests are expected to generalize beyond the immediate context of the words contained on the tests, then words ought to be treated as random variables even if the sample of words was not drawn randomly from a population of words (see Briggs & Wilson, 2007; Clark, 1973). Because we are interested in using word, person, and word-by-person covariates to explain variability in decoding accuracy and because we are interested in generalizing our results to words and people beyond those included in the current study, the most appropriate analysis is an explanatory model with random effects of both persons and words (a.k.a. a crossed random-effects model).

In a crossed random-effects model, responses (i.e., unit of analysis) are considered cross-classified in nature because they originate from the intersection of each student responding to each word in the same set of pseudowords (Baayen et al., 2008). Had each student responded to a different set of pseudowords, then a response would be nested within a person who would be nested within a set of pseudowords. Instead, our response data were considered cross-classified because our lower-level units (i.e., words) belong to upper-level units (i.e., persons and words) that are not nested within each other (Rabash & Browne, 2007; Raudenbush & Bryk, 2002). In addition to the responses being crossed by persons and items, persons were also nested in classrooms that are nested in schools. This nesting means a student was a member of only one classroom, and that classroom was a member of only one school. Clustering at those levels were taken into account by adding random effects for each. Figure 1 illustrates the 4-level data structure in graphical form.

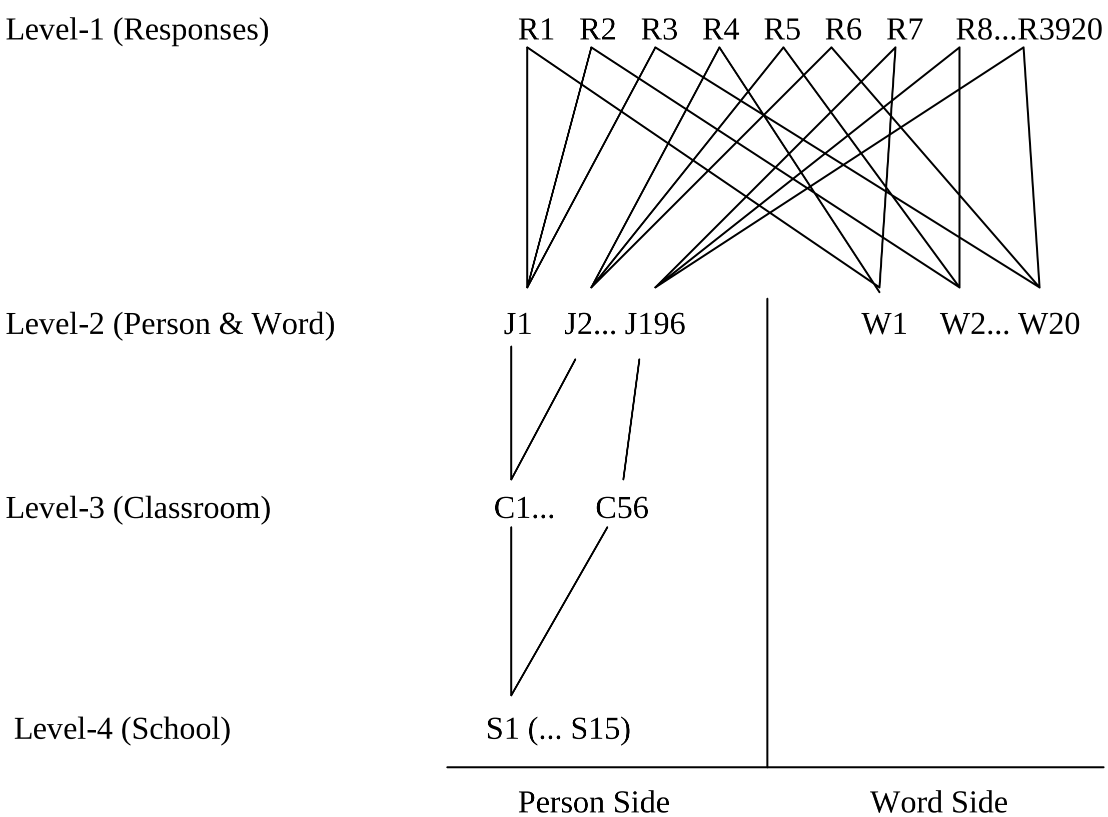

Figure 1. Graphical display of 4-level data structure

The first step in analyzing our data was to estimate a Base model in which no covariates were included in the model. The purpose of this initial model was to partition variance into the respective levels of influence on the outcome. This information was useful because it allowed us to calculate the proportion of variance in the outcome that existed at each level (i.e., the intraclass correlation [ICC]). To determine which Base model was most appropriate for the data, five models were estimated and compared. The first possible Base model (Full Base model) included all potential sources of variance (i.e., word, person, classroom, and school). Then for the sake of parsimony, four reduced models were estimated: one in which the parameter for random word variance was omitted (Base without word), one in which the parameter for random person variance was omitted (Base without person), one in which the parameter for random school variance was omitted (Base without school), one in which the parameter for random classroom variance was omitted (Base without classroom). Comparing the Full Base model to the reduced models allowed us to determine whether each level of variance was necessary for accurately describing the data. Equation 1 displays the Full Base model.

Level-1 (Responses$_{jkmi}$) $\quad Logit(\pi_{jkmi}) = \lambda_{0jkmi}$ (1)

Level-2 (Person$_{jkm}$ & Word$_i$) $\quad \lambda_{0jkmi} = \gamma_{00km} + r_{01jkm} + r_{02i}$, $r_{01jkm} \sim N(0, \sigma^2_{r01})$ & $r_{02i} \sim N(0, \sigma^2_{r02})$

Level-3 (Classroom$_{km}$) $\quad \gamma_{00km} = \phi_{000m} + u_{km}$, $u_{km} \sim N(0, \sigma^2_u)$

Level-4 (School$_m$) $\quad \phi_{000m} = \omega_{0000} + \varepsilon_m$, $\varepsilon_m \sim N(0, \sigma^2_\varepsilon)$,

where π_{jkmi} is the probability of a correct response from person *j* in classroom *k* in school *m* on word *i*, λ_{0jkmi} is the intercept and represents the logit of the probability of a correct response from an average person ($r_{01jkm} = 0$) in an average classroom ($u_{km} = 0$) in an average school ($\varepsilon_m = 0$) decoding a word of average difficulty ($r_{02i} = 0$). At Level-2, λ_{0jkmi} is reformulated into the mean logit of the probability of getting a correct response in classroom *k* in school *m* (γ_{00km}), plus a student-specific residual (r_{01jkm}) and a word-specific residual (r_{02i}) on a word of average difficulty. At Level-3, γ_{00km} is reformulated as the mean logit of the probability of a correct response in school m (ϕ_{000m}) plus a classroom-specific residual (u_{km}) on a word of average difficulty. Finally at Level-4, the school-average logit of the probability (ϕ_{000m}) is reformulated into a grand-mean logit of the probability (ω_{0000}) plus a school-specific residual (ε_m) on a word of average difficulty. The magnitude of variance at for each random component was recorded for later reduction-in-variance calculations. The leveled equations make the structure of the data clear, but this set of equations can be simplified into a single equation by way of substitution. See Equation 2 for the combined model that includes an intercept, a set of person-related random effects, and a word random effect.

$$Logit(\pi_{jkmi}) = \omega_{000} + \left(r_{01jkm} + u_{km} + \varepsilon_m\right) + \left(r_{02i}\right) \tag{2}$$

ICCs can be calculated from the combined model represented in Equation 2. Raudenbush and Bryk (2002) describe how to calculate the ICC in models for binary data that use the logit link, where the residuals (τ_{jkmi}) are assumed to follow the logistic distribution (0, $\frac{\pi^2}{3}$). More specifically, Cho and Rabe-Hesketh (in press) describe how to calculate ICCs for item and person models, where person variance is conditional on item difficulties and item variance is conditional on person abilities. Because this is a cross-classified model, the ICCs must be interpreted slightly differently than in nested models because the person ICC is conditional on word difficulties and the word ICC is conditional on the person abilities. For example, the person ICC represents the proportion of variance in decoding accuracy that is due to between-person differences conditional on word difficulties. The ICC is useful for determining the magnitude of influence exerted by each source of variance. Assuming each source of variance is necessary for best describing the data, Equation 3a shows the ICC calculation for person variance; Equation 3b is for classroom variance; Equation 3c is for school variance; Equation 3d is for word variance.

$$\rho(P) = \frac{\sigma_{r01}^2}{\sigma_{r01}^2 + \sigma_u^2 + \sigma_\varepsilon^2 + \frac{\pi^2}{3}} \tag{3a}$$

$$\rho(C) = \frac{\sigma_u^2}{\sigma_{r01}^2 + \sigma_u^2 + \sigma_\varepsilon^2 + \frac{\pi^2}{3}} \tag{3b}$$

$$\rho(S) = \frac{\sigma_\varepsilon^2}{\sigma_{r01}^2 + \sigma_u^2 + \sigma_\varepsilon^2 + \frac{\pi^2}{3}} \tag{3c}$$

$$\rho(W) = \frac{\sigma_{r02}^2}{\sigma_{r02}^2 + \frac{\pi^2}{3}} \tag{3d}$$

In addition to the Base model, we proposed three models to address the four research questions. Table 1 lists each research question with its associated model number and covariates.

