The Role of Imageability in Word Learning Efficiency and Transfer among First and Second Graders At-Risk for Reading Disabilities

By

Laura M. Steacy

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

May, 2015

Nashville, Tennessee

Approved by:

Donald L. Compton, Ph.D.

Douglas Fuchs, Ph.D.

Lynn S. Fuchs, Ph.D.

Christopher J. Lemons, Ph.D.

Sonya K. Sterba, Ph.D.

To my parents

ACKNOWLEDGEMENTS

This research would not have been possible without the help and support of numerous people. Thank you to my advisor, Don Compton, for his unwavering support throughout the process. From the beginning, he never failed to inspire critical thinking and a passion for research. I am grateful for the opportunity to learn from him. I am also grateful to Doug and Lynn Fuchs, who helped to form the foundation for this project and from whom I learned so much about intervention research. Thank you to Chris Lemons and Sonya Sterba for sharing their expertise and for helping throughout this research. Thank you also to two mentors from my Master's program, John Kirby and Lesly Wade-Woolley, for supporting me through the early stages of my academic career and for providing me with a strong foundation in research. I could not have reached this goal without them. Thank you to Leilani Dela Cruz, an incredible teacher and a lifelong friend, who shared her artistic talents in creating the materials for this project.

This research would not have been possible without the amazing team of people who assisted in the collection of data across these two projects: Ashley Love, Esther Lindstrom, Alyson Collins, Bobette Bouton, Erika Hsu, Michael Jackson, Johny Daniel, Margaret Schiller, and Morgan Polans. I'm grateful for all of their hard work. Thank you especially to Eunsoo Cho, Jenny Gilbert, and Amy Elleman. Eunsoo was the best mentor and office mate I could have asked for. I'm grateful for her invaluable advice throughout the program and beyond. Thank you to Jenny for her statistical expertise and for her selfless commitment to helping others. Thank you to Amy for her guidance and support over the past five years.

Thank you to my family for supporting me through all of this. Thank you for always helping me to put things into perspective and for supporting my choices without question. Thank you especially to my sister, Lisa, who is a constant and unwavering support. Finally, thank you to my former students, for all of the lessons I learned from them and for inspiring this research.

TABLE OF CONTENTS

Page

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF TABLES vi

LIST OF FIGURES vii

Chapter

I. Introduction 1

Development of the orthographic lexicon 2
The role of lexical input 4
Child and word characteristics related to orthographic lexicon development 5
Child characteristics 5
Word regularity 6
Word imageability 6
Attempts to intervene for imageability 8
Research questions 8

II. Experiment 1 10

Method 10
Participants 10
Measures. 10
Experimental design 14
Feedback conditions 14
Mastery 16
Procedure 16
Data analysis 16
Results 18
Crossed-random effects models 20
Discussion: Experiment 1 29

III. Experiment 2 32

Method 32
Participants 32
Measures 32
Experimental design 34
Conditions 35
Procedure 36

Data analysis..37
Results 37
Crossed-random effects models..39
Discussion: Experiment 2 45

IV. General Discussion 47

The role of imageability in the development of the orthographic lexicon.....................................47
The role of imageability in transfer of learning 48
The role of imageability in learning efficiency...49
Lexical involvement in word learning 50
Intervening for imageability 51
Limitations and future directions...52

Appendix

A. Hypotheses...54

B. Crossed-Random Effects Models...57

C. Imageability training pictures..59

D. Vocabulary training definitions 60

REFERENCES 61

LIST OF TABLES

Table Page

1. Demographic Statistics for Experiment 19
2. Child Level Descriptive Statistics for Experiment 1 20
3. Zero Order Correlations Between Child Variables: Experiment 1 20
4. Fixed Effects and Variance Estimates for Target Word Recognition for Cohort 1 23
5. Experiment 1: Fixed Effects and Variance Estimates for Number of Exposures Required for Mastery 28
6. Demographic Statistics for Experiment 2 38
7. Child Level Descriptive Statistics for Experiment 2 39
8. Zero Order Correlations between child variables: Experiment 2 39
9. Fixed Effects and Variance Estimates for Target Word Recognition for Experiment 2 ...41
10. Experiment 2: Fixed Effects and Variance Estimates for Number of Exposures Required for Mastery 45

LIST OF FIGURES

Figure Page

1. Connectionist Model of Reading 5

2. Word Factorial Design for Experiment 1....................13

3. Experimental design for Experiment 1 14

4. Crossed-random effects models for Experiment 1 and Experiment 2 17

5. Word Lists for Experiment 2 34

6. Experimental Design for Experiment 2 35

7. Interaction between Initial Word Reading (TOWRE Sight Word Efficiency) and Condition in Experiment 2....................44

CHAPTER I

INTRODUCTION

Comprehension of text is the ultimate goal of all reading instruction. It is impossible for children to reach this goal, however, without accurate and automatic word reading skills (Adams, 1994; Foorman & Torgesen, 2001; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001). Given that a large number of children struggle to read at a basic level (NAEP, 2011), the need to narrow the gap between poor readers and their typically developing peers persists (Griffin, Burns, & Snow, 1998). Identifying instructional techniques that will allow developing readers to quickly begin to read within authentic and instructional texts is critical. Although it is generally accepted that instructional time should be devoted to teaching children high frequency words as a means of reaching this goal (Ehri, 1995; Solity & Vousden, 2009; Vousden, 2008), little research has focused on which words to teach, the most effective and efficient instructional techniques, and how this learning transfers to reading novel words. Thus a focus on efficient and effective instructional methods that promote word learning and transfer in developing readers is warranted, particularly for children who struggle to learn to read. Although learning to read words has traditionally been thought of as a task requiring children to link phonology and orthography, there is evidence to suggest that meaning plays an important role in learning to read words as well (Keenan, & Betjemann, 2008; Nation, & Cocksey, 2009; Ricketts, Nation, & Bishop, 2007). The purpose of the two experiments in this study was to explore how child- and word-level features impact word learning (i.e. word reading) and transfer within a controlled experimental design as well as implications for instruction. Specifically, the two studies focused on the role of imageability in word learning efficiency and transfer among first and second grade children.

Development of the Orthographic Lexicon

According to Perfetti (1992), as children learn to read, they create an autonomous lexicon that develops over time in two distinct ways: (a) the number of lexical entries increases, and (b) the quality of the representations improves. Words in the autonomous lexicon are recognized automatically with minimal influence of background knowledge or context-based expectancies (see Perfetti, 1992; Stanovich, 1991), allowing fluent and reliable retrieval of word representations from the orthographic lexicon, activating phonological, syntactic, morphological, and semantic information. The number of lexical entries can increase in a variety of ways as children learn more words through oral and print exposure and their vocabularies grow. Firstly, as children learn decoding rules, the number of lexical entries increases. Secondly, the number of entries increases due to exposure to specific words that children acquire at the whole word level (Perfetti, 1992). According to Perfetti, the quality of the representations improves as the bond between phonemic and orthographic representations increase, thereby making the representation fully specified. Thus, two lexical acquisition systems are at work as a child learns to read—the addition of word-specific entries and the expansion of orthographic-phonological connections. These acquisition systems are mutually facilitative: “The more powerful the context-sensitive decoding rules (or analogic capabilities), the more entries the learner can acquire. And the more entries, the more powerful the decoding rules” (Perfetti, 1992, pp. 161–162).

Lexical development has been examined within samples of students both with and without reading difficulties. The results of these studies suggest that there is variability in how students acquire word representations. It comes as no surprise that the course of lexical development is different for students with reading difficulties compared to their typically developing peers. According to Ehri (1995), students typically develop sight word representations (i.e., fully specified lexical entries in the autonomous lexicon) by continually strengthening connections between phonemes and graphemes in words. Skilled readers successfully map all graphemes in a word onto all phonemes in pronunciations, thus binding the two systems to form a consolidated representation that allows automatic activation of the phonological and semantic representations associated with the specific letter string. Poor readers, on the other hand, tend to form partial connections between graphemes and phonemes in words that are less consolidated and do not lead to automatic access to the phonological and semantic representations. Thus the lexical representations of poor readers are considered less specified and of lower quality. This leads poor readers to rely on other sources of information (e.g., local text content), that are considerably less efficient, to facilitate word recognition (see Stanovich, 1980).

Given that poor readers have difficulty developing highly specified word representations that can be added to the autonomous orthographic lexicon, a focus on efficiency is particularly important for the lowest readers. The number of exposures required to create a representation varies among individuals and is an important focus for early development (Ehri, 1994). Ehri offers that word representations are developed through a “connection-forming process” whereby developing readers gradually form connections at the specific word level based on their knowledge of grapheme-phoneme correspondences. By focusing on the sub-lexical features of the words on multiple occasions, they gradually build up their lexicons. Typically developing students can learn to read words with relatively few exposures. For example, Dutch students trained on unfamiliar words learned word-specific features surprisingly fast, with students being able to differentiate between standard spelling of a word and its homophonic alternate spelling in as little as four to six exposures (Reitsma, 1983). Students with reading disabilities, however, typically require many more exposures than their typically developing peers and retain less complete representations of words in their lexicons (Ehri, 1997; Ehri & Saltmarsh, 1995).

The role of Lexical Input

Increasing evidence suggests that readers rely on word meaning (e.g. semantics, familiarity, or imageability) to aid in the development of word-specific links between orthography and phonology, thus allowing the word to be added to the autonomous orthographic lexicon. This evidence comes from multiple sources. Nation and Cocksey (2009) have reported that item-level familiarity (also known as lexical phonology) was a significant predictor of word reading in young developing readers, with the association being stronger for irregular words, but that deeper semantic knowledge did not predict word-reading success above and beyond familiarity with the phonological form. However, Taylor, Punkett, and Nation (2011) have reported that word learning of an artificial orthography in adults was enhanced by pre-exposure to item definitions but not item lexical phonology. Furthermore, this semantic benefit was specific to items containing low-frequency-inconsistent vowels.

Support for the importance of semantics also comes from connectionist models of reading (e.g. Plaut, 1998; Plaut and Shallice, 1993; Figure 1). Evidence suggests that the addition of a semantic processor (represented as item-specific knowledge) to a model containing phonological and orthographic processors improves both nonword and irregular word recognition (see Plaut, McClelland, Seidenberg, & Patterson, 1996). Ricketts, Nation, and Bishop (2007) also found that item-specific vocabulary knowledge accounted for unique variance in irregular word reading in developing readers. Furthermore, having item-specific vocabulary knowledge for a word has been shown to be a significant predictor of orthographic learning within a self-teaching model of reading development (Wang, Nickels, Nation, & Castles, 2013). Keenan and Betjemann (2006) have speculated that item-specific semantic activation may help to “fill voids” in phonological-orthographic processing in individuals with poor mappings, such as children with reading difficulties (p. 193). Furthermore, evidence from studies on context-free single word reading demonstrate that semantic word features such as concreteness, imageability, and meaningfulness all play a role in lexical decision tasks (see Keenan & Betjemann, 2007). Despite these findings, semantics, in particular imageability, has received little attention within in the word reading literature and the role of semantics at the word level requires more attention (Keenan & Betjemann, 2007).

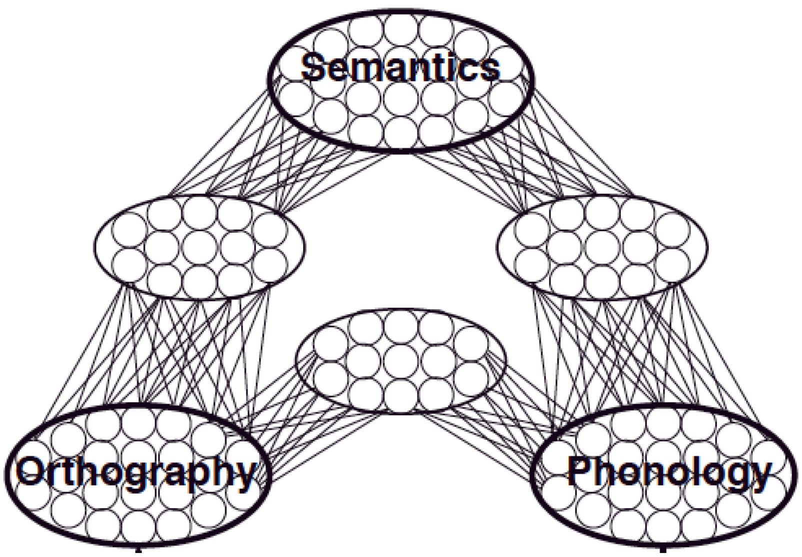

Figure 1. Connectionist model of reading.

Child and Word Characteristics Related to Orthographic Lexicon Development.

***Child characteristics*.** Several important cognitive skills have been related to word recognition accuracy. Important factors include phonological awareness (Lambrecht Smith et al., 2008; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997; Wagner et al., 1997), rapid automatized naming (RAN; for reviews see Kirby, Georgiou, Martinussen, & Parrila, 2010; Norton & Wolf, 2012; Wolf, Bowers, & Biddle, 2000), letter knowledge (e.g. Lambrecht Smith et al., 2008), and vocabulary (e.g. Ouellette, 2006). Given these important relationships, we include these skills in our experiments to examine the extent to which child-level measures of phonological awareness, RAN, vocabulary, and word reading skill predict the efficiency of learning new words and transfer to novel nonwords.

