Journal of Educational Psychology

Exploring Individual Differences in Irregular Word Recognition Among Children With Early-Emerging and Late-Emerging Word Reading Difficulty
Laura M. Steacy, Devin M. Kearns, Jennifer K. Gilbert, Donald L. Compton, Eunsoo Cho, Esther R. Lindstrom, and Alyson A. Collins


CITATION
Exploring Individual Differences in Irregular Word Recognition Among Children With Early-Emerging and Late-Emerging Word Reading Difficulty

Laura M. Steacy, Vanderbilt University
Jennifer K. Gilbert, Metropolitan Nashville Public Schools, Nashville, Tennessee

Devin M. Kearns, University of Connecticut
Donald L. Compton, Eunsoo Cho, Esther R. Lindstrom, and Alyson A. Collins, Vanderbilt University

Models of irregular word reading that take into account both child- and word-level predictors have not been evaluated in typically developing children and children with reading difficulty (RD). The purpose of the present study was to model individual differences in irregular word reading ability among 5th grade children (N = 170), oversampled for children with RD, using item-response crossed random-effects models. We distinguish between 2 subtypes of children with word reading RD, those with early emerging and late-emerging RD, and 2 types of irregular words, "exception" and "strange." Predictors representing child-level and word-level characteristics, along with selected interactions between child- and word-characteristics, were used to predict item-level variance. Individual differences in irregular word reading were predicted at the child level by nonword decoding, orthographic coding, and vocabulary; at the word level by word frequency and a spelling-to-pronunciation transparency rating; and by the Reader group Imageability and Reader group Irregular word type interactions. Results are interpreted within a model of irregular word reading in which lexical characteristics specific to both child and word influence accuracy.

Keywords: individual differences, irregular word reading, reading difficulty

An essential development in learning to read is the acquisition of automatic word reading skills (defined in this study as the ability to pronounce written words in isolation) that are impenetrable to factors such as knowledge and expectation (Perfetti, 1992; Stanovich, 1991). Automaticity of word reading allows fluent and reliable retrieval of word representations from the orthographic lexicon, activating phonological, syntactic, morphological, and semantic information to be used by the reader to form faithful representations of text (e.g., Kintsch & Rawson, 2005; Perfetti, Landi, & Oakhill, 2005). As children learn to read, the orthographic lexicon expands via an increase in the absolute number of orthographically addressable entries, referred to as "word-specific" representations (Castles & Nation, 2006; Ehri, 2014; Perfetti & Stafura, 2014). Word-specific representations are considered to be less dependent on phonological processes because these representations have been supplanted by specific connections linking spelling directly to pronunciations (Perfetti, 1992; Share, 1995).

The addition of word-specific representations in developing readers likely depends on both child- (Vellutino, Fletcher, Snowling, & Scanlon, 2004) and word-level (Balota et al., 2007; Seidenberg, Waters, Barnes, & Tanenhaus, 1984) factors. Children add word-specific entries to the orthographic lexicon, to a large extent, by employing a phonologically based recoding mechanism. Phonological recoding is the process of translating a printed word to speech by employing grapheme-phoneme correspondence rules. Thus, phonological recoding functions as a self-teaching mechanism (Jorm & Share, 1983; Share, 1995; Share & Stanovich, 1995) that builds the orthographic lexicon through item-specific learning (see Nation, Angell, & Castles, 2006).
2007; Wang, Nickels, Nation, & Castles, 2013). However, the addition of word-specific orthographic representations in English is likely modulated by child-level factors beyond phonological-based recoding skill (i.e., phonological awareness and nonword decoding skills) that include, but are not limited to, vocabulary knowledge, orthographic processing, rapid automatized naming, and print experience (see Cunningham, Perry, & Stanovich, 2001; Harm & Seidenberg, 2004; Keenan & Betjemann, 2008; Nation & Snowling, 1998; Plaut, McClelland, Seidenberg, & Patterson, 1996). In addition, word-level features (e.g., frequency, length, regularity, and imageability) likely affect the ease with which words are added to the orthographic lexicon as well (see Coltheart, Laxon, & Keating, 1988; Wang, Nickels, Nation, & Castles, 2013; Waters, Bruck, & Seidenberg, 1985; Waters, Seidenberg, & Bruck, 1984).

Because English orthography is semiopaque (i.e., spelling-tosound relationships are not consistent), words are often characterized as being regular or irregular. Although this binary distinction between regular and irregular words is not totally accurate, English orthography is best described as quasi-regular (see Plaut, 1999; Seidenberg, 2005), it is relevant within a self-teaching framework describing the development of the orthographic lexicon in children (Wang, Castles, & Nickels, 2012). From a self-teaching perspective, irregular words are challenging for developing readers by requiring them to grapple with the quasi-regular system governing English orthography (Plaut et al., 1996). As a result, studies have shown that it is harder for children to acquire orthographic representations of irregular words compared with regular words, and these orthographic representations may be less precise until they become fully automatized (see Wang et al., 2012; Wang, Castles, Nickels, & Nation, 2011; Wang et al., 2013).

The purpose of the present study was to develop a comprehensive model of irregular word reading skill in developing readers that simultaneously takes into account both child- and word-level predictors, emphasizes performance differences between typically developing (TD) children and children with reading disability (RD), and focuses on the role of lexical influences on word reading accuracy. This study was motivated by results from computational models (Plaut et al., 1996), experimental learning studies (Wang et al., 2011, 2012, 2013), and comparison studies (Waters, Seidenberg, & Bruck, 1984) of irregular word reading suggesting an important interplay between word regularity, word frequency, child phonological skills (i.e., phonemic awareness and nonword decoding) and child semantic knowledge. For instance, Wang et al. (2013) concluded

> when phonological decoding can be only partially successful, as was the case with irregular words in this study, orthographic learning was assisted by factors such as vocabulary knowledge. On the other hand, when phonological decoding is not compromised, vocabulary knowledge does not appear to provide additional assistance to acquiring orthographic representations. (p. 14)

Plaut et al. further expand on the relationship between child phonological skills, child semantic knowledge, word frequency, and word regularity by stating

> the reading system learns gradually to be sensitive to the statistical structure among orthographic, phonological, and semantic representations and these representations simultaneously constrain each other in interpreting a given input . . . As a result, words with a relatively weak semantic contribution (e.g., abstract or low-imageability words) exhibit a stronger frequency by consistency interaction—in particular, naming latencies and error rates are disproportionately high for items that are weak on all three dimensions: abstract (low-imageability), low-frequency, exception words. (pp. 99–101)

Thus, these studies suggest an important interplay between word-level and child-level factors in explaining irregular word reading variance.

Our study is unique in the sense that it combines child-, word-, and child by word interactions into a single model of irregular word reading. Very few studies have combined both child- and word-level predictors into a single model predicting item-level variance in word reading ability. Those that have include models of nonword reading (Gilbert, Compton, & Kearns, 2011), morphologically complex word reading (Goodwin, Gilbert, & Cho, 2013; Kearns et al., 2014), and multisyllabic (Kearns, 2015) word reading. A comprehensive model of irregular word reading has the potential to provide important insights into the relationships between child and word factors that affect word reading development. Such a model also affords the opportunity to explore potentially important child by word interactions that might ultimately inform studies designed to explore how to improve the word reading abilities of struggling readers. We proceed by providing a brief review of the literature that influenced the selection of word and child factors included in the model.

### Word-Level Predictors

In the present study we attempted to include a comprehensive set of word-level factors known, or suspected, to predict individual differences in irregular word reading among developing readers. The predictors included regularity, frequency, length, imageability, orthographic neighborhood size, and spelling-to-pronunciation transparency.

### Regularity

A sizable literature exists examining the effects of regularity on word reading accuracy in typically developing and RD children (see Metsala, Stanovich, & Brown, 1998). Words such as made and best are considered regular because their pronunciations are predictable on the basis of simple spelling-sound rules (Rastle & Coltheart, 1999; Venezky, 1999), and all words with similar rime patterns (-ade, -ext) rhyme. Words, such as have and give, are considered irregular because their pronunciations violate simple spelling-sound correspondences, and they have no rhymes with similar pronunciations (Glushko, 1979). This class of irregular words has often been referred to as “exception” words. Waters, Seidenberg, and colleagues (Seidenberg, Waters, Barnes, & Tannenhaus, 1984; Waters et al., 1985; Waters et al., 1984; Waters & Seidenberg, 1985) further subdivided irregular words by including a strange word class. Strange words (e.g., yacht) have irregular pronunciations like exception words, but unlike exception words they also contain spelling patterns that occur in very few or no other English words (i.e., strange words do not share rime patterns with other words). Because strange words have deviations in both orthographic and phonological representation it has been suggested that they may be processed differently from exception words (Seidenberg et al., 1984).
Studies have documented differential effects of word class (i.e., regular, exception, and strange) and word frequency as a function of reading skill in developing readers (Waters, Seidenberg, & Bruck, 1984). In general, TD readers made more errors in recognizing exception words than regular and strange words, whereas children with RD made more errors in recognizing strange and exception words compared to regular words (Waters et al., 1985). In the present study we coded our irregular words as either exception or strange.