Table 1
Research Questions with Associated Model Number and Covariates

Research Question	Model Number	Covariates
1. What is the effect of word-specific GPC knowledge on decoding accuracy?	1	Person-by-word: GPCK
2. What is the relative importance of phonemic awareness, naming speed, and working memory in predicting decoding accuracy after controlling for GPC knowledge? How much person variance do these factors account for?	2	Person-by-word: GPCK Person: PA, RN, WM
3. What is the relative importance of average rime-neighbor frequency, rime-neighborhood size, interaction of frequency and size, and grapheme (vowel versus consonant) complexity in predicting decoding accuracy after controlling for GPC knowledge? How much word variance do these factors account for?	2	Person-by-word: GPCK Word: R_FREQ, R_SIZE, FREQ*SIZE, VC
4. Are there interactions between the person and word characteristics?	3	Person-by-word: GPCK Person: PA, RN, WM Word: R_FREQ, R_SIZE, FREQ*SIZE, VC Interaction: WM*(S/F), PA*VC

In addressing the research questions, we first estimated a model including only word-specific GPC knowledge (Model 1) to address research question 1. Model 1 is described in Equation 4.

Level-1 (Responses$_{jkmi}$) $$Logit(\pi_{jkmi}) = \lambda_{0jkmi} + \lambda_{1jkmi} GPCK_{jkmi} \quad (4)$$

Level-2 (Person$_{jkm}$ & Word$_i$) $$\lambda_{0jkmi} = \gamma_{00km} + r_{01jkm} + r_{02i}\,,\; r_{01jkm} \sim \mathrm{N}(0,\ \sigma^2_{r01}) \;\&\; r_{02i} \sim \mathrm{N}(0,\ \sigma^2_{r02})$$

$$\lambda_{1jkmi} = \lambda_{10km} + r_{11jkm} + r_{12i}\,,\; r_{11jkm} \sim \mathrm{N}(0,\ \sigma^2_{r11}) \;\&\; r_{12i} \sim \mathrm{N}(0,\ \sigma^2_{r12})$$

Level-3 (Classroom$_{km}$) $$\gamma_{00km} = \phi_{000m} + u_{km}$$

Level-4 (School$_m$) $$\phi_{000m} = \omega_{0000} + \varepsilon_m$$

where λ_{10km} is the increment/decrement to the logit of the probability of an average person in an average classroom in an average school getting a correct response on an average word when all the GPCs are known in the word compared to when the GPCs are not known

ω_{0000} is the logit of the probability of an average person in an average classroom in an average school getting a correct response on an average word when all the GPCs are not known in the word

At level-2, λ_{10km} is reformulated into the mean difference in logits of the probability between knowing and not knowing word-specific GPCs in classroom *k* in school *m* plus a person-specific (r_{11jkm}) and a word-specific residual (r_{12i}). The covariance between the intercept and GPCK slope variance terms is estimated for both persons ($\sigma_{r01*11}{}^2$) and words ($\sigma_{r02*12}{}^2$) to account for the possible relations between them. The variance, covariance structures for persons and words, respectively, are:

$$\begin{bmatrix} r_{01jkm} \\ r_{11jkm} \end{bmatrix} \sim MN\begin{pmatrix} \sigma^2_{r01} & \sigma^2_{r01*11} \\ \sigma^2_{r01*11} & \sigma^2_{r11} \end{pmatrix}$$

$$\begin{bmatrix} r_{02i} \\ r_{12i} \end{bmatrix} \sim MN\begin{pmatrix} \sigma^2_{r02} & \sigma^2_{r02*12} \\ \sigma^2_{r02*12} & \sigma^2_{r12} \end{pmatrix}$$

Model 1 provides information about the effect of word-specific GPC knowledge on decoding accuracy. From this model, we can calculate the probability of accurate decoding when all the GPCs within words are known and when all the word-specific GPCs are not known.

These estimates are theoretically appealing as GPC knowledge is assumed to be a prerequisite for decoding (Byrne & Fielding-Barnsley, 1991; Ehri, 1998; Share, 1995, 1999), but this relationship has yet to be addressed at the word-level.

We next added predictors to Model 1 to determine which person and word factors affected the outcome above and beyond word-specific GPC knowledge (Model 2). See Equation 5 for Model 2.

Level-1 (Responses$_{jkmi}$) $\quad Logit(\pi_{jkmi}) = \lambda_{0jkmi} + \lambda_{1jkmi} GPCK_{jkmi}$ (5)

Level-2 (Person$_{jkm}$ & Word$_{i}$)

$$\lambda_{0jkmi} = \gamma_{00km} + \gamma_{10} PA_j + \gamma_{20} RN_j + \gamma_{30} WM_j + \gamma_{40} R\_FREQ_i + \gamma_{50} R\_SIZE_i + \gamma_{60} FREQ_i * SIZE_i + \gamma_{70} VC_i + r_{01jkm} + r_{02i}$$

$$\lambda_{1jkmi} = \lambda_{10km} + r_{11jkm} + r_{12i}$$

Level-3 (Classroom$_{km}$) $\quad \gamma_{00km} = \phi_{000m} + u_{km}$

Level-4 (School$_{m}$) $\quad \phi_{000m} = \omega_{0000} + \varepsilon_m$

where γ_{10} is the average fixed effect of PA

γ_{20} is the average fixed effect of RN

γ_{30} is the average fixed effect of WM

γ_{40} is the average fixed effect of R_FREQ

γ_{50} is the average fixed effect of R_SIZE

γ_{60} is the average fixed effect of the interaction between FREQ & SIZE

γ_{70} is the average fixed effect of vowel vs. consonant complexity (VC)

Model 2 estimated the person effects of phonemic awareness, rapid naming, and working memory on a student's probability of correctly pronouncing a pseudoword, controlling for word-specific GPC knowledge. It also estimated the effects of average rime-neighborhood frequency, rime-neighborhood size, the interaction of frequency and size, and grapheme complexity on students' decoding accuracy, again controlling for word-specific GPC knowledge. VC_i was a dummy variable which equaled 1 if the pseudoword contained a complex vowel grapheme and 0 if it contained a complex consonant grapheme. Results were examined by first determining the relative importance of each covariate by looking at the *z*-value associated with coefficients γ_1, γ_2, γ_3, γ_4, γ_5, γ_6, and γ_7. Then, we calculated the proportion of person variance accounted for by the three person predictors above and beyond what was explained by word-specific GPC knowledge. Reduction in variance was calculated by ($r_{01jkm\text{MODEL1}}$ - $r_{01jkm\text{MODEL2}}$)/ $r_{01jkm\text{MODEL1}}$. Similarly, we calculated the proportion of word variance accounted for by the four word predictors. Reduction in variance was calculated by ($r_{02i\text{MODEL1}}$ - $r_{02i\text{MODEL2}}$)/ $r_{02i\text{MODEL1}}$. Model 2 included the main effects of person and word covariates and thus it served as a reference model to which the interaction model (Model 3) was compared.

As the final step in our data analysis, we estimated a model (Model 3) to test two person-by-word interactions. One interaction was between working memory and rime frequency and the other was between phonemic awareness and (vowel versus consonant) grapheme complexity (see Equation 6). The generic variable (S/F) appears in Equation 6 because the interaction term was created from the index of frequency (rime-neighborhood size or average rime-neighbor frequency) that accounted for the most word variance as determined by Model 2. The significance of the interactions was determined by the *p*-value associated with the coefficients λ_2 and λ_3 in Equation 6.

Level-1 (Responses$_{jkmi}$) (6)

$$Logit(\pi_{jkmi}) = \lambda_{0jkmi} + \lambda_{1jkmi}GPCK_{jkmi} + \lambda_2 WM_{jkm} * (S/F)_i + \lambda_3 PA_{jkm} * VC_i$$

Level-2 (Person$_{jkm}$ & Word$_i$)

$$\lambda_{0jkmi} = \gamma_{00km} + \gamma_{10}PA_j + \gamma_{20}RN_j + \gamma_{30}WM_j + \gamma_{40}R\_FREQ_i + \gamma_{50}R\_SIZE_i + \gamma_{60}FREQ_i * SIZE_i + \gamma_{70}VC_i + r_{01jkm} + r_{02i}$$

$$\lambda_{1jkmi} = \lambda_{10km} + r_{10jkm} + r_{11i}$$

Level-3 (Classroom$_{km}$) $$\gamma_{00km} = \phi_{000m} + u_{km}$$

Level-4 (School$_m$) $$\phi_{000m} = \omega_{0000} + \varepsilon_m$$

All models were fit using the *lmer* function in the R package lme4 (Bates & Maechler, 2009). *Lmer* employs Laplace approximation, a marginal maximum likelihood approach in which item parameters are estimated first and then person parameters are estimated based on the estimated item parameters. Laplace approximation is computationally efficient as it approximates the integrand as opposed to the integral. Although approximating the integrand tends to underestimate variance when responses are binary in nature and sample sizes are small (Joe, 2008), a simulation study by Cho and Rabe-Hesketh (in press) showed that accuracy was compromised in the case of 100 persons and 10 items but was not compromised in the case of 100 persons and 50 items; both with similar ICCs as expected in the current study. The current study had double the persons and items as the problematic case in the Cho and Rabe-Hesketh study. Neither standard errors nor *p*-values for random effects are produced by the *lmer* function because *lmer* uses likelihood methods which do not make the necessary symmetric parameter-distribution assumptions that are necessary for accurately calculating standard errors and *p*-values (Bates; 2006, May 19 & July 15). Finally, all continuous person and word variables were standardized before they were entered into the models.