***Word regularity*.** Orthographic regularity refers to the degree to which the pronunciations of phonemes within a word reflect common spelling-sound correspondences (Metsala, Brown, & Stanovich, 1998). There has been consistent support for the important role regularity plays for both typically developing students and students with or at-risk for reading disabilities. Evidence from experimental designs of word reading (Balota & Ferraro, 1993, Seidenberg, Waters, Barnes, & Tanenhaus, 1984; Stanovich & Bauer, 1978; Waters & Seidenberg, 1985), eye-movement studies (Sereno & Rayner, 2000), and a meta-analysis (Metsala, Brown, & Stanovich, 1998) suggest that the impact of regularity on word reading for low frequency words is consistent across samples, ages, and ability groups. Furthermore, both children and adults are more likely to apply grapheme phoneme correspondences when reading nonwords, have a tendency to read regular words with more accuracy, and regularize irregular words (Coltheart & Leahy, 1992). Despite the importance of regularity in word reading development, there remains considerable variance to explain at the word level, suggesting that there may be other important word characteristics to consider (Griffiths & Snowling, 2002; Steacy et al., submitted; Wang, Nickels, Nation, & Castles, 2013).

***Word imageability*.** Imageability is a word specific feature that refers to the ease with which a word can elicit a mental image in the reader (Paivio, Yuille, & Madigan, 1968). A high imageability word is *farm*, while a low imageability word is *which*. This word feature has received some attention given that it is a good predictor of word reading (Walker & Hulme, 1999). For instance, Strain & Herdman (1999) found that when adults read words low on the imageability scale, mean response time increases, particularly for irregular words. Furthermore, when high and low imageability regular and irregular words are read, adults make more errors on low imageability irregular words, particularly when individuals are identified as having medium, as opposed to high, reading skill. Even for the highly skilled readers, more errors were made on low imageability irregular words than highly imageable words. There were no significant differences between skill groups or word types on highly imageable words. Furthermore, imageability did not seem to play an important role in reading regular words for either skill group (Strain & Herdman, 1999).

In other studies of the role of imageability in word reading, there has been a focus placed on the role of imageability in learning nonwords. Laing and Hulme (1999) approached this issue using word abbreviations that children were encouraged to pair with spoken English words. This study illustrates that when children are primed with the phonetic structure of words, the abbreviations are easier to learn. Furthermore, high imageability abbreviations (e.g.,ltr for ladder; lzn for listen) were learned more easily than low imageability abbreviations. In a further study, Duff and Hulme (2012) found that when students learned words that varied in imageability and spelling sound consistency, imageability impacted word reading accuracy, particularly in later trials and for words that were irregular. Furthermore, imageability has been reported to be a particularly important word feature for poor readers (Coltheart, Laxon, & Keating, 1988). Also, having item-specific vocabulary knowledge for words has been shown to be a significant predictor of orthographic learning (Wang, Nickels, Nation, & Castles, 2013) and having passage specific prior knowledge of the content of a text decreases the number of word reading errors in poor readers (Priebe, Keenan, & Miller, 2012).

Attempts to Intervene for Imageability

Although research indicates that imageability can play an important role in word reading, there have been very few attempts to intervene and support students in their learning of low imageability words. One interesting study by Duff and Hulme (2012), attempted to make nonwords imageable using sentences (experiment 2). They compared two conditions, one with only phonological information for nonwords and one with both phonological and semantic information. They found no added benefit for semantic knowledge over and above the benefit of phonological information. In another study, Wang, Nickels, Nation, and Castles (2013) used a vocabulary training condition that included picture supports. They found that vocabulary knowledge of the words was only advantageous for irregular words. This study did not, however, focus only on low imageability or abstract words. One of our goals in the present study was to address this gap in the literature for real words.

Research Questions

The proposed study addresses important gaps within the word learning literature by presenting results from two experiments exploring the efficiency and transfer of word learning in developing readers. Specifically, the two studies focus on the role of imageability in word learning efficiency and transfer among first and second grade children. These studies add to the literature in six distinct ways: (1) it is the first study of imageability effects that focuses on students with or at-risk for reading disabilities, (2) it addresses the role of both imageability and regularity in word learning and rate of mastery, (3) it focuses on real words rather than nonwords, (4) it compares the impact of two different corrective feedback techniques, (5) it includes maintenance of learning measures and transfer measures, and (6) it examines the malleability of imageability for instruction. We extend the literature on word reading efficiency and transfer by answering the following research questions over two studies: (a) what role do regularity and imageability play in initial word learning for at-risk first and second graders?, (b) what roles do imageability and regularity play in the efficiency of word learning mastery?, (c) does maintenance of word learning differ depending on regularity and imageability controlling for pretest?, (d) does maintenance of word learning differ depending on regularity and imageability controlling for posttest?, (e) does transfer to orthographically similar nonwords differ depending on regularity and imageability?, (f) after word learning, do skills transfer to spelling?; does this relationship differ depending on the regularity and/or imageability of the words?, (g) does imageability training impact posttest performance on low imageability words over and above word only and/or vocabulary training?, (h) does word regularity impact posttest performance on low imageability words?, (i) do children in the imageability condition outperform others on posttest spelling of target words? , (j) does imageability training impact the number of exposures necessary for mastery?, and (k) does imageability training impact the number of exposures necessary for mastery differentially for students who start with poor word reading skills? A detailed description of our hypotheses for each of these questions can be found in Appendix A.

CHAPTER II

EXPERIMENT 1

Method

Participants. For Experiment 1 forty-seven at-risk children were drawn from one rural school district in the Southeastern region of the United States. Fifteen first and second grade teachers nominated their lowest five to eight students in reading and were asked to send consent forms home with these students. We pretested all students using measures of word reading, pseudoword reading, phonemic awareness, rapid naming, rapid letter naming, rapid sound naming, and vocabulary. Students were also pretested for their knowledge of the study target words to ensure sufficient room for growth. We excluded students who knew all or nearly all words.

Measures.

Word reading. The word reading task from the Wechsler Individual Achievement Test (WIAT-III; Wechsler, 2009) is a measure of both word reading accuracy and speed. From this task, we extracted two scores. First, students were given an overall accuracy score for all items read correctly and fluently before they reached the ceiling of four consecutive incorrect items. Second, students were given a timed score, reflecting the number of items read correctly in 30 seconds. The split-half reliability reported in the manual for word reading in both first and second grade is .98.

Pseudoword decoding. The pseudoword decoding task from the WIAT-III (Wechsler, 2009) is a measure of nonsense word reading, which is designed to measure speed and accuracy of decoding skill. The first score this test yields is the total number of items read correctly before the ceiling of four consecutive incorrect items was met. The second score is the number of items decoded correctly in 30 seconds. The split-half reliability reported in the manual for pseudoword reading in both first and second grade is .97.

Picture vocabulary. The picture vocabulary test of the Woodcock Johnson-III (Woodcock, McGrew, & Mater, 2001) is a measure of oral language development and vocabulary knowledge. This is primarily an expressive language task in which students were asked to identify pictured objects. The items become increasingly difficult as the test goes on. The ceiling rule for this test is the six highest-numbered items on a page incorrect. The manual reports a split-half reliability of .70 for six year olds and .71 for seven year olds.

Word identification. The Word Identification task from the Woodcock Reading Mastery Test (WRMT-III; Woodcock, 2011) requires children to read isolated real words aloud. The test was discontinued when the test-taker incorrectly identified four consecutive items. Woodcock (2011) reported a split-half reliability of .96 and .94 for first and second grade, respectively.

Word attack. The Word Attack task from the WRMT-III (Woodcock, 2011) requires children to read isolated pseudowords aloud (e.g., ree, ip, and weaf). The test was discontinued after four consecutive errors. Reported split-half reliabilities were .96 and .92 for first and second grade, respectively (Woodcock, 2011).

Phonemic awareness (PA). The phonemic awareness task we used in this study was an elision task that required students to delete phonological units from words. The items became increasingly difficult as the test went on. The task began with syllable deletion and transitioned to deletion of initial, final, and medial phonemes. The task was developed based on the elision task from the Rosner Test of Auditory Awareness Skills (TAAS; Rosner, 1979).

Rapid letter naming. For the rapid letter naming task (Fuchs et al., 2001), we asked students to name an array of 52 letters (all 26 letters upper and lower case) in random order in one minute. The score for this task was the number of letters correctly identified in one minute.

Rapid sound naming. The rapid letter sound naming task (Fuchs et al., 2001), requires students to rapidly name letter sounds. After four practice items in which the tester modelled how to name the sounds of the letters, students were given one minute to name an array of 26 letter sounds in random order. The score for these tasks was the number of sounds correctly identified in one minute.

Rapid automatized naming (RAN). For this naming task, developed by Denckla and Rudel (1976), we gave students one practice item with five letters. We then presented them with a page of these five letters in random order and we asked them to name the letters as quickly as possible.

Target word reading. The target word reading task was a researcher developed task that required students to read a list of 32 words. These words were selected by the researcher to meet the goals of this study. The words were chosen based on a factorial design in which words were selected based on regularity and imageability. Eight words were selected to be highly imageable regular words, eight were selected to be highly imageable irregular words, eight were selected to be regular, low imageability words, and eight were selected to be irregular, low imageability words. These words are provided in Figure 2. Every attempt was made to control for frequency, number of letters, and initial phoneme when selecting the words. The target word reading measure was administered at pretest, posttest, and maintenance.

	Regular	Irregular
High Imageability	brain birth coast farm ground hunt space wife	bowl eye foot guard laugh soup world young
Low Imageability	choice cost lie plus real stuff trust went	broad build false learn lose none once worse

Figure 2. Word factorial design for Experiment 1.

Nonword reading. The nonword reading task was a researcher developed measure that examines transfer of word learning to words that were orthographically similar to the trained target words. These words had the same rime unit as the target words, with different onsets (e.g. *sowl* for *bowl*; *klain* for *brain*). This task was administered at posttest only.

Spelling. The spelling task required students to spell selected target words. Students were asked to spell 16 target words, four from each cell, which varied on orthographic regularity and imageability. This task was administered at posttest only.

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). These data were collected from 22 graduate students in special education. They were asked to rate the difficulty of bringing about a mental image for the 32 words used in this study. They were asked to rate this difficulty on the same seven-point scale used by Paivio, Yuille, and Madigan (1968).

Experimental Design. The design of this experiment was based on a design used in other word learning studies conducted by Martin-Chang and Levy (2005; 2006) and Martin-Chang, Levy, and O'Neill (2007). This experiment was conducted over the course of a two-week period. Students first participated in a brief pretesting session (30 minutes) on the first day, three training sessions, with four exposures per word each day on three consecutive school days (15-20 minutes), one posttesting session on the fifth day (15-20 minutes), and maintenance testing one week after posttesting (five minutes). Students were rank ordered based on pretest performance and then randomly assigned to either whole word feedback or phoneme blending feedback. To prevent order effects in word learning words were randomized across student and session. This is a within subject design, where students act as their own control group (see Kirk, 1982). This experimental design is outlined in Figure 3. The feedback conditions are outlined below.

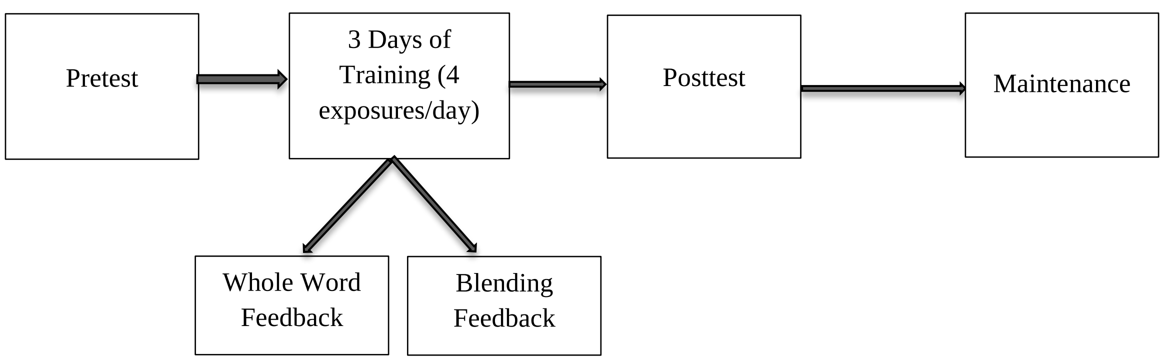

Figure 3. Experimental design for Experiment 1.

Feedback conditions. Students were randomly assigned to one of two feedback conditions. The first condition used a whole word (WW) feedback method, which focused on presenting words as singular units. This approach is supported by the dual route theory, which characterizes sight word reading and decoding as two independent skills (for review see Castles, 2006). Furthermore, according to the psycholinguistic grain size theory (Ziegler & Goswami, 2005), students master word-level skills before syllable-level skills, onset-rime skills, and phoneme-level skills. The WW technique has been widely used and has had some empirical support. Martin-Chang and Levy (2006) explored this technique in comparison to learning words in context. Students in Grade 3 (24 average readers and 24 students with poor reading skills) were assigned to either a whole word or context condition. In the WW condition, when children made reading errors, they were provided with the correct word. They were taught the same 48 new words in each group. Students with poor reading skills who participated in the WW condition retained more words read in isolation and read words faster at posttest than students who participated in the context condition. Further support for the WW technique is provided by the literature on auditory feedback. Students provided with auditory feedback at the word-level perform slightly better on posttest measures of word reading than students who receive phonologically segmented auditory feedback (Spaai, Ellerman, & Reitsma, 2004).