Frequency

Research examining individual differences in word reading among developing readers report fairly robust interactions between word frequency and reading skill. Compared with TD children, RD children tend to show greater effects of word frequency on word reading accuracy (see Kuperman & Van Dyke, 2013; Waters et al., 1984, 1985). Results suggest skilled readers are better able than struggling readers to identify a large pool of high frequency words without interference from irregular spelling-sound correspondences (Waters et al., 1985). We include an estimate of printed word frequency for our irregular words as estimated by Zeno, Ivens, Millard, and Duvvuri (1995).

Length

In developing readers word length is related to both the speed and accuracy of word reading, particularly in more transparent orthographies (e.g., Bijeljac-Babic, Millogo, Farioli, & Grainger, 2004; De Luca, Barca, Burani, & Zoccolotti, 2008). In addition, stronger word length effects have been reported for children with RD (Martens & de Jong, 2006; Ziegler, Perry, & Coltheart, 2003; Zoccolotti et al., 2005), perhaps reflecting impairment in the application of larger orthographic units in parallel when reading unfamiliar words (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001; Hautala, Aro, Eklund, Lerkakanen, & Lytynen, 2013). Given that we were interested in differences between TD and RD children on irregular word reading we included number of letters as a predictor.

Imageability

Word imageability has also been shown to be a strong predictor of word reading accuracy in developing readers (Laing & Hulme, 1999). Imageability is a word feature that captures the ease with which a word can elicit a mental image in the reader (Paivio, Yuille, & Madigan, 1968). To date, some studies have reported a main effect for imageability on word reading accuracy with no interaction for regularity (Duff & Hulme, 2012; Laing & Hulme, 1999; Monaghan & Ellis, 2002) whereas others have reported that the effects of imageability have been isolated to irregular words (e.g., Strain & Herdman, 1999; Strain, Patterson, & Seidenberg, 2002). Others have found that imageability is particularly important for poor readers (Coltheart, Laxon, & Keating, 1988; Steacy et al., 2013). Some have suggested that imageability affects irregular word reading by eliciting an itemspecific mental image associated with the word that helps to facilitate the formation of word-specific orthographic-to-phonological connections in irregular words and leads to the establishment of more stable word representations (see Keenan & Betjemann, 2008). Because of the conflicting evidence regarding the interaction between imageability and regularity, we have included it as a predictor in our model of irregular word reading.

Orthographic Neighborhood Size

The most frequently used measure to investigate the influence of similar orthographic representations on word reading performance is sensitivity to orthographic neighborhood size. The orthographic neighborhood size of a given word represents all of the existing words that can be created by substituting one of its letters for another one (Coltheart, Davelaar, Jonasson, & Besner, 1977). Studies have found that orthographic neighborhood size is related to word reading and has more of an effect on familiar than unfamiliar words (Laxon, Gallagher, & Masterson, 2002; Laxon, Masterson, & Moran, 1994). Laxon, Coltheart, and Keating (1988) reported that poor readers exhibited larger orthographic neighborhood size effects for words than good readers. For this reason we include a measure of orthographic neighborhood size as a word-level predictor.

Spelling-to-Pronunciation Transparency

It is clear that irregular words vary significantly in terms of the distance between regularized decoding pronunciation and the actual phonological representation in the lexicon (known as partial decoding). For example, the distance from decoding pronunciation to the phonological representation for irregular words like yacht and suede appear larger than words like pint and touch and this may affect the ability of developing readers to efficiently add irregular items to their orthographic lexicons (see Plaut et al., 1996). Yet, this distance has not been used as a word-level characteristic to explain irregular word reading variance in developing readers. In our study, we build off of the set for variability paradigm (Elbro et al., 2012; Tunmer, & Chapman, 2012) by asking experts to rate the ease of determining the lexical representation of the irregular words based on faithful application of decoding rules. The assumption was that the smaller the distance the simpler it would be to make the match between partial decoding and the actual lexical representation, and thus the easier it would be to add the word to the orthographic lexicon. We included a spelling-to-pronunciation transparency rating as a word-level predictor in the model.

Child-Level Predictors

In addition to word features, child skills likely to play an important role in predicting irregular word reading accuracy were included in the model. We include a comprehensive set of childlevel factors known to predict individual differences in the word reading of developing readers. The predictors included reader group, phonological awareness, nonword decoding, vocabulary, orthographic coding, rapid automatized naming, and a measure of print exposure.

Reader Group

There is a growing body of literature to suggest that some students who exhibit typical reading skill growth in the early years develop difficulty with word reading later in elementary school as the demands of reading increase (Catts, Compton, Tomblin, & Bridges, 2012; Leach, Scarborough, & Rescorla, 2003; Lipka, Lesaux, & Siegel, 2006). One potential contributor to these unexpected reading difficulties may be the increasing demands placed on the reader at
the word level, as students are expected to read more complex text and they are faced with new words many of which are lower in frequency and contain irregular spelling patterns and thus require the formation of complex connections between phonology, orthography, and semantics (Catts et al., 2012; Leach et al., 2003). In the present study, we were particularly interested in whether differences exist across reader groups representing TD, late-emerging RD, and early emerging RD on irregular word reading. Late-emerging RD refers to students who have been identified with reading difficulties later in elementary school while early emerging RD refers to students that have been identified in first grade.

**Phonological Awareness**

Both cross-sectional and longitudinal studies indicate that phonological awareness (PA) has a high correlation with and accounts for unique variance in word reading (Kirby, Parrila, & Pfeiffer, 2003; Melby-Lervåg, Lyster, & Hulme, 2012). Deficits in phonological processing skills have been causally linked to poor word-identification skills through a mechanism that disrupts the development of decoding skills (Brady & Schankweiler, 1991; Bruck, 1992; Siegel & Faux, 1989; Stanovich & Siegel, 1994; Torgesen, 2000; Vellutino et al., 1996). Deficits in the ability to recognize and manipulate the phonemes of words are believed to disrupt the acquisition of spelling-to-sound translation routines that form the basis of early decoding-skill development (Bus & van Ijzendoorn, 1999; van Ijzendoorn & Bus, 1994; Rack, Snowling, & Olson, 1992). For this reason, we include PA as a child-level predictor in the model.

**Nonword Decoding**

Nonword reading is considered a proxy of decoding skill. Even though irregular words are not decodable there is a close relationship between decoding skill and irregular word reading with the association being stronger in TD compared to RD children (Griffiths & Snowling, 2002). Griffiths and Snowling (2002) speculated that, “even the ability to read words that do not conform to regular grapheme–phoneme correspondences depends on having access to segmental phonological representations” (pp. 40–41). Thus, we incorporated a measure of nonword decoding in the irregular word reading model.

**Vocabulary**

Keenan and Betjemann (2008) have speculated that semantic activation may help to “fill voids” in phonological-orthographic processing in individuals with poor mappings, such as children with RD (p. 193). A growing literature implicates the role of lexical knowledge (e.g., semantic or lexical phonology) in the learning of irregular words by developing readers (see Harm & Seidenberg, 2004; Plaut et al., 1996; Ricketts, Nation, & Bishop, 2007). For this reason we included a measure of vocabulary ability in the model.

**Orthographic Coding**

Evidence suggests that orthographic coding measures a skill distinct from phonological decoding, print exposure, and other reading related skills (e.g., Cunningham & Stanovich, 1991; Hagiliassisi, Pratt, & Johnston, 2006). Orthographic coding measures have been shown to be a unique predictor of children’s word reading after controlling for PA and RAN (Cunningham, Perry, & Stanovich, 2001). Thus we include orthographic coding as a child-level measure.

**Rapid Automated Naming**

Rapid automated naming (RAN) of familiar stimuli such as objects, colors, digits, or letters repeatedly accounts for unique variance in both concurrent and future reading and spelling achievement (for a review see Kirby et al., 2010). Research indicates that these relationships hold even after controlling for socioeconomic status, IQ, and PA (e.g., Kirby et al., 2010; Lervåg & Hulme, 2009) and that students with good and poor reading skills differ in their performance on RAN tasks (Bowers, 1995). We therefore include RAN as a child-level predictor.

**Print Exposure**

A number of researchers have reported evidence in support of an association between reading experience (indexed by measures of print exposure) and reading skill (Cunningham & Stanovich, 1991; McBride-Chang, Manis, Seidenberg, Custodio, & Doi, 1993; cf. Barker, Torgesen, & Wagner, 1992). Griffiths and Snowling (2002) reported that in developing readers print exposure was significantly associated with irregular word reading but not nonword decoding. A measure of print exposure was included as a child-level predictor.

**Research Questions**

In the present study we ask three related research questions regarding irregular word reading in 5th grade students. The first explores whether differences exist in irregular word reading between classes of children identified as late-emerging RD and those identified as early emerging RD and TD. We predict an order effect in which the TD group outperforms both RD groups and the late-emerging RD group outperforms the early emerging RD group, suggesting that the early emerging RD group has more severe word reading difficulties. The second explores the relative role of child- and word-level characteristics as predictors of item-level variance on irregular word reading accuracy, with a particular interest in the role of child- and word-level lexical influence. We predict that both child- and word-level features associated with lexical processing will make unique and significant contributions to item-level irregular word reading performance. At the child level we hypothesize that vocabulary will uniquely predict irregular word reading performance after controlling for nonword decoding, phonemic awareness, rapid automated naming, print exposure, and orthographic processing. At the word level we hypothesize that imageability and a rating of spelling-to-pronunciation ease will uniquely predict irregular word reading performance after controlling for word frequency, word length, orthographic neighborhood size, and orthographic distinctiveness (i.e., exception vs. strange word). The final research question is exploratory in nature, investigating the importance of child-level by word-level interactions in explaining irregular word reading variance, with a specific focus on interactions between RD status (i.e., late-emerging RD, early emerging RD, and TD) and word-level characteristics (i.e., frequency, imageability, spelling-to-pronunciation transparency rating, and whether an irregular word is strange). Although these analyses are exploratory, we were
particularly interested in whether differences existed between RD groups in terms of the relative importance of word- and child-level variables associated with lexical processing (i.e., imageability, spelling-to-pronunciation transparency rating, and vocabulary knowledge) in predicting item-level irregular word reading.

