Instruction on sight words, or words read from memory without the need for decoding, are a common focus for beginning readers. However, young readers often have difficulty reading previously learned sight words correctly across different contexts. In this study, the author compared sight word instruction using incremental rehearsal (IR) to teach single sight words in isolation or embedded within contextual phrases to determine whether the addition of contextual information may aid students in sight word acquisition. A modified alternating treatment design was used to compare the effects of teaching sight words in isolation or within contextual phrases to 4 first-grade students. Sight word learning was assessed using both oral reading data and eye movement data. Results indicated that students did not benefit more from teaching of sight words in contextual phrases compared to sight words taught in isolation.

INDEX WORDS: Incremental rehearsal, sight words, context, eye movements
THE EFFECT OF CONTEXT IN SIGHT WORD ACQUISITION OF YOUNG READERS

by

PATRICK ROLAND MORIN

B.S., The University of Tennessee at Chattanooga, 2012

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2015
THE EFFECT OF CONTEXT IN SIGHT WORD ACQUISITION OF YOUNG READERS

by

PATRICK ROLAND MORIN

Major Professor: Scott Ardoin
Committee: Amy Reschly
            Kristin Sayeski

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
August 2015
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2 METHOD</td>
<td>9</td>
</tr>
<tr>
<td>3 DATA ANALYSIS PROCEDURES</td>
<td>20</td>
</tr>
<tr>
<td>4 RESULTS</td>
<td>21</td>
</tr>
<tr>
<td>5 DISCUSSION</td>
<td>24</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>28</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Mean Percentage of Words Read Correctly During Assessment Sessions in IRC and IRI Conditions .................................................................31

Table 2: Mean Percentage of Words Read Correctly Across Unknown Intervention, Unknown Control, and Known Control Target Words ........................................................................32

Table 3: Mean Percentage of Words Read Correctly Across Semi-Known and Unknown Intervention Words ........................................................................................................33

Table 4: Mean Total Dwell Time Across Semi-Known and Unknown Intervention Target Words ..........................................................................................34
LIST OF FIGURES

Page

Figure 1: Number of words read correctly within each assessment session, comparing performance in context (IRC) and isolation (IRI) conditions..................................................35

Figure 2: Percentage of words read correctly by participants on isolation (IRI), context (IRC), known control and unknown control assessments. .................................................................36

Figure 3: Mean total dwell time across intervention words within context and isolation assessments. ..........................................................................................................................37

Figure 4: Mean total dwell time across target words within intervention, known control, and unknown control assessments. "Intervention" results refer to previously unknown intervention words whereas "semi-known" refer to semi-known intervention words......38
CHAPTER 1
INTRODUCTION

First-grade achievement scores in reading are consistently among the most predictive measures of reading outcomes in secondary school, even when controlling for cognitive ability (Cunningham & Stanovich, 1997). As a result, much emphasis is placed on promoting basic decoding skills of young readers, recognizing that such skills form the foundation of adequate reading proficiency (Griffin, Burns, & Snow, 1998). However, many beginning readers who can decode words accurately nonetheless lack the automaticity required for adequate comprehension. These readers demonstrate limited vocabularies of "sight words," or words that are read frequently enough to allow them to be read from memory (i.e., without the need for decoding; Ehri, 2005). As students gain exposure to and practice with a given sight word, their reading behavior comes under direct stimulus control of the entire word (i.e., the full configuration of letters comprising the word), rather than individual letters or "chunks" that are decoded separately (Kupzyk, Daly, & Andersen, 2011; MacQuarrie, Tucker, Burns, & Hartman, 2002). With this degree of automaticity, they can devote less effort to decoding text and focus instead on word meaning and contextual information (LaBerge & Samuels, 1974).

Sight word interventions focus on increasing students' total sight word vocabulary to improve automaticity and increase the extent to which they can generalize reading to other contexts (i.e., read the same sight words embedded in different passages). One frequently employed method for training sight words is to present words to students using flash-cards. Unfortunately, many traditional flash-card methods are limited by suboptimal student growth,
student disengagement with the task, and little modeling and corrective feedback (Kupzyk et al., 2011). Fortunately, recent evidence-based flash-card interventions overcome these weaknesses by involving the presentation of stimulus items, student responses, and instructor feedback. In this manner, a three-term-contingency (presentation, response, consequence) is part of each learning trial, and strengthens the student’s future responding to each stimulus item (Catania, 2007). These methods allow the instructor to control the learning process by delivering prompts, providing reinforcement for accurate responding, and giving consistent corrective feedback (Kupzyk et al., 2011).