The second technique for sight-word instruction, the phonological-analysis (PhA) method, focused on analyzing the sounds in each word. In this condition, feedback provided the student with the word and the word segmented into its sounds. This technique is supported by Ehri's (1995) stages of sight word development, which suggests phonological decoding and sight word reading are intertwined. Such a technique is also supported by the self-teaching hypothesis, which predicts that increases in phonological awareness and phonological encoding are the central means by which orthographic representations are acquired (for review see Share, 2008). This approach is further supported by studies of phonological segmentation training resulting in improved word reading skills (for review see Nelson et al., 2003) and by studies that found direct teaching of two levels of phonological units, onset-rime and grapheme-phoneme, lead to similar gains in word reading skills (e.g., Walton & Walton, 2002). Furthermore, van Daal and Reitsma

(1990) found that feedback at the PhA level showed similar gains to the WW method when the outcome measure was word reading.

Mastery. Mastery was defined as four out of five exposures correct. This decision was made due to the fact that students had four exposures per day. Setting mastery at four out five correct required students to demonstrate either all exposures correct on one day or mastery across days allowing for some error.

Procedure. Test examiners were graduate research assistants who had been trained on tests until procedures were implemented with 90% fidelity. All tests were double-scored and double-entered; discrepancies were resolved by a third examiner. 20% of exposure data was double entered. The average kappa across exposures was approaching 1.0 and agreement exceeded 99%. Average fidelity of test administration procedures (based on a random selection of 20% of the taped assessment sessions) exceeded 92% for all tests and word exposure trials. Study data were entered and managed using REDCap electronic data capture tools hosted at Vanderbilt University (Harris et al., 2009).

Data analysis. Crossed-random effects models were used to answer the research questions outlined above (Van den Noortgate, De Boeck, & Meulders, 2003). A diagram of these models is provided in Figure 4.These models are based on item response theory and allow for responses to be predicted by person and item level effects. These models are cross-classification multilevel models that allow variance to be partitioned across person and words. For these models, words and persons are assumed to be random samples from a population of words and a population of persons. Since words are not nested within persons, these models are not strictly hierarchical models. Words and persons are on the same level and crossed in the design. Responses are nested within persons and within words.

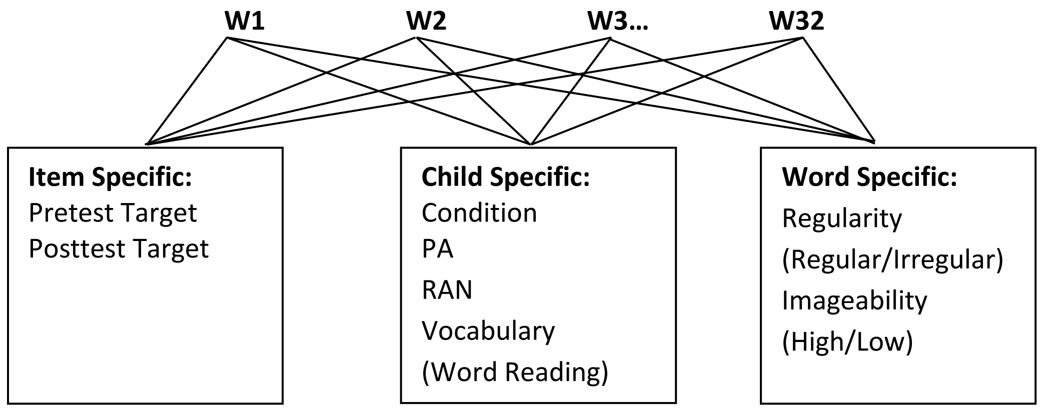

Figure 4. Crossed-random effects models for Experiment 1 and Experiment 2.

The crossed-random effects models were built gradually in a stepwise fashion using model comparisons to determine the model that best fit the data. The first model (unconditional model) contained only random effects for the child and word, with no predictors. Next, an item level predictor of either pretest or posttest was added to the model. Child and word level fixed effects were added to the model and the model was compared to the base model with the item-level predictor. After this step, random slopes were tested one at a time to determine those that needed to be included in the model. Random slopes were added for predictors based on theoretical decisions about which predictors might require a random slope. The following steps were used to establish the random effects structure: (a) fixed effects were added together, (b) appropriate random slopes for fixed effects were then added based on theory, and (c) final models were estimated based on iterations of these steps and model comparisons. A detailed description of the equations used for these models is 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 $p_{jik} = \frac{1}{1+\exp(-\hat{\gamma}_{000}+\hat{\gamma}_v)}$, where *v* is the covariate of interest. The WW group was the referent category and predicted probabilities are given for an average item and an average child in the WW 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 word variance from the base model containing only random effects for child and word and an item specific word reading predictor at either pre or posttest. The formulas were $\frac{\sigma^2_{r01(base)} - \sigma^2_{r01(model\ n)}}{\sigma^2_{r01(base)}}$ and $\frac{\sigma^2_{r02(base)} - \sigma^2_{r02(model\ n)}}{\sigma^2_{r02(base)}}$, 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 variance explained using the fixed slopes, random intercept models, a method supported by recent simulations by LaHuis, Hartman, Hakoyama, and Clark (2014).

Results

Demographic data for the participants in this study (*N*=47) are provided in Table 1. Table 2 provides child-level performance across measures disaggregated by condition with associated mean comparisons (ANOVAs). As noted in the table, there were no significant differences across groups on any of the pretest measures. Table 3 provides the zero order correlations amongst the child level predictors of posttest word recognition. There were significant correlations between all child level predictors of word reading. The child level predictors were correlated with posttest reading performance with correlations ranging from .19 to .56.

Table 1

Demographic Statistics for Experiment 1

Variable		*n*	%	Mean	(SD)
		N = 47			
Age (years)				7.08	(.60)
Gender					
	Male	28	59.57		
	Female	19	40.43		
Grade					
	1	29	61.70		
	2	18	38.30		
Group					
	Whole Word	25	53.19		
	Blending	22	46.81		
Race					
	African American	3	6.38		
	Hispanic	3	6.38		
	Caucasian	40	85.12		
	Biracial	1	2.12		

Note: Age was calculated based on age at the outset of the study.

Table 2
Child Level Descriptive Statistics for Experiment 1

Variable	Whole Word M	Whole Word (SD)	Blending M	Blending (SD)	All children M	All children (SD)	Pairwise comparisons[a]
	n = 25		*n* = 22		*n* = 47		
Word Identification	10.92	(6.09)	10.95	(6.07)	10.94	(6.01)	WW= PhA
Word Attack	3.76	(3.47)	4.59	(3.58)	4.15	(3.51)	WW= PhA
RLN	56.44	(17.86)	52.50	(16.77)	54.60	(17.28)	WW= PhA
RSN	38.24	(11.86)	37.77	(15.20)	38.02	(13.38)	WW= PhA
PA	9.00	(3.74)	7.77	(3.96)	8.43	(3.85)	WW= PhA
RAN	42.00	(9.56)	51.09	(29.33)	46.26	(21.48)	WW= PhA
VOC	17.56	(2.86)	18.05	(3.50)	17.79	(3.15)	WW= PhA

Note: RLN=Rapid Letter Naming; RSN=Rapid Sound Naming; PA = Phonological Awareness; RAN=Rapid Automatized Naming; VOC= Vocabulary; WW = Whole word feedback; PhA = Phonological Analysis Feedback.
[a] Mean comparisons were conducted using ANOVA with Bonferonni post-hoc pairwise comparisons.

Table 3
Zero Order Correlations between child variables: Experiment 1

	1	2	3	4
1 Pretest Word Reading	–			
2 PA	**.56**	–		
3 RAN	**-.46**	**-.37**	–	
4 VOC	**.40**	.22	-.19	–

Note: p < .001 for all variables in bold. PA = Phonological Awareness; RAN=Rapid Automatized Naming; VOC= Vocabulary

Crossed-random effects models. There were two sets of research questions of interest in this study. The first related to predicting the probability of a correct response on a variety of posttest measures and the other was related to the number of exposures required for word learning mastery. Five models were constructed to answer the first set of questions. The final conditional models for each posttest outcome are presented in Table 4 (Models 1 – 5). As noted in the table, the first model predicted posttest recognition of the target words, controlling for item specific pretest performance, condition, PA, RAN, vocabulary, grade, and three planned comparisons for words with low imageability irregular words as the referent group. The unconditional model for posttest performance included only a random effect for child and word had a mean intercept of γ_{000} = 1.28, corresponding to a predicted probability of a correct response of .78 for the average child on the average word. Variability around that estimate was evident for both the child ($\sigma^2 r_{010j}$ = 5.63) and words ($\sigma^2 r_{020i}$ = 1.81). The next model in the model building process, which included a random effect for child and word and item-specific pretest as a predictor, significantly improved model fit over the base model ($\Delta\chi^2_1$= 39.39, p < .0001). Further execution of the model building process, as outlined above, resulted in a model (Model 1) that included a random effect for word and child, a fixed effect for item-specific pretest, fixed effects for all child level and word level predictors, and a random slope for grade across words. This final conditional model fit the data significantly better than the model with only random effects for child and word and fixed effects for item-specific pretest performance ($\Delta\chi^2_{11}$= , p < .0001). The fixed effects for this model indicate that significant predictors of item level posttest performance were item-specific pretest performance, PA, RAN, and the contrast between high and low imageability irregular words (see Model 1, Table 4). These effects correspond to a probability of .18 for an average child reading a low imageability word correctly at posttest if she did not read it correctly at pretest. A student who did not read the word correctly at pretest but performed one standard deviation above the mean on PA and was average on all other measures had a probability of .34 of reading a low imageability irregular word correctly at posttest. Conversely, a student who performed one standard deviation below the mean had only a probability of .08 of reading the word correctly. Likewise, a student who performed one standard deviation above the mean on RAN had a probability of .25 while a student who performed one standard deviation below the mean on RAN had a probability of .12 of reading a low imageability irregular word correctly. Since low imageability irregular words were the referent group in the model, this indicates an imageability effect for irregular words, with high imageability words being easier to read than low imageability words. A child with average scores on all child-level variables would have a probability of .48 of reading high imageability irregular words correctly.

Table 4

Fixed Effects and Variance Estimates for Target Word Recognition for Cohort 1

	Model 1 Posttest Model			Model 2 Nonword Transfer Model			Model 3 Spelling Model			Model 4 Maintenance Model (Posttest)			Model 5 Maintenance Model (Pretest)		
Fixed Effects Parameter	Est.	(SE)	z	Est.	(SE)	z	Est.	(SE)	z	Est.	(SE)	z	Est.	(SE)	z
Intercept (γ_{000})	-1.52	(.99)	-1.53	**-3.59**	**(.55)**	**-6.48**	**-6.10**	**(1.34)**	**-4.54**	**-3.32**	**(.74)**	**-4.50**	**-2.28**	**(.95)**	**-2.39**
Item covariates															
λ_1 pretest	**1.65**	**(.32)**	**5.16**	—	—	—	—	—	—	—	—	—	**1.96**	**(.31)**	**6.24**
λ_2 posttest	—	—	—	**1.63**	**(.27)**	**5.97**	**2.68**	**(.79)**	**3.41**	**3.10**	**(.27)**	**11.59**	—	—	—
Child covariates															
γ_{003} Condition	.56	(.48)	1.16	**.60**	**(.21)**	**2.82**	.51	(.39)	1.31	.60	(.38)	1.56	.69	(.48)	1.43
γ_{004} PA	**.23**	**(.07)**	**3.15**	**.11**	**(.03)**	**3.67**	**.23**	**(.06)**	**3.84**	**.13**	**(.06)**	**2.23**	**.19**	**(.07)**	**2.62**
γ_{005} RAN	**-.02**	**(.01)**	**-2.01**	**-.07**	**(.02)**	**-4.39**	**-.07**	**(.03)**	**-2.37**	-.01	(.01)	-1.13	-.02	(.01)	-1.63
γ_{006} Vocabulary	.06	(.08)	.71	**.09**	**(.04)**	**2.47**	.01	(.07)	.19	.02	(.06)	.30	.03	(.08)	.34
γ_{007} Grade	1.02	(.63)	1.63	-.51	(.31)	-1.66	.34	(.55)	.64	**1.27**	**(.49)**	**2.58**	**1.53**	**(.61)**	**2.49**
Word covariates															
γ_{008} High Imag./Irregular	**1.44**	**(.57)**	**2.52**	.62	(.42)	1.45	-.31	(1.14)	-.27	.64	(.38)	1.70	**1.30**	**(.59)**	**2.19**
γ_{009} High Imag./Regular	.75	(.56)	1.34	**.91**	**(.42)**	**2.15**	.51	(1.13)	.45	.17	(.37)	.65	.37	(.59)	.64
γ_{010} Low Imag./ Regular	1.01	(.57)	1.79	**1.63**	**(.42)**	**3.88**	1.40	(1.13)	1.25	.59	(.38)	1.57	.81	.59	1.37
	% var explained			% var explained			% var explained			% var explained			% var explained		
Intercepts															
Person	65.06			78.35			81.29			82.54			65.09		
Word	30.64			38.51			21.21			75.50			31.71		

Note: PA = Phonological awareness (elision); RAN = Rapid automatic naming; Imag. = Imageability
$p < .05$ for variables in bold.