This study extends the literature in several important ways. Our study is the first to employ item-response crossed random-effects models (Bates, Maechler, & Bolker, 2013) to explain variability in children’s irregular word reading at the item level using a comprehensive set of predictors. In doing so we are able to model the unique contributions of child, word, and Child Word interactions simultaneously. In addition, we include two word-level measures, imageability and a spelling-to-pronunciation transparency rating, which have not previously been incorporated into models of irregular word reading. Finally, this study adds to the literature by distinguishing between three important reader classes based on word reading skill development over time, those who have early emerging RD, late-emerging RD, and TD. In proposing a comprehensive model of irregular word reading we were particularly interested in the relative importance of various lexical processing measures at the word-level (imageability and spelling-to-pronunciation transparency ratings) and child-level (e.g., vocabulary knowledge) across RD groups as predictors of item-level irregular word reading. This allows us to interpret our item-level analyses within the broader literature examining irregular word reading in TD and RD children, as well as search for “malleable factors” that might be exploited to improve the irregular word reading abilities of struggling readers.

Method

Participants

Participants were drawn from a multiyear cohort longitudinal study examining response-to-intervention decision rules (see Compton et al., 2010) and prevention efficacy (see Gilbert et al., 2013) in first grade children. The subjects in this sample are identical to those reported in Kearns (2015) and thus the sampling procedures are the same. For this study children were assessed at the end of first through fourth grades on measures of word identification and comprehension. Latent transition analyses were used to assign each child to either RD (late-emerging RD or early emerging RD) or TD classes as a function of time. In this study late-emerging RD membership was defined as a child who transitioned from an initial classification of TD to RD over time; early emerging RD as a child who was assigned to the RD class at the end of first grade and remained in the class across time; and TD as a child who was assigned to the TD class at the end of first grade and remained in that class over time. (A very small number of children in the sample transitioned from RD to TD over time but this group was not included in this study.) In the case of word reading (-W), LTA allowed the identification of classes representing TD, early emerging RD-W, late-emerging RD-W and for reading comprehension (-C) classes representing TD, early emerging RD-C, and late-emerging RD-C. Results from the two LTA models were combined to further identify early emerging RD-CW and late-emerging RD-CW classes. Thus, seven latent classes were identified: TD, early-emerging RD-W, -C, -CW, late-emerging RD-W, -C, and -CW. Counts for the various reading classes identified through LTA in fourth grade as a function of cohort are displayed in Table 1. We then selected from the larger fourth grade sample a subsample of children to be assessed in the fall of fifth grade. These target children consented to three 1-hr testing sessions measuring reading, language, knowledge, executive function, and attention. Our sampling strategy attempted to consent all early emerging RD and late-emerging RD children in a given cohort and then to randomly select TD children to assess. Table 1 provides the number of children who were consented and administered the fifth-grade battery. Since this study specifically targeted word-reading skills we only selected early emerging RD and late-emerging RD classes in which word reading difficulties were present: early emerging RD-W (n 1), early emerging RD-CW (n 18), late-emerging RD-W (n 15), and late-emerging RD-CW (n 30) along with TD children (n 109). The wordonly and the mixed word and comprehension difficulty groups were combined for the early emerging RD (n 19) and lateemerging RD (n 45) groups. Three children had missing data on some measures and were removed from the analyses. The final sample included 170 students in the following classes: TD (n 108), late-emerging RD (n 43), and early emerging RD (n 19). More detail regarding the Latent Transition Analyses and a description of the specific sampling plan for each of the cohorts is provided in Appendix A.

Measures

LTA measures (Grades 1–4).

<table>
<thead>
<tr>
<th>Reading</th>
<th>Cohort 1</th>
<th>Cohort 2</th>
<th>Cohort 3</th>
<th>Total sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>class</td>
<td>4th grade/5th grade</td>
<td>4th grade/5th grade</td>
<td>4th grade/5th grade</td>
<td>4th grade/5th grade</td>
</tr>
<tr>
<td>TD</td>
<td>172/38</td>
<td>64/30</td>
<td>86/41</td>
<td>322/109</td>
</tr>
<tr>
<td>ERD-W</td>
<td>1/1</td>
<td>1/0</td>
<td>1/1</td>
<td>3/1</td>
</tr>
<tr>
<td>ERD-C</td>
<td>28/11</td>
<td>23/10</td>
<td>13/5</td>
<td>64/26</td>
</tr>
<tr>
<td>ERD-CW</td>
<td>15/8</td>
<td>10/4</td>
<td>10/6</td>
<td>35/18</td>
</tr>
<tr>
<td>LERD-W</td>
<td>10/6</td>
<td>6/3</td>
<td>8/6</td>
<td>24/15</td>
</tr>
<tr>
<td>LERD-C</td>
<td>36/27</td>
<td>35/12</td>
<td>30/17</td>
<td>101/56</td>
</tr>
<tr>
<td>LERD-CW</td>
<td>23/16</td>
<td>18/4</td>
<td>16/10</td>
<td>57/30</td>
</tr>
</tbody>
</table>

Note. All children who did not move stayed eligible for the study. ERD Early identified reading difficulty; LERD Late emerging reading difficulty; TD Typically developing. Bolded 5th grade numbers represent the sample used in the present study.

Total Sample of 4th Grade Students, 5th Grade Students Sampled, and Counts of the Various Reading Classes Derived From Latent Transition Analysis as a Function of Cohort

The table above shows the distribution of students across different reading classes from fourth to fifth grade. The counts are given for each class, and the total sample is also provided. The note at the bottom indicates that all children who did not move stayed eligible for the study, and bolded numbers represent the sample used in the present study.
**Word identification.** Word identification was measured with the Word Identification subtest from the Woodcock Reading Mastery Tests—Revised/Normative Update (Woodcock, 1998). For this task, children were asked to read words aloud one at a time. The test was not timed but children were encouraged to move to the next item after a 5-s silence. Correct pronunciations were counted as correct and the total score was the sum of correct items. Basal and ceiling rules were applied. The examiner’s manual reports the split half reliability for fifth grade students as .91 (Woodcock, 1998).

**Passage comprehension.** General reading comprehension was measured with the Passage Comprehension subtest from the Woodcock Reading Mastery Tests—Revised/Normative Update (Woodcock, 1998). For this test, children are asked to silently read 1 to 2 sentence prompts in which a single word had been removed. Children were asked to provide the omitted word. Basal and ceiling rules were applied. Split-half reliability, provided by the Technical Manual, is .91 for 9-year olds and .89 for 10-year olds (Woodcock, 1998).

**Child measures.**

**Irregular word reading.** Irregular word reading was the dependent variable of interest. Irregular word reading ability was assessed using an experimental word list developed by Adams and Huggins (1985). The list contained 50 words with irregular spelling-to-sound correspondences that varied in word frequency. According to the authors, the frequencies of the words ranged from 134.1 per million to .12 per million according to Carroll, Davies, and Richman’s (1971) dispersion-adjusted norms (U-scale). The 50 words were selected from a set of 80 words, used in pilot testing with 80 children, so as to exclude words that were of inordinate ease or difficulty given their frequency and words that appeared to be beyond the children’s listening or speaking vocabularies.

Words were arranged to increase in difficulty and students attempted to read all words. The list contained words that prove very difficult to recode phonologically because of the complex orthographic patterns they comprise (e.g., ocean, heights, tongue, guitar, recipe). The list contained both exception and strange words based on Seidenberg et al. (1984). The list was originally developed for second through fifth graders. One item was dropped from the analyses due to an administration error. Scores ranged from 0 to 49 for this sample. The internal-consistency reliability for this task was .94.

**Reader group.** Children were classified into one of three reading groups based on the LTA analyses (see above): early emerging RD (early emerging RD-W & early emerging RD-CW), lateemerging RD (late-emerging RD-W & late-emerging RD-CW), and TD children. To contrast group performance two dummy codes were created comparing the early emerging RD group to the late-emerging RD group (designated as early emerging RD) and comparing the TD group to the late-emerging RD group (designated as TD).

**Orthographic choice.** The orthographic choice task in this study was a shortened version of the one used by Olson, Kliegl, Davidson, and Foltz (1985). Whereas the original test had 80 items, the current version had only 40 (only the odd-numbered items from the original test were retained). Children were presented two sheets of paper, each containing two columns of test items. They were asked to circle the real word in each pair. Each item comprised the correctly spelled word and a pseudohomophone foil (e.g., rain and ran). Children completed and received feedback on 4 practice items prior to beginning the test. The total score was the sum of the correct items. The minimum score in this sample was 21 and the maximum score was 40. Coefficient alpha for our sample was .76.