CHAPTER III

RESULTS

Descriptive Statistics

Frequency indices were calculated for each pseudoword contained in the TNT assessment. The correlation between average rime-neighbor frequency and rime-neighborhood size was .39. Table 2 contains descriptive statistics for the frequency indices of the pseudowords. Means, medians, and standard deviations in the table indicate both of the frequency variables have skewed distributions; however, this does not pose a problem because models with binary outcomes make no distributional assumptions about independent variables (Tabachnick & Fidell, 2007). Descriptive statistics of the assessment scores are provided in Table 3. Correlations among the student assessment scores are in Table 4.

Table 2

Description of the Frequency Indices of Pseudowords on TNT (N = 20)

	Mean	*Median*	*SD*	*Min.*	*Max.*
Average rime-neighbor frequency	69.22	34.20	100.52	0.00	406.59
Rime neighborhood size	4.90	2.50	5.06	0.00	16.00

Table 3
Description of Student Sample (N = 196)

	Mean	Median	SD	Min.	Max.
GPC total	23.17	23	2.14	12	27
Phonemic awareness	14.05	16	4.98	1	20
Rapid naming	58.07	55	17.10	31	130
TNT total	9.16	9	4.25	0	19
Working memory	5.64	6	4.14	0	16

Table 4
Correlations Among Student Assessment Scores (N = 196)

	1	2	3	4	5
1. GPC total	—				
2. Phonemic awareness	.31	—			
3. Rapid Naming	-.11	-.18	—		
4. TNT total	.52	.36	-.23	—	
5. Working memory	.22	.33	-.24	.23	—

Note. Correlations ≥ |.18| are significant at the $\alpha = .05$ level.

Base Model Results

Before building models to answer each of the research questions, we estimated five potential Base models that included no covariates. For the sake of parsimony, the Full Base model that included all potential sources of variance was compared to four reduced models, one for each omitted random parameter: word, person, classroom, and school. Table 5 contains model fit indices for all five models and the relevant model comparison tests. Although the likelihood ratio test is typically used to compare nested models that share common random variance components, accurate results can be obtained when comparing models that differ in numbers of random effects by halving the *p*-value produced by the LR test (Molenberghs & Verbeke, 2004; Wilson & De Boeck, 2004).

Table 5

Model Fit Indices and Relevant Comparisons Among Base Models (N = 196)

		Model Fit Indices			Model Comparison	
Model	Number of parameters	Deviance	AIC	BIC	Referent Model	χ^2 Test of Equivalence
Full Base	5	4587	4597	4628	n/a	n/a
Base w/o word	4	5069	5077	5102	Full Base	481.81, $p < .01$
Base w/o person	4	4731	4739	4764	Full Base	144.56, $p < .01$
Base w/o school	4	4593	4601	4626	Full Base	5.95, $p < .05$
Base w/o classroom	4	4593	4601	4626	Full Base	6.35, $p < .05$

In all cases, the likelihood ratio tests indicated the reduced Base models fit significantly worse than the Full Base model. Because the Full Base model fit the data better than the reduced Base models, all random effects were retained in the remaining models. Estimates of the fixed and random effects for the Full Base model are in Table 6. We used Equations 3a-3d to calculate the ICCs for each source of variance. The ICCs for persons, classrooms, and schools, conditional on word difficulty, were 14.36%, 4.54%, and 7.05%, respectively. The proportion of variance between words was 20.03%, conditional on student ability.

Table 6

Results from Full Base Model (N = 196)

Fixed Effect				
Variable	Coefficient	*SE*	*z*	*p*
Intercept, ω_{000}	-0.31	0.27	-1.14	0.25
Random Effects				
Source	Variance	*SD*		
Person (intercept), σ^2_{r01}	0.64	0.80		
Classroom (intercept), σ^2_{u}	0.20	0.45		
School (intercept), σ^2_{τ}	0.31	0.56		
Word (intercept), σ^2_{r02}	0.82	0.91		

Note. Coefficients are in log-odds units. Level-1 variance assumed to be $\pi^2/3$.

Though the coefficients in the tables are in log-odds units, they were converted to probabilities in the text for ease of interpretation using the formula: $1/(1+\exp[-\gamma_0])$. Results from the Full Base model revealed that the average predicted probability for an average person ($r_{01jkm} = 0$) in an average classroom ($u_{km} = 0$) in an average school ($\varepsilon_m = 0$) decoding a word of average difficulty ($r_{02i} = 0$) was .42, 1/(1+exp[-{-0.31}]). However, variability around this average existed among persons and words (see σ^2_{r01} and σ^2_{r02} in Table 6). One way of describing the magnitude of variation of correct responses across persons and words is to compute the 95% plausible value range (Hoffman & Rovine, 2007; Raudenbush & Bryk, 2002). The 95% plausible value range for persons was .13 to .78 ($1/[1+\exp\{-\omega_{000} \pm 1.96 * \sqrt{\sigma^2_{r01}}\}]$) for the predicted probability of getting an average word correct. For words, the range was .11 to .81 ($1/[1+\exp\{-\omega_{000} \pm 1.96 * \sqrt{\sigma^2_{r02}}\}]$) for the predicted probability of an average person. The variability here highlights the advantage of using a person-and-word analysis to model the variation in responses across persons and words instead of choosing one to ignore in the case of aggregated analyses. In a person-analysis, for instance, correct responses are aggregated over items and each person is assigned a total score. What the Base model shows for these data is that the there is variability in responses across words (σ^2_{r02} in Table 6) that we can model and attempt to explain – information that would be lost in an aggregate, person analysis.

Results for Research Questions

Research question 1

The aim of research question 1 was to determine if the probability of correctly decoding a pseudoword was different depending on a person's knowledge or lack of knowledge of all the GPCs in that word. Therefore, Model 1 contained the GPCK variable that was specific to each person and each word. Model 1 results are in Table 7. The *z*-statistic for the GPCK variable indicated that there was a significant (at the $\alpha = .05$ level) difference in the probability of a correct response when word-specific GPCs were known versus not known. Because GPCK is a dichotomous variable, the intercept in the model represents the case in which word-specific GPCs were not known (e.g., when GPCK = 0). In that case, the average predicted probability of correct decoding was .35, 1/(1+exp[-{-0.64}]), for an average person and an average word. On the other hand, the predicted probability of correct decoding when word-specific GPCs were known was .46, 1/(1+exp[-{-0.64 + .48}]), for an average person and an average word. This effect of word-specific GPC knowledge varied substantially across students and words (see σ^2_{r11} and σ^2_{r12} in Table 7); the 95% plausible value range for the predicted difference of correctly decoding a word when a student knew versus did not know the GPCs ranged from 6.27% to 80.64% ($1/[1+\exp\{-\lambda_1 \pm 1.96^*\sqrt{\sigma^2_{r11}}\}]$) on an average word. In other words, some students were greatly affected by GPC knowledge, and it hardly mattered at all for others. The effect of GPC knowledge also varied by word as the predicted difference ranged from 39.91% to 79.63%

$(1/[1+\exp\{-\lambda_1 \pm 1.96*\sqrt{\sigma^2_{r12}}\}])$ for an average person. Findings suggest that although full knowledge of word-specific GPCs is an important predictor of accurate decoding, it is neither a necessary nor sufficient condition for every person to accurately decode every word; other factors are at work.

Table 7

Results from Model 1 (N = 196)

Fixed Effect				
Variable	Coefficient	*SE*	*z*	*p*
Intercept, ω_{000}	-0.64	0.29	-2.21	.03
GPCK, λ_1	0.48	0.16	2.97	.00
Random Effects				
Source	Variance	*SD*		
Person (intercept), σ^2_{r01}	1.11	1.05		
Person (GPCK), σ^2_{r11}	0.47	0.69		
Classroom (intercept), σ^2_{u}	0.20	0.45		
School (intercept), σ^2_{τ}	0.26	0.51		
Word (intercept), σ^2_{r02}	0.92	0.96		
Word (GPCK), σ^2_{r12}	0.20	0.45		

Note. Coefficients are in log-odds units.