The next model focused on the posttest nonword transfer task. This model was used to predict posttest nonword reading, controlling for posttest target word performance (see Model 2, Table 4). The unconditional model, which only included random effects for child and word, indicated a predicted probability of a correct response of .13 for the average child on the average word. The model that included item-specific posttest word reading resulted in an improved model fit ($\Delta\chi^2_1$= 49.71, *p* < .0001). The final model for this research question included only a random effect for child and word and fixed effects for item-specific pretest performance, condition, PA, RAN, vocabulary, grade, and three planned comparisons for words with low imageability irregular words as the referent group. This main effect model fit the data significantly better than the model with only random effects for child and word and an item-specific predictor for posttest word reading ($\Delta\chi^2_8$= 62.78, *p* < .0001). Significant child-level predictors included: posttest items, condition, PA, RAN, and vocabulary. Whereas students who read low imageability irregular words incorrectly at posttest had a probability of .03 of reading the transfer nonword correctly, students who read the words correctly at posttest had a probability of .12 of reading the transfer nonword correctly. Furthermore, students in the PA feedback condition who were average on all other measures and read the target word correctly at posttest had a probability of .20 of reading transfer nonwords correctly, which represents a .08 higher probability than students in the WW condition. Students who scored one standard deviation above the mean on PA increased their probability of reading the nonwords to .18 while students who scored one standard deviation below the mean had a probability of .08 of reading the transfer nonwords correctly. Likewise, students who scored one standard deviation above the mean on RAN, who read the target word correctly at posttest, and were at the mean on all other measures, had a probability of .38 of a correct response. Students who scored one standard deviation below the mean had a probability of .03 of a correct response. Students who scored one standard deviation above the mean on picture vocabulary had a probability of .15 of a correct response and students who scored one standard deviation below the mean had a probability of .10 of a correct response. There was a regularity effect for the nonwords. Nonwords with target words that were regular were associated with a higher probability of a correct response. High imageability regular target words were associated with a probability of .26 of a correct response while low imageability regular words were associated with a probability of .42 of a correct response when performance on all other measures was at the mean and target words were read correctly at posttest.

The next model of interest was a posttest spelling model that examined the transfer of reading target words correctly to spelling. The unconditional model for this question indicated a predicted probability of .06 for the average word for the average child. The model that included item-specific posttest word reading resulted in an improved model fit ($\Delta\chi^2_1$= 23.71, p < .0001). The final model from the model building process included a random effect for child and word, a random effect for the item level posttest predictor across words and children, and fixed effects for child-level and word-level predictors (see Model 3, Table 4). This model fit the data significantly better than the model with only random effects for child and word and an item-specific predictor for posttest word reading ($\Delta\chi^2_8$= 28.40 , p < .0001). These models indicated a predicted probability of a correct spelling response of .03 for low imageability irregular words if students read the target word correctly at posttest. Furthermore, these models indicate that the only child level predictors associated with response were PA and RAN. Students who scored one standard deviation above the mean (and at the mean on all other measures) on PA had a probability of .07 of a correct response, while students who scored one standard deviation below the mean had a probability of just .01 of a correct response. Students who scored one standard deviation above the mean on RAN had a probability of .13 of a correct spelling response while students who scored one standard deviation below the mean had a probability of just .007.

The next models of interest were used to answer the research question regarding maintenance of word learning (see Models 4 & 5, Table 4). Two models were developed to answer this question, one controlling for item specific pretest performance and the other controlling for item specific posttest performance. The overall base model, which includes a random effect for child and word, indicates that the probability of reading a word correctly at maintenance testing for the average child and word is .75. The probability of a student reading a low imageability irregular word correct at maintenance if the student read it correctly at posttest was .45. The final model included a random effect for child and word, a random effect for the item level posttest predictor across words and children, and fixed effects for child-level and word-level predictors (see Model 4, Table 5). The model that included item-specific posttest word reading resulted in an improved model fit ($\Delta\chi^2_1$= 236.59, $p < .0001$). This model fit the data significantly better than the model with only random effects for child and word and item-specific posttest word reading ($\Delta\chi^2_8$= 28.23, $p < .0001$). For students who scored one standard deviation above the mean on PA, the probability rose to .57, while for students who scored one standard deviation below the mean on PA had a probability of .33 of reading it correctly at maintenance. Likewise, when controlling for pretest, the probability of a first grade student reading a low imageability irregular word correctly at maintenance if they did not read the word correctly at pretest was .09. The model that included item-specific pretest word reading resulted in an improved model fit ($\Delta\chi^2_1$= 53.27, $p < .0001$). The final model from the model building process included a random effect for child and word and fixed effects for child-level and word-level predictors (see Model 5, Table 4). This model fit the data significantly better than the model with only random effects for child and word ($\Delta\chi^2_8$= 35.99, p < .0001). If a student scored one standard deviation above the mean on PA and average on all other measures, the probability of a correct response increased to .17 while the probability decreased to .05 if a student was one standard deviation below the mean. For students in second grade, the probability of a correct response was .32 if they did not read the word correctly at pretest. Finally, if the word was high imageability and irregular, the probability of a correct response increased to .27 if a student did not read the word correctly at pretest, the student was in first grade, and was at the mean for all child-level measures.

Table 5
Experiment 1: Fixed Effects and Variance Estimates for Number of Exposures Required for Mastery

Fixed Effects Parameter	Mastery Model (Model 6) Est.	(SE)	t
Intercept (γ_{000})	**10.56**	**.47**	**22.59**
Item covariate			
λ_1 posttest	**-2.13**	**.15**	**-14.08**
Child covariates			
γ_{002} Condition	-.52	.40	-1.30
γ_{003} PA	**-.21**	**.06**	**-3.84**
γ_{004} RAN	**.03**	**.01**	**3.07**
γ_{005} Vocabulary	-.11	.06	-1.76
Word covariate			
γ_{006} High Imag./Irregular	-.99	.54	-1.85
γ_{007} High Imag./Regular	-.80	.54	-1.49
γ_{008} Low Imag./Regular	**-1.64**	**.54**	**-3.06**
	% var explained		
Intercepts			
Person	44.70		
Word	29.82		

Note: PA = Phonological awareness (elision); RAN = Rapid automatic naming; Imag. = Imageability
$p < .05$ for variables in bold

The final research question in this experiment concerned the number of exposures necessary for mastery of each item. All students had 12 exposures to the words across three days. The mastery model predicted a continuous outcome, number of exposures required for mastery, with mastery defined as four out of five exposures correct. In the mastery model, item-level posttest performance was controlled for because not all children achieved mastery on all items in 12 exposures. The unconditional model, which included only a random effect for word and child, indicated that an average of 8.28 exposures were required for mastery across items and children. In the model that controlled for posttest performance only, the number of exposures necessary for mastery was reduced to 7.31 if the word was read correctly at posttest. The model that included item-specific posttest word reading resulted in an improved model fit ($\Delta\chi^2_1$= 226.51, $p <$ .0001). The fixed effects model (see Model 6, Table 5) fit the data significantly better than a model with just a random effect for words and children and an item-specific predictor of posttest word reading ($\Delta\chi^2_7$= 44.43, p < .0001). The referent group was again irregular low imageability words. This model indicated a mean number of exposures to mastery of 8.43 for first grade students on low imageability irregular words. Students who scored one standard deviation above the mean on PA required 7.63 exposures, while students who scored one standard deviation below the mean on PA required 9.23 exposures. Students who scored one standard deviation above the mean on RAN required 9.07 exposures while students who scored one standard deviation below the mean required 7.79 exposures. There was a clear regularity effect for low imageability words. The number of exposures required to master low imageability regular words was 6.79, nearly two exposures less than low imageability irregular words.

Discussion: Experiment 1

The results for the first study indicate that student word learning differed depending on child-level and word-level characteristics. These relationships were observed for both overall learning and rate of mastery and are consistent with other findings that both imageability and regularity impact word recognition mastery and efficiency (Duff & Hulme, 2012; Laing & Hulme, 1999).

We found a clear effect for word imageability for posttest word reading favoring high imageability words in addition to the child-level predictors that were expected. This is an interesting finding because there was no effect of general vocabulary skill but there was an item-specific influence of imageability. This imageability effect was observed by comparing the words in the high imageability irregular word cell to the words in the low imageability irregular word cell. These results indicate that a word such as *laugh* had a higher probability of being read correctly at posttest than a word such as *false*. Consistent with this finding, we found a regularity effect for mastery of low imageability words. When low imageability regular words are compared to low imageability irregular words, regular words required fewer exposures to mastery than irregular words. It was evident that there was still considerable variance left to explain at the word level across models. There may be other important word features missing from the model that we were not able to include due to small sample size. Although we could not include these in the models, we attempted to restrict the range of other important word characteristics (e.g. frequency) in the selection of the words.

We were interested in how the word learning in this study transferred to orthographically similar nonwords. The results from these models indicated that there were main effects for both child- and word-level characteristics. There was a significant effect for condition, with children in the phonological analysis feedback condition having a higher probability of reading nonwords correctly at posttest than students in the whole word feedback condition. There were also the expected main effects for PA and RAN. Additionally, there was a significant main effect for vocabulary on only the nonword transfer task. We speculate that vocabulary may be serving as a proxy for general intelligence because of the complexity of the orthographic task, with students required to apply their knowledge of previously learned words to novel nonwords, a task that proved very difficult for most children. The list of nonwords was orthographically complex, including both regular and irregular words.

The spelling models indicated that there was a significant main effect of PA and RAN on posttest spelling. There was no main effect for word-level characteristics. However, these models were likely underpowered and require further exploration. We performed two different analyses for maintenance of learning after one week. The first model controlled for posttest word knowledge while the second model controlled for pretest word knowledge. The first model indicated that for words that were correctly identified at posttest, there was only a significant main effect for PA and grade, with students in second grade more likely to maintain the words. When we controlled for pretest, however, there was also a significant imageability effect. Results across the two models indicate that irregular words that were highly imageable were more likely to be learned and maintained than irregular words that were low on the imageability scale.

The final models for the first study examined the number of exposures that were necessary for mastery (defined as four out of five attempts correct) across word groups. These models controlled for posttest word reading, which means that the results are conditioned on posttest performance. This was done because not all students mastered all items but we wanted to include all students in the analyses. For the mastery model, there was a significant main effect for PA and RAN and a regularity effect. The regularity effect indicated that it took fewer exposures for students to master low imageability regular words than it did for them to master low imageability irregular words. These results were expected given that students could apply decoding rules to the regular words but not the irregular words.

In Experiment 2, we were interested in examining the potential for manipulating imageability to impact efficiency of word learning in an unselected sample of first grade students. We believe the unselected sample is representative of the typical distribution of first graders in the district. We took only the low imageability words (*N*=16) from Experiment 1 and assigned students to one of three instructional conditions: (1) word only (WO), (2) vocabulary (VOC), or (3) imageability (IMAG).

CHAPTER III

EXPERIMENT 2

Method

Participants. For Experiment 2 seventy-eight first grade children were drawn from one urban school district in the Southeastern region of the United States. Six first grade teachers were asked to send consent forms home with all students. We pretested all students using measures of word reading, pseudoword reading, phonemic awareness, rapid automatized naming, rapid letter naming, rapid sound naming, and vocabulary. Students of all ability levels were included in the study.

Measures.

Word reading. The word reading task in this study was the Sight Word Efficiency task from the Test of Word Reading Efficiency (TOWRE; Torgesen, Wagner, & Rashotte, 2012). For this task, students were asked to read a series of words in order of increasing difficulty for 45 seconds. The maximum score is 108 and the authors report an alternate forms reliability of .91.

Pseudoword decoding. To test pseudoword decoding skill, we used the Pseudoword Decoding Efficiency task from the TOWRE (Torgesen, Wagner, & Rashotte, 2012). For this task, students are asked to reading a list of pseudowords in order of increasing difficulty for 45 seconds. The maximum possible score is 66 and the authors report an alternate forms reliability of .92.

Picture vocabulary. The picture vocabulary test of the Woodcock Johnson-III (Woodcock, McGrew, & Mater, 2001) is a measure of oral language development and vocabulary knowledge. This is primarily an expressive language task in which students were asked to identify pictured objects. The items become increasingly difficult as the test goes on.

The ceiling rule for this test is the six highest-numbered items on a page incorrect. The manual reports a split-half reliability of .70 for six year olds and .71 for seven year olds.

Phonemic awareness (PA). The phonemic awareness task used in this study was the Elision task from the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, 2013). This task requires students to delete phonological units from words, starting with syllables, and increasing in difficulty to deleting initial, final, and medial phonemes. The authors report test-retest reliability of .93.

Rapid letter naming. For the rapid letter naming task (Fuchs et al., 2001), we asked students to name an array of 52 letters (all 26 letters upper and lower case) in random order in one minute. The score for this task was the number of letters correctly identified in one minute.

Rapid sound naming. The rapid letter sound naming task (Fuchs et al., 2001), requires students to rapidly name letter sounds. After four practice items in which the tester modelled how to name the sounds of the letters, students were given one minute to name an array of 26 letter sounds in random order. The score for this task was the number of sounds correctly identified in one minute.

Rapid automatized naming (RAN). To test for rapid automatized naming, we used the letter naming task from the CTOPP (Wagner, Torgesen, and Rashotte, 2013). For this task, students were asked to name a series of letters as fast as they could without making mistakes. The total score was the number of seconds students took to name all of the letters. The authors report a test-retest reliability of .85.

Target word reading. The target word reading task was a researcher developed task that required students to read a list of 32 words, which was the same list administered in Experiment

1. This list was scored in two steps, first for all 32 words, and then for only the 16 words targeted in the intervention (low imageability; see Figure 5).

	Regular	**Irregular**
Low Imageability	choice cost lie plus real stuff trust went	broad build false learn lose none once worse

Figure 5. Word lists for Experiment 2.

Spelling. The spelling task required students to spell only the target words that were taught during the intervention. This task was administered at posttest only.

***Regularity*.** Words were considered irregular if they were not consistent with typical letter sound correspondences. Regularity was determined based on the criteria outlined by Rastle and Coltheart (1999).