**Phonological awareness.** Phonological awareness was measured with the Ellison subtest of the Comprehensive Test of Phonological Processing (CTOPP, Wagner, Torgesen, & Rashotte, 1999). In this test, children were presented a word, asked to repeat the word, and then asked to say the word without a specified syllable for the first 3 items and without a specified phoneme for the remaining 17 items. Items are ordered by increasing difficulty, and the examiner discontinued administration after three consecutive incorrect items. In addition to the 20 test items (for 5 of which examiners provided performance feedback), 6 practice items were administered. The total score was the sum of correct items. Scores ranged from 1 to 20 in this sample. Coefficient alpha provided by the manual for age 10 was .91 and age 11 was .86.

**Nonword decoding.** The Woodcock Reading Mastery Test—Revised—Normative Update: Word Attack (Woodcock, 1998), a norm referenced test, evaluates students’ ability to pronounce pseudowords presented in list form. It contains 45 nonsense words ordered from easiest to most difficult. Students were asked to read (decode) the words aloud, one at a time. The developerrecommended basal and ceiling rules were applied to minimize boredom and frustration. Scores ranged from 0–39 in this sample. Split-half reliability exceeded .90 for the sample.

**Rapid automatic naming.** Rapid automatized naming was assessed using the Rapid Letter Naming subtest of the CTOPP (Wagner, Torgesen, & Rashotte, 1999). Two versions of the test were given. On both, six letters were randomly printed in four rows of nine letters. After ensuring each child could identify the letters, the child was asked to name the letters as fast as possible. The total score was the number of seconds it took the child to name the letters on both tests. Scores ranged from 21–80 in this sample. Test–retest reliability was .72 for children of ages 8–17 years per the test manual.

**Vocabulary.** Receptive vocabulary was measured using the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007). Students were asked to select one of four pictures when presented with a spoken word. Scores ranged from 103 to 195 for this sample. The alternate-forms reliability coefficient, Cronbach’s alpha, and split half reliabilities for 10 year olds are .87, .96, and .93, respectively, as per the manual.

**Book Title Questionnaire.** Print exposure was measured with the Book Title Questionnaire (Beall, 2011). This self-report questionnaire was adapted from Cunningham and Stanovich (1991) in two ways: books that were made into movies were removed and book titles were updated to include more recent popular books. Fifty real book titles along with 14 foil titles were presented. Children were asked simply to check the titles that they knew were real books. They were informed that some of the titles were not real, and therefore encouraged not to guess but only to mark the titles they knew were real books. The total score was the number of real titles checked minus the number of foils checked. Samplebased coefficient alpha for our sample was .94 based on all items (real and foil).

**Word measures.**

**Frequency.** The metric used for word frequency was the standard frequency index (SFI) from the Educator’s Word Frequency Guide (Zeno, Ivens, Millard, & Duuvuri, 1995). SFI represents a
logarithmic transformation of the frequency of word type per million tokens within a corpus of more than 60,000 samples of texts from various sources. These sources range from textbooks to popular literature. The range of SFI within the corpus is 3.5 to 88.3. Words in our sample ranged from 33 to 61.60.

**Imageability.** Imageability is a word-specific feature referring to the ease with which a word can elicit a mental image in the reader (Paivio, Yuille, & Madigan, 1968). Existing tables of imageability ratings were missing many of the words on the Adams and Huggins (1985) list and therefore we collected our own data. These data were collected from 68 undergraduate students in an introductory special education class. They were asked to rate the difficulty of bringing about a mental image for the 49 words on the irregular word reading list. The instructions included the following prompt from the original paper by Paivio, Yuille, and Madigan (1968):

> The words that arouse mental images most readily for you should be given a rating of 7; words that arouse images with the greatest difficulty or not at all should be rated 1; words that are intermediate in ease or difficulty of imagery, of course, should be rated appropriately between the two extremes.

Average ratings for these words ranged from 1.75–6.87. Cronbach’s alpha for imageability for this sample was .93.

**Orthographic neighborhood size.** To account for the orthographic similarity of the target words to other words, we used orthographic neighborhood size measured by Coltheart’s N (ON; Coltheart, Davelaar, Jonasson, & Besner, 1977). This metric is the number of words that can be produced by changing a letter in a word of the same length. These data were obtained from the English Lexicon Project (Balota et al., 2007).

**Spelling-to-pronunciation transparency rating.** To address how easy it was for students to arrive at the correct pronunciation of each irregular word by applying typical decoding rules, we asked expert raters to rate this difficulty on a 6-point scale. Our expert raters were professors and graduate students with a firm background in phonics and decoding. We collected these data from 22 expert raters. Experts were asked to rate words on a 6-point scale and they were given the following prompt:

> Below you will find a list of irregular words. We would like you to pretend that the letter string is unfamiliar to you and apply a decoding strategy to the letter string and rate the ease of matching your recoded form of the letter string to the actual word pronunciation. Rate the difficulty of making the match between recoded form and pronunciation on a scale from 1 to 6, with 1 being very easy and 6 being very difficult.

Cronbach’s alpha for the spelling-to- pronunciation transparency rating for this sample was .84.

**Strange versus exception words.** We assessed whether the words included on the irregular word list were strange or exception words according to the criteria outlined by Seidenberg, Waters, Barnes, & Tanenhaus, 1984. Words were coded as a ‘1’ if they were strange and as a ‘0’ if they were exception. For multisyllabic words, Seidenberg et al. criteria were applied to each syllable separately. The authors of this paper coded the words for strange versus exception. There were 20 strange words on the list and 29 exception words.

**Word length.** To account for differences in word length across items the number of letters in each irregular word was used as a word-level covariate.

**Procedure**

Test examiners were graduate research assistants who had been trained on tests until procedures were implemented with 90% fidelity. Most students were tested in three 1-hr sessions, although a minority were tested in two 1.5-hr sessions or in one 3-hr session. All tests were given individually, audio recorded for reliability/validity purposes, and scored by the original examiner. Children received small school-related prizes or a $5 gift card for participating in each testing session. All tests were double-scored and double-entered; discrepancies were resolved by a third examiner. Average fidelity of test administration procedures (based on a random selection of 20% of the taped assessment sessions) exceeded 94% for all tests. Study data were entered and managed using REDCap electronic data capture tools hosted at Vanderbilt University (Harris et al., 2009).

**Analyses**

A series of crossed-random effects models (DeBoeck, 2008; Van den Noortgate, De Boeck, & Meulders, 2003) were used to answer the research questions outlined above because these models allowed us to include both child- and word-level predictors in the same model as well as address interactions between the two. These item response theory based models are cross-classification multilevel models that allow variance to be partitioned across the person and item level and allow for responses to be predicted by both person and item level effects. We conducted these analyses using Laplace approximation available through the lmer function (Bates & Mächler, 2009) from the lme4 library in R (R Development Core Team, 2012). The analyses included 170 children and 49 words. For these models, words and persons are assumed to be random samples from a population of words and a population of persons. Because words are not nested within persons, these models are not strictly hierarchical models, but instead cross-classified. Words and persons are on the same level and crossed in the design and responses are nested within persons and within words (see Appendix A, Figure 1A). Power for these analyses has been addressed through simulation studies (see Cho, Partchev, & De Boeck, 2012). Various methods for examining model parameters indicate little difference in fixed effect estimates across methods with precision being relatively robust to sample size and number of items. For random effects, although some methods (e.g., the alternating imputation posterior method) may present larger bias when the models are used with smaller samples, these same models also tend to result in smaller mean standard errors (Cho et al., 2012). To address power within our own sample we conducted a simulation study to determine the minimal detectable effect size defining power at .80 (.05). Because crossed-random effects models do not yield traditional effect size estimates, our simulation estimated
with the fixed effects model using a likelihood ratio test and the null hypothesis that the more parsimonious model best fit the data. Finally, we estimated a model based on the results of the first two steps. This multistep procedure follows Bates’ (2011) recommendations and allows us to find a final model that provides the best fit for the data. The intercept and slope random effects were assumed to have zero covariance.

We built a series of models to answer our research questions using Model 1 as a base. Model 1 included only reader group and addressed the first research question regarding group differences in irregular word reading skill. Model 2 addressed the second research question, concerned with child- and word-level covariates. This model included child reading-related skills (003–007) and word characteristics (008–011). Model 3 was an interaction model including hypothesized interactions, designed to answer research question three. In this model, we added interactions for reader group with frequency, imageability, spelling-to-pronunciation transparency rating, strange versus irregular words (014–012), as well as interactions for frequency with spelling-to-pronunciation transparency rating, strange versus exception, and imageability (022–024). Details about the structure of the data for these analyses, the equations for the models, and the covariates included for each model are provided in Appendix B.

The minimal $R^2$ change detectable when a covariate was added to the model to predict either child or word variance and then this minimal variance change was converted into an $F$ statistic which is interpretable using guidelines provided by Cohen (1988). Using this method, our sample of words and children allows us to detect a minimal variance change on words equivalent to 8.26% and on children equivalent to 3.27%. These reductions in variance correspond to $F$ statistics of .09 for words and .03 for children, representing small effects. Therefore, our models, with a sample size of 170 children and 49 items (totaling 8,330 observations), are powered to detect small effects based on Cohen’s criteria for multiple $R^2$ (Cohen, 1988).