Another notable improvement in flash-card interventions has involved the interspersal of previously known stimulus items among unknown stimulus items. In a study comparing the efficacy of three flash-card interventions in teaching 25 third-grade and 26 fourth-grade students words from the Esperanto International Language, MacQuarrie et al. (2002) found that interspersal improved students’ acquisition and retention of sight words more than procedures simply involving the presentation of unknown stimuli. Specifically, when compared to Drill Sandwich (DS) and traditional flash-card methods, Incremental Rehearsal (IR) – described in more detail below – led to significantly higher rates of retention than either of the other techniques.

The IR method provides students with frequent opportunities to practice both known and unknown items, resulting in more practice per word than many traditional flash-card interventions. Using IR procedures, an unknown item is presented, followed by the presentation of a known item. The unknown item is presented again, followed by the first known item, then the next known item. The instructor continues to add additional known items until all known items have been presented. Next, the last known item is removed from the order and the first
unknown item administered is placed into the first known position. This process continues with the remaining unknown items until, upon completion, all unknown items have been presented and worked into the stack of known items. This method gives a learner the benefit of practicing both previously known and unknown items many times over the course of an intervention session, with a greater number of practice opportunities than other flash-card drill methods (MacQuarrie et al., 2002) The benefits of IR have been demonstrated repeatedly with different populations and stimuli (Coddin, Archer, & Connell, 2010; Volpe, Burns, DuBois, & Zaslofsky, 2011). For instance, Burns (2005) studied the effects of utilizing IR to effectively teach single-digit multiplication facts to third-grade students identified as learning disabled in math computation.

Teaching in Context

Although extensive research exists supporting the efficacy of sight word instruction, many students have difficulties recognizing previously taught sight words within text even after correctly reading words during sight word based flash-card interventions (Tan & Nicholson, 1997). To address this problem, practitioners have developed sight word phrase cards, which present target sight words within simple sentences or short contextual phrases (e.g., "The cat is big!", "he had to"). For example, several websites (e.g., http://www.theschoolbell.com/Links/Dolch/Directions/phrases.html, https://www.teacherspayteachers.com/Product/Dolch-Word-Sight-Words-Fluency-Phrase-Cards-167074) provide practitioners with the opportunity to purchase or create these cards which present Dolch sight words within short contextual phrases. Unfortunately, research examining the efficacy of these materials does not currently exist. Furthermore, researchers have yet to
develop and evaluate alternative procedures that might better promote generalization of gains from traditional flash-card intervention sessions.

Although researchers have yet to determine how best to promote generalization of sight word reading across contexts (i.e., different texts), some previous studies have examined the effects of teaching sight words in isolation versus in context (i.e., within simple phrases). Of note, however, none of these studies have investigated how intervention methods involving contextual presentation impact the accuracy of students' sight word reading. However, given that they may provide preliminary insight into the effects of such instructional methods, three relevant studies are described here briefly.

In a study examining three intervention procedures for teaching students the meaning of target words, Tan and Nicholson (1997) compared: (a) a control condition which involved only an oral presentation of the meaning of the words, (b) a single-word condition involving presentation and training of single-word stimuli, and (c) a phrase-training condition that involved presentation of phrases with an imbedded target word. If students could not pronounce target words in the single-word condition, the student was shown a two-word phrase including the target word (i.e., "lemonade drink"). These context clues were, however, given only until the student understood the meaning of the word, then students were expected to read the word presented in isolation. After students learned to pronounce each target word within 1 s, the student was asked to read a randomly ordered list of target words, then a passage including the target words, and finally each student responded to 12 comprehension questions regarding the passage. Students in the single-word and phrase-training conditions substantially outperformed the control students in measures of comprehension. Furthermore, students in the control condition were also less accurate in passage reading, even with words they read correctly in the
word list assessment. These findings suggest that students in the phrase- and single-word-training groups experienced significant automaticity benefits from the sight word training for both target words and non-target words. Unfortunately, analyses were not conducted to examine differences in students' word reading accuracy across conditions.

López, Thompson, and Walker-Dalhouse (2011) conducted one of the few studies examining the use of contextual information in sight word reading of young students. As part of the study, they examined the developmental trajectory of fluency rates among three differentially-skilled reading groups reading stimuli presented within contextual and isolation presentations. The authors utilized sight word lists as a measure of reading in isolation and short paragraphs as a measure of reading within context. Results indicated that skilled readers utilized context consistently and read more proficiently with contextual information than without contextual information. Similarly, by the end of the first-grade year, average readers performed moderately better with context than isolation. Less-skilled readers did not, however, benefit from the use of context, likely due to the inability to clearly read and utilize the contextual information. These findings lend further evidence to the possibility that young readers may benefit from contextual presentation of sight words during learning trials, though students may benefit differentially based on their ability to adequately read and utilize the surrounding text.