Experimental Design. The design of this experiment was similar to the design used in Experiment 1. Each student participated in the experiment for six consecutive days whenever possible, with everyone completing the experiment within 10 days and modifications being made to the experimental design when circumstances (i.e., attendance and school holidays) demanded it. Each group of students started the study on Friday and participated in a pretest battery of cognitive and reading related tests for approximately 30 minutes. Next, based on their pretest scores on the pretest target word reading test, they were randomly assigned to one of three conditions: (1) imageability (IMAG), (2) vocabulary (VOC), and (3) word-only (WO). Then, they participated in four days of training from Monday to Thursday for approximately 15-20 minutes per day. Students then completed a short posttest battery on Friday for approximately 10-15 minutes. We only administered the target word reading task and the spelling task at post-test. This was a within subject design, where students act as their own control group (see Kirk, 1982). This design is illustrated in Figure 6.

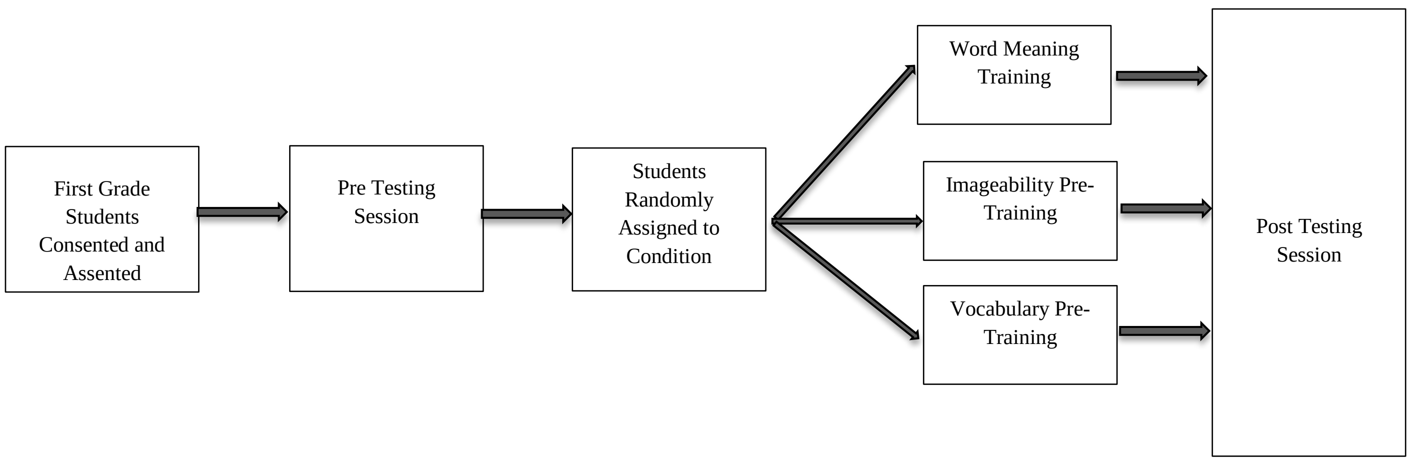

Figure 6. Experimental design for Experiment 2.

Conditions. On the first day of training (Monday), students in the imageability and the vocabulary conditions were exposed to either pictures (imageability) or verbal definitions for the words (vocabulary). Students in the word only condition read storybooks with their tutors during that time. On the three remaining training days (Tuesday – Thursday), based on condition students spent 5 minutes reviewing pictures, reviewing definitions, or reading storybooks before being exposed to each target word 4 times each day. To prevent order effects words were randomized across student and session. The number of times the students heard the words was held constant across the imageability and vocabulary conditions. Corrective feedback was given at the whole word level across conditions.

Imageability (IMAG). Materials for the imageability condition were created with a children's illustrator. The illustrator helped us to make the low imageability words imageable and she created the pictures for the training materials. On the first day of training, students in the imageability condition were exposed to each picture twice and were encouraged to think of the picture when they heard that word. They were then asked to identify each picture when provided the word and asked to produce the word when presented with a picture. These activities were done using an imageability flipbook and the game board provided in Appendix C. The game board receptive and expressive activities were done at the beginning of the word learning sessions on days 2-4 of training. During word learning, students were encouraged to think of the picture for each word.

Vocabulary (VOC). On the first day of training, students in the vocabulary condition were exposed to each definition twice and were encouraged to think of the meaning when they heard that word. These definitions are provided in Appendix D. Students were then asked to do an oral multiple choice task in which they were asked to identify each meaning when presented with the word and asked to produce the word when presented with the meaning. The multiple choice receptive and expressive activities were done at the beginning of the word learning sessions on days 2-4 of training. During word learning, students were encouraged to think of the meaning for each word.

Word only (WO). In the word only condition, students read story books for 15 minutes on the first day and then for 5 minutes at the beginning of each word learning session to equalize instructional time. During word learning, students were only given corrective feedback at the whole word level, with no encouragement to think of the meaning or a picture for the words.

Procedure. Test examiners were graduate research assistants who had been trained on tests until procedures were implemented with 90% fidelity. All tests were double-scored and double-entered; discrepancies were resolved by a third examiner. 20% of exposure data was double entered. The average kappa across exposures was approaching 1.0 and agreement exceeded 99%. Average fidelity of test administration procedures (based on a random selection of 20% of the taped assessment sessions) exceeded 94% for all tests and training sessions. Study data were entered and managed using REDCap electronic data capture tools hosted at Vanderbilt University (Harris et al., 2009).

Data Analysis. The analytic technique used in Experiment 2 was similar to the technique used in Experiment 1 (see above). The same model building process was used and the structure of the models was the same as the model illustrated in Figure 4. The models, corresponding random effects structure, and corresponding research questions for each model are presented in Appendix B. The referent for these models was the imageability group.

Results

Demographic data for the participants in this study (*N*=78) are provided in Table 6. Table 7 provides child-level performance across measures disaggregated by condition with associated mean comparisons. As noted in the table, there were no significant differences across groups on any of the pretest measures. Table 8 provides the zero order correlations amongst the child level predictors of posttest word recognition. There were significant correlations between child level predictors of word reading. The child level predictors were correlated with posttest reading performance with correlations ranging from .11 to .69.

Table 6

Demographic Statistics for Experiment 2

	N = 78			
Variable	*n*	%	Mean	(SD)
Age (years)			7.08	(.60)
Gender				
Male	42	47.73		
Female	33	42.31		
Unreported	3	3.85		
Group				
Word Only	26	33.33		
Vocabulary	26	33.33		
Imageability	26	33.33		
Race				
African American	14	17.95		
Hispanic	8	10.26		
Caucasian	51	65.38		
Unreported	8	10.26		

Note: Age was calculated based on age at the outset of the study.

Table 7
Child Level Descriptive Statistics for Experiment 2

Variable	Word Only M	Word Only (SD)	Vocabulary M	Vocabulary (SD)	Imageability M	Imageability (SD)	All children M	All children (SD)	Pairwise comparisons[a]
	n = 25		*n* = 22		*n* = 22		*n* = 47		
SWE	10.92	(6.09)	10.95	(6.07)	34.12	(16.44)	10.94	(6.01)	WO=VOC=IMAG
PDE	3.76	(3.47)	4.59	(3.58)	12.08	(8.48)	4.15	(3.51)	WO=VOC=IMAG
RLN	56.44	(17.86)	52.50	(16.77)	58.96	(18.21)	54.60	(17.28)	WO=VOC=IMAG
RSN	38.24	(11.86)	37.77	(15.20)	34.62	(9.29)	38.02	(13.38)	WO=VOC=IMAG
PA	9.00	(3.74)	7.77	(3.96)	17.42	(7.63)	8.43	(3.85)	WO=VOC=IMAG
RAN	42.00	(9.56)	51.09	(29.33)	33	(12.31)	46.26	(21.48)	WO=VOC=IMAG
VOC	17.56	(2.86)	18.05	(3.50)	18.31	(3.06)	17.79	(3.15)	WO=VOC=IMAG

Note: SWE=TOWRE Sight Word Efficiency; PDE=TOWRE Pseudoword Decoding; RLN=Rapid Letter Naming; RSN=Rapid Sound Naming; PA = Phonological Awareness; RAN=Rapid Automatized Naming; VOC= Vocabulary
[a] Mean comparisons were conducted using ANOVA with Bonferonni post-hoc pairwise comparisons.

Table 8
Zero Order Correlations between child variables: Experiment 2

	1	2	3	4
1 Pretest Word Reading	–			
2 PA	**.69**	–		
3 RAN	**-.43**	**-.35**	–	
4 VOC	**.37**	**.41**	-.11	–

Note: p < .001 for all variables in bold. PA = Phonological Awareness; RAN=Rapid Automatized Naming; VOC= Vocabulary

Crossed-random effects models. Similar to the first study, there were two sets of research questions of interest in this study. The first related to predicting the probability of a correct response on posttest target word reading and spelling. The other research questions were related to the number of exposures required for mastery. Two models were constructed to answer the first set of questions. These models are presented in Table 9. As noted in the table, the first model was used to predict posttest recognition of the target words, controlling for item-specific pretest word reading accuracy. Child-level variables related to posttest performance were: item-specific pretest target word reading, condition, PA, vocabulary, and RAN. The only word-level predictor was regularity. The unconditional model for posttest performance included only a random effect for child and word and had a mean intercept of γ_{000} = 3.43, corresponding to a predicted probability of a correct response of .97 for the average child on the average word. Variability around that estimate was evident for both the child ($\sigma^2 r_{010j}$ = 10.38) and items ($\sigma^2 r_{020i}$ = 1.71). The next model in the model building process, which included an item-specific predictor for pretest word recognition, significantly improved model fit over the base model ($\Delta\chi^2_1$= 5.13, p =.02). Further execution of the model building process outlined in Experiment 1 resulted in a model that included a random effect for word and child, and fixed effects for all child level and word level predictors. This model fit the data significantly better than the model with only random effects for child and word and an item-specific predictors for posttest word reading ($\Delta\chi^2_6$=46.40 , p < .0001). The fixed effects for this model indicate that significant predictors of item level posttest performance were PA, RAN, and word regularity. These effects correspond to a probability of .93 for an average child reading a low imageability word correctly at posttest if she did not read it correctly at pretest. A student who did not read the word correctly at pretest but performed one standard deviation above the mean on PA who was average on all other measures had a probability of .98 of reading a low imageability irregular word correctly at posttest. Conversely, a student who performed one standard deviation below the mean had only a probability of .73 of reading the word correctly. Likewise, a student who performed one standard deviation above the mean on RAN had a probability of .97 while a student who performed one standard deviation below the mean on RAN had a probability of .83 of reading a low imageability irregular word correctly. Furthermore, there was a significant main effect for regularity. This indicates that a student with average scores on all child-level predictors who did not read the word correctly at pretest had an increased probability of .99 of reading a regular low imageability word at posttest, compared to a predicted probability of .93 for low imageability irregular words.

Table 9

Fixed Effects and Variance Estimates for Target Word Recognition for Experiment 2

	Model 7 Posttest Model			Model 8 Spelling Model		
Fixed Effects Parameter	Est.	(SE)	z	Est.	(SE)	z
Intercept (γ_{000})	**2.56**	**(.67)**	**3.83**	**-2.80**	**(.70)**	**-3.98**
Item covariates						
λ_1 pretest	.58	(.34)	1.69	—	—	—
λ_2 posttest	—	—	—	**1.03**	**(.37)**	**2.77**
Child covariates						
γ_{003} WO vs. IMAG Cond.	.17	(.78)	.21	-.32	(.45)	-.72
γ_{004} VOC vs. IMAG Cond.	-.89	(.78)	-1.14	-.19	(.45)	-.42
γ_{005} PA	**.22**	**(.06)**	**3.48**	**.16**	**(.03)**	**5.12**
γ_{006} RAN	**-.08**	**(.03)**	**-3.08**	-.03	(.02)	-1.59
γ_{007} Vocabulary	.06	(.12)	.47	.05	(.07)	.75
Word covariate						
γ_{008} Regularity	**1.66**	**(.53)**	**3.15**	**2.20**	**(.77)**	**2.84**
	% var explained			% var explained		
Intercepts						
Person	50.84			55.70		
Word	48.28			38.46		

Note: PA = Phonological awareness (elision); RAN = Rapid automatic naming; IMAG. = Imageability; VOC = Vocabulary; WO = Word Only; Cond.=Condition

p < .05 for variables in bold

The same model building process was used for the spelling model. The model that included item-specific posttest word reading resulted in an improved model fit ($\Delta\chi^2_1$= 11.05, $p <$ .0001) when compared to the base model with only a random effect for word and child. The final spelling model included both child-level and word-level predictors, controlling for posttest item-specific target word reading. The model building process resulted in a main effect model that included only a random effect for child and word. This model fit the data significantly better than the model with only random effects for child and word ($\Delta\chi^2_6$= 47.90, $p < .0001$). The results of this model indicated a significant main effect for item-specific posttest target word reading, PA, and word regularity. The unconditional model indicated that the overall predicted probability of a correct response on the spelling task was .27. When controlling for posttest item-specific word reading, that probability increased to .31. For irregular low imageability words, students who read the word correctly at posttest and who were average on all other child level measures had a predicted probability of .15 of spelling the word correctly. Students who scored one standard deviation above the mean on the PA task had a probability of .34 of spelling low imageability irregular words while students one standard deviation below the mean had a probability of .05 of spelling it correctly. The regularity effect indicated that students average on all other measures who read target words correctly at posttest had a predicted probability of .61 of spelling low imageability regular words correctly.