The crossed-random effects models (see Models 0–3 in Table 7) were built gradually in a stepwise fashion using model comparisons to determine the model that best fit the data. The unconditional model (Model 0) was fit first by adding a person-specific random effect ($r_{ij0}$) and an item-specific random effect ($r_{ij0}$) because we expected random variation related to each of these variables. The binary outcome ($p_{ij}$, the probability of a correct response from person $j$ on item $i$) was assumed to follow the Bernoulli distribution and random effects were assumed to be normally distributed. We used an unconditional model with only random effects for persons and items to determine the variability associated with persons and items. The next model (Model 1) contained fixed effects for TD ($u_{00}$) and early emerging RD ($u_{02}$) with the late-emerging RD group acting as the referent group. We chose late-emerging RD as the referent group because we were interested in comparing them to the TD and early emerging RD groups. This was done as the next step because we expected differences across groups. Model 1 was the base model for all subsequent analyses because we expected group differences to account for a large amount of variance attributable to the sampling procedure and how groups were established a priori. This model also allowed us to determine how well the subsequent models explained irregular word reading variance. For Model 1 and all subsequent models, we determined the structure of random effects (i.e., random slopes) using a three-step process. (A random slope here refers to allowing a child characteristic to vary randomly across words and allowing a word characteristic to vary randomly across children.) First, we added all fixed effects together. Next, we added random slopes for each fixed effect and completed model comparisons. We compared this model to the base model with the likelihood ratio test and selected the best-fitting model.

Finally, we estimated a model based on the results of the first two steps. This multistep procedure follows Bates’ (2011) recommendations and allows us to find a final model that provides the best fit for the data. The intercept and slope random effects were assumed to have zero covariance.

We built a series of models to answer our research questions using Model 1 as a base. Model 1 included only reader group and addressed the first research question regarding group differences in irregular word reading skill. Model 2 addressed the second research question, concerned with child- and word-level covariates. This model included child reading-related skills (003–007) and word characteristics (008–011). Model 3 was an interaction model including hypothesized interactions, designed to answer research question three. In this model, we added interactions for reader group with frequency, imageability, spelling-to-pronunciation transparency rating, strange versus irregular words (014–012), as well as interactions for frequency with spelling-to-pronunciation transparency rating, strange versus exception, and imageability (022–024). Details about the structure of the data for these analyses, the equations for the models, and the covariates included for each model are provided in Appendix B.

We examined the effect of each word level and child level covariate by calculating the probability of a correct response with the addition of the covariate to the intercept, following the formula $p_{ij} = \frac{\exp\left(\hat{b}_{0} + \hat{b}_{j} + \hat{b}_{i} + \beta_{1}\hat{w}_{1} + \beta_{2}\hat{w}_{2} + \ldots + \beta_{n}\hat{w}_{n}\right)}{1 + \exp\left(\hat{b}_{0} + \hat{b}_{j} + \hat{b}_{i} + \beta_{1}\hat{w}_{1} + \beta_{2}\hat{w}_{2} + \ldots + \beta_{n}\hat{w}_{n}\right)}$, with $v$ representing the covariate of interest. The late-emerging RD group was the referent group and predicted probabilities are given for an average item and an average child in the late-emerging RD group where all other covariates are at their mean values for our sample. We calculated the variability explained by calculating the reduction in child and item variance from the base model containing only the TD and early emerging RD covariates. Two formulas were used, one for the child level and one for the item level. These formulas were $\frac{\sigma_{j}^{2} - \sigma_{j}^{2}}{\sigma_{j}^{2}}$ and $\frac{\sigma_{\hat{w}_{i}^{2} - \sigma_{\hat{w}_{i}^{2}}}}{\sigma_{\hat{w}_{i}^{2}}}$, respectively, where $n$ represents the model to which the base model was compared (Bryk & Raudenbush, 1992). For the final models that included random slopes in addition to the random intercepts, we calculated the variance explained using only the fixed effects models, a method supported by current simulations by LaHuis, Hartman, Hakoyama, and Clark (2014).

**Results**

Demographic data for participants (N 170) are presented in Table 2. There were more females than males in the sample, and 10 children were retained (repeated a grade). The sample represents the demographics of the local district with respect to the percent of African American (48.82% sample/47% district) and Caucasian (37.06% sample/35% district) children. The sample had a lower percentage of Hispanic children compared to the district (2.94% sample/16% district) because of the initial sampling requirement across the 3 cohorts that children enrolled in first grade English Language Learner instruction be eliminated from the sample.

Table 3 provides the child-level performance across measures disaggregated by reader group with associated mean comparisons.

![Figure 1. Interaction between reader type (TD typically developing, LERD Late Emerging Reading Difficulty) and word imageability.](image)
formed both students in the late-emerging RD and the early emerging RD groups. Students in the late-emerging RD group outperformed students in the early emerging RD group on orthographic processing, phonological awareness, rapid automatized naming, and overall irregular word reading. No significant differences were found between early emerging RD and late-emerging RD groups on measures of vocabulary and book title recognition. The table also provides disaggregated performance on standardized measures for word identification, passage comprehension, and nonword decoding, with a consistent ordering effect of TD late-emerging RD early emerging RD. Table 4 provides the descriptive statistics for the word characteristics across all 49 words in the analyses. The average word frequency using the SFI metric was 50.51, the average spelling-to-pronunciation transparency rating across all 49 words was 3.59 on the 6-point scale, the average imageability rating was 5.82 (7-point scale), the average word length was 5.22 letters, and the average orthographic neighborhood size was 2.51.

Table 5 provides the zero-order correlations among the child level predictors of irregular word reading. There were significant correlations between all child level predictors of word reading with the exception of the Book Title Questionnaire variable. The Book Title Questionnaire was not significantly correlated with any of the other child level predictors, including average performance on the irregular word reading task. All of the remaining child-level predictors were highly correlated with irregular word reading with correlations ranging from .47 (rapid automatized naming) to .85 (nonword decoding). Table 6 provides zero order correlations for all word level predictors. In this table, we included the aggregated irregular word reading accuracy for words across children. Word level characteristics most highly correlated with irregular word reading difficulty were frequency \( r = .71 \) and the spelling-to-pronunciation transparency rating \( r = .50 \). There were also significant correlations among the word level predictors. The spelling-to-pronunciation transparency rating was significantly correlated with frequency \( r = .45 \), with more difficult words being less frequent. Similarly, words that were strange were less frequent \( r = .38 \) and had higher spelling-to-pronunciation transparency ratings \( r = .38 \). Length and imageability were significantly correlated \( r = .35 \), with longer words typically being more imageable. Orthographic neighborhood size was significantly correlated with strange \( r = .40 \) and length \( r = .42 \), meaning that strange words had smaller orthographic neighborhood sizes and were typically shorter.

Child- and Word-Level Characteristics Related to Irregular Word Recognition

The unconditional model (Model 0) containing only person and item random effects had an intercept logit estimate of \( \alpha_0 = 0.64 \), corresponding to a predicted probability of a correct irregular word reading response of .66 for the average child and the average item (see Table 7). There was variability around that average for both children \( \sigma^2_{u_0} = 4.58 \) and items \( \sigma^2_{v_0} = 5.87 \).
To address whether differences exist in irregular word reading between classes of children identified as late-emerging RD and those identified as early emerging RD and TD we added the two dummy variables representing the late-emerging RD-TD comparison (TD) and the late-emerging RD-early emerging RD comparison (ERD). Because we expected there to be differences between reader groups, this model was completed first and acted as the base model to which all subsequent models were compared. Results of the reader group (Model 1) are displayed in Table 7. The model with TD (108) and ERD (n = 19) fixed effects improved model fit over the unconditional model as expected, $\chi^2(171.41, p = .0001$. In an effort to construct the best fitting model, an iterative process was used in which random slopes for TD (108) and ERD (108) were then added to the model. This model fit the data significantly better than the previous model, $\chi^2(194.44, p = .0001$. Decisions about the inclusion of these random slopes were based on the mixed chi-square distribution LRT (Stram & Lee, 1994). The variance associated with TD and ERD indicated that only TD required a random slope. A comparison of the two models was conducted and the model with a random slope for ERD did not significantly improve model fit ($\chi^2(20, p = .65$) and therefore the most parsimonious model was used. This model was used as the base model for all subsequent model comparisons. The Model 1 person random effect variance was 1.46 and was the variance against which the subsequent models were compared. The intercept (108 0.697) for Model 1 indicated a mean probability of a correct response of .33 for the average LERD child, .04 for the average ERD child, and .86 for the average TD child. Fixed effects suggest that when only reader group is entered into the model the probability of correctly reading an irregular word was greater for TD compared with LERD as well as greater for LERD compared with ERD.