The negative correlations between the variance components of the intercept and the GPCK slope for both persons and words (σ_{r01*11}^2 = -.73 and σ_{r02*12}^2 = -.45, respectively) also suggest there are factors affecting correct decoding other than word-specific GPC knowledge. On the person side, the negative correlation between the variability around the probability of a correct response and the variability around the effect of word-specific GPC knowledge on a correct response suggests that when there was much variability among students in getting a word correct, there was not much variability across students in the effect of GPC knowledge on that word. This means that the effect of word-specific GPC knowledge behaved as predicted in that if students knew the GPCs in a word, they were mostly predicted to decode it correctly and if students did not know the GPCs in a word, they were mostly predicted to decode it incorrectly.

Conversely, when there was not much variability around the probability of students getting a correct response for a word, there was great variability in the effect of word-specific GPC knowledge. What this latter case means is that if students mostly got a word correct or mostly got it incorrect, then the way GPC knowledge affected that correct response was inconsistent. In the mostly correct situation, students were getting the word right regardless of word-specific GPC knowledge. Similarly in the mostly incorrect situation, students were getting the word wrong regardless of word-specific GPC knowledge. Keep in mind that the situations described here pertain to performance on an average word. The same pattern of results held true for the word side as well. The additional person and word factors assessed in research questions 2 and 3 were entered to help explain what other than word-specific GPC knowledge is important for explaining variability accurate decoding.

Research questions 2 and 3

The aim of research question 2 was to determine the relative importance of person and word factors in their contribution to decoding accuracy while controlling for a student's knowledge of a word's GPCs. Model 2 included word-specific GPC knowledge plus the following person variables: phonemic awareness, rapid naming, and working memory. To answer research question 3, word variables were also included in Model 2: average rime-neighbor frequency, rime-neighborhood size, frequency*size, and grapheme (vowel versus consonant) complexity. Results are in Table 8. First, we discuss results pertaining to research question 2. Table 8 reveals that controlling for word-specific GPC knowledge, phonemic awareness (γ_{10} = 0.28, z = 3.69, p < .01) and rapid naming skill (γ_{20} = -0.17, z = -2.29, p < .05) were significant predictors of decoding accuracy, while working memory was not (γ_{30} = 0.03, z = 0.33, p > .05).

The null effect of working memory may be attributable to the short length of the words included in the sample. Because working memory has a limited capacity and is affected by the amount of information it must attend to (Siegel, 1994), we suspect working memory might play a role in decoding words that are longer than the 4-letter pseudowords included in the current study. In favor of this explanation, Abu-Rabia et al. (2003), who found an effect of working memory on decoding, measured decoding skill with an assessment that contained pseudowords up to five syllables long. On the other hand, Conners et al. (2001) found an effect of working memory on one-syllable words. The authors, however, did not consider any covariates other than age in the analysis, which means working memory may have served as a proxy for other related reading skills. Whether or not short words mask the effect of working memory, our finding of its null effect adds to the growing body of mixed conclusions regarding the relationship between working memory and reading difficulties (see review by Savage, Lavers, & Pillay, 2007). The effect of working memory notwithstanding, the other person factors functioned as expected.

As hypothesized, phonemic awareness and rapid naming had significant influence on decoding. Phonemic awareness was the most influential. Compton, DeFries, and Olson (2001) also found phonemic awareness to be more influential than rapid naming on measures of decoding accuracy. Controlling for word-specific GPC knowledge, a one standard deviation increase in phonemic awareness was associated with a 56.87% increase in the predicted probability of accurate decoding for an average word for a person with average skill in rapid naming and working memory. In our sample of at-risk readers, phonemic awareness was even more influential than word-specific GPC knowledge (z = 3.69 and z = 2.75, respectively), which is exactly what Byrne and Fielding-Barnsley (1993) found in their first grade sample of typical readers. Compton (2000), however, found the opposite pattern for GPC knowledge and phonemic awareness. The discrepancy in findings may be due to the difference in phonemic awareness tasks (phoneme segmentation and blending versus phoneme identification) or the difference in predictive versus concurrent measurement. Still, by all accounts, both GPC knowledge and phonemic awareness are critical to early literacy development. Rapid naming also appears to be an influential factor, though not quite as robust (z = -2.29) as phonemic awareness. Controlling for all other variables, a one standard deviation increase in the speed of rapid letter naming was associated with a 45.66% increase in the predicted probability of the correct decoding of an average word.

As a block, the person variables accounted for 16.61% of the person variance compared to the model with only word-specific GPC knowledge (Model 1). Though phonemic awareness and rapid naming clearly affect decoding accuracy, the amount of unexplained variance indicates there may be additional person factors that affect decoding accuracy that were not accounted for in the model.

Table 8

Results from Model 2 (N = 196)

Fixed Effect

Variable	Coefficient	*SE*	*z*	*p*
Intercept, ω_{000}	-0.09	0.31	-0.29	.77
GPCK, λ_1	0.44	0.16	2.75	.01
Phonemic awareness, γ_{10}	0.28	0.07	3.69	.00
Rapid naming, γ_{20}	-0.17	0.08	-2.29	.02
Working memory, γ_{30}	0.03	0.08	0.33	.74
R_FREQ, γ_{40}	0.76	0.21	3.52	.00
R_SIZE, γ_{50}	-0.10	0.17	-0.60	.55
FREQ*SIZE, γ_{60}	-0.48	0.22	-2.13	.03
Grapheme complexity, γ_{70}	-0.70	0.29	-2.38	.02

Random Effects

Source	Variance	*SD*
Person (intercept), σ^2_{r01}	0.93	0.96
Person (GPCK), σ^2_{r11}	0.45	0.67
Classroom (intercept), σ^2_u	0.19	0.44
School (intercept), σ^2_τ	0.14	0.37
Word (intercept), σ^2_{r02}	0.77	0.88
Word (GPCK), σ^2_{r12}	0.21	0.37

Note. Coefficients are in log-odds units. R_FREQ = average rime-neighbor frequency. R_SIZE = rime-neighborhood size.

As for the word factors in research question 3, it appears that although words with more frequent rimes were decoded at a higher probability than words with less frequent rimes (γ_{40} = 0.76, z = 3.52, p < .01) (also found by Treiman et al., 1990), that effect was moderated by rime-neighborhood size (γ_{60} = -0.48, z = -2.13, p < .05). Figure 2 illustrates this interaction. This finding corroborates that of Calhoon and Leslie (2002) and Leslie and Calhoon (1995) that young children are affected by a word's rime-neighborhood size when the word is found infrequently in text. Results from Model 2 also reveal words with complex vowel graphemes had a lower predicted probability of correct decoding than complex consonant graphemes (γ_{70} = -0.70, z = -2.38, p < .05). Controlling the frequency of the word, when the word-specific GPCs were known, the predicted probability for an average person to correctly decode a word with a complex vowel was .41 but was .59 for a word with a complex consonant. When word-specific GPCs were not known, the predicted probabilities were .31 and .48, respectively, for complex vowels and complex consonants. This finding is aligned with that of Treiman et al. that vowels are the most problematic type of grapheme in terms of decoding. The present finding is particularly interesting because knowledge of the word's GPCs was controlled in the analysis. That means that although students knew the phonemes associated with the complex vowel and consonant graphemes, they still had more difficulty decoding words that contained the complex vowel grapheme than the complex consonant grapheme. Altogether, the four word covariates explained 16.50% of the word variance after accounting for word-specific GPC knowledge (Model 1). The proportion of unexplained word variance suggests that additional word factors should be included in future analyses.