In one of the only studies measuring the effectiveness of sight word phrase cards, Cates and Rhymer (2006) explored the effects of an explicit timing intervention implemented with four second- and third-grade students. In both the baseline and experimental conditions, students received identical sight word phrase card intervention, modeling, and feedback for several minutes each day. However, during the experimental condition, students also were shown a stopwatch during intervention and were told that the teacher was recording the amount of time
required to correctly respond. Although the explicit timing condition resulted in a greater number of accurate responses, the baseline condition also resulted in significant increases in accurate response rate. These results provide some evidence that sight word phrase interventions may be effective in increasing sight word fluency.

Eye Tracking

One possible reason for the lack of studies comparing the effects of interventions training sight words in context versus isolation is the difficulty in evaluating effects specific to individual words when presented in phrases/sentences/context. Recent improvements in technology, however, have allowed for measurement of the time required by students to read individual words presented within a sentence. For example, Ardoin, Binder, Zawoyski, Foster, and Blevins (2013) utilized eye tracking technology to evaluate the effects of fluency-based interventions on specific target words. In this study, students repeatedly practiced high- and low-frequency target words either in a single passage or across multiple passages. Intervention effects were evaluated using a novel passage containing the same target words. By employing eye tracking procedures, the researchers were able to measure student performance specific to the target words practiced in the intervention passage(s) and demonstrate that intervention effects were relegated only to low-frequency target words. Results indicated no differences in the total dwell time (TDT) on high-frequency words read by intervention students and control students. Students in the two intervention conditions, however, fixated on low-frequency words which they had previously practiced for significantly less time than students in the control condition. Similarly, research by Foster, Ardoin, and Binder (2013) and Zawoyski, Ardoin, and Binder (2014) indicates that students’ TDT on low-frequency target words decreases significantly in response to repeated
practice of a passage but continues to exceed that on high-frequency words during students' first reading of the passage.

Although readers generally perceive their reading behavior to be fluid and smooth across words, reading behaviors vary depending on the frequency of words and the level of complexity of the text. Eye movement data can thus provide valuable information regarding the level of difficulty students are encountering when reading materials. In analyses of eye movement data, a number of eye movement parameters and reading behaviors provide information to researchers as to student reading patterns (Rayner, 1998). One measure frequently employed when evaluating reading behavior is fixations. Fixations occur between eye movements and are characterized by the eyes remaining relatively still for approximately 200-300 ms. The sum of all fixation durations on a target stimulus before the eye moves to the next stimulus is known as gaze duration. The total amount of time in milliseconds a student spends gazing or fixating on a stimulus is known as total dwell time (TDT).

The Present Study

The purpose of the current study was to compare the generalization effects of teaching sight words in isolation and within context. In order to make this comparison, four students were taught sight words using a modified version of the IR method introduced by MacQuarrie et al. (2002) to teach sight words either in isolation or within context. Generalization effects were then evaluated by having students read the newly learned sight words in novel sentences. In an extension of research by MacQuarrie et al. (2002), eye tracking procedures were used to measure students’ total time reading the sight words after receiving each of the modified IR sight word interventions. Four students read the sentences aloud while having their eye movements recorded, comparing the generalization effects of sight words taught in isolation (Incremental
Rehearsal Isolation [IRI]) versus sight words taught within context (Incremental Rehearsal Context [IRC]). Given the established effectiveness of IR for teaching sight word acquisition, the current study utilized this methodology as the basis of the isolation and contextual conditions.
CHAPTER 2

METHOD

Participants and Setting

Participants were four first-grade students (ages 6–7 years) who were identified by their teachers as students who could potentially benefit from sight word intervention. Participants included Ted (Black male), Paula (White female), Dennis (Black male) and Colin (White male). All students attended a public elementary school in the southeastern United States. The school was a Title 1 elementary school serving approximately 500 students. Prior to or during participation in the study, the students did not have any special education identification or receive any supplemental reading instruction from their school. Intervention sessions were conducted in an empty classroom or in the participants’ classroom hallway. Assessment sessions were all conducted in an empty classroom located near the students' home room.

Materials

Screening. Students were administered lists of sight words from the kindergarten, first-, second-, third-, and fourth-grade Dolch and Fry word lists (Dolch, 1936; Fry, 1980). These sight word lists were administered in order to identify sets of known, semi-known, and unknown words for each participant. For one participant, Ted, an insufficient number of unknown words were identified; thus, two additional sight word lists of similar difficulty were administered. Sight words used during the screening stage were printed in black Comic Sans font and arranged in columns on a single 8.5-in x 11-in sheet of paper.
**Intervention stimuli.** Participants' designated intervention words or sentences containing intervention words were printed in black 44-point Calibri font on 3-in x 5-in index cards. The back