Similar to Experiment 1, the second set of research questions in this experiment concerned the number of exposures required for mastery. These models are presented in Table 10. The unconditional model to address this question contained only a random effects for child and word. The mean intercept for this model was γ_{000} = 5.95, representing the average number of exposures required for mastery for the average child on the average word. Variability around that estimate was evident for both the child ($\sigma^2 r_{010j}$ = 2.15) and items ($\sigma^2 r_{020i}$ = 0.72). The next model in the model building process, which included an item specific predictor for posttest word recognition, significantly improved model fit over the base model ($\Delta\chi^2_1$= 227.17, p <.001). The model building process resulted in a final model that included a random effect for word and child, a random slope for PA across words, and a random slope for TOWRE across words. This model fit the data significantly better than the model with only random effects for child and word ($\Delta\chi^2_{13}$= 153.94 , p < .0001). Significant main effects include posttest target word reading, TOWRE Sight Word Efficiency, and the WO vs. imageability condition comparison. The average number of exposures necessary for mastering low imageability irregular words for students in the imageability condition was 5.36 for students who read the word correctly at posttest. For students who scored one standard deviation above the mean on TOWRE, the average number of exposures for irregular words was 4.28 while the average number of exposures for students who scored one standard deviation below the mean was 6.44. Students who were in the word only condition who were average on all other measures required, on average, 6.08 exposures to low imageability irregular words. The interaction model indicated that there was a significant interaction between condition and initial word reading skill on the TOWRE task. Students who started the intervention with low word reading skills benefited more from the imageability training than the word only training. This interaction is illustrated in Figure 7.

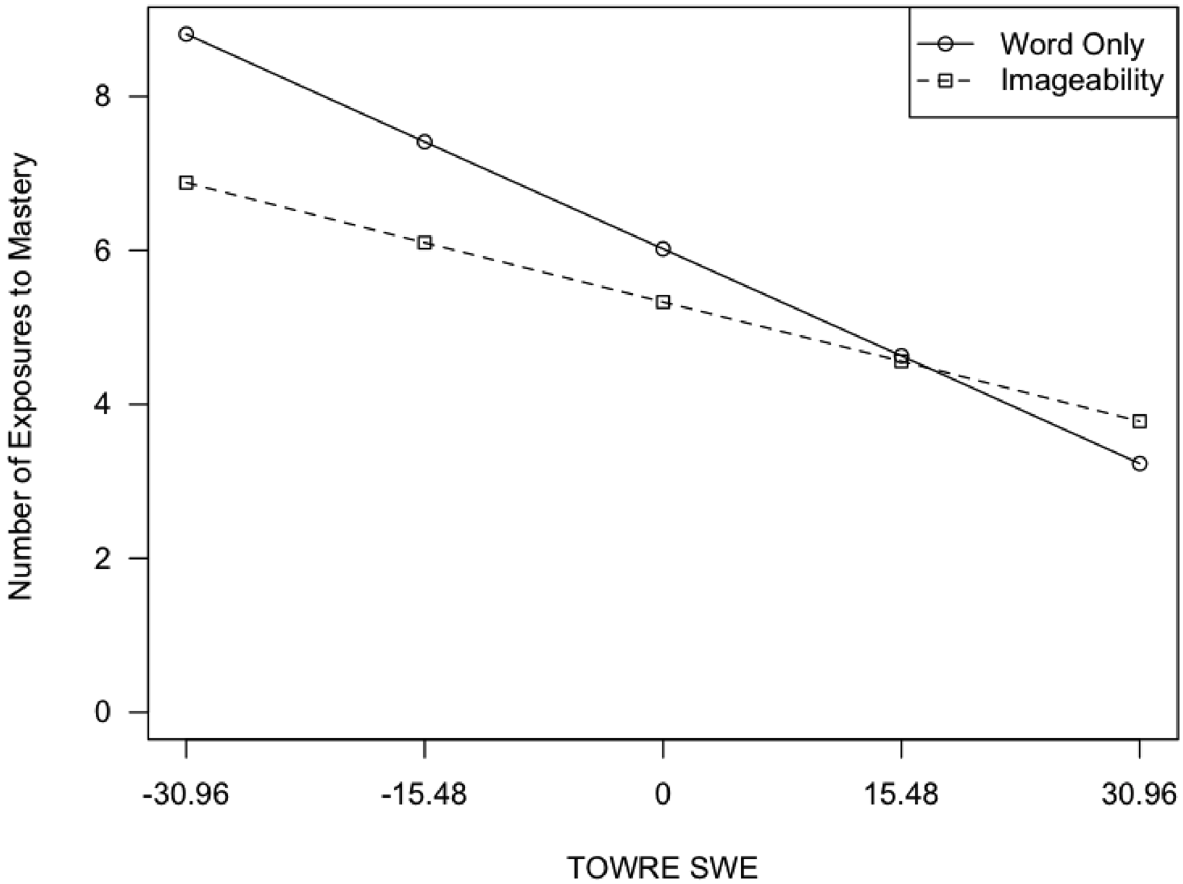

Figure 7. Interaction between initial word reading (TOWRE Sight Word Efficiency) and condition in Experiment 2.

Table 10

Experiment 2: Fixed Effects and Variance Estimates for Number of Exposures Required for Mastery

Fixed Effects Parameter	Model 9 Mastery Model Est.	(SE)	t	Model 10 Interaction Model Est.	(SE)	t
Intercept (γ_{000})	**7.51**	**(.29)**	**25.85**	**7.49**	**(.29)**	**26.02**
Item covariate						
λ_1 posttest	**-2.15**	**(.18)**	**-11.87**	**-2.16**	**(.18)**	**-11.98**
Child covariates						
γ_{002} TOWRE	**-.07**	**(.01)**	**-4.98**	**-.05**	**(.02)**	**-3.22**
γ_{003} WO vs. IMAG Condition	**.72**	**(.29)**	**2.50**	**.69**	**(.28)**	**2.44**
γ_{004} VOC vs. IMAG Condition	.14	(.29)	.49	.16	(.28)	.58
γ_{005} PA	-.04	(.02)	-1.79	-.04	(.02)	-1.86
γ_{006} RAN	.01	(.01)	.98	.01	(.01)	.98
γ_{007} Vocabulary	.00	(.04)	.00	.01	(.04)	.33
Word covariate						
γ_{008} Regularity	-.22	(.15)	-1.51	-.22	(.15)	-1.51
Interactions						
γ_{009} TOWRE*WO vs. IMAG	—	—	—	**-.04**	**(.02)**	**-2.24**
γ_{010} TOWRE*VO vs. IMAG	—	—	—	-.01	(.02)	-.76
	% var explained			% var explained		
Intercepts						
Person	59.66			61.00		
Item	27.81			27.86		

Note: PA = Phonemic awareness (elision); RAN = Rapid automatized naming; IMAG. = Imageability; VOC = Vocabulary; WO = Word Only.
$p < .05$ for variables in bold.

Discussion: Experiment 2

The results of Experiment 2 were consistent with the findings from Experiment 1 and other studies that have explored the role of imageability in word learning. The first set of models, which were used to predict the probability of reading and spelling words correctly at posttest, indicated that there were several child- and word-level measures associated with performance. The first model, which predicted posttest reading of the target words, indicated that PA and RAN were significant predictors of item-specific posttest performance. Furthermore, there was the expected regularity effect that was also found in Experiment 1. Students were more likely to read words at posttest if words contained regular spelling patterns. We observed similar effects for spelling, with a main effect for posttest item specific word recognition, PA, and regularity. Again, these results indicate that students had a higher probability of spelling regular low imageability words correctly at posttest than they did for spelling low imageability irregular words. We did not observe a main effect for condition on posttest reading or spelling performance. That is, the probability of a student reading or spelling the words correctly at posttest was not dependent on which of the three instructional groups they were in.

The same regularity effect was not observed for rate of mastery. We found that there was a main effect for pretest general word reading (TOWRE) and a main effect for condition (WO vs. IMAG), with students in the imageability group outperforming students in the word only condition. There was also a significant interaction effect for initial general word reading and condition meaning that students who started with low performance on general word reading benefited more from the imageability condition than they did from the word only condition. For the interaction between initial general word reading and the IMAG vs. VOC comparison, the same pattern was not observed.

CHAPTER IV

GENERAL DISCUSSION

The results from these experiments both confirm and extend previous findings from the literature on the role of imageability in word reading. We found that imageability was an important factor for both establishing a precise final lexical representation and for promoting efficient word learning. These findings are consistent with others who have found that imageability ratings impact word learning differently across time (Duff & Hulme, 2012; Laing & Hulme, 1999). Furthermore, the results from Experiment 2 offer some preliminary findings that suggest that imageability instruction may allow for adding an additional word feature to a word's lexical representation.

The role of Imageability in the Development of the Orthographic Lexicon

The first important finding from these experiments is that imageability seems to play a role in the development of word specific representations. Results from both studies suggest that imageability impacts the strength of a word representation across exposures. The findings for the posttest measure of word reading in the first experiment indicated that imageability was particularly important for irregular words. These results are similar to those of Duff and Hulme (2012), who found that imageability did not matter for earlier trials but did matter for later trials, with low imageability irregular words being the hardest of all words to learn and the low imageability regular words being harder to learn than high imageability regular words. Our finding is also in line with the work of Strain, Patterson, and Seidenberg (1995) who found that imageability facilitates recognition of low-frequency irregular words and the findings of Strain and Herdman (1999), who found that reaction time for reading low imageability irregular words was much greater than reaction time for reading high imageability irregular words in adults.

According to Perfetti, “the major essential development in learning to read is the acquisition of individual word representations” (pp.154) and “a word, once acquired, may be represented strictly as a specific unit” (pp. 155). The impact of imageability over the two studies could be interpreted as supporting children (particularly those with poor word reading skills) in creating these specific word representations. The results for the posttest word reading measure for the second study also fit with this interpretation. Although we did not find a main effect for the imageability training for posttest word reading, we did find a main effect for the imageability training in predicting the number of word exposures required for mastery. Perhaps the visual and auditory supports in the imageability training serve as an added support in the word learning process, allowing students to establish stronger representations for words with fewer exposures than students in the word only condition. According to Perfetti’s theory, we would expect that after several exposures, students would no longer require such supports once they created a word specific representation. It is possible that 12 exposures were adequate for students to create a word specific representation even in the word only condition, thereby resulting in no difference between the groups at posttest.

The Role of Imageability in Transfer of Learning

We also addressed the development of the orthographic lexicon and the strength of these lexical representations through several transfer and maintenance tasks across Experiments 1 and 2. The nonwords we included in the first experiment were words that shared a rime unit with the target words but differed in onset. This task allowed us to test the students’ ability to apply their orthographic learning of the target words to these novel nonwords that shared orthographic units (i.e., rime units). Our findings from the first experiment suggested that while there was no imageability effect for transfer, there was an effect for condition, with students in the phonological analysis condition more likely to read the nonwords correctly. It was expected that students who received blending feedback would be more likely to decode these words. As discussed earlier, we speculate that vocabulary may be serving as a proxy for general intelligence in this case. A second transfer task in the both experiments was the spelling task. Perfetti (1992) also noted that as lexical knowledge increases there is increasing convergence of reading and spelling. Thus, as the quality of a lexical entry increases a single representation serves both reading and spelling. The spelling task, though underpowered, did not indicate an advantage for highly imageable words. These transfer measures are a unique contribution to the literature and offer some insight into the strength of the lexical representations after 12 exposures. It is possible that we did not find an effect of imageability on transfer to nonwords because imageability may be specific to the item. While imageability may help to strengthen the representation for a specific word, this knowledge likely does not transfer to other orthographically similar words. Despite the speculated effect that imageability has on the strength of the representation, it is still likely that the representation is not strong enough to see this kind of transfer.

The Role of Imageability in Learning Efficiency

A focus for both experiments was the efficiency of word learning and the role of child and word characteristics in the number of exposures children require for mastery. Using a criterion of four out of five exposures correct, we measured the number of exposures children required to master words. As we noted earlier, Ehri (1995) has demonstrated in the past that students at-risk for reading disabilities require more exposures to master words. Though our criterion for mastery was different than that for experiments done by Reitsma (1983) and Ehri (1995), our findings are in line with their general findings. Our studies extend previous work by exploring both child and word characteristics related to this continuous outcome. We found several factors that contribute to mastery at both the child- and word-levels. Interestingly, we found in the first experiment that when we controlled for pretest performance, low imageability regular words were easier to master than low imageability irregular words. In our second study when we intervened for imageability, we found an effect of condition, with students receiving imageability training requiring fewer exposures for mastery than students who were in the word only condition. In the second experiment, we did not find an effect for regularity on mastery, although results were approaching significance.

Lexical Involvement in Word Learning

The results of our study and others suggest that imageability is particularly important for irregular words. Irregular words present unique challenges to students because they cannot rely exclusively on their knowledge of decoding rules to access irregular words. The results of our study seem to support others that associate lexical knowledge (e.g. semantic or lexical phonology) with learning to read irregular words. The importance of lexical knowledge has been demonstrated from several areas of the literature. The significant role of imageability in these two studies is consistent with the findings from studies on connectionist models of word recognition (e.g. Harm & Seidenberg, 2004; Plaut et al., 1996), which have been more successful at reading irregular words when a semantic processor is included in the model in addition to the orthographic and phonological processors. These findings are also consistent with studies that have found that either item-specific vocabulary knowledge (Ricketts, Nation, & Bishop, 2007) or familiarity with lexical phonology (McKague, Pratt, & Johnston, 2001; Nation & Cocksey, 2009; Taylor et al., 2011) are good predictors of reading irregular words correctly. These results may provide further support for Keenan and Betjemann's (2008) speculations that item-specific semantic activation may help to "fill voids" in phonological-orthographic processing in individuals with poor mappings, such as children with reading difficulties (p. 193). Our results seem to support a developmental word-reading model in which orthographic-to-phonological pathways become at least partially dependent on lexical input, with this influence being increasingly important for irregular words (Nation & Snowling, 1998; Ricketts et al., 2007; Tumner & Chapman, 2012). We speculate that imageability is related to the lexical input in word reading, which makes high imageability words easier to read.

Intervening for Imageability

Our attempt to increase imageability through imageability training in Experiment 2 is one of several attempts to address imageability experimentally. As mentioned above, Duff and Hulme (2012) used nonwords to manipulate both phonological and semantic knowledge of their target words. They compared two conditions, one with only phonological information for the nonwords and one with both phonological and semantic information. They found no added benefit for semantic knowledge over and above the benefit of phonological information. Similarly, in our second experiment, we did not find a significant difference between the vocabulary and imageability groups. As noted earlier, the number of times students heard the words was equalized across these two conditions. It is possible that hearing the phonological representation was enough to reduce the number of exposures required for mastery. We speculate that the overall pool of words students had to choose from was reduced by the pretraining in both imageability and vocabulary, thus helping students to identify the words with fewer exposures. These findings are consistent with the findings of Wang, Nickels, Nation, and Castles (2013), who found that item-specific vocabulary knowledge was a predictor of orthographic learning only for irregular words. Although they did not focus only on the imageability, they trained on vocabulary using visual supports and thus have the same issue separating the benefits of vocabulary and imageability training as we do. It is clear that there remains a question within the literature regarding whether students only require phonological experience with words or whether visual supports and/or vocabulary definitions contribute something unique to the word learning process. Furthermore, the fact that we found an interaction favoring the imageability group for poor readers suggests that imageability supports may be beneficial only for our poorest readers. Further exploration of these questions is required.

Limitations and Future Directions

The results from this study suggest that lexical input plays a role in overall word learning and efficiency of learning. The results also suggest that imageability is a potentially malleable word feature for instruction. There were several limitations, however, that should be considered when interpreting these results. First, this study focuses on only 32 words sampled from a large corpus of regular and irregular words. We are uncertain about how representative these words are of the entire corpus and thus we should exercise caution when generalizing these results to other words. Likewise, the sampling of children for both the at-risk sample and the representative sample was limited to the scope of the study. Both samples were relatively small and the results should be interpreted with this in mind. From an instructional perspective, there are several areas of interest for future studies. First, future studies could include a condition that pairs pictures with words rather than doing imageability training prior to word exposures. The experimental design and comparisons in the present study did not allow us to answer this question but it may be beneficial for instruction. Next, a study exploring other feedback conditions is warranted. The feedback conditions in Experiment 1 may not have been the best forms of feedback and studies examining alternative feedback methods may be helpful. Finally, a study that examines irregular word reading instruction in more detail is warranted. For example, a study that encourages students to decode irregular words and look more carefully at the words at a subword level could be informative for both instruction and theory.

APPENDIX A

Our hypotheses for all research questions are listed below.

1) What role do regularity and imageability play in initial word learning for at-risk first and second graders?

   Based on previous research, we anticipated that imageability would play an important role, particularly for irregular words.

2) What role does imageability and regularity play in the efficiency of word learning mastery?

   After controlling for posttest target word performance, we hypothesized that both imageability and regularity would play a role in the number of exposures required for mastery. We anticipated the regular, highly imageable words would be the easiest to learn and low imageability, irregular words would be the hardest to learn.

3) Does maintenance of word learning differ depending on regularity and imageability?

   To answer this question we controlled for both preteset and posttest word reading. We anticipated that highly imageable words would be easier to maintain than low imageability words.

4) Does transfer to orthographically similar nonwords differ depending on regularity and imageability?

   The outcome measure for this model was posttest target nonword reading. We anticipated that the highly imageable words would have a stronger lexical representation according to Perfetti's theory and thus students would be more likely to apply their knowledge to orthographically similar nonwords.

5) After word learning, do skills transfer to spelling? Does this relationship differ depending on the regularity and/or imageability of the words?

We anticipated that regular words would be easier to spell than irregular words and highly imageable words would be easier to spell that low imageability words. This hypothesis was based on our thinking that students would have a stronger lexical representation for high imageability and regular words.

6) Does imageability training impact posttest performance on low imageability words over and above a word only and/or vocabulary training?

   We hypothesized that the visual supports provided in the imageability condition would support word learning and would result in higher posttest performance.

7) Does word regularity impact posttest performance on low imageability words?

   We hypothesized that the probability of reading a low imageability regular word would be higher than the probability of reading a low imageability irregular word.

8) Do children in the imageability condition outperform other groups on posttest spelling of the target words?

   We expected that students in the imageability condition would gain stronger lexical word representations, which they would be able to apply to spelling the words at posttest. For this reason, we expected a significant main effect for group for posttest spelling.

9) Does imageability training impact the number of exposures necessary for mastery?

   We hypothesized that the imageability training would reduce the number of exposures required for mastery. For this reason, we hypothesized that the students in the imageability condition would outperform students in the word only condition. The comparison between the vocabulary condition and imageability condition was exploratory and we did not have a firm hypothesis based on theory or previous research.

10) Does imageability training impact the number of exposures necessary for mastery differentially for students who start with poor word reading skills?

We hypothesized that imageability training would be most effective for students who started with poor word reading skills. Given that typically developing students are able to add lexical representations with more ease, we expected that they would not require the additional support of the imageability training.

APPENDIX B

Table 1B

Crossed-Random Effects Models Used in Experiments 1 & 2

Base Models	
Posttest	Level 1 (Responses$_{jik}$): $Logit\left(\pi_{jik}\right) = \lambda_{0jk}$ Level 2 (Person$_j$ & Word$_i$): $\lambda_{0jk} = \gamma_{00} + r_{01j} + r_{02i},$ $r_{01j} \sim N(0, \sigma^2{}_{r01})\&\, r_{02i} \sim N(0, \sigma^2{}_{r02})$
Mastery	Level 1 (Responses$_{ji}$): $Y_{ij} = \lambda_{0jik} + e_{ijk}$ Level 2 (Person$_j$ & Word$_i$): $\lambda_{0jk} = \gamma_{00} + r_{01j} + r_{02i},$ $r_{01j} \sim N(0, \sigma^2{}_{r01})\&\, r_{02i} \sim N(0, \sigma^2{}_{r02})\&\, e_{ijk} \sim N(0, \sigma^2{}_{e})$
Main Effects Models	
Posttest, Spelling, Maintenance (Models 1-5, 7-8)	Level 1 (Responses$_{ji}$): $Logit\left(\pi_{ji}\right) = \lambda_{0jk} + \sum_{c=1}^{C} \gamma_c P_{cij}$ Level 2 (Person$_j$ & Word$_i$): $\lambda_{0jk} = \gamma_{00} + \sum_{a=1}^{A} \gamma_a M_{ai} + \sum_{b=1}^{B} \gamma_b N_{bj} + r_{01j} + r_{02i},$ $r_{01j} \sim N(0, \sigma^2{}_{r01})\&\, r_{02i} \sim N(0, \sigma^2{}_{r02})$
Mastery (Models 6, 9)	Level 1 (Responses$_{ji}$): $Y_{ij} = \lambda_{0jk} + \sum_{c=1}^{C} \gamma_c P_{cij} + e_{ijk}$ Level 2 (Person$_j$ & Word$_i$): $\lambda_{0jk} = \gamma_{00} + \sum_{a=1}^{A} \gamma_a M_{ai} + \sum_{b=1}^{B} \gamma_b N_{bj} + r_{01j} + r_{02i},$ $r_{01j} \sim N(0, \sigma^2{}_{r01})\&\, r_{02i} \sim N(0, \sigma^2{}_{r02})\&\, e_{ijk} \sim N(0, \sigma^2{}_{e})$
Interaction Models	
Mastery (Model 10)	Level 1 (Responses$_{ji}$): $Y_{ij} = \lambda_{0jk} + \sum_{c=1}^{C} \gamma_c P_{cij} + e_{ijk}$ Level 2 (Person$_j$ & Word$_i$): $\lambda_{0jk} = \gamma_{00} + \sum_{a=1}^{A} \gamma_a M_{ai} + \sum_{b=1}^{B} \gamma_b N_{bj} + r_{01j} + r_{02i} + \gamma_d M_{1i} M_{2i} + \gamma_d M_{1i} M_{3i},$ $r_{01j} \sim N(0, \sigma^2{}_{r01})\&\, r_{02i} \sim N(0, \sigma^2{}_{r02})\&\, e_{ijk} \sim N(0, \sigma^2{}_{e})$

Note. π_{ji} = probability of a correct response from person j on word i, k = item, γ_{00} = intercept, M_a = item covariate, N_b = person covariate, P_c = person-by-item 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 2B

Research Questions and Associated Covariates

Research Question	Model	Outcome	Covariates
Experiment 1			
What role do regularity and imageability play in initial word learning for at-risk first and second graders?	1	Post	Level 1 (person x word): pretest target words Level 2 (person): Con, PA, RAN, VOC, GR Level 2 (word): Hi/Irreg, Hi/Reg, Low/Reg
Does transfer to orthographically similar nonwords differ depending on regularity and imageability?,	2	Nonwords	Level 1(person x word): post target words Level 2 (person): Con, PA, RAN, VOC, GR Level 2 (word): Hi/Irreg, Hi/Reg, Low/Reg
After word learning, do skills transfer to spelling? Does this relationship differ depending on the regularity and/or imageability of the words?	3	Spelling	Level 1(person x word): post target words Level 2 (person): Con, PA, RAN, VOC, GR Level 2 (word): Hi/Irreg, Hi/Reg, Low/Reg
Does maintenance of word learning differ depending on regularity and imageability controlling for posttest?	4	Main.	Level 1(person x word): post target words Level 2 (person): Con, PA, RAN, VOC, GR Level 2 (word): Hi/Irreg, Hi/Reg, Low/Reg
Does maintenance of word learning differ depending on regularity and imageability controlling for pretest?	5	Main.	Level 1 (person x word): pretest target words Level 2 (person): Con, PA, RAN, VOC, GR Level 2 (word): Hi/Irreg, Hi/Reg, Low/Reg
What role does imageability and regularity pay in the efficiency of word learning mastery?	6	Mastery	Level 1(person x word): post target words Level 2 (person): Con, PA, RAN, VOC, GR Level 2 (word): Hi/Irreg, Hi/Reg, Low/Reg
Experiment 2			
Does imageability training impact posttest performance on low imageability words over and above a word only and/or vocabulary training?	7	Post	Level 1 (person x word): pretest target words Level 2 (person): WR, PA, RAN, VOC, WO, VOCAB Level 2 (word): Regularity
Does word regularity impact posttest performance on low imageability words?	7	Post	Level 1 (person x word): pretest target words Level 2 (person): WR, PA, RAN, VOC, WO, VOCAB Level 2 (word): Regularity
Do children in the imageability condition outperform other on posttest spelling of the target words?	8	Spelling	Level 1(person x word): post target words Level 2 (person): WR, PA, RAN, VOC, WO, VOCAB Level 2 (word): Regularity
Does imageability training impact the number of exposures necessary for mastery?	9	Mastery	Level 1(person x word): post target words Level 2 (person): WR, PA, RAN, VOC, WO, VOCAB Level 2 (word): Regularity
Does imageability training impact the number of exposures necessary for mastery differentially for students who start with poor word reading skills?	10	Mastery	Level 1(person x word): post target words Level 1 (person x person interaction): WR* IMAG, WR*VOCAB Level 2 (person): WR, PA, RAN, VOC, WO, VOCAB Level 2 (word): Regularity

Note. Con= condition, PA=phonological awareness, RAN=rapid automatized naming, VOC=vocabulary, GR=grade, Hi=high imageability, Low=low imageability, Irreg.=irregular, Reg. = regular, WR=word reading, WO=word only, VOCAB=vocabulary training

APPENDIX C

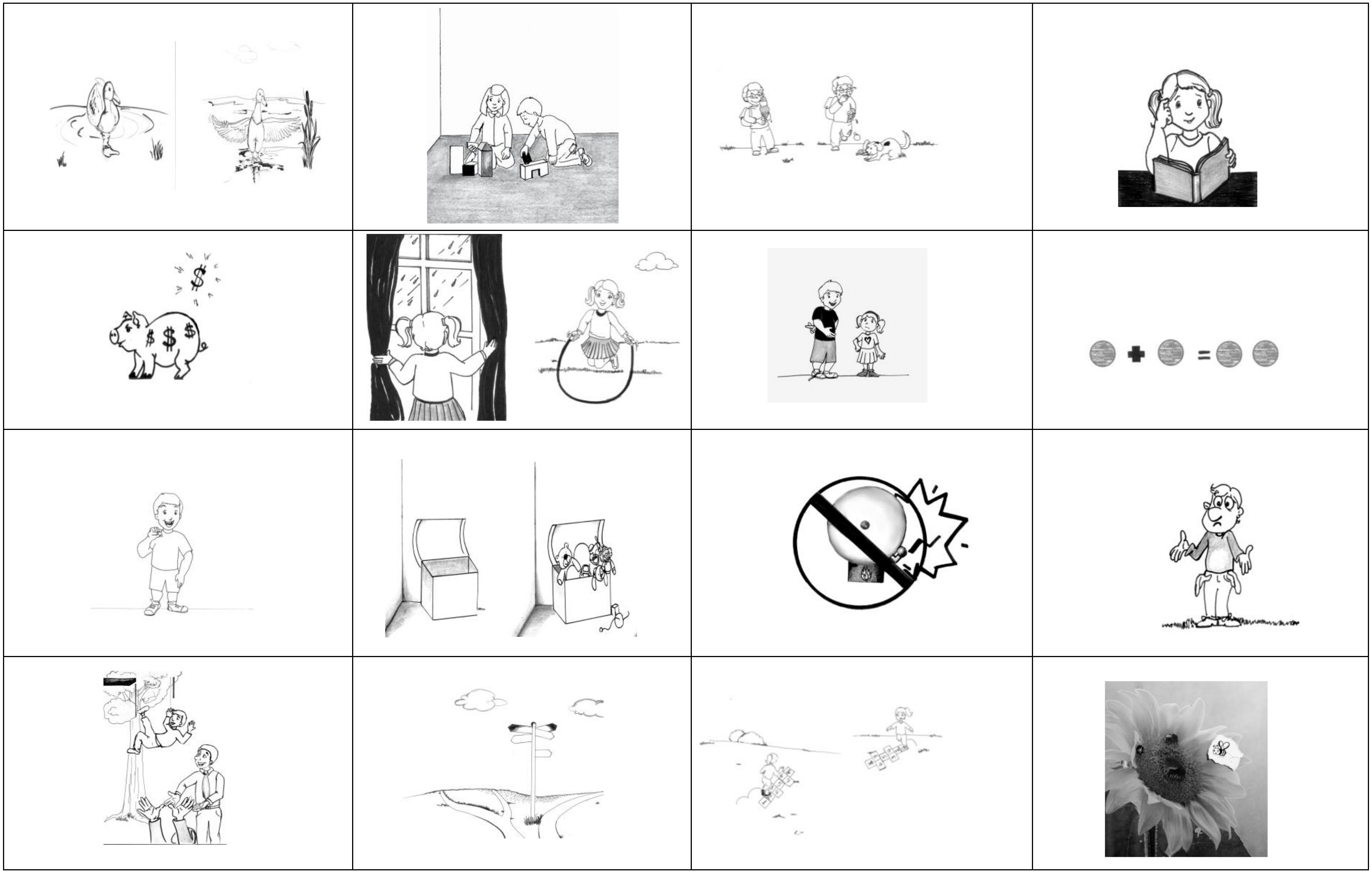

Figure 1C. Imageability training pictures.

Appendix D

The word is…	***It means…***	Read sentence	***When you hear the word…***
went	to leave	Sally **went** to the store.	**went,** remember that **went** means to leave.
trust	to have faith in	I **trust** my friends.	**trust,** remember that **trust** means to have faith in.
plus	adding together	I added two **plus** two.	**plus,** remember that **plus** means adding together.
worse	more bad	I feel **worse** than I did yesterday.	**worse,** remember that **worse** means more bad.
learn	to find out about something	We have so many things to **learn** at school.	**learn,** remember the word **learn** means to find out about something.
cost	the amount paid for something	Before you buy something, you find out the **cost**.	**cost,** remember the word **cost** means the amount paid for something.
choice	the act of choosing	You can have a **choice** between two things.	**choice,** remember the word **choice** means the act of choosing.
stuff	things that people need or use	We have lots of **stuff** in our garage.	**stuff,** remember that **stuff** means things that people need or use.
lie	to say something that is not true	I was taught to never tell a **lie**.	**lie,** remember that **lie** means to say something that is not true.
none	not any	I had lots of bananas and now I have **none.**	**none**, remember that **none** means not any.
broad	stretching far and wide	The eagle's wings are **broad.**	**broad**, remember the word **broad** means stretching far and wide.
false	not true	If something is untrue, it is **false**.	**false,** remember the word **false** means not true.
lose	to not be able to find	I hope I don't **lose** my favorite toy.	**lose**, remember that **lose** means to not be able to find.
build	to put together	Birds **build** nests for their homes.	**build,** remember that **build** means to put together.
once	one time only	Dinosaurs were **once** the largest creatures living on Earth.	**once,** remember that **once** means one time only.
real	not imaginary	I saw a **real** panda at the zoo.	**real**, remember that **real** means not imaginary.

Figure 1D. Vocabulary training definitions

REFERENCES

Adams, M. J. (1994). *Beginning to read: Thinking and learning about print* The MIT Press, Cambridge, MA.

Balota, D. A., & Ferraro, F. R. (1993). A dissociation of frequency and regularity effects in pronunciation performance across young-adults, older adults, and individuals with senile dementia of the Alzheimer-type. *Journal of Memory and Language*, *32*, 573–592 .

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models for social and behavioural research: Applications and data analysis methods*. Newbury Park, CA: Sage Publications.

Castles, A. (2006). The dual route model and the developmental dyslexias. *London Review of Education*, *4*, 49–61.

Coltheart, V., & Leahy, J. (1992). Children's and adults' reading of nonwords: Effects of regularity and consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 718–729.

Coltheart, V., Laxon, V. J., & Keating, C. (1988). Effects of word imageability and age of acquisition on children's reading. *British Journal of Psychology*, *79*, 1–12 .

Denckla, M. B., & Rudel, R. G. (1976). Rapid ‘automatized’naming (RAN): Dyslexia differentiated from other learning disabilities. *Neuropsychologia*, *14*, 471–479 .

Duff, F. J., & Hulme, C. (2012). The role of children's phonological and semantic knowledge in learning to read words. *Scientific Studies of Reading*, *16*, 504–525 .

Ehri, L. (1994). Development of the ability to read words: Update. In R. Ruddell, M. Ruddell, & H. Singer (Eds.), Theoretical models and processes of reading (4th ed. pp. 323–358). Newark, DE: International Reading Association.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18,* 116–125.

Ehri, L. C. (1997). Sight word learning in normal readers and dyslexics. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia* (pp. 163–189). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Ehri, L. C., & Saltmarsh, J. (1995). Beginning readers outperform older disabled readers in learning to read words by sight. *Reading and Writing*, *7*(3), 295-326.

Foorman, B. R., & Torgesen, J. (2001). Critical elements of classroom and small-group instruction promote reading success in all children. *Learning Disabilities Research & Practice, 16*(4), 203-212.

Fuchs, D., Fuchs, L. S., Thompson, A., Otaiba, S. A., Yen, L., Yang, N. J., ... & O'Connor, R. E. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology*, *93*(2), 251–267.

Griffin, P., Burns, M. S., & Snow, C. E. (Eds.). (1998). *Preventing reading difficulties in young children*. National Academies Press.

Griffiths, Y. M., & Snowling, M. J. (2002). Predictors of exception word and nonword reading in dyslexic children: The severity hypothesis. *Journal of Educational Psychology*, *94*(1), 34-43.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conrad, J. G. (2009). Research electronic data capture (REDCap): A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Information, 42,* 377-381.

Keenan, J. M., & Betjemann, R. S. (2007). Comprehension of single words: The role of semantics in word identification and reading disability. *Single-word reading: Behavioral and biological perspectives*, 191-210.

Kirby, R., Georgiou, G., Martinussen, R., & Parrila, R. (2010). Naming speed and reading: A review of the empirical and theoretical literature. Reading Research Quarterly, 45, 341–362 .

Kirk, R. E. (1982). *Experimental design*. John Wiley & Sons, Inc..

LaHuis, D. M., Hartman, M. J., Hakoyama, S., & Clark, P. C. (2014). Explained variance measures for multilevel models. *Organizational Research Methods*, *17*(4), 433-451.

Laing, E., & Hulme, C. (1999). Phonological and semantic processes influence beginning readers' ability to learn to read words. *Journal of Experimental Child Psychology*, *73,* 183–207.

Lambrecht Smith, S., Scott, K. A., Roberts, J., & Locke, J. L. (2008). Disabled readers' performance on tasks of phonological processing, rapid naming, and letter knowledge before and after kindergarten. *Learning Disabilities Research and Practice, 23,* 113–124.

Martin-Chang, S. L., & Levy, B. A. (2005). Fluency transfer: Differential gains in reading speed and accuracy following isolated word and context training. *Reading and Writing, 18,* 343–376 .

Martin-Chang, S., & Levy, B. (2006). Word reading fluency: A transfer appropriate processing account of fluency transfer. *Reading & Writing*, *19*, 517–542.

Martin-Chang, S. L., Levy, B. A., & O'Neil, S. (2007). Word acquisition, retention, and transfer: Findings from contextual and isolated word training. *Journal of Experimental Child Psychology, 96*, 37–56.

Metsala, J. L., Stanovich, K. E., & Brown, G. D. (1998). Regularity effects and the phonological deficit model of reading disabilities: A meta-analytic review. *Journal of Educational Psychology, 90*, 279–293.

Nation, K. & Cocksey, J. (2009). The relationship between knowing a word and reading it aloud in children's word reading development. *Journal of Experimental Child Psychology, 103*(3), 296-308.

National Center for Education Statistics (2011). *The Nation's Report Card: Reading 2011* (NCES 2012–457). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Nelson, J. R., Benner, G. J., & Gonzalez, J. (2003). Learner characteristics that influence the treatment effectiveness of early literacy interventions: A meta-analytic review. *Learning Disabilities and Practice, 18,* 255–267.

Norton, E. S. and M. Wolf (2012). "Rapid Automatized Naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities." Annual Review of Psychology 63: 427–452 .

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

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of experimental psychology*, *76*, 1–25.

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

Plaut, D. C. (1998). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science, 23*, 543–568.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: computational principles in quasi-regular domains. *Psychological review*, *103*(1), 56.

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive neuropsychology*, *10*(5), 377-500.

Priebe, S. J., Keenan, J. M., & Miller, A. C. (2012). How prior knowledge affects word identification and comprehension. *Reading and writing*, *25*, 131–149.

Rastle, K., & Coltheart, M. (1999). Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human perception and performance*, *25*(2), 482.

Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest, 2*(2), 31-74.

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

Rosner, J. (1979). *Test of Auditory Analysis Skills (TAAS)*. In: *Helping children overcome learning difficulties: A step-by-step guide for parents and teachers* (pp. 77–80). New York: Academic Therapy.

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

Seidenberg, M. S., Waters, G. S., Barnes, M. A., & Tanenhaus, M. K. (1984). When does irregular spelling or pronunciation influence word recognition?. *Journal of Verbal Learning and Verbal Behavior, 23*, 383–404.

Sereno, S. C., & Rayner, K. (2000). Spelling-sound regularity effects on eye fixations in reading. *Perception & Psychophysics, 62*, 402–409.

Share, D. L. (2008). Orthographic learning, phonological recoding, and self-teaching. In R. V. Kail (Ed.), *Advances in child development and behavior* (pp. 31–82). San Diego, CA: Elsevier.

Solity, J., & Vousden, J. (2009). Real books vs reading schemes: a new perspective from instructional psychology. *Educational Psychology*, *29*, 469–511.

Spaai, G. G., Ellermann, H. H., & Reitsma, P. (1991). Effects of segmented and whole-word sound feedback on learning to read single words. *Journal of Educational Research, 84*, 204–213.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading research quarterly*, 32-71.

Stanovich, K. E., & Bauer, D. W. (1978). Experiments on the spelling-to-sound regularity effect in word recognition. *Memory & Cognition, 6*, 410–415.

Steacy, L. M., Kearns, D. K, Gilbert, J. K., Compton, D. L., Cho, E., Lindstrom, E. R., & Collins, A. A. (submitted, 2014). Exploring individual differences in irregular word recognition among children with early-emerging and late-emerging word reading difficulty. *Journal of Educational Psychology*.

Strain, E., Patterson, K., & Seidenberg, M. S. (1995). Semantic effects in single-word naming. Journal of Experimental Psychology: Learning, Memory, and Cognition, 21(5), 1140.

Strain, E., & Herdman, C. M. (1999). Imageability effects in word naming: an individual differences analysis. Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 53, 347–359.

Taylor, J. S. H., Plunkett, K., & Nation, K. (2011). The influence of consistency, frequency, and semantics on learning to read: An artificial orthography paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(1), 60.

Torgesen, J. K., Wagner, R., & Rashotte, C. (2012). *Test of Word Reading Efficiency 2*. Austin, TX: Pro-Ed.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second-to fifth-grade children. *Scientific studies of reading*, *1*(2), 161–185.

van Daal, V. P., & Reitsma, P. (1990). Effects of independent word practice with segmented and whole-word sound feedback in disabled readers. *Journal of Research in Reading*, 13, 133–48.

Van den Noortgate, W., De Boeck, P., & Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics*, *28*(4), 369-386.

Vousden, J. I. (2008). Units of English spelling- to- sound mapping: a rational approach to reading instruction. *Applied Cognitive Psychology*, *22*, 247–272.

Wagner, R. K., Torgesen, J. K., Rashotte, C., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing 2*. Austin, TX: Pro-Ed.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J.,& Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology*, *33*, 468–479.

Walker, I., & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1256–1271.

Walton, P. D., & Walton, L. M. (2002). Beginning reading by teaching in rime analogy: Effects on phonological skills, letter-sound knowledge, working memory, and word-reading strategies. *Scientific Studies of Reading*, 6, 79–115.

Wang, H. C., Nickels, L., Nation, K., & Castles, A. (2013). Predictors of orthographic learning of regular and irregular words. *Scientific Studies of Reading*, *17*, 369–384.

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading: Time-course and decision criteria. *Memory & Cognition*, *13*, 557–572.

Wechsler, D. (2009). *Wechsler Individual Achievement Test–Third Edition.* San Antonio, TX: The Psychological Corporation.

Wolf, M., Bowers, P. G., & Biddle, K. (2000). Naming-speed processes, timing, and reading: A conceptual review. *Journal of Learning Disabilities, 33*, 397–407.

Woodcock, R. W. (2011). *Woodcock reading mastery tests: WRMT-III. Manual*. Pearson.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement.* Rolling Meadows, IL: Riverside Publishing.

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