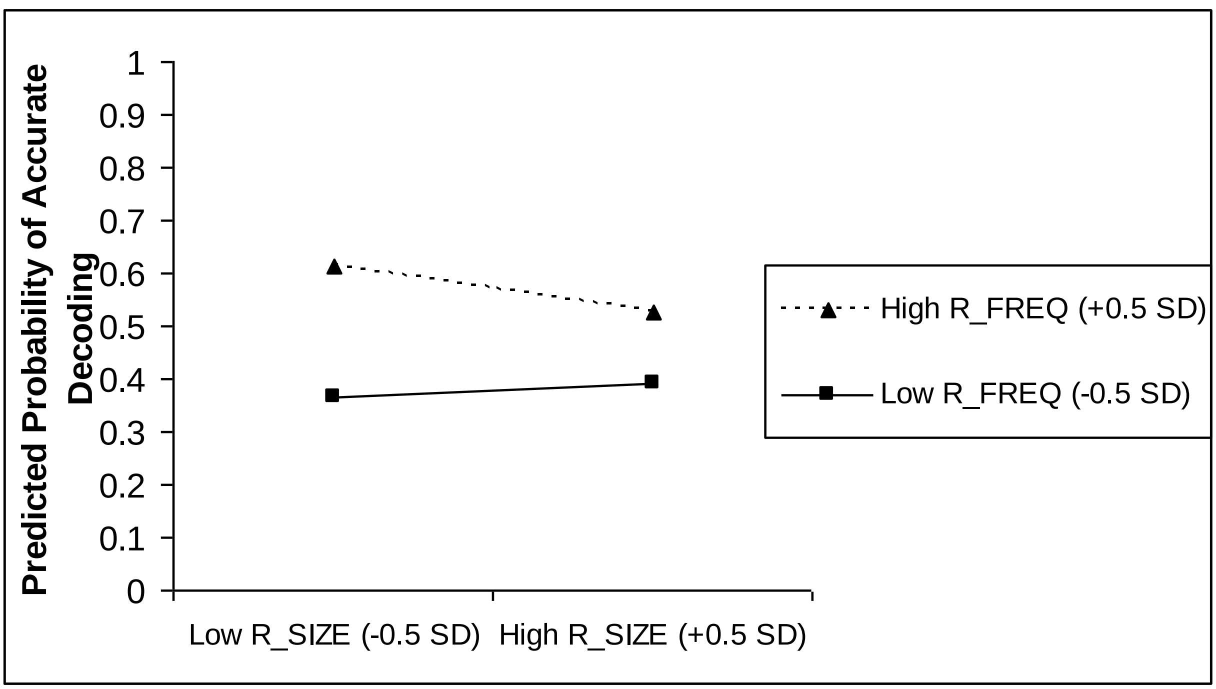

Figure 2. Graph of interaction between average rime-neighbor frequency and rime-neighborhood size

In general, findings from Model 2 corroborate findings from past studies that have used more traditional, total-score analyses in terms of person (Byrne & Fielding-Barnsley, 1991; Compton, 2000, 2003; Manis et al., 2000; Treiman et al., 1990) and word effects (Calhoon & Leslie, 2002; Leslie & Calhoon, 1995; Treiman et al.). The interaction of average rime-neighbor frequency and rime-neighborhood size extends the work of Calhoon and Leslie (2002) and Leslie and Calhoon (1995) by suggesting that it is not only the frequency of the word that matters but also the frequency of the word's neighbors. Further research on the importance of average rime-neighbor frequency in real word decoding may need to be conducted to determine if its effect is specific to pseudowords, which have no frequency of their own.

Research question 4

Two person-by-word interactions were assessed under research question 4. The first interaction was between working memory and frequency of the rime unit. We assume that students are more likely to decode words in a grapheme-by-grapheme manner when the words contain infrequent rime units (Treiman et al., 1990). Conversely, students may be able to use some type of analogy strategy when decoding words with frequent rime units. Because the grapheme-by-grapheme decoding process may be more taxing on working memory than analogy decoding (Siegel & Faux, 1989), students with poorer working memory may have more difficulty with infrequent rimes than students with adequate working memory. Therefore, we hypothesized that the accuracy of decoding might be more discrepant between words with infrequent rimes than frequent rimes for students with poorer working memory than would be the case with students with better working memory.

We chose to use average rime-neighbor frequency as opposed to rime-neighborhood size to create the working memory by frequency interaction because the former was a significant predictor of decoding accuracy in Model 2. Results for the interaction model are in Table 9. As the table shows, there was not a significant interaction of working memory and rime frequency ($\lambda_2 = 0.01$, $z = 0.26$, $p > .05$). Similar to the speculation of the null effect of working memory in research question 3, the interaction of working memory and rime frequency may exist in lengthier words in which working memory is more heavily taxed.

The second interaction was between phonemic awareness and grapheme (vowel versus consonant) complexity. Past research has shown a trend for students with poorer phonological skills to have more difficulty with complex vowel graphemes than complex consonant graphemes (Marinus & de Jong, 2008). The interaction entered into Model 3 was evaluated in the presence of person and word random effects so that the variability of both was taken into account. Results in Table 9 confirm that there was indeed a significant interaction between phonemic awareness and grapheme complexity ($\lambda_3 = 0.17$, $z = 2.18$, $p < .05$). Model 3 was re-estimated without the non-significant Working Memory X Average Rime-Neighbor Frequency interaction so that a likelihood ratio test could provide secondary evidence that a model including the phonemic awareness by grapheme complexity interaction fit the data better than a model without it. Results of the likelihood ratio test comparing Model 2 and Model 3 with only the Phonemic Awareness X Grapheme Complexity interaction confirmed the necessity of the interaction, $\chi^2 = 4.75$, $p < .05$. Figure 3 depicts the interaction.

Table 9

Results from Model 3 (N = 196)

Fixed Effect

Variable	Coefficient	*SE*	*z*	*p*
Intercept, ω_{000}	-0.08	0.31	-0.26	.79
GPCK, λ_1	0.44	0.16	2.77	.01
Phonemic awareness, γ_{10}	0.20	0.08	2.44	.01
Rapid naming, γ_{20}	-0.17	0.08	-2.29	.02
Working memory, γ_{30}	0.03	0.08	0.33	.74
R_FREQ, γ_{40}	0.77	0.22	3.58	.00
R_SIZE, γ_{50}	-0.11	0.17	-0.66	.51
FREQ*SIZE, γ_{60}	-0.49	0.22	-2.19	.03
Grapheme complexity, γ_{70}	-0.72	0.29	-2.43	.01
Working memory*R_FREQ, λ_2	0.01	0.04	0.26	.80
Phonemic awareness* grapheme complexity, λ_3	0.17	0.08	2.18	.03

Random Effects

Source	Variance	*SD*
Person (intercept), σ^2_{r01}	0.92	0.98
Person (GPCK), σ^2_{r11}	0.46	0.68
Classroom (intercept), σ^2_{u}	0.19	0.44
School (intercept), σ^2_{τ}	0.14	0.37
Word (intercept), σ^2_{r02}	0.79	0.89
Word (GPCK), σ^2_{r12}	0.20	0.45

Note. Coefficients are in log-odds units. R_FREQ = average rime-neighbor frequency. R_SIZE = rime-neighborhood size.

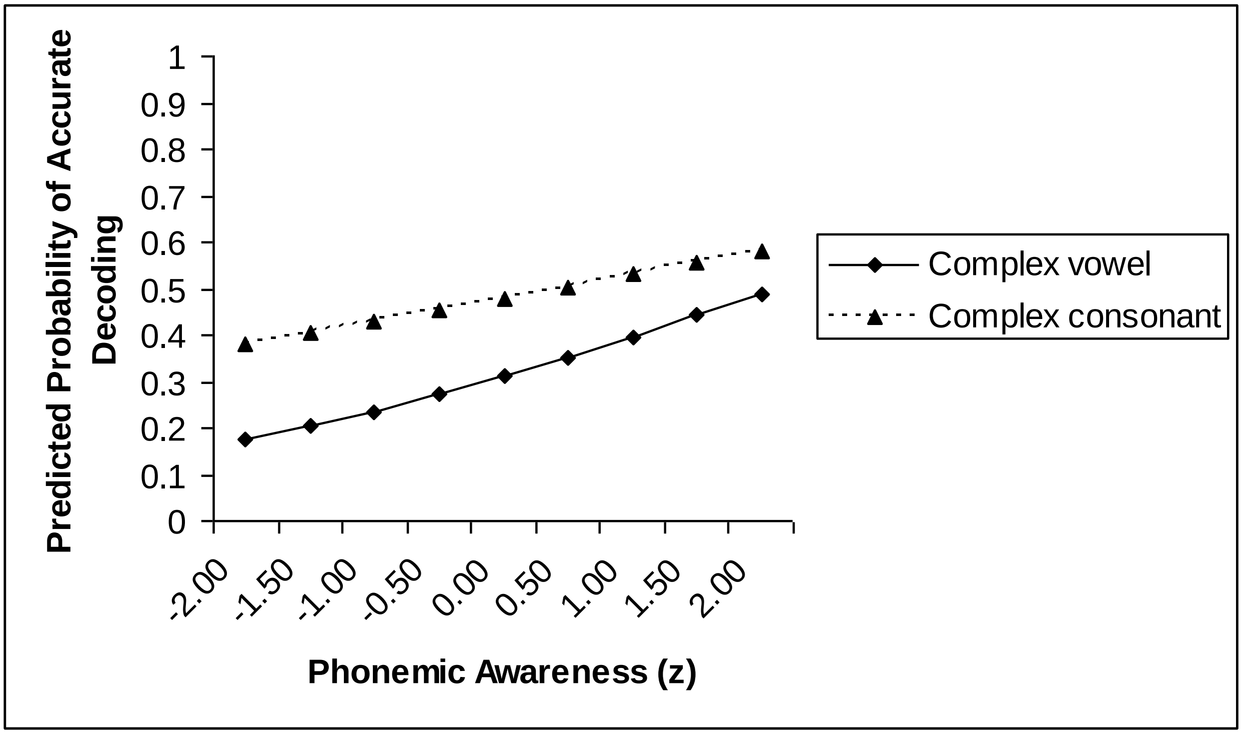

Figure 3. Graph of interaction between phonemic awareness and grapheme complexity

The graph illustrates that (for an average word) students in general had a lower predicted probability of correctly decoding a word with a complex vowel than a word with a complex consonant, but that the difference between the two types of words was most prominent for students with lower phonemic awareness scores.