<table>
<thead>
<tr>
<th>Variable</th>
<th>$M$ (SD)</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (SFI)</td>
<td>50.51 (6.84)</td>
<td>33</td>
<td>61.60</td>
</tr>
<tr>
<td>Spelling-to-pronunciation transparency rating</td>
<td>3.59 (.78)</td>
<td>1.95</td>
<td>5.27</td>
</tr>
<tr>
<td>Imageability</td>
<td>5.82 (1.34)</td>
<td>1.67</td>
<td>6.88</td>
</tr>
<tr>
<td>Length</td>
<td>5.22 (1.04)</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>Orthographic neighborhood size</td>
<td>2.51 (3.10)</td>
<td>0</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 5
Zero Order Correlations Between Child Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>TD (108) M (SD)</th>
<th>ERD (n = 19) M (SD)</th>
<th>LERD (n = 43) M (SD)</th>
<th>All children (N = 170) M (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PA</td>
<td>13.84 (4.49)</td>
<td>6.58 (3.92)</td>
<td>9.19 (3.56)</td>
<td>11.85 (5.00)</td>
</tr>
<tr>
<td>OC</td>
<td>36.20 (2.34)</td>
<td>26.79 (4.42)</td>
<td>33.40 (3.14)</td>
<td>34.44 (4.11)</td>
</tr>
<tr>
<td>RAN</td>
<td>35.30 (7.97)</td>
<td>48.68 (15.82)</td>
<td>40.79 (9.31)</td>
<td>38.18 (10.39)</td>
</tr>
<tr>
<td>VOC</td>
<td>158.48 (17.36)</td>
<td>130.21 (21.69)</td>
<td>139.07 (19.46)</td>
<td>150.41 (21.34)</td>
</tr>
<tr>
<td>BTQ</td>
<td>12.06 (6.97)</td>
<td>16.47 (10.22)</td>
<td>11.84 (8.38)</td>
<td>12.49 (7.83)</td>
</tr>
<tr>
<td>IRWR</td>
<td>34.78 (5.95)</td>
<td>9.42 (9.74)</td>
<td>21.09 (5.19)</td>
<td>28.48 (10.92)</td>
</tr>
<tr>
<td>WJ-WID</td>
<td>101.11 (7.56)</td>
<td>75.89 (12.50)</td>
<td>87.02 (3.43)</td>
<td>94.73 (11.69)</td>
</tr>
<tr>
<td>WJ-WA</td>
<td>106.10 (8.23)</td>
<td>78.95 (15.94)</td>
<td>90.47 (5.68)</td>
<td>99.11 (13.19)</td>
</tr>
<tr>
<td>WJ-PC</td>
<td>100.18 (7.26)</td>
<td>70.63 (13.54)</td>
<td>87.16 (8.23)</td>
<td>93.58 (12.92)</td>
</tr>
</tbody>
</table>

Note. OC Orthographic choice; PA Phonological Awareness; RAN Rapid Automatized Naming; VOC Vocabulary (PPVT); BTQ Book title questionnaire; IRWR Irregular Word reading; WJ Woodcock Johnson; WID Word Identification; PC Passage Comprehension; WA Word Attack. *Mean comparisons were conducted using ANOVA with Bonferroni post hoc pairwise comparisons. a Standard scores ($M$ 100, $SD$ 15).

Table 6
Zero Order Correlations Between Word Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Irregular word reading</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. PA</td>
<td>.60</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. RAN</td>
<td>.47</td>
<td>.39</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4. OC</td>
<td>.79</td>
<td>.40</td>
<td>.43</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>5. VOC</td>
<td>.65</td>
<td>.46</td>
<td>.27</td>
<td>.37</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>6. BTQ</td>
<td>.11</td>
<td>.13</td>
<td>.04</td>
<td>.06</td>
<td>.07</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>7. NWD</td>
<td>.85</td>
<td>.59</td>
<td>.51</td>
<td>.71</td>
<td>.49</td>
<td>.09</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. p < .001 for all variables in bold. PA Phonological Awareness; RAN Rapid Automatized Naming; OC Orthographic choice; VOC Vocabulary (PPVT); BTQ Book Title Questionnaire; NWD Nonword Decoding.
irregular word reading, controlling for reader group. To address this question, we used a child and word covariate model (Model 2) with results provided in Table 7. When we constructed Model 2, main effects for all child and word level covariates were first added to Model 1 simultaneously. This model fit the data significantly better than the reader group model, \( \chi^2 = 220.72, p = .0001 \). Using an iterative model building process, a random slope for each of the covariates was then added to the model. The best fitting model contained a random item slope for TD and a random item slope for nonword decoding. Adding a random slope across words for nonword decoding significantly improved model fit over the model with only main effects, \( \chi^2 = 14.31, p = .001 \). The intercept (0.081) indicated a mean probability of a correct response of .55 for the average LERD child, all other things being equal. Model 2 explained 74% of the person variance present in Model 1 and 60% of the item variance.

Note that the main effects reported here are conditional on the other covariates in the model.

For child characteristics, we found significant effects for three covariates (orthographic choice, vocabulary, and nonword decoding). Below, we list probabilities for each covariate for an average late-emerging RD child. For orthographic choice (0.186), an orthographic choice score 1 SD above the sample mean corresponds to a .73 probability of a correct response, whereas an orthographic choice score 1 SD below the mean would correspond to a .37 probability of a correct response. For vocabulary (0.027), a PPVT score 1 SD above the mean corresponded to a .67 probability of a correct response, compared with .41 for 1 SD below the mean. Finally, for nonword decoding (0.073), a score 1 SD above the mean increased the probability of a correct response to .71 whereas a score 1 SD below the mean decreased the probability to .38.

For word characteristics, we found a significant effect of frequency, 0.268, indicating that the probability of correct response for an average late-emerging RD child on a word with frequency 1 SD above the sample mean was .88, whereas a word with frequency 1 SD below the mean was .17. For the spelling-to-pronunciation transparency rating, the effect was marginally significant in the main effect model (z = 1.92) and significant in the interaction model. Therefore, we interpret that effect here. The probability of a correct response for an average late-emerging RD child on a word with a rating 1 SD above the sample mean (meaning rated as more difficult) was .42 whereas a word with a rating 1 SD below the sample mean was .67.
The third research question concerned interactions between child and word-level predictors in explaining irregular word reading item variance. This exploratory model (Model 3) estimated 11 interaction terms to answer these questions. This model explained 74% of the person level variance and 63% of the item level variance. There were two significant child by word interactions, TD status by imageability and ERD status by strange versus exception word class. The first significant interaction, TD status by imageability and ERD status by strange versus exception word class, indicates that students in the TD group significantly outperformed students in the late-emerging RD group on words that were highly imageable but not on words that were low on the
Imageability Scale (see Figure 1). The second significant interaction, ERD status by strange word class (.021 .645), indicated that students in the late-emerging RD group had a significantly higher probability of reading a word correctly if it was classified as strange than did a child in the early emerging RD group. This pattern was not true for words that were classified as exception words (see Figure 2).

Discussion

In this study, we argue that irregular words present unique challenges for developing readers (Waters et al., 1984; Taylor et al., 2011; Wang et al., 2013). Using a self-teaching mechanism that builds the orthographic lexicon through item-specific learning, simply applying GPC rules to irregular words will not result in the correct pronunciation (a phenomenon referred to as partial decoding), and this likely disrupts or slows the addition of items to the orthographic lexicon. This has led many to suggest that the addition of irregular words to the orthographic lexicon in English is modulated by child- and word-level factors beyond phonologically based recoding skill and print experience (Nation & Snowling, 1998; Plaut et al., 1996; Wang et al., 2013; Waters et al., 1985; Waters et al., 1984). We maintain that comprehensive models of irregular word reading are needed to assess which child-level variables, word-level variables, and child by word interactions are associated with item-level variance. Furthermore, we assert that models such as the one developed in this study are necessary to begin the search for potentially malleable factors that can improve the ability of children, with particular attention to those with RD, to recognize irregular words. Overall, results suggest that there are multiple sources that explain irregular word reading variance including child-level characteristics, word-level characteristics, and interactions between child- and word-level characteristics. The following sections discuss results within the context of our three research questions with a focus on a model of irregular word reading in which lexical characteristics specific to both child and word influence accuracy.

Child- and Word-Level Characteristics Related to Irregular Word Reading

We found several significant predictors at the child level. The first predictor was reader group, with a significant difference in irregular word reading identified between TD students and those with late-emerging RD and between late-emerging RD and early emerging RD students (Model 1). These group differences remained after child- and word-level predictors were entered into the model (Model 2), but were no longer significant once the interaction terms were added to the child- and word-level predictors. Overall, results indicate that the general word reading difficulties observed in the late-emerging RD and early emerging RD groups also impact their irregular word reading skill, but once controlling for important child-level factors (e.g., decoding, orthographic knowledge, vocabulary) along with word by child interactions, differences between all three groups on irregular word reading skill were minimized. Thus, differences between early emerging RD and late-emerging RD groups on irregular word reading appear explainable by the lower performance of the early emerging RD group on important cognitive skills associated with reading performance.

Cognitive variables that were significantly related to irregular word reading within the entire sample of children included nonword decoding, vocabulary, and orthographic knowledge. Given the overwhelming evidence for the correlation between decoding skills and word reading, a relationship with irregular word reading was not unexpected. The significant role of decoding on irregular word reading is consistent with speculation by Griffiths and Snowling (2002) that the ability to read irregular words is at least partially dependent on having access to “segmental phonological representations” (p. 41). A second significant predictor was orthographic knowledge measured using the orthographic choice task. Students with greater orthographic knowledge had a higher probability of reading irregular words correctly. This finding is consistent with others who have found that orthographic knowledge is related to general word reading skill (e.g., Cunningham, Perry, & Stanovich, 2001). There was a similar pattern for vocabulary; students with a higher receptive vocabulary score had a higher probability of successfully recognizing irregular words, even when controlling for reader group, decoding, OC, and RAN. This pattern was consistent across Models 2 and 3, and supports our hypothesis, and a developmental word-reading model, in which orthographic-to-phonological pathways necessary to establish irregular word representations may be at least partially dependent on lexical input (Nation & Snowling, 1998; Plaut et al., 1996; Ricketts et al., 2007; Tunnmer & Chapman, 2012). It is noteworthy that we did not find reading experience, as measured by the BTQ, to be a significant predictor of irregular word reading. The nonsignificant correlation between irregular word reading and the BTQ suggests that it was not a sensitive measure. We interpret these results as a measurement issue with the BTQ and do not suggest that print exposure is unimportant for the acquisition of irregular words (for a discussion of the measurement issues related to the BTQ and print exposure see Foorman, 1994), and instead we speculate that the titles contained on the list did not capture important individual differences in print exposure.

