Published in final edited form as:

J Educ Psychol. 2011 May 1; 103(2): 489–507. doi:10.1037/a0023001.

Word and Person Effects on Decoding Accuracy: A New Look at an Old Question

Jennifer K. Gilbert, Donald L. Compton, and Devin M. Kearns
Vanderbilt University

Abstract

The purpose of this 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. The sample was comprised of 196 English-speaking grade 1 students. A researcher-developed pseudoword list was used as the primary outcome measure. Because grapheme-phoneme correspondence (GPC) knowledge was treated as person and word specific, we are able to conclude that it is neither necessary nor sufficient for a student to know all GPCs in a word before accurately decoding the word. And controlling for word-specific GPC knowledge, students with lower phonemic awareness and slower rapid naming skill have lower predicted probabilities of correct decoding than counterparts with superior skills. By assessing a person-by-word interaction, we found that students with lower phonemic awareness have more difficulty applying knowledge of complex vowel graphemes compared to complex consonant graphemes when decoding unfamiliar words. Implications of the methodology and results are discussed in light of future research.

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). Though item- rather than stage-based in nature, the self-teaching hypothesis also hails the importance of decoding in its underlying assumption that children's practice of translating written text into phonological units (i.e., decoding) acts as a self-teaching mechanism whereby orthographic representations (knowledge structures comprised of a word's letter pattern, pronunciation, and/or meaning) are established and subsequent fluent word recognition flourishes (Ehri, 1992; Share, 1995). 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 reading unfamiliar words (Brown & Deavers, 1999; Ehri, 1991; Ehri & Wilce, 1985). Furthermore, Cunningham, Stanovich, and Wilson (1990) established that decoding ability is still highly correlated with word recognition ($r = .66$) even among adults. These two latter points, that decoding is used when readers are faced with unknown words at the same time those readers recognize familiar words by sight and that decoding skill is related to word reading in adults, suggest that reading acquisition can be considered

Correspondence concerning this article should be addressed to Jennifer K. Gilbert, Department of Special Education, Peabody Box 228, 110 Magnolia Circle – MRL 205, Nashville, TN 37203. jennifer.k.gilbert@vanderbilt.edu.

Publisher's Disclaimer: The following manuscript is the final accepted manuscript. It has not been subjected to the final copyediting, fact-checking, and proofreading required for formal publication. It is not the definitive, publisher-authenticated version. The American Psychological Association and its Council of Editors disclaim any responsibility or liabilities for errors or omissions of this manuscript version, any version derived from this manuscript by NIH, or other third parties. The published version is available at www.apa.org/pubs/journals/edu

an item-based progression in which each word must be learned one-by-one rather than a stage-based process in which individuals pass through different stages of reading where a single strategy (e.g., decoding or sight recognition) is used to read all words (Share, 1995). Still by either account, decoding plays an essential role in the development of fluent word reading.

Approaches to decoding are not uniform, however. There is evidence that individuals 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; Ehri & Robbins, 1992; Goswami, 1986, 1998; Gottardo, Chiappe, Siegel, & Stanovich, 1999; Ziegler & Goswami, 2005). Both strategies suggest that decoding is affected by both person (e.g., grapheme-phoneme correspondence [GPC] knowledge) and word (e.g., number of similar words) characteristics. In the following sections, prior research on both types of effects is discussed.

Person Effects

GPC knowledge is an undisputed prerequisite to successful decoding (Ehri, 1998; Share, 1995, 1999). In a recent study by Shmidman and Ehri (2010), children were better able to remember the GPCs that were taught with a corresponding letter-embedded mnemonic than the GPCs taught with a dissociated mnemonic; in turn, young children were able to use those better-remembered GPCs to decode unfamiliar letter strings. Treiman et al. (1990) also investigated the role of GPC knowledge in decoding. The authors showed significantly high correlations between GPC knowledge and pseudoword reading for first-grade children. Moreover, when re-analyzing data at the word-level, the authors found that the correlation was somewhat higher among low-frequency¹ pseudowords (.73) than high-frequency pseudowords (.61). This finding lends support to the hypothesis that accurate decoding depends on both person and word factors. And although GPC knowledge is necessary for decoding, it is not necessarily sufficient for every person. For instance, Gottardo et al. (1999) found that less-skilled readers compared to younger reading-level controls and age-matched controls were less successful at using GPCs to decode words. It could also be that the effect of GPC knowledge depends on the word being decoded. In this study, we test the strong hypothesis that word-specific GPC knowledge is necessary (as stage-based theories would suggest) and/or sufficient (as item-based theories would suggest) for decoding new words. Assuming for the moment that it is neither, we consider word and person effects that could explain variability in decoding above and beyond knowledge of a word's GPCs.

In addition to GPC knowledge, phonemic awareness is necessary for decoding (Juel, Griffith, & Gough, 1986). Byrne and Fielding-Barnsley (1991) demonstrated a unique contribution of phonemic awareness to decoding in an experiment where preschool children were provided either phonemic awareness instruction or exposed to story reading. The same result emerged in the 1-year follow-up study (Byrne & Fielding-Barnsley, 1993). 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 was used in the current investigation as a potential contributor of decoding accuracy 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.

¹Frequency was defined in terms of the rime unit.

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 (Compton, 2003; Manis, Doi, & Bhadha, 2000). 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 associated with slower cognitive processing assessed by rapid automatized naming (Kail & Hall, 2001; Perfetti, 1986), which has been shown to detrimentally affect decoding accuracy potentially due to the interference with the full processing of letters in a word (Bowers, Sunseth, & Golden, 1999). Prior research suggests 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).

Findings from Compton (2000) provide evidence that GPC knowledge, phonemic awareness, and rapid naming speed each contributes significantly to decoding skill when considered simultaneously. The author found that each of these skills 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) could 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 phonological awareness was included in the model. Furthermore, Connors, Atwell, Rosenquist, and Sligh (2001) and Abu-Rabia, Share, and Mansour (2003) have found evidence that phonological working memory is a reliable difference between good and poor decoders, controlling for several other factors. On the contrary, Savage, Lavers, and Pillay (2007) conducted a review of the relation between working memory and reading and found a lack of convincing evidence that working memory plays a key role in word reading acquisition. Mixed results regarding working memory prompted its inclusion in the present study. In the current investigation, working memory along with phonemic awareness and rapid naming speed were entered as student-level predictors in a model in the presence of word-specific GPC knowledge. The ability to control for word-specific GPC knowledge represents an extension to the current decoding literature because it provides information about child characteristics that influence decoding after taking into account whether the child actually knows the GPCs in a word. For instance, better phonemic awareness may not increase the probability of correctly decoding a word if a child does not know the GPCs within the word. In addition to the person characteristics that affect decoding, word characteristics influence decoding as well.

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 as opposed to accuracy. 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*). 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, an error analysis by Graham (1980) revealed that elementary-school students had more difficulty with vowels than any other grapheme type. With regard to consonant letters, Olson, Forsberg, Wise, and Rack (1994) reported that children made more errors when attempting to read words with consonant clusters than words with single consonants. Snowling (1981) noted that this deficit in complex consonants was particularly noticeable for children with dyslexia when compared to children with typical reading skills. Though past research has demonstrated the difficulty of complex vowels and consonants compared to their single counterparts, in the current study we compared the effect of grapheme complexity between vowels and consonants by systematically varying vowel grapheme complexity and consonant grapheme complexity in a list of pseudowords.