CHAPTER IV

DISCUSSION

General Conclusions

The aim of the current study was to extend the literature on decoding by bringing together two lines of research, namely person and word factors that affect decoding, using a crossed random-effects model. Results generally confirm what has been found in prior, total-score analyses: Accurate decoding is influenced by both person and word characteristics. One advantage of examining both types of effects in the same statistical model is the ability to explore person-by-word covariates and interactions between person and word factors. GPC knowledge, for example, was treated as person-and-word specific so that the relation between a student's knowledge of a word's GPCs and the student's correct decoding of that word could be examined. As expected, students who knew all the GPCs within a word had a higher predicted probability of decoding that word accurately compared to students who did not know all the word's GPCs. This finding provides further support of the stage theories of reading development that insist GPC knowledge is crucial in learning to read (Chall, 1983; Ehri, 1991; Frith, 1985; Gough & Hillinger, 1980; Mason, 1980; Perfetti, 1992).

We found it interesting that the positive relation between GPC knowledge and decoding does not necessarily hold for every word. This means it is neither necessary nor sufficient for all the GPCs in a word to be known before accurate decoding takes place. Some students in our study knew all the GPCs within a word yet decoded it inaccurately and some students did not know all the GPCs within a word yet were able to decode it accurately. For instance, there were

33 cases in which students knew the phonemes associated with the graphemes *v*, *ou*, and *n* when given the AGCPK test but were unable to correctly decode the pseudoword *voun* on the TNT. On the other hand, there were 23 cases in which students did not know all the phonemes associated with graphemes in the pseudoword *voun* yet were still able to correctly decode the pseudoword. This finding provides evidence in favor of the psycholinguistic grain size theory that states children are able to use orthographic information of different sizes (e.g., single letters, pairs of letters, clusters of letters, etc.) when decoding unfamiliar words (Ziegler & Goswami, 2005). This finding also suggests that GPC knowledge need not be explicit to be useful. It is clear from the cases in which students were unable to produce the correct phoneme when shown a grapheme in isolation (on the AGPCK task) but were able to correctly produce the phoneme when it was embedded in a nonword (on the TNT task) that in some cases implicit GPC knowledge is sufficient to decode unfamiliar words. This is not to suggest that GPCs should not be taught explicitly but that the nature of the relation between GPC knowledge and accurate decoding may be more complex than our current, dichotomous measure of GPC knowledge allows us to fully explore. Perhaps a continuum of GPC quality (i.e., stability, impenetrability) would be a more appropriate measure (Perfetti, 1992).

When students were able to correctly decode a word despite not knowing all the GPCs within that word, it may be that they were using other word features (e.g., familiar rime unit) to aid their decoding. Based on a count of the raw data, when students did not know all the GPCs in words from larger neighborhoods (M = 8.90 neighbors) the proportion of an accurate decoding was .35. This proportion fell to .23 when the word was from a smaller neighborhood (M = 0.90 neighbors). That means when students did not know the GPCs in a word, they were more likely to get the word correct if the word had an abundance of rime-neighbor information that the student could use (implicitly or explicitly) to accurately decode the unknown word. Students were apparently able to use more than just rime information, though. More than half (45/72) of the cases in which students were able to decode small-neighborhood words despite not knowing the GPCs within the word was a result of students pronouncing the pseudoword *goob* as if they were pronouncing the real word *good* but then changing the ending sound. These students were not able to produce the /*uu*/ phoneme when shown the grapheme *oo* in the AGPCK task, but they were able to use the /*uu*/ phoneme to make an accurate pronunciation of *goob* (which could have been correctly pronounced with the /*oo*/ or /*uu*/ phoneme). This indicates that children are sensitive to more than just rime similarities in words and can use various pieces of orthographic information to decode new words.

In addition to word-specific GPC knowledge, students who had higher phonemic awareness scores were more likely to decode pseudowords accurately than students with lower scores. Of all the student factors, phonemic awareness was the most influential. This is not surprising considering the convincing prior research that has dubbed GPC knowledge and phonemic awareness co-requisites of successful decoding (Byrne & Fielding-Barnsley, 1991; Bradley & Bryant, 1983; Ehri & Sweet, 1991; Hatcher et al., 1994). In fact, phonemic awareness appears to have an even greater influence on decoding accuracy than word-specific GPC knowledge. The supremacy of phonemic awareness over GPC knowledge was reported by Byrne and Fielding-Barnsley (1993) and Hatcher et al. (1994) as well.

In terms of person factors, there also appears to be a relation between rapid naming skill and accurate decoding above and beyond the relation between phonemic awareness and decoding. The unique contribution of rapid naming, along with a low correlation between phonemic awareness and rapid naming (-.18), lends support to the double-deficit hypothesis in which Wolf and Bowers (1999, 2000) claim that phonological and rapid naming deficits make relatively independent contributions to reading difficulty. Although the person effects mentioned here certainly impact accurate decoding, the skills people bring to the decoding situation are only half the story; the types of words being decoded also matter.

According to word acquisition theories, the process of acquiring unfamiliar words is a word-by-word endeavor rather than a developmental stage of reading (Share, 1995; Perfetti, 1992). Our results corroborate those theories by demonstrating that variability of accurate decoding exists across words and that word characteristics significantly affect decoding. One of those characteristics is shared rime information (McClelland & Rumelhart, 1981). We found that pseudowords containing frequently-encountered rime units (e.g., *nish*) were easier to decode than pseudowords with infrequently shared rimes (e.g., *bosh*). Although the main effect of rime-neighborhood size was not a significant factor in decoding, there was an interaction between rime-neighborhood size and average rime-neighbor frequency indicating that children are affected by a word's rime-neighborhood size when the word is found infrequently in text (Leslie and Calhoon, 1995; Calhoon & Leslie, 2002). In addition, words with complex consonant graphemes (e.g., *jeth*) were easier to decode than words with complex vowel graphemes (e.g., *dail*), which confirms the trend found by Marinus and de Jong (2008). Whereas prior research has shown that complex vowel and consonant graphemes are generally more difficult than their respective simple grapheme types (Marinus & de Jong; Olson et al., 1994), our comparison between the two complex types provides information about which type is the more difficult one to decode (i.e., word with complex vowel graphemes). In general, it appears that a word's rime information may have a stronger impact on decoding than individual grapheme information.

Implications

Results from the current study have implications for reading instruction and research. Regarding instruction, our results suggest that students with poorer phonemic awareness have lower predicted probabilities of decoding words accurately regardless of their GPC knowledge. For example, there were fewer cases in which students with below average phonemic awareness decoded words correctly even when they knew the associated GPCs (29.06%) than cases in which students with above average phonemic awareness knew the GPCs and decoded correctly (42.72%). Siegel and Faux (1989) reported a similar phenomenon. In their study, students with a reading disability compared to students without a disability were less likely to correctly decode pseudowords that contained the same letters that were in real words that they could read. Therefore, students with poor phonemic awareness skills may need additional instruction in applying GPCs once the GPCs have been acquired (see Cunningham, 1990, for a similar argument with regard to phonemic awareness). This conclusion is in agreement with the recommendation by the National Reading Panel (National Institute for Literacy, n.d.) that best practice in early reading instruction should include the concurrent instruction of phonemic awareness and phonics (i.e., GPCs). Our conclusion is also in line with Hatcher et al. (1994) who found that training in phonology plus phonics was more effective for poor readers than training in either skill alone. Another possible explanation is that students with poorer phonemic awareness may require more time and practice to learn GPCs to a degree that the GPCs can be used when decoding unfamiliar words (Bransford & Schwartz, 1999). In this vein, Compton et al. (2005) found that students with reading disabilities who exhibited the greatest growth in treatment-aligned measures were the most likely to display gain in other areas of reading. This may suggest a certain level of mastery is required before growth in one reading skill can promote growth in related reading skills, and that struggling readers should be provided more time to master GPCs before being asked to apply them when decoding new words.

We mentioned previously that although students are able to use rime-unit familiarity to help them decode unfamiliar words, there was also some evidence in our data to suggest that students are able to use other parts of the word as well. This appears particularly true for words containing low-frequency rimes. One potential instructional implication is to group words that have infrequent rime units with other words using a common feature other than the rime unit. For instance, the word *seem* has no rime neighbors in K-3 texts (Zeno et al., 1995), but there are several words beginning with see- that could be helpful to a student trying to decode *seem*: see, seed, seek, seen, seep, and sees. To confirm the conclusion drawn here, additional research is needed on the extent to which children can use onset (all consonants prior to a vowel in a single-syllable word) plus vowel familiarity to help them decode new words.