At the word level, we found the expected main effect for frequency. We also found that our expert rating of spelling-to-pronunciation difficulty was marginally significant in the main effect model and was significant in the interaction model. This finding is consistent with our hypothesis that the ease of phonological recoding from an orthographic stimulus is a good predictor of success in reading an irregular word correctly. This pattern seems to suggest that as the phonological output from decoding more closely represents the actual lexical phonological representation, the higher...
the probability of a correct response (Elbro et al., 2012; Tunmer, & Chapman, 2012; Venezky, 1999). Contrary to suggestions by Seidenberg et al. (1984), we did not find a significant main effect for the irregular word type (i.e., exception vs. strange) after controlling for other word-level features, suggesting that these two classes of words may be processed similarly for the entire sample, however differences between late-emerging RD and early emerging RD groups were present. It may also be that the rating of spelling-to- pronunciation is a more precise way to characterize the degree of irregularity than dichotomous coding of exception or strange.

**Irregular Word Reading in Students With Late-Emerging RD**

The value of distinguishing between students who are TD and those with early emerging RD and late-emerging RD was also of interest in this study. The use of item-based random effects models allowed us to probe for interactions between reader group and word-level characteristics. We were particularly interested in whether children with late-emerging RD differed in how they process irregular words compared with TD children and children with early emerging RD. One of the prevailing hypotheses in the field is that the reading difficulties associated with late-emerging RD may be attributable to the increasing demands placed on the reader at the word level, as students are faced with new words many of which are lower in frequency and contain irregular spelling patterns, and thus require the formation of complex connections between phonology, orthography, and semantics (Catts et al.; Leach et al., 2003).

Two interactions emerged between word-level characteristics and reader group that tend to support the view that lexical feedback may be somewhat limited during irregular word learning in children with late-emerging RD. The first involved an interaction between the TD and late-emerging RD groups and imageability in which increases in word imageability were associated with greater irregular word reading skill in TD, but not late-emerging RD children. This relationship is in the opposite direction typically reported in the literature in which less skilled readers benefit more from words which are higher in imageability (Coltheart, Laxon, & Keating, 1988; Steacy et al., 2013). However, it should be noted that the interaction between early emerging RD and late-emerging RD on imageability was approaching significance. Results suggest that after controlling for everything else in the model, children with late-emerging RD may have specific difficulties exploiting words with higher imageability. Without knowing whether late-emerging RDs are familiar with each of the words, we can only speculate that the students in the late-emerging RD may not activate and use lexical properties such as imageability in the same way as the TD students and that this makes it unlikely that they will retrieve images associated with the word to aid in the formation of an orthographic representation. Furthermore, we interpret our results to suggest that having a word in the semantic lexicon that is highly imageable may increase the probability that a student will accurately identify an irregular word for students who are typically developing and students in the early emerging RD group. Future studies would benefit from assessing item-level semantics, lexical phonology, and imageability to allow the teasing apart of these influences on irregular word reading in children with late-emerging RD. It would also be quite helpful to use the set for variability paradigm as an item-level variable to assess whether individual differences in children’s ability to bridge the gap between the regularized decoding pronunciation and the actual phonological representation is an important predictor of item level variance.

The second interaction involved early emerging RD and lateemerging RD groups and word-level orthographic complexity (i.e., exception vs. strange words). There was little difference between the groups on exception word reading but significant differences on strange word reading favoring the late-emerging RD group. Results do not seem to point to specific problems in the formation of complex connections between phonology and orthography in the late-emerging RD group as the interaction between late-emerging RD and TD groups and orthographic complexity was not significant. Instead, results seem to favor a model in which the early emerging RD group has greater problems forming the complex connections between phonology and orthography that are contained in strange words. Seidenberg et al. (1984) have offered that strange words are processed differently than exception words. It is the case that the early emerging RD group has greater deficits in phonological processing compared to the late-emerging RD group and this may make it more difficult for early emerging RD children to form precise representations between the orthography and phonology represented in strange words. Overall, the interactions seem to suggest that lexical differences (e.g., semantic and lexical phonology) better explain differences between lateemerging RD and TD children as compared to more traditional phonological-orthographic processing skills. However, we regard these interaction effects to be exploratory in nature because of the relatively small sample size, particularly the early emerging RD group. Differences in results between our study and those reported by Griffiths and Snowling (2002) are likely attributable to the inclusion of both RD and TD children, the use of both child- and word-level characteristics as predictors, and the added power garnered by itemlevel analyses.

**Limitations**

There are several limitations that should be considered in relation to this study. The first is the sample sizes of both the students and the words. Although we note that we should have adequate power to detect an effect across the 49 words and 170 children, readers should exercise caution when generalizing these results to other students and words. Furthermore, an item-specific measure of familiarity would be useful for these analyses to control for whether students have phonological representations of these words in their lexicons. Finally, a parallel measure of regular words sampled in the same way as the Adams and Huggins (1985) irregular word list used in this study would also be informative and provide a contrast between regular and irregular words.

**Conclusion**

In conclusion, this study has identified multiple sources affecting individual differences in irregular word reading among 5th grade children oversampled for RD. Results indicate that variance in irregular word reading was predicted at the child level by decoding skill, orthographic coding, and vocabulary; at the word level by word frequency and a spelling-to-pronunciation transparency rating; and by the Reader group Imageability and Reader group Orthographic complexity interactions. For irregular words, partial decoding likely requires having item-specific knowledge (either semantic and/or lexical phonology) to help to fill voids resulting from the mismatch
between orthographic-phonology mapping and lexical phonology in emergent readers (Keenan & Betjemann, 2008). Our results support a developmental model of word reading in which orthographic-to-phonological pathways become at least partially dependent on lexical input (input of semantic knowledge), with this influence being increasingly important for irregular words (Nation & Snowling, 1998; Ricketts et al., 2007; Tunmer & Chapman, 2012). Overall, results from our study support the role of lexical influence on irregular word reading with vocabulary skill and spelling-to-pronunciation ease having a main effect and word imageability acting as a moderator. Thus we concluded that lexical representations appear to be important in irregular word reading and advocate for further work examining the role of lexical processing on irregular word reading in children with RD. Although the design of this study does not allow for causal inferences, allowing word- and child-attributes to compete for variance in the same model provides an opportunity to consider new, and possibly untested, approaches to effectively teach irregular word reading skills to struggling readers (Compton, Miller, Elleman, & Steacy, 2014). Specifically, we encourage future studies examining whether item-level training of semantics and lexical phonology can increase the speed at which with children with RD add irregular words to their orthographic lexicons.

References


Bates, D., Maechler, M., & Bolker, B. (2013). lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.


At the end of fourth grade single indicator hidden Markov models were fit separately for the four time points representing word reading and reading comprehension development. These models were considered a first-order Markov process where the transition matrices are specified to be equal over time (i.e., measurement invariance across time; Langeheine & van de Pol, 2002). Hidden Markov models are a form of latent class analysis, known as latent transition analysis (LTA), where class indicators (categorical variables indicating RD and TD groups) are measured over time and individuals are allowed to transition between latent classes. LTA addresses questions concerning prevalence of discrete states and incidence of transitions between states and produces parameter estimates corresponding to the proportion of individuals in each latent class initially, as well as the probability of individuals changing classes with time. LTA models were generated using mixture modeling routines contained in Mplus 5.0 (Muthén & Muthén, 1998–2012). Model estimation was carried out using a maximum likelihood estimator with robust standard errors. Detailed discussions of LTA can be found in Collins and Wugalter (1992) and Reboussin, Reboussin, Liang, and Anthony (1998). A cut point of the 25th-percentile (based on a national norming sample) was used at each time point to represent reading difficulty estimated with the likelihood ratio chi-square. The likelihood ratio compares the observed response proportions with the response proportions predicted by the model (Kaplan, 2008). As with most SEM-based models the null hypothesis for chi-square model tests is that the specified model holds for the given population, and therefore accepting the null hypothesis implies that the model is plausible. All models across cohorts were found to fit the data adequately.

Cohort 1 (Compton et al., 2010)

Participants in the first cohort were selected from 56 firstgrade classrooms in 14 schools within an urban district located in middle Tennessee. Seven study schools were Title I institutions. We assessed every formally consented child (n = 712) with three 1-min study identification measures: WIF screen, rapid letter naming (RLN), and rapid letter sound (RLS). With WIF screen, children are presented with a single page of 50 high-frequency words randomly sampled from 100 high-frequency words from the Dolch preprimer, primer, and first-grade level lists (Fuchs, Fuchs, & Compton, 2004). They have 1 min to read words. With RLN and RLS, the speed at which children name an array of the 26 letters and the sounds of the

Appendix A

Sampling Plan for Each of the Cohorts

Any cut point of a dimensional variable to designate reading difficulty is arbitrary (Fletcher et al., 1999), however the 25th percentile has consistently been employed in the literature.
letters is measured. For all three measures, scores were prorated if a child named all items in less than 1 min. We used these data to divide the 712 children into high-, average-, and low-performing groups with the use of latent class analysis and then randomly selected study children from each group. We oversampled low-performing children to increase the number of struggling readers in the prediction models. Four hundred and eighty-five children were included: 310 low study entry (LSE), 83 average study entry (ASE), and 92 high study entry (HSE). Follow-up testing was performed at the end of first through fourth grade. At follow-up in the spring of fourth grade, 200 of the original 485 children (41% of the sample) had moved from the district and were unavailable for assessment.