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 with fewer shared features. Bowey (1990) and Booth and Perfetti (2002) 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, Calhoun and Leslie (2002) and Leslie and Calhoun (1995) investigated the effects of rime-neighborhood size, or the number of monosyllabic words sharing the same rime unit, and word frequency on word and nonword reading across grades 1–3. 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 these investigations did not account for both person and word variance, results may be biased in that standard errors for covariates may have been underestimated (Janssen, Schepers, & Peres, 2004). Thus, we examined the effects of orthographic rime-neighborhood size, average orthographic rime-neighbor frequency², the interaction of size and frequency, and grapheme complexity while accounting for all necessary sources of variance to determine the effect of each on decoding accuracy, all while controlling for word-specific GPC knowledge. Because pseudowords have no frequency of their own, the average orthographic rime-neighbor frequency was calculated for each word as a proxy for pseudoword frequency; it was calculated by taking the average frequency of all the rime neighbors. Once again, the ability to control for word-specific GPC knowledge represents an extension to the decoding literature, this time with regard to word effects. For instance, examining the difference in accuracy between reading a word with a complex vowel versus a complex consonant is most informative in the case that a child knows the GPCs associated with both words. That way the difference represents the difficulty in decoding words with two different types of graphemes rather than difficulty in learning the two types of graphemes.

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 (person and word) factors have been explored together in one model. As a solution, we combined factors found to be significant in prior research

²Because 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 real word that contains the same rime unit as the pseudoword itself.

into one model so that effects could be established in the presence of other potential influences. 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.

Crossed random-effects models allow 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 persons and 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 literature because 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) allowed a detailed and accurate investigation of the factors that affect decoding because the unit of analysis was the intersection of each person and each word, and inter-word variance and inter-person variance were taken into account (Janssen et al., 2004).

The second point of rationale is that most prior research has utilized aggregated data either across persons or across words. One potential 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 decoding accuracy score, which does not provide specific information about the pseudowords that were being decoded. Information for questions 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 possible. 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. Furthermore, the current data analysis approach extends the literature by considering the effect of GPC knowledge specific to each word and each student. For example, Compton (2000) and Treiman et al. (1990) found that children who knew more GPCs were predicted to decode more words than children who knew fewer GPCs. The authors used the total number of known GPCs to predict the total number of correctly decoded words. By using a crossed random-effects model, we were able to ask a finer-grained question: How does a child's knowledge (vs. lack of knowledge) of

the GPCs within a word affect the probability of that child correctly decoding that particular word? A few case studies suggest that the strong relation between GPC knowledge and decoding may not hold at the word level (Ehri & Wilce, 1985; Fletcher-Flinn & Thompson, 2000; Thompson & Fletcher-Flinn, 2007).

In light of the finer-grained question, we 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 effect of word-specific GPC knowledge, additional variables were entered into a model to further explain variability in decoding accuracy while controlling for word-specific GPC knowledge.

Four research questions were asked in accordance with the aim of the proposed study, and they are listed in Table 1 with associated model numbers and covariates. The first question is: What is the effect of word-specific GPC knowledge on decoding accuracy? As a follow-up to this first question, we want to know whether word-specific GPC knowledge is necessary and/or sufficient for accurate decoding. Second, what are the effects of phonemic awareness, rapid naming speed, and working memory in accounting for person variance in decoding accuracy after controlling for word-specific GPC knowledge and word characteristics? Third, what are the effects 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 and person characteristics? Fourth, are there significant person-by-word interactions? First, we considered an interaction term between working memory and rime frequency. Second, we considered an interaction between phonemic awareness and grapheme (vowel versus consonant) complexity.

Our hypothesis regarding the first research question was that there will 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. We hypothesize for the second question that all student factors entered in the model will have significant influence on decoding accuracy in accordance with prior research, although we are somewhat less certain about working memory than the other factors. For the third question, we expect that there will be a significant interaction between rime size and rime frequency such that words with lower average-neighbor frequency will be less affected by neighborhood size than words with higher average-neighbor frequency (Calhoon & Leslie, 2002; Leslie & Calhoon, 1995). Furthermore, we expected that pseudowords with complex vowels will be significantly more difficult to decode than pseudowords with complex consonants, as suggested by trends in past research. With regard to the fourth research question, we expect that both interactions will be significant. For the first interaction, we expect students with poorer working memory will perform worse on pseudowords with infrequent rimes than frequent rimes, while students with better working memory will 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). As for the second interaction, we expect that children with poor phonological skills will have more difficulty

with complex vowel clusters than complex consonant clusters, which follows trends reported by Marinus and de Jong (2008) and Siegel and Faux (1989).

Method

Participants and Procedures

Participants were part of a larger study investigating the efficacy of supplemental reading tutoring for English-speaking first-grade students who were deemed at risk for later reading difficulties. In the larger study, we asked 56 teachers from 15 schools to nominate the 6–10 poorest readers in their class. Of that pool, 287 students were consented and screened. The screening battery included measures of accuracy of decoding and word reading accuracy, 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 the factor score were chosen for further testing. It should be noted that although the lowest 250 students were designated as at risk for later reading difficulties based on the screening battery and by teacher nomination, the average standard scores on standardized measures of word reading, decoding, fluency, and comprehension for students in the present sample were well within the normal range with scores between 96.25–101.40. Of the additional measures given to the at risk group, only the Backward Digit Recall subtest of the Working Memory Test Battery (Pickering & Gathercole, 2001) was relevant to the current study. Though this subtest is the only measure in the current study to contain numeric rather than alphabetic information, use of a working memory task containing verbal information (Sentence Span) would likely be more confounded with verbal ability than is the case with a task containing numeric information.

Students who completed all pretests ($N = 215$) were randomly assigned to one of four groups. Three were intervention groups in which students were provided one-on-one reading instruction that was supplemental to the school's reading curriculum; the control group received only the school's reading curriculum. All intervention groups received one-on-one instruction for 30–45 minutes, three times per week for 19 weeks in sight word recognition, grapheme-phoneme correspondences, decoding, spelling, and fluency. In addition to these core word reading components, intervention for one of the groups included comprehension instruction and intervention for another included math instruction. The designated classroom reading program for the school district was the Treasures basal reading/language arts program (MacMillan/McGraw-Hill, 2008). The program prescribes explicit instruction and ample practice in the following areas: oral vocabulary, phonemic awareness, phonics, spelling, vocabulary, comprehension, fluency, grammar, and writing. Included in the program are suggestions for whole-class and small-group instruction, and an assortment of reading materials depending on students' reading achievement. It should be noted that treatment was not considered in the current set of analyses because we were concerned with how children's skills affect decoding regardless of how they acquired those skills. Plus, a model including only treatment as an independent variable revealed that neither the coefficient for the treatment variable ($p = 0.67$) nor the LR test comparing the treatment model (including random treatment effects at the classroom and school levels) to a null model was significant, $\chi^2(5) = 7.35$, $p = 0.20$; both pieces of information suggest results of this study would not be affected by treatment status.

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., 1999), and the experimenter-created Targeted Nonword Test (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 (47.96% females). 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 was classified as English-language learners. Finally, 20 students (10.75%) had a special education label, and 15 (8.06%) were repeating first grade.

Measures

Decoding—Decoding accuracy was assessed with the TNT, which is a 20-item experimenter-created measure designed to explore the relation between decoding accuracy and other factors. Thus, the pseudowords on the TNT were *targeted* because they were created to answer specific research questions. 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-vowel-consonant) phoneme pattern; Calhoon and Leslie (2002) also followed this format. Students were given 1 of 4 randomly ordered stimuli sheets, and each item was scored as correct or incorrect with no partial credit given.