The crossed random-effects model allowed us to ask questions regarding which types of words might be especially hard for which types of readers. In the current study, for example, the interaction between phonemic awareness and grapheme complexity suggests that the extra instruction recommended for students with poor phonemic awareness should specifically include practice in applying complex vowel GPCs when decoding new words. In our data, among the students with below average phonemic awareness, fewer read *keab* (12.82%) correctly than *sath* (58.21%) despite knowing the respective word-specific GPCs. Modeling the relationships between types of educational materials and types of students could help match students with the most appropriate instruction, a goal shared by researchers who assess person-level aptitude-by-treatment interactions.

Another implication for future research is using the crossed random-effects model to test the self-teaching hypothesis (Jorm & Share, 1983; Share, 1995). According to the self-teaching hypothesis, individuals acquire orthographic representations of words through a phonological recoding self-teaching mechanism. Phonological recoding (a.k.a. decoding) is the process of translating a printed word to speech via access to the phonological loop. In the phonological loop, written symbols are associated with, or mapped onto, phonological information. Therefore, each time individuals are successful in decoding a word, they (implicitly) teach themselves word-specific and general orthographic knowledge.

In past studies of the self-teaching hypothesis, researchers have almost exclusively used total-score analyses to assess the relation between accurate decoding and word-specific orthographic learning. Generally, researchers record the percentage of targets that were correctly decoded by the participant during the training and the number of correct target choices the participant made on the orthographic choice task, and these two total scores are correlated. This correlation describes the relation between the percentage of targets that were correctly decoded and the percentage of correct choices in target spellings. The problem with the total-score correlation is that it is unable to properly address word-specific orthographic learning because it does not address the precise relationship between the accuracy of decoding a particular target and later choosing that target over its homophone or other visually similar choices.

To properly assess the effect of decoding accuracy on word-specific orthographic learning, a person-and-word analysis must be conducted. In a crossed random-effects model, decoding could be treated as a person-by-word covariate so that one could get a more detailed understanding of how a person's accuracy in decoding a particular word during training predicts whether or not that person will identify that word's correct spelling on an orthographic choice task. This information would provide a critical test of the self-teaching mechanism by investigating the importance of accurate decoding in word-specific orthographic learning and which, if any, other person or word factors contribute to that orthographic learning.

Limitations

Findings and conclusions from the present study should be interpreted in light of the study limitations. First, we relied exclusively on pseudowords to draw conclusions about decoding in general. The universal unfamiliarity of pseudowords is favorable for detecting true decoding skill that is not dependent on vocabulary knowledge or sight word recognition. Though decoding pseudowords is often used as a proxy to study how children decode unfamiliar words in general, there are limitations for doing so. Children have no prior knowledge of the pronunciation or meaning of pseudowords before they are asked to decode them. In cases of decoding unfamiliar *real* words, children are often familiar with the phonology and the meaning of a word even if they are not familiar with the word's orthography (i.e., written form). As an example, children are able to say the word *dog* and understand the definition of *dog* long before they recognize that the letter pattern *d-o-g* represents the sound and meaning of a word they already know. This discrepancy between real and pseudowords means that there may be predictors of real word decoding that are deemed superfluous when predicting pseudoword decoding. For example, vocabulary knowledge might be expected to influence real word decoding but not pseudoword decoding. Nation and Snowling (1998), for instance, found that children with poor vocabulary knowledge were less accurate when reading exception words than students with more extensive vocabulary knowledge even when both groups were matched on phonological skill and nonword decoding accuracy.

One solution for eliminating the real word-pseudoword discrepancy in examining decoding skill is to have children decode real words that are unfamiliar in written form but familiar in oral form. This could be accomplished by compiling a list of words that are both (a) likely to be in the oral vocabulary of the sample being assessed and (b) amenable to pseudohomophone spellings. For instance, the words *boat*, *fish*, and *clock* are familiar in sound and meaning to most all first grade students. To assess first graders' decoding skills, one could present the students with the pseudohomophones *bote*, *phish*, and *klok* so that the written forms are unfamiliar, but the phonology and meaning of the words are familiar.

Another limitation was that working memory was assessed approximately seven months prior to the assessment of the other skills. This occurred because students were involved in a larger intervention project that had a set testing schedule. Though it would have been ideal to assess working memory closer to the time the other skills were measured, we have no reason to expect that students' relative standing on working memory would have been different at the end of the year than at the beginning of the year. This means that although students' working memory scores may have increased throughout the school year, their rank order was not expected to change. Therefore, we believe that the relation between end-of-year working memory and end-of-year decoding would be equivalent to what we found between beginning-of-year working memory and end-of-year decoding.

APPENDIX

	Complex vowel grapheme	Complex consonant grapheme
infrequent rime unit	goob naif feem voun keab	naff fidd bosh jeth tesh
frequent rime unit	teek soom pean dail mout	dess guck nish vell sath

REFERENCES

Abu-Rabia, S., Share, D., & Mansour, M. S. (2003). Word recognition and basic cognitive processes among reading-disabled and normal readers in Arabic. *Reading and Writing, 16,* 423-442.

Adams, R. J., Wilson, M., & Wu, M. (1997). Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics, 22,* 47-76.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412.

Baddeley, A. (1986). Working memory. In D. E. Broadbent, J. L. McGaugh, N. J. Mackintosh, M. I. Posner, E. Tulving, & L. Weiskrantz (Series Eds.)., *Oxford Psychology Series No. 11.* Oxford: Clarendon Press.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158-173.

Bates, D. (2006, May 19). [R] lmer, p-values and all that. Message posted to https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html

Bates, D. (2006, July 15). [R] storing the estimates from lmer. Message posted to http://www.mail-archive.com/r-help@stat.math.ethz.ch/msg67236.html

Bates, D., & Maechler, M. (2009). The lme4: Linear mixed-effects models using S4 classes. URL: http://lme4.r-forge.r-project.org/.

Booth, J. R., & Perfetti, C. A. (2002). Onset and rime structure influences naming but not early word identification in children and adults. *Scientific Studies of Reading, 6,* 1-23.

Bowers, P. G., Sunseth, K., & Golden, J. (1999). The route between rapid naming and reading progress. *Scientific Studies of Reading, 3,* 31-53.

Bowey, J. A. (1990). Orthographic onsets and rimes as functional units of reading. *Memory & Cognition, 18*, 419-427.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read - a causal connection. *Nature, 301*(3), 419-421.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A. Iran-Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 24, pp. 61-100). Washington, D.C.: American Educational Research Association.

Briggs, D. C., & Wilson, M. (2007). Generalizability in item response modeling. *Journal of Educational Measurement*, *44*, 131-155.

Brown, G. D. A., & Deavers, R. P. (1999). Units of analysis in nonword reading: Evidence from children and adults. *Journal of Experimental Child Psychology, 73,* 208-242.

Byrne, B., & Fielding-Barnsley, R. (1991). Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology, 83,* 451-455.

Byrne, B., & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology, 85,* 104-121.

Byrne, B., Fielding-Barnsley, R., Ashley, L., & Larsen, K. (1997). Assessing the child's and the environment's contribution to reading acquisition: What we know and what we don't know. In B. A. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (pp. 265-285). Mahwah, NJ: Lawrence Erlbaum Associates.

Calhoon, J. A., & Leslie, L. (2002). A longitudinal study of the effects of word frequency and rime-neighborhood size on beginning readers' rime reading accuracy in words and nonwords. *Journal of Literacy Research, 34,* 39-58.

Carlisle, J. F., & Katz, L. A. (2006). Effects of word and morpheme familiarity on reading of derived words. *Reading and Writing, 19,* 669-693.

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill.

Cho, S.-J., & Rabe-Hesketh, S. (in press). Alternating imputation posterior estimation of models with crossed random effects. *Computational Statistics and Data Analysis.*

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior, 12,* 335-359.

Compton, D. L. (2000). Modeling the growth of decoding skills in first-grade children. *Scientific Studies of Reading, 4,* 219-259.

Compton, D. L. (2003). Modeling the relationship between growth in rapid naming speed and growth in decoding skill in first-grade children. *Journal of Educational Psychology, 95,* 225-239.

Compton, D. L., DeFries, J. C., & Olson, R. K. (2001). Are RAN- and phonological awareness-deficits additive in children with reading disabilities? *Dyslexia, 7,* 125-149.

Compton, D. L., Olinghouse, N. G., Elleman, A., Vining, J., Appleton, A. C., Vail, J., et al. (2005). Putting transfer back on trial: Modeling individual differences in the transfer of decoding-skill gains to other aspects of reading acquisition. *Journal of Educational Psychology, 97,* 55-69.

Conners, F. A., Atwell, J. A., Rosenquist, C. J., & Sligh, A. C. (2001). Abilities underlying decoding difference in children with intellectual disability. *Journal of Intellectual Disability Research, 45,* 292-299.

Cunningham, A. E. (1990). Explicit versus implicit instruction in phonemic awareness. *Journal of Experimental Child Psychology, 50,* 429-444.