**Cohorts 2 and 3 (Gilbert et al., 2013)**

The sampling procedures for Cohorts 2 and 3 were identical and are therefore combined here. Initially we asked first-grade teachers to identify the lowest half of their class in terms of reading skill. Children in Cohort 2 were drawn from 9 schools (5 Title I) in 37 first-grade classrooms and children in Cohort 3 from 9 schools (5 Title I) and 32 first-grade classrooms within an urban district located in middle Tennessee. We screened 628 of the identified students with three 1-minute measures: two WIF lists and a RLN. Scores were prorated if a student named all items in less than one minute. To identify an initial pool of students potentially at elevated risk for poor reading outcomes we applied latent class analysis (Nylund, Asparouhov, & Muthen, 2007) to the three screening measures. The purpose of such an analysis was to obtain model-based latent (unobserved) categories of students who are performing similarly on the three screening measures. Models were developed and evaluated using Mplus version 6 (Muthén & Muthén, 1998–2010). A clear category of at-risk students was identified for Cohort 2 having 223 and Cohort 3 having 214 at-risk students. Students not populating the at-risk category were excluded from further follow-up. A portion of the at-risk first grade children were randomly assigned to 14 weeks of small group tutoring or a business as usual control group. Follow-up testing was performed at the end of first through fourth grade. At follow-up in the spring of fourth grade, 66 of the original 223 children (30% of the sample) in Cohort 2 and 50 of the original 223 children (23% of the sample) in Cohort 3 had moved from the district and were unavailable for assessment. A chi-square test was performed to examine the relation between first-grade tutoring (treatment and control) and fourth grade LERD status (ERD, LERD, and TD). Results indicate no relationship between first grade treatment and reading class assignment in fourth grade, $\chi^2(2, N = 172) = .2828, p = .868$.

**Appendix B**

**Data Structure and Equations**

![Figure B1](image.png)

*Figure B1.* Data structure for crossed-random effects models (Gilbert et al., 2011).

**(Appendices continue)**

Table B1

<table>
<thead>
<tr>
<th>Crossed-Random Effects Model Equations</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Base models</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unconditional (Model 0)</th>
<th>Level 1 (Responses): $\logit(\pi_k) = \beta_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base model</td>
<td>Level 1 (Responses): $\logit(\pi_k) = \beta_k$</td>
</tr>
</tbody>
</table>
Model 1

Child- and word-level predictors

Level 1 (Responses, \( y_{ij} \)): \( \text{Logit}(y_{ij}) = \beta_0 + \beta_1 \cdot \text{ERRD} + \beta_2 \cdot \text{TD} + \epsilon_{ij} \)

Level 2 (Person \( j \) & Word \( i \)): \( 0_{jk} + \text{ERD} \cdot N(j,0,2r_{01}) + \text{TD} \cdot N(0,2r_{02}) + \epsilon_{01} + \epsilon_{02} \cdot i + \eta_{01} + \eta_{02} \cdot j + \eta_{00} \)

Main effects model

Level 1 (Responses, \( y_{ij} \)): \( \text{Logit}(y_{ij}) = \beta_0 + \beta_1 \cdot \text{ERRD} + \beta_2 \cdot \text{TD} + \epsilon_{ij} \)

Level 2 (Person \( j \) & Word \( i \)): \( 0_{jk} + \text{ERRD} \cdot N(j,0,2r_{01}) \cdot \text{TD} \cdot N(0,2r_{02}) + \beta_3 \cdot \text{MA} + \beta_4 \cdot \text{OC} + \beta_5 \cdot \text{RAN} + \beta_6 \cdot \text{VOC} + \beta_7 \cdot \text{BTQ} + \beta_8 \cdot \text{NWD} + \beta_9 \cdot \text{FREQ} + \beta_{10} \cdot \text{SPR} + \beta_{11} \cdot \text{STR} + \beta_{12} \cdot \text{IMAG} + \beta_{13} \cdot \text{LENGTH} + \beta_{14} \cdot \text{ON} + \epsilon_{01} + \epsilon_{02} \cdot i + \eta_{01} + \eta_{02} \cdot j + \eta_{00} \)

Interaction model

Level 1 (Responses, \( y_{ij} \)): \( \text{Logit}(y_{ij}) = \beta_0 + \beta_1 \cdot \text{ERRD} \cdot \text{FREQ} + \beta_2 \cdot \text{TD} \cdot \text{FREQ} + \beta_3 \cdot \text{ERRD} \cdot \text{IMAG} + \beta_4 \cdot \text{TD} \cdot \text{IMAG} + \beta_5 \cdot \text{ERRD} \cdot \text{SPR} + \beta_6 \cdot \text{TD} \cdot \text{SPR} + \beta_7 \cdot \text{ERRD} \cdot \text{STR} + \beta_8 \cdot \text{TD} \cdot \text{STR} + \beta_9 \cdot \text{ERDFREQ} + \beta_{10} \cdot \text{TDFREQ} + \beta_{11} \cdot \text{ERDIMAG} + \beta_{12} \cdot \text{TDIMAG} + \beta_{13} \cdot \text{ERDSPR} + \beta_{14} \cdot \text{TDSPR} + \beta_{15} \cdot \text{ERDSTR} + \beta_{16} \cdot \text{TSTR} + \epsilon_{ij} \)

Level 2 (Person \( j \) & Word \( i \)): \( 0_{jk} + \text{ERRD} \cdot \text{FREQ} \cdot N(j,0,2r_{01}) + \text{TD} \cdot \text{FREQ} \cdot N(0,2r_{02}) + \beta_3 \cdot \text{MA} + \beta_4 \cdot \text{OC} + \beta_5 \cdot \text{RAN} + \beta_6 \cdot \text{VOC} + \beta_7 \cdot \text{BTQ} + \beta_8 \cdot \text{NWD} + \beta_9 \cdot \text{FREQ} + \beta_{10} \cdot \text{SPR} + \beta_{11} \cdot \text{STR} + \beta_{12} \cdot \text{IMAG} + \beta_{13} \cdot \text{LENGTH} + \beta_{14} \cdot \text{ON} + \epsilon_{01} + \epsilon_{02} \cdot i + \eta_{01} + \eta_{02} \cdot j + \eta_{00} \)

Note. \( y_{ij} \) probability of a correct response from person \( j \) on work \( i \); \( k \) item; \( \alpha_0 \) intercept; \( M_a \) item covariate; \( N_b \) person covariate. Main effects models are shown with random intercepts only for simplicity but random slopes were included in some models, as described in the text.

Table B2

<table>
<thead>
<tr>
<th>Research question</th>
<th>Model no.</th>
<th>Covariates</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is the role of child-level skills and word-level characteristics as predictors of item-level variance on irregular word reading accuracy, with a particular interest in the role of child- and word-level lexical influence?</td>
<td>2</td>
<td>Level 2 (person): ERD, TD, PA, OC, RAN, VOC, BTQ, NWD Level 2 (word): FREQ, SPR, STR, IMAG, LENGTH, ON</td>
</tr>
<tr>
<td>Do differences exist in irregular word reading between classes of children identified as late-emerging RD and those identified as early-emerging RD and TD?</td>
<td>1</td>
<td>Level 2 (person): ERD, TD</td>
</tr>
<tr>
<td>What is the importance of child-level by word-level interactions in explaining irregular word reading variance, with a specific focus on interactions between RD status (i.e., late-emerging RD, early-emerging RD, and TD) and word-level characteristics (i.e., frequency, imageability, spelling pronunciation transparency rating, and whether an irregular word is strange)?</td>
<td>3</td>
<td>Level 2 (Person Word interaction): ERDFREQ, TDFREQ, ERDIMAG, TDIMAG, ERDSPR, TDSPR, ERDSTR, LERDSTR Level 2 (person): ERD, TD, PA, OC, RAN, VOC, BTQ, NWD Level 2 (word): FREQ, SPR, STR, IMAG, LENGTH, ON</td>
</tr>
</tbody>
</table>

Note. ERD early emerging reading difficulty; LERD late emerging reading difficulty; RD reading difficulty; PA phonological awareness; OC orthographic choice; RAN rapid automatized naming; VOC vocabulary; BTQ Book Title Questionnaire; NWD nonword decoding; FREQ frequency; SPR spelling to pronunciation rating; STR strange vs. exception; IMAG imageability; ON orthographic neighborhood size; ERDFREQ Early emerging reading difficulty Frequency; TDFREQ Late emerging reading difficulty Frequency; ERDIMAG Early emerging reading difficulty Imageability; TDIMAG Late emerging reading difficulty Imageability; ERDSPR Early emerging reading difficulty Spelling to pronunciation rating; LERDSPR Late emerging reading difficulty Spelling to pronunciation rating; LERDSTR Early emerging reading difficulty Strange vs. exception; ERDSTR Late emerging reading difficulty Strange vs. exception.

Received December 5, 2014
Revision received November 4, 2015
Accepted December 22, 2015