Words on the TNT varied systematically in two ways. The first source of planned variability was rime frequency, defined in terms of the orthographic 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 frequency 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 and a simple consonant grapheme (e.g., *soom*) and half contained a complex consonant grapheme cluster and a simple vowel grapheme (e.g., *jeth*). A single dummy variable (CV) 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. 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 TNT measure, the cluster order was reversed to help clarify the complexity versus position issue.

The frequency and complexity dimensions created 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. Appendix A 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 nonwords were consistent (or regular) in that none of the rime units had real words with illegal pronunciations.

Interrater agreement 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 agreement, based on 20% of the tests, was .92 for total test score and ranged from .78–1.00 for each individual item. All items except one had a reliability exceeding .87. The Kruder-Richardson 20 internal consistency coefficient for the measure was .80. 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. The correlation between the total number of correct responses on the TNT and the Word Attack subtest of the Woodcock Reading Mastery Tests (Woodcock, 1987) was .60 and its correlation with the Phonemic Decoding Efficiency subtest of the Comprehensive Test of Phonological Processing (CTOPP, Wagner et al., 1999) was .67, revealing the TNT has some evidence of construct validity as a test of decoding.

Phonemic awareness—Phonemic awareness was assessed with the Sound Matching subtest of the CTOPP (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.

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.

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. (1995), 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 determined 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. Interrater agreement for this sample exceeded .96.

GPCK scores are specific to each student and each pseudoword given 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. 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. This variable was scored dichotomously because, as we stated previously, our goal was to test the strong hypothesis that word-specific GPC knowledge is necessary and/or sufficient for decoding new words. Most all theories of reading development suggest that GPC knowledge precedes accurate decoding, and we wanted our variable to reflect that assumption. 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.

Data Analysis

Van den Noortgate, De Boeck, and Meulders (2003) proposed a method of assessing item-by-person 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 an unconditional means model (Base model). A likelihood ratio (LR) test comparing models with and without person and word random effects serves as further evidence for the inclusion/exclusion of those random effects.

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, a person- and item-specific residual can be considered simultaneously such that the model accounts for both types of dependencies in the responses.

Treating words as a random factor allows the variability in decoding 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., units 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 set of pseudowords that would be nested within a person. Instead, our response data were considered cross-classified because our lower-level units (i.e., responses) 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.

The first step in analyzing our data was to estimate a Base model in which no covariates were included in the model (a.k.a. an unconditional means model). The purpose of this initial model was to partition variance into the respective levels of influence on the outcome and 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 unconditional means models were estimated and compared. The first 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 _{<i>jkmi</i>})	$Logit(\pi_{jkmi}) = \lambda_{0jkmi}$	
Level-2 (Person _{<i>jk</i>} & Word _{<i>i</i>})	$\lambda_{0jkmi} = \gamma_{00km} + r_{01jkm} + r_{02i}, r_{01jkm} \sim N(0, \sigma_{r01}^2) \& r_{02i} \sim N(0, \sigma_{r02}^2)$	
Level-3 (Classroom _{<i>km</i>})	$\gamma_{00km} = \phi_{000m} + u_{km}, u_{km} \sim N(0, \sigma_u^2)$	
Level-4 (School _{<i>m</i>})	$\phi_{000m} = \omega_{0000} + \varepsilon_m, \varepsilon_m \sim N(0, \sigma_\varepsilon^2),$	(1)

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 ($r_{02i} = 0$). The term average here represent the average for our particular sample and not average in the universal sense. 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 (ω_{0000}) 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 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. The following equation for the combined model includes an intercept, a set of person-related random effects, and a word random effect:

$$Logit(\pi_{jkmi}) = \omega_{0000} + (r_{01jkm} + u_{km} + \varepsilon_m) + (r_{02i}).$$

ICCs can be calculated from the combined model. 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 (2011) 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; 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. Appendix B contains equations 2a–2d for calculating each ICC specific to the full base model. Assuming each source of variance is necessary for best describing the data, Equation 2a shows the ICC calculation for person variance; Equation 2b is for classroom variance; Equation 2c is for school variance; Equation 2d is for word variance.

In addition to the Base model, we proposed three models to address the four research questions. Again, Table 1 lists each research question with its associated model number and covariates. In addressing the research questions, we first estimated a model including only word-specific GPC knowledge (Model 1) to address research question 1. Once again, the

effect of word-specific GPC knowledge was assessed with a dummy code at Level 1. This means that the intercept represented 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 in the word are not known. The coefficient for the word-specific GPC knowledge variable represented 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 in the word are known compared to when the GPCs are not known. Both the intercept and the slope were allowed to vary randomly over persons and words. In addition, the covariance between the intercept and GPCK slope terms was estimated for both persons (σ_{r01*11}) and words (σ_{r02*12}) to account for the possible relations between them.

Model 1 was estimated to provide information about the effect of word-specific GPC knowledge on decoding accuracy. From this model, we calculated the probability of accurate decoding when all the GPCs within words were known and when at least one word-specific GPC was 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. If word-specific GPC knowledge is necessary for accurate decoding, then the predicted probability of a correct response should be 0 when at least one of word's GPCs is not known. On the other hand, if word-specific GPC knowledge is sufficient for accurate decoding, then the predicted probability of a correct response should be 1 when all the word's GPCs are known.

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). At Level 2, the following person factors were entered: phonemic awareness, rapid naming, and working memory. At the same time, the following word factors were entered at Level 2: average rime-neighborhood frequency, rime-neighborhood size, the interaction of frequency and size, and grapheme complexity. Grapheme complexity was assessed with a dummy variable that equaled 1 if the pseudoword contained a complex vowel grapheme and 0 if it contained a complex consonant grapheme. The aim of Model 2 was to estimate the person and word effects on a student's probability of correctly decoding a pseudoword, controlling for word-specific GPC knowledge. Results were examined by first describing the influence of each covariate by interpreting the coefficients and the *p*-values associated with each. 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 $(r_{01jkmMODEL1} - r_{01jkmMODEL2}) / r_{01jkmMODEL1}$. Similarly, we calculated the proportion of word variance accounted for by the four word predictors. Reduction in variance was calculated by $(r_{02iMODEL1} - r_{02iMODEL2}) / r_{02iMODEL1}$. 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. The significance of the interactions was first determined by the *p*-value associated with the coefficients associated with the interaction terms and then by a model comparison between models with and without the significant interaction term.

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 (2011) 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 most problematic cases in Cho and Rabe-Hesketh's simulation were the ones with samples of 100 persons, 10 items, or ICCs of .50. The current study had double the persons and items as the problematic case in the study, and ICCs were expected to be much lower than .50 (results from the unconditional model reveal ICCs for the current sample). Neither standard errors nor *p*-values for random effects are produced by the *lmer* function because *lmer* uses likelihood methods which do not make 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.

Results

Descriptive Statistics

Frequency indices were calculated for each pseudoword on 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). Table 3 contains descriptive information for word difficulties which ranged from means of .15–.72. Descriptive statistics of the student assessment scores are provided in Table 4. Correlations among the student assessment scores are in Table 5.

Base Model Results

Before estimating 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 6 contains model fit indices for all five models and the relevant model comparison tests. Although the LR 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).

In all cases, the LR 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 all the conditional models. Estimates of the fixed and random effects for the Full Base model are in Table 7. We used Equations 2a–2d 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. The sizeable ICCs for both persons and words indicate that exploration of person and word factors is sensible for these data.

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, $1/(1+\exp[-\{-0.31\}])$. Average person, average classroom, average classroom, and average word are of course relative to our sample of first-grade students reading words on the experimenter-created TNT. However, variability around this average existed among persons and words (see σ_{r01} and σ_{r02} in Table 7). One way of describing the magnitude of variation of correct responses across persons and words is to compute the 95% plausible value range, which indicates a range of values in which 95% of the cases are expected to fall (Hoffman & Rovine, 2007; Raudenbush & Bryk, 2002). The 95% plausible value range for persons was .13 to .78 ($1/[1+\exp\{-\omega_{000} \pm 1.96*\sqrt{\sigma_{r01}}\}]$) for the predicted probability of getting an average word correct. For words, the range was .11 to .81 ($1/[1+\exp\{-\omega_{000} \pm 1.96*\sqrt{\sigma_{r02}}\}]$) 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 there is variability in responses across words (σ_{r02} in Table 7) that we can model and attempt to explain – information that would be lost in an aggregate, person-level analysis.

Results for Research Questions

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. Secondly, we wanted to establish whether word-specific GPC knowledge was necessary and/or sufficient for accurate decoding. Therefore, Model 1 contained the GPCK variable that was specific to each person and each word. Model 1 results are in Table 8. The z-statistic for the GPCK variable ($z = 2.97$) indicated that there was a significant difference (at the $\alpha = .01$ level) in the probability of a correct response when word-specific GPCs were known versus not known, with the known case having a significantly higher probability. The effect of word-specific GPC knowledge varied substantially across students and words (see σ_{r11} and σ_{r12} in Table 8); 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% ($1/[1+\exp\{-\lambda_1 \pm 1.96*\sqrt{\sigma_{r11}}\}]$) on an average word. That is, some students' decoding was greatly affected by whether they knew the GPCs in words (those near 80.64%), and it hardly mattered at all for others (those near 6.27%). The effect of GPC knowledge also varied by word as the 95% plausible value range of the predicted difference ranged from 39.91% to 79.63% ($1/[1+\exp\{-\lambda_1 \pm 1.96*\sqrt{\sigma_{r12}}\}]$) for an average person. Again, this suggests that word-specific GPC knowledge has a greater effect on some words than others. Although our theoretical interest was between complete versus incomplete GPC knowledge, some of the variability in the effect of word-specific GPC knowledge could be attributable to the coding of that variable. Scores of 0 represented knowledge of 0, 1, or 2 GPCs in a word, and this scoring introduces variability by placing knowledge of no GPCs and knowledge of some GPCs into the same category of not knowing all the GPCs. Coding decisions notwithstanding, findings suggest that knowledge of word-specific GPCs is an important predictor of accurate decoding.

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. The .35 predicted probability is significantly different from 0 ($z = -2.21$, $p = .03$), suggesting that full word-specific GPC knowledge is not necessary for

accurate decoding. 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. Although there is no significance test to determine whether a predicted probability of .46 is statistically different from 1.00, we are quite confident that this finding means that word-specific GPC knowledge is not sufficient for accurate decoding. So, we conclude that although knowing the GPCs in a word is associated with a higher probability of decoding the word than not knowing the GPCs, having that knowledge is neither a necessary nor sufficient condition for every person to accurately decode every word; other factors are at work.

The negative covariance terms³ between the intercept and the GPCCK slope for both persons and words (correlations of $r_{\tau 01*11} = -.73$ and $r_{\tau 02*12} = -.45$, respectively) also suggest there are factors affecting correct decoding other than word-specific GPC knowledge. The negative covariances suggest that as the probability of accurate decoding was increased, the effect of GPC knowledge was decreased. Alternatively, as the probability of correct decoding decreased, the effect of GPC knowledge increased. So, it appears that the effect of GPC knowledge is bigger for harder words than for easier words and bigger for more able decoders than less able decoders. The 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 in accurate decoding.

The aim of research question 2 was to determine the effects of person factors in their contribution to decoding accuracy while controlling for a student's knowledge of a word's GPCs and for word characteristics. So, Model 2 included word-specific GPC knowledge plus person (phonemic awareness, rapid naming, and working memory) and word (average rime-neighbor frequency, rime-neighborhood size, Frequency X Size, and grapheme complexity) covariates. The results in Table 9 reveal that controlling for word-specific GPC knowledge and the various word characteristics, phonemic awareness ($\gamma_{10} = 0.28$, $z = 3.69$, $p < .001$) and rapid naming skill ($\gamma_{20} = -0.17$, $z = -2.29$, $p = .02$) were significant (at least at an α level of .05) predictors of decoding accuracy, while working memory was not ($\gamma_{30} = 0.03$, $z = 0.33$, $p = .74$).

As hypothesized, phonemic awareness and rapid naming had significant influence on decoding. 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. Rapid naming also appears to be an influential factor ($z = -2.29$), though not quite as robust 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.

To answer research question 3 addressing the effects of word factors, we examined the significance tests associated with the following word variables in Model 2: average rime-neighbor frequency, rime-neighborhood size, Frequency X Size, and grapheme (vowel

³Although significance tests are not available for random effects, a model without covariance terms between the intercept and slope for persons and words and a model with them revealed a worse fit when the terms were not included, $\chi^2(2) = 7.92$, $p = .01$. Thus, covariances between intercept and GPCCK were estimated for persons and words in Models 1–3.

versus consonant) complexity. Again, results are in Table 9. Controlling for word-specific GPC knowledge and person characteristics, it appears that although words with more frequent rimes were decoded at a higher probability than words with less frequent rimes ($\gamma_{40} = 0.76, z = 3.52, p < .001$) (also found by Treiman et al., 1990), that effect was moderated by rime-neighborhood size ($\gamma_{60} = -0.48, z = -2.13, p = .03$). Figure 2 illustrates this interaction. This finding corroborates that of Calhoun and Leslie (2002) and Leslie and Calhoun (1995) that young children are most affected by a word's rime-neighborhood size when the word is found infrequently in text. Results from Model 2 also reveal that grapheme complexity had a significant influence. Words with complex vowel graphemes had a lower predicted probability of correct decoding than complex consonant graphemes ($\gamma_{70} = -0.70, z = -2.38, p = .02$). 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.

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 10. As the table shows, there was not a significant interaction of working memory and rime frequency ($\lambda_2 = 0.01, z = 0.26, p = .80$). 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 10 confirm that there was indeed a significant interaction between phonemic awareness and grapheme complexity ($\lambda_3 = 0.17, z = 2.18, p = .03$). 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 X 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, $\chi^2(1) = 4.75, p = .03$. Figure 3 depicts the interaction. 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.

Discussion

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 (Byrne & Fielding-Barnsley, 1991; Compton, 2000, 2003; Manis et al., 2000; Treiman et al., 1990) and word characteristics (Calhoon & Leslie, 2002; Leslie & Calhoon, 1995; Treiman et al., 1990). 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, when all the GPCs within a word were known, there was a higher predicted probability of accurate decoding than when all the word's GPCs were not known. This finding provides further support for the stage theories of reading development that posit that 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, though, that the positive relation between GPC knowledge and decoding does not necessarily hold for every person or 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; others 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 AGPCK 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 that children are able to use orthographic information of different grain 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 may not need to be explicit to be useful. 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), it is clear that in some cases implicit GPC knowledge may be sufficient to decode unfamiliar words. We are not suggesting that GPCs should not be taught explicitly but only 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 representation quality (i.e., stability, impenetrability) would be a more appropriate measure (Perfetti, 1992).

When students are able to correctly decode a word despite not knowing all the GPCs within that word, it may be that they are 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 students 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 resulted from 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. To confirm this conclusion, 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.

In addition to and controlling for word-specific GPC knowledge, students who had higher phonemic awareness scores were more likely to decode pseudowords accurately than students with lower scores. 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), but this is the first study to show that phonemic awareness influences decoding after controlling for knowledge of word-specific GPCs. The word-specific nature of our GPC variable allows us to conclude that even when two individuals know the GPCs in a word, the one with the poorer phonemic awareness is less likely to read the word correctly than the one with better phonemic awareness. This suggests poor phonemic awareness may be associated with difficulty in the application of known GPCs.

As for other person effects, there also appears to be a unique relation between rapid naming skill and accurate decoding that exists in the presence of phonemic awareness, working memory, and word-specific GPC knowledge. This unique contribution, along with a low correlation between phonemic awareness and rapid naming ($-.18$), supports previous studies that claim that phonological and rapid naming deficits make relatively independent contributions to reading difficulty (Wolf & Bowers, 1999, 2000). Working memory, however, was found not to have a significant relation to decoding in the present study.

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 masked 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 et al., 2007). The effect of working memory notwithstanding, the other person factors functioned as expected. Although the person effects mentioned here certainly impact accurate decoding, the skills that students 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 (Ehri, 1992; 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 after controlling for word-specific GPC knowledge and person characteristics, 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 more affected by a word's average-rime neighbor frequency when a word has fewer rime neighbors than more rime neighbors (Leslie & Calhoun, 1995; Calhoun & Leslie, 2002). In other words, if a word contains a rime that is not shared by many other words, it is more important for those few neighbors to show up frequently in text whereas it matters less how frequently neighbors show up in text if a word has a rime with many neighbors. One case in point is the much larger discrepancy in the proportion of students who read *feem* (.56) versus *voun* (.28) in comparison to *vell* (.58) and *guck* (.62). The former pair had only one neighbor each but *feem*'s neighbor had a higher frequency (116 versus 10); the latter pair were from large neighborhoods with 14 and 12 neighbors, respectively, and although they differed in average rime-neighbor frequency by 1104 and 183, respectively, the difference in the proportion of students reading them was much smaller than was the case with the pair of smaller neighborhood words. Further research on the importance of average rime-neighbor frequency in real word decoding should be conducted to determine if its effect is specific to pseudowords, which have no frequency of their own.

Another word effect was that words with complex consonant graphemes (e.g., *jeth*) were easier to decode than words with complex vowel graphemes (e.g., *mout*), which confirms the trend found by Marinus and de Jong (2008) and is aligned with that of Treiman et al. (1990) and Graham (1980) that vowels are the most problematic type of grapheme in terms of decoding. 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. Whereas prior research has shown that complex vowel and consonant graphemes are generally more difficult than their respective simple grapheme types (Marinus & de Jong, 2008; 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., words with complex vowel graphemes). And 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.

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 students with poor phonemic awareness had particular difficulty in applying complex vowel GPCs when decoding new words. In our data, among the students with below average phonemic awareness, fewer read *dail* (15.79%) correctly than *tesh* (40.68%) despite knowing the respective word-specific GPCs, compared to a nearly equal probability (39.02% and 38.46%, respectively) for those two words for students with above average phonemic awareness. Siegel and Faux (1989) also found that less-skilled readers had more difficulty applying the GPCs that they knew to pseudowords. Modeling the relationships between types of educational materials and types of students in this way could help match students with the most appropriate instruction, a goal shared by researchers who assess person-level aptitude-by-treatment interactions.

The methodology used in this study could also be used to test the self-teaching hypothesis (Jorm & Share, 1983; Share, 1995). According to the self-teaching hypothesis, individuals acquire orthographic representations of words through phonological recoding. 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 correlate the percentage of targets that were correctly decoded by the participant during training and the number of correct target choices the participant made on the orthographic choice task. One disadvantage of the total-score correlation is the inability to address the relationship between accurate decoding and orthographic learning at the word level.

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 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.

Findings and conclusions from the present study should be interpreted in light of three 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 (Nation and Snowling, 1998) but not pseudoword decoding.

Second, all of the pseudowords included on the TNT were orthographically consistent in that none had enemies. One reason this effect was controlled was that we were concerned about the number of pseudowords first-grade children would be able to decode before losing motivation and concentration. This was of particular concern because the TNT was given as part of a larger battery of tests lasting 1 hour in duration. Because the TNT words had no orthographic enemies, our results can be generalized only to the larger population of monosyllabic English words that also have no enemies. Future research would certainly benefit from including words that varied in orthographic or phonological consistency and exploring the effect of consistency on decoding accuracy.

Finally, our sample of words was limited to those found in the English language. Because the English language is somewhat of an "outlier" (Share, 2008, p. 584), it is quite possible that person and word characteristics that affect decoding accuracy in English may not have the same influence in other languages. For example, vowels in the German language are less variable than vowels in the English language so the effect of vowel complexity we found in

this study may not apply to the German language (Näslund, 1999). Also, because the Italian language is far more transparent than the English language (Näslund, 1999), the effect of phonemic awareness on decoding accuracy may not be as influential to decoding accuracy in Italian as it is in English. As a result of the differences in orthographies across languages, the findings from this study should not be generalized beyond the English language.

Despite these limitations, we believe this study adds useful information to the literature on decoding and provides a promising method for conducting future research. For example, future research on the relation between GPC knowledge and decoding could incorporate a continuous measure of word-specific GPC knowledge, one in which children receive partial credit for knowing some of the GPCs in words. This would allow researchers to ask questions about the correspondence between the percentage of GPCs known in a word and the probability of accurately decoding the word. GPCs could also be coded to facilitate questions about the position of the known GPCs such that the relative influence of knowing a word's initial, middle, or final GPC on correct decoding could be explored. For future theory building, this type of model could be used to explore the word-specific nature of decoding acquisition while also considering the child's stage of reading acquisition. And for future work on reading instruction, researchers interested in designing targeted reading interventions could explore additional person-by-word interactions to determine if certain types of children have particular difficulty with certain types of words. Perhaps these and other questions that have lingered in the field of reading could benefit from a fresh look using the new method discussed here.

Acknowledgments

This research would not have been possible without the funding of Grants R305B040110 and R324G060036 from the U. S. Department of Education, Institute of Education Sciences; and Core Grant HD056109 from the National Institute of Child Health and Human Development to Vanderbilt University. In addition, we would like to acknowledge Douglas and Lynn S. Fuchs for allowing us to use participants from their larger research project. We would also like to thank Sun-Joo Cho for reviewing the statistical models and Bobette Bouton and the numerous graduate research assistants who helped with data collection. This research was conducted to fulfill the requirements of the doctoral dissertation.

References

- Abu-Rabia S, Share D, Mansour MS. Word recognition and basic cognitive processes among reading-disabled and normal readers in Arabic. *Reading and Writing*. 2003; 16:423–442.
- Adams RJ, Wilson M, Wu M. Multilevel item response models: An approach to errors in variables regression. *Journal of Educational and Behavioral Statistics*. 1997; 22:47–76.
- Baayen RH, Davidson DJ, Bates DM. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*. 2008; 59:390–412.
- Baddeley, AD. *Working Memory*. New York, NY: Oxford University Press; 1986.
- Baddeley A, Gathercole S, Papagno C. The phonological loop as a language learning device. *Psychological Review*. 1998; 105:158–173. [PubMed: 9450375]
- Bates D. [R] lmer, p-values and all that. 2006 May 19. Message posted to <https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html>.
- Bates D. [R] storing the estimates from lmer. 2006 July 15. Message posted to <http://www.mail-archive.com/r-help@stat.math.ethz.ch/msg67236.html>.
- Bates D, Maechler M. The lme4: Linear mixed-effects models using S4 classes. 2009 URL: <http://lme4.r-forge.r-project.org/>.
- Booth JR, Perfetti CA. Onset and rime structure influences naming but not early word identification in children and adults. *Scientific Studies of Reading*. 2002; 6:1–23.
- Bowers PG, Sunseth K, Golden J. The route between rapid naming and reading progress. *Scientific Studies of Reading*. 1999; 3:31–53.

- Bowey JA. Orthographic onsets and rimes as functional units of reading. *Memory & Cognition*. 1990; 18:419–427. Retrieved from http://www.psychonomic.org/backissues/6939/mc/vol18-4/PDFs/MC009_V18No4.pdf.
- Bradley L, Bryant PE. Categorizing sounds and learning to read - a causal connection. *Nature*. 1983; 301(3):419–421.
- Briggs DC, Wilson M. Generalizability in item response modeling. *Journal of Educational Measurement*. 2007; 44:131–155.
- Brown GDA, Deavers RP. Units of analysis in nonword reading: Evidence from children and adults. *Journal of Experimental Child Psychology*. 1999; 73:208–242. [PubMed: 10357873]
- Byrne B, Fielding-Barnsley R. Evaluation of a program to teach phonemic awareness to young children. *Journal of Educational Psychology*. 1991; 83:451–455.
- Byrne B, Fielding-Barnsley R. Evaluation of a program to teach phonemic awareness to young children: A 1-year follow-up. *Journal of Educational Psychology*. 1993; 85:104–121.
- Byrne, B.; Fielding-Barnsley, R.; Ashley, L.; Larsen, K. Assessing the child's and the environment's contribution to reading acquisition: What we know and what we don't know. In: Blachman, BA., editor. *Foundations of reading acquisition and dyslexia: Implications for early intervention*. Mahwah, NJ: Lawrence Erlbaum Associates; 1997. p. 265-285.
- Calhoon JA, Leslie L. 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*. 2002; 34:39–58.
- Carlisle JF, Katz LA. Effects of word and morpheme familiarity on reading of derived words. *Reading and Writing*. 2006; 19:669–693.
- Chall, JS. *Stages of reading development*. New York, NY: McGraw-Hill; 1983.
- Cho S-J, Rabe-Hesketh S. Alternating imputation posterior estimation of models with crossed random effects. *Computational Statistics and Data Analysis*. 2011; 55:12–25.
- Clark HH. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:335–359.
- Compton DL. Modeling the growth of decoding skills in first-grade children. *Scientific Studies of Reading*. 2000; 4:219–259.
- Compton DL. Modeling the relationship between growth in rapid naming speed and growth in decoding skill in first-grade children. *Journal of Educational Psychology*. 2003; 95:225–239.
- Conners FA, Atwell JA, Rosenquist CJ, Sligh AC. Abilities underlying decoding difference in children with intellectual disability. *Journal of Intellectual Disability Research*. 2001; 45:292–299. [PubMed: 11489050]
- Cunningham, AE.; Stanovich, KE.; Wilson, MR. Cognitive variation in adult college students differing in reading ability. In: Carr, TH.; Levy, BA., editors. *Reading and Its Development: Component Skills and Approaches*. San Diego, CA: Academic Press, Inc; 1990. p. 129-159.
- De Boeck P. Random item IRT models. *Psychometrika*. 2008; 73:533–559.
- De Boeck, P.; Wilson, M., editors. *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer; 2004.
- Ehri, LC. Development of the ability to read words. In: Barr, R.; Kamil, ML.; Mosenthal, PB.; Pearson, PD., editors. *Handbook of reading research*. Vol. Vol. 2. Mahwah, NJ: Lawrence Erlbaum Associates; 1991. p. 383-417.
- Ehri, LC. Reconceptualizing the development of sight word reading and its relationship to recoding. In: Gough, PB.; Ehri, LC.; Treiman, R., editors. *Reading Acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992. p. 107-143.
- Ehri, LC. Grapheme-phoneme knowledge is essential for learning to read words in English. In: Metsala, JL.; Ehri, LC., editors. *Word recognition in beginning literacy*. London, U.K.: Erlbaum; 1998. p. 3-41.
- Ehri LC, Robbins C. Beginners need some decoding skill to read words by analogy. *Reading Research Quarterly*. 1992; 27:12–26. Retrieved from: <http://www.jstor.org/stable/747831>.
- Ehri LC, Sweet J. Fingerprint-reading of memorized text: What enables beginners to process the print? *Reading Research Quarterly*. 1991; 26:442–462.

- Ehri LC, Wilce LS. Movement into reading: Is the first stage of printed word learning visual or phonetic? *Reading Research Quarterly*. 1985; 20:163–179.
- Fletcher-Flinn CM, Thompson GB. Learning to read with underdeveloped phonemic awareness but lexicalized phonological recoding: a case study of a 3-year-old. *Cognition*. 2000; 74:177–208. [PubMed: 10617781]
- Fletcher-Flinn CM, Thompson GB. Dissociation between deficits in explicit procedures and implicit processes in the visual-spatial and the phonological systems during reading acquisition. *Cognitive Neuropsychology*. 2007; 24:471–484. [PubMed: 18416502]
- Fox J-P. Multilevel IRT using dichotomous and polytomous response data. *British Journal of Mathematical and Statistical Psychology*. 2005; 58:145–172. [PubMed: 15969844]
- Frederiksen JR, Kroll JF. Spelling and sound: Approaches to the internal lexicon. *Journal of Experimental Psychology: Human Perception and Performance*. 1976; 2:361–379.
- Frith, U. Beneath the surface of developmental dyslexia. In: Patterson, K.; Marshall, J.; Coltheart, M., editors. *Surface dyslexia*. London, U.K.: Erlbaum; 1985.
- Goswami U. Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology*. 1986; 42:73–83.
- Goswami, U. The role of analogies in the development of word recognition. In: Metsala, JL.; Ehri, LC., editors. *Word recognition in beginning literacy*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1998. p. 41-63.
- Gottardo A, Chiappe P, Siegel LS, Stanovich KE. Patterns of word and nonword processing in skilled and less-skilled readers. *Reading and Writing: An Interdisciplinary Journal*. 1999; 11:465–487.
- Gough PB, Hillinger ML. Learning to read: An unnatural act. *Bulletin of the Orton Society*. 1980; 30:179–196.
- Gough PB, Tunmer WE. Decoding, reading, and reading disability. *Remedial and Special Education*. 1986; 7(1):6–10.
- Graham S. Word recognition skills of learning disabled children and average students. *Reading Psychology*. 1980; 2:23–33.
- Hatcher PJ, Hulme C, Ellis AW. Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*. 1994; 65:41–57.
- Hoffman L, Rovine MJ. Multilevel models for the experimental psychologist: Foundations and illustrative examples. *Behavior Research Methods*. 2007; 39:101–118. Retrieved from <http://brm.psychonomic-journals.org/content/39/1/101.full.pdf+html>. [PubMed: 17552476]
- Hogaboam TW, Perfetti CA. Reading skill and the role of verbal experience in decoding. *Journal of Educational Psychology*. 1978; 70:717–729.
- Hulme, C.; Mackenzie, S. *Working memory and severe learning difficulties*. Hove, U.K.: Lawrence Erlbaum; 1992.
- Janssen, R.; Schepers, J.; Peres, D. Models with item and item group predictors. In: De Boeck, P.; Wilson, M., editors. *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer; 2004. p. 189-212.
- Joe H. Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics and Data Analysis*. 2008; 52:5066–5074.
- Jorm AF, Share DL. Phonological recoding and reading acquisition. *Applied Psycholinguistics*. 1983; 4:103–147.
- Jorm AF, Share DL, Maclean R, Matthews RG. Phonological recoding skills and learning to read: a longitudinal study. *Applied Psycholinguistics*. 1984; 5:201–207.
- Juel C, Griffith PL, Gough PB. Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology*. 1986; 78:243–255.
- Kail R, Hall LK. Distinguishing short-term memory from working memory. *Memory & Cognition*, 29. 2001:1–29. Retrieved from <http://mc.psychonomic-journals.org/content/29/1/1.full.pdf+html>.
- Kim Y-S, Petscher Y, Foorman BR, Zhou C. The contributions of phonological awareness and letter-name knowledge to letter-sound acquisition – a cross-classified multilevel approach. *Journal of Educational Psychology*. 2010; 102:313–326.

- Kearns DK. Assessment of Grapheme-Phoneme Correspondence Knowledge. 2006 Unpublished test.
- Leslie L, Calhoun A. Factors affecting children's reading of rimes: Reading ability, word frequency, and rime-neighborhood size. *Journal of Educational Psychology*. 1995; 87:576–586.
- Luo W, Kwok O. The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research*. 2009; 44:182–212.
- MacMillan/McGraw-Hill. Treasures: A Reading/Language Arts Program. New York: 2008.
- Maier KS. A Rasch hierarchical measurement model. *Journal of Educational and Behavioral Statistics*. 2001; 26:307–330.
- Manis FR, Doi LM, Bhadha B. Naming speed, phonological awareness, and orthographic knowledge in second graders. *Journal of Learning Disabilities*. 2000; 33:333–374.
- Marinus E, de Jong PF. The use of sublexical clusters in normal and dyslexic readers. *Scientific Studies of Reading*. 2008; 12:253–280.
- Mason JM. When do children begin to read: An exploration of four year old children's letter and word reading competencies. *Reading Research Quarterly*. 1980; 15:203–227. Retrieved from <http://www.jstor.org/stable/pdfplus/747325.pdf>.
- McClelland JL, Rumelhart DE. An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*. 1981; 88:375–407.
- Moerbeek M. The consequences of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research*. 2004; 39:129–149.
- Molenberghs, G.; Verbeke, G. An introduction to (generalized (non)linear mixed models. In: De Boeck, P.; Wilson, M., editors. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York, NY: Springer; 2004. p. 111-153.
- Näslund JC. Phonemic and graphemic consistency: Effects on decoding for German and American children. *Reading and Writing: An Interdisciplinary Journal*. 1999; 11:129–152.
- Nation K, Snowling MJ. Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language*. 1998; 39:85–101.
- National Institute for Literacy. (n.d.). Put reading first: Kindergarten through grade 3. Retrieved from <http://www.nationalreadingpanel.org/>.
- No Child Left Behind Act of 2001. 2002 Pub. L. No. 107-110, 115 Stat. 1425.
- Olson, R.; Forsberg, H.; Wise, B.; Rack, J. Measurement of word recognition, orthographic, and phonological skills. In: Lyon, GR., editor. *Frames of Reference for the Assessment of Learning Disabilities: New Views on Measurement*. Baltimore, MD: Paul H. Brookes Publishing; 1994. p. 243-278.
- Perfetti, CA. Cognitive and linguistic components of reading ability. In: Foorman, BR.; Siegel, AW., editors. *Acquisition of reading skills: Cultural constraints and cognitive universals*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1986. p. 11-40.
- Perfetti, CA. The representation problem in reading acquisition. In: Gough, PB.; Ehri, LC.; Treiman, R., editors. *Reading acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992. p. 145-174.
- Pickering, SJ.; Gathercole, SE. *Working Memory Test Battery for Children - Manual*. London: Psychological Corporation UK; 2001.
- Preisser DA, Hodson BW, Paden EP. Developmental phonology: 18–9 months. *Journal of Speech and Hearing Disorders*. 1988; 53:125–130. [PubMed: 3361854]
- Rabash, J.; Browne, WJ. Non-hierarchical multilevel models. In: de Leeuw, J.; Meijer, E., editors. *Handbook of Multilevel Analysis*. New York, NY: Springer; 2007. p. 303-336.
- Raudenbush, SW.; Bryk, AS. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed.. Thousand Oaks, CA: Sage Publications, Inc; 2002.
- Reitsma P. Printed word learning in beginning readers. *Journal of Experimental Child Psychology*. 1983; 36:321–339.
- Share DL. Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*. 1995; 55:151–218. [PubMed: 7789090]
- Share DL. Phonological recoding and orthographic learning: A direct test of the self-teaching hypothesis. *Journal of Experimental Child Psychology*. 1999; 72:95–129. [PubMed: 9927525]

- Share DL. On the Anglocentricities of current reading research and practice: The perils of overreliance on an “outlier” orthography. *Psychological Bulletin*. 2008; 134:584–615. [PubMed: 18605821]
- Shmidman A, Ehri L. Embedded picture mnemonics to learn letters. *Scientific Studies of Reading*. 2010; 14:159–182.
- Siegel LS, Faux D. Acquisition of certain grapheme-phoneme correspondences in normally achieving and disabled readers. *Reading and Writing: An Interdisciplinary Journal*. 1989; 1:37–52.
- Snowling MJ. Phonemic deficits in developmental dyslexia. *Psychological Research*. 1981; 43:219–234. [PubMed: 7302091]
- Stanovich, KE. Word recognition: Changing perspectives. In: Barr, R.; Kamil, ML.; Mosenthal, PB.; Pearson, PD., editors. *Handbook of reading research*. Vol. 2. Mahwah, NJ: Lawrence Erlbaum Associates; 1991. p. 418-452.
- Tabachnick, BG.; Fidell, LS. *Using Multivariate Statistics*. 5th ed.. Boston, MA: Pearson Education, Inc; 2007.
- Tennessee/Metro Nashville Public Schools Academic Standards/Reading – Grade 1. Retrieved from <http://www.mnps.org/AssetFactory.aspx?did=3606>.
- Treiman R, Goswami U, Bruck M. Not all nonwords are alike: Implications for reading development and theory. *Memory & Cognition*. 1990; 18:559–567. Retrieved from http://www.psychonomic.org/backissues/6954/mc/vol18-6/PDFs/MC001_V18No6.pdf.
- Van den Noortgate W, De Boeck P, Meulders M. Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics*. 2003; 28:369–386.
- Van den Noortgate, W.; Paek, I. Person regression models. In: De Boeck, P.; Wilson, M., editors. *Explanatory item response models: A generalized linear and nonlinear approach*. New York, NY: Springer; 2004. p. 167-187.
- Wagner, R.; Torgesen, J.; Rashotte, C. *Comprehensive test of phonological processing*. Austin, TX: PRO-ED; 1999.
- Wilson, M.; De Boeck, P. Descriptive and explanatory item response models. In: De Boeck, P.; Wilson, M., editors. *Explanatory Item Response Models: A Generalized Linear and Nonlinear Approach*. New York, NY: Springer; 2004. p. 43-74.
- Wolf M, Bowers PG. The double-deficit hypothesis for the developmental dyslexias. *Journal of Educational Psychology*. 1999; 91:415–438.
- Wolf M, Bowers PG. Naming-speed processes and developmental reading disabilities: An introduction to the special issue on the double-deficit hypothesis. *Journal of Learning Disabilities*. 2000; 33:322–324. [PubMed: 15493094]
- Woodcock, R. *Woodcock Reading Mastery Tests*. Circle Pines, MN: American Guidance Service; 1987.
- Zeno, SM.; Ivens, SH.; Millard, RT.; Duvvuri, R. *The educator’s word frequency guide (CD-ROM)*. New York, NY: Touchstone Applied Science Associates; 1995.
- Ziegler JC, Goswami U. Reading acquisition, developmental dyslexia, and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, 131. 2005:3–29.
- Ziegler JC, Stone GO, Jacobs AM. 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*. 1997; 29:600–618. Retrieved from: <http://www.psychonomic.org/backissues/126/B245.pdf>.

APPENDIX A

	Complex vowel grapheme	Complex consonant grapheme
infrequent rime unit	goob	naff
	naif	fidd

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

APPENDIX B

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

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

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

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

Level-1 (Responses)

R1 R2 R3 R4 R5 R6 R7 R8...R3920

Level-2 (Person & Word)

J1 J2... J196

W1 W2... W20

Level-3 (Classroom)

C1... C56

Level-4 (School)

S1 (... S15)

Person Side

Word Side

Figure 1.
Graphical display of 4-level data structure.

Figure 2.
Graph of interaction between average rime-neighbor frequency and rime-neighborhood size.
4

⁴Although the rime frequency variable ranged from -0.69 to 3.36 and the rime size variable ranged from -0.97 to 2.20 , the graph depicts a restricted range based on realistic combinations of those values from our sample of words.

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

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	Level 1 (person-by-word): GPCK
2. What are the effects of phonemic awareness, naming speed, and working memory in predicting decoding accuracy after controlling for GPC knowledge and word characteristics? How much person variance do these factors account for?	2	Level 1 (person-by-word): GPCK Level 2 (person): PA, RN, WM Level 2 (word): R_FREQ, R_SIZE, FREQ*SIZE, VC
3. What are the effects 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 and person characteristics? How much word variance do these factors account for?	2	Level 1 (person-by-word): GPCK Level 2 (person): PA, RN, WM Level 2 (word): R_FREQ, R_SIZE, FREQ*SIZE, VC
4. Are there interactions between the person and word characteristics?	3	Level 1 (person-by-word): GPCK Level 1 (person-by-word interaction): WM*(S/F), PA*VC Level 2 (person): PA, RN, WM Level 2 (word): R_FREQ, R_SIZE, FREQ*SIZE, VC

Table 2

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

	Mean	Median	SD	Minimum	Maximum
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

Means and SDs of Word Difficulties (N = 20)

Word	M	SD
guck	.62	.49
fidd	.55	.50
pean	.28	.45
jeth	.60	.49
goob	.34	.48
naif	.15	.36
voun	.28	.45
nish	.72	.45
dess	.43	.50
naff	.60	.50
feem	.56	.50
vell	.59	.50
mout	.44	.50
bosh	.35	.48
keab	.16	.37
sath	.55	.50
tesh	.36	.49
dail	.28	.45
soom	.61	.49
teek	.69	.46

Table 4

Description of Student Sample (N = 196)

	Mean	Median	SD	Minimum	Maximum
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 5

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 $\geq |.18|$ are significant at the $\alpha = .05$ level.

Table 6

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

Model	# of parameters	Model Fit Indices			Model Comparison	
		Log-likelihood	AIC	BIC	Referent Model	χ^2 Test of Equivalence
Full Base	5	-2293	4597	4628	n/a	n/a
Base w/o word	4	-2534	5077	5102	Full Base	481.81, $p < .001$
Base w/o person	4	-2366	4739	4764	Full Base	144.56, $p < .001$
Base w/o school	4	-2296	4601	4626	Full Base	5.95, $p = .01$
Base w/o classroom	4	-2297	4601	4626	Full Base	6.35, $p = .01$

Table 7

Results from Full Base Model (N = 196)

<i>Fixed Effect</i>				
Variable	Coefficient	SE	z	p
Intercept, ω_{000}	-0.31	0.27	-1.14	0.25
<i>Random Effects</i>				
Source	Variance			
Person (intercept), σ_{r01}	0.64			
Classroom (intercept), σ_u	0.20			
School (intercept), σ_t	0.31			
Word (intercept), σ_{r02}	0.82			

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

Table 8

Results from Model 1 (N = 196)

<i>Fixed Effects</i>					
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	
<i>Random Effects</i>					
Source	Variance				
Person (intercept), σ_{r01}	1.11				
Person (GPCK), σ_{r11}	0.47				
Person (intercept, GPCK), σ_{r01*11}	-0.53				
Classroom (intercept), σ_u	0.20				
School (intercept), σ_r	0.26				
Word (intercept), σ_{r02}	0.92				
Word (GPCK), σ_{r12}	0.20				
Word (intercept, GPCK), σ_{r02*12}	-0.20				

Note. Coefficients are in log-odds units. GPCK = word-specific grapheme-phoneme correspondence knowledge.

Table 9

Results from Model 2 (N = 196)

<i>Fixed Effects</i>				
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	<.001
Rapid naming, γ_{20}	-0.17	0.08	-2.29	.02
Working memory, γ_{30}	0.03	0.08	0.33	.74
Rime freq., γ_{40}	0.76	0.21	3.52	<.001
Rime 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
<i>Random Effects</i>				
Source	Variance			
Person (intercept), σ_{r01}	0.93			
Person (GPCK), σ_{r11}	0.45			
Person (intercept, GPCK), σ_{r01*11}	-0.44			
Classroom (intercept), σ_u	0.19			
School (intercept), σ_t	0.14			
Word (intercept), σ_{r02}	0.77			
Word (GPCK), σ_{r12}	0.21			
Word (intercept, GPCK), σ_{r02*12}	-0.30			

Note. Coefficients are in log-odds units. GPCK = word-specific grapheme-phoneme correspondence knowledge. R_FREQ = average rime-neighbor frequency. R_SIZE = rime-neighborhood size.

Table 10

Results from Model 3 (N = 196)

<i>Fixed Effects</i>				
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
Rime freq, γ_{40}	0.77	0.22	3.58	<.001
Rime 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*Rime freq, λ_2	0.01	0.04	0.26	.80
Phonemic awareness* grapheme complexity, λ_3	0.17	0.08	2.18	.03
<i>Random Effects</i>				
Source	Variance			
Person (intercept), $\sigma_{\rho 01}$	0.92			
Person (GPCK), $\sigma_{\rho 11}$	0.46			
Person (intercept, GPCK), $\sigma_{\rho 01*11}$	-0.47			
Classroom (intercept), σ_{η}	0.19			
School (intercept), σ_{τ}	0.14			
Word (intercept), $\sigma_{\rho 02}$	0.79			
Word (GPCK), $\sigma_{\rho 12}$	0.20			
Word (intercept, GPCK), $\sigma_{\rho 02*12}$	-0.30			

Note. Coefficients are in log-odds units. GPCK = word-specific grapheme-phoneme correspondence knowledge. R_FREQ = average rime-neighbor frequency. R_SIZE = rime-neighborhood size.