De Boeck, P. (2008). Random item IRT models. *Psychometrika, 73,* 533-559.

De Boeck, P. & Wilson, M. (Eds.). (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer.

Ehri, L. C. (1991). Development of the ability to read words. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 383-417). Mahwah, NJ: Lawrence Erlbaum Associates.

Ehri, L. C. (1998). Grapheme-phoneme knowledge is essential for learning to read words in English. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 3-41). London: Erlbaum.

Ehri, L. C., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading & Writing Quarterly, 14,* 135-163.

Ehri, L. C., & Sweet, J. (1991). Fingerpoint-reading of memorized text: What enables beginners to process the print? *Reading Research Quarterly, 26,* 442-462.

Ehri, L. C., & Wilce, L. S. (1985). Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly, 20,* 163-179.

Fox, J.-P. (2005). Multilevel IRT using dichotomous and polytomous response data. *British Journal of Mathematical and Statistical Psychology, 58,* 145-172.

Fredericksen, J. R., & Kroll, J. F. (1976). Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 2,* 361-379.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall, & M. Coltheart (Eds.), *Surface dyslexia.* London: Erlbaum.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology, 42,* 73-83.

Goswami, U. (1998). The role of analogies in the development of word recognition. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 41-63). Hillsdale, NJ: Lawrence Erlbaum Associates.

Goswami, U., Ziegler, J. C., & Richardson, U. (2005). The effects of spelling consistency on phonological awareness: A comparison of English and German. *Journal of Experimental Child Psychology, 92,* 345-365.

Gough, P. B., & Hillinger, M. L. (1980). Learning to read: An unnatural act. *Bulletin of the Orton Society, 30,* 179-196.

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

Hatcher, P. J., Hulme, C., & Ellis, A. W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, *65*, 41-57.

Hoffman, L., & Rovine, M. J. (2007). Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods, 39,* 101-118.

Hogaboam, T. W., & Perfetti, C. A. (1978). Reading skill and the role of verbal experience in decoding. *Journal of Educational Psychology, 70,* 717-729.

Hulme, C., & Mackenzie, S. (1992). *Working memory and severe learning difficulties.* Hove, U.K.: Lawrence Erlbaum.

Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 189-212). New York: Springer.

Joe, H. (2008). Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis, 52,* 5066-5074.

Jorm, A. F., & Share, D. L. (1983). Phonological recoding and reading acquisition. *Applied Psycholinguistics, 4,* 103-147.

Jorm, A. F., Share, D. L., Maclean, R., & Matthews, R. G. (1984). Phonological recoding skills and learning to read: a longitudinal study. *Applied Psycholinguistics*, *5*, 201-207.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78,* 243-255.

Kail, R., & Hall, L. K. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition, 29,* 1-29.

Kim, Y.-S., Petscher, Y., Foorman, B. R., & Zhou, C. (2010). The contributions of phonological awareness and letter-name knowledge to letter-sound acquisition – a cross-classified multilevel approach. *Journal of Educational Psychology, 102,* 313-326.

Kearns, D. K. (2006). *Assessment of Grapheme-Phoneme Correspondence Knowledge.* Unpublished test.

Leslie, L., & Calhoon, A. (1995). Factors affecting children's reading of rimes: Reading ability, word frequency, and rime-neighborhood size. *Journal of Educational Psychology, 87,* 576-586.

Luo, W., & Kwok, O. (2009). The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research, 44*, 182-212.

Maier, K. S. (2001). A Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics, 26,* 307-330.

Manis, F. R., Doi, L. M., & Bhadha, B. (2000). Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities, 33,* 333-374.

Marinus, E., & de Jong, P. F. (2008). The use of sublexical clusters in normal and dyslexic readers. *Scientific Studies of Reading, 12,* 253-280.

Mason, J. M. (1980). When do children begin to read: An exploration of four year old children's letter and word reading competencies. *Reading Research Quarterly, 15,* 203-227.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*, 375-407.

Moerbeek, M. (2004). The consequences of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research, 39,* 129-149.

Molenberghs, G., & Verbeke, G. (2004). An introduction to (generalized (non)linear mixed models. In P. De Boeck & M. Wilson (Eds.), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach* (pp. 111-153). New York: Springer.

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

National Institute for Literacy. (n.d.). *Put reading first: Kindergarten through grade 3.* Retrieved February 1, 2010, from http://www.nationalreadingpanel.org/

Olson, R., Forsberg, H., Wise, B., & Rack, J. (1994). Measurement of word recognition, orthographic, and phonological skills. In G. R. Lyon, *Frames of Reference for the*

Assessment of Learning Disabilities: New Views on Measurement (pp. 243-278). Baltimore, MD: Paul H. Brookes Publishing.

Perfetti, C. A. (1986). Cognitive and linguistic components of reading ability. In B. R. Foorman & A. W. Siegel (Eds.), *Acquisition of reading skills: Cultural constraints and cognitive universals* (pp. 11-40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Perfetti, C. A. (1992). The representation problem in reading acquisition. In P. B. Gough, L. C. Ehri, & R. Treiman (Eds.), *Reading acquisition* (pp. 145-174). Hillsdale, NJ: Lawrence Erlbaum Associates.

Piasta, S. B., & Wagner, R. K. (2010). Learning letter names and sounds: Effects of instruction, lettery type, and phonological processing skill. *Journal of Experimental Child Psychology*, *105*, 324-344.

Pickering, S. J., & Gathercole, S. E. (2001). *Working Memory Test Battery for Children.* London: Psychological Corporation UK.

Preisser, D. A., Hodson, B. W., & Paden, E. P. (1988). Developmental phonology: 18-29 months. *Journal of Speech and Hearing Disorders, 53,* 125-130.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "the language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41,* 416-426.

Rabash, J., & Browne, W. J. (2007). Non-hierarchical multilevel models. In J. de Leeuw & E. Meijer (Eds.), *Handbook of Multilevel Analysis* (pp.303-336). New York: Springer.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods* (2^{nd} ed.). Thousand Oaks, CA: Sage Publications, Inc.

Reitsma, P. (1983). Printed word learning in beginning readers. *Journal of Experimental Child Psychology, 36,* 321-339.

Savage, R., Lavers, N., & Pillay, V. (2007). Working memory and reading difficulties: What we know and what we don't know about the relationship. *Educational Psychology Review, 19,* 185-221.

Share, D. L. (1995). Phonological recoding and self-teaching: *sine qua non* of reading acquisition. *Cognition, 55,* 151-218.

Share, D. L. (1999). Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology, 72,* 95-129.

Siegel, L. S. (1994). Working memory and reading: A life-span perspective. *International Journal of Behavioral Development, 17,* 109-124.

Siegel, L. S., & Faux, D. (1989). Acquisition of certain grapheme-phoneme correspondences in normally achieving and disabled readers. *Reading and Writing, 1*, 37-52.

Snowling, M. J. (1981). Phonemic deficits in developmental dyslexia. *Psychological Research, 43*, 219-234.

Stanovich, K. E. (1991). Word recognition: Changing perspectives. In R. Barr, M. L. Kamil, P. B. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. 2, pp. 418-452). Mahwah, NJ: Lawrence Erlbaum Associates.

Swanson, H. L. (2003). Age-related differences in learning disabled and skilled readers' working memory. *Journal of Experimental Child Psychology, 85*, 1-31.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics* (5th ed.). Boston, MA: Pearson Education, Inc.

Tennessee/Metro Nashville Public Schools Academic Standards/Reading – Grade 1. Retrieved February 15, 2010, from http://www.mnps.org/AssetFactory.aspx?did=3606

Treiman, R., Goswami, U., & Bruck, M. (1990). Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition, 18*, 559-567.

Van den Noortgate, W., De Boeck, P., & Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics, 28*, 369-386.

Van den Noortgate, W., & Paek, I. (2004). Person regression models. In P. De Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 167-187). New York: Springer.

Wagner, R., Torgesen, J., & Rashotte, C. (1999). *Comprehensive test of phonological processing.* Austin, TX: PRO-ED.

Wilson, M., & De Boeck, P. (2004). Descriptive and explanatory item response models. In P. De Boeck & M. Wilson (Eds.), *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach* (pp. 43-74). New York: Springer.

Wolf, M., & Bowers, P. G. (1999). The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology, 91*, 415-438.

Wolf, M., & Bowers, P. G. (2000). Naming-speed processes and developmental reading disabilities: An introduction to the special issue on the double-deficit hypothesis. *Journal of Learning Disabilities, 33*, 322-324.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). The educator's word frequency guide (CD-ROM). New York: Touchstone Applied Science Associates.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin, 131*, 3-29.

Ziegler, J. C., Stone, G. O., & Jacobs, A. M. (1997). What is the pronunciation for *–ough* and the spelling for /u/? A database for computing feedforward and feedback consistency in English. *Behavior Research Methods, Instruments, & Computers, 29,* 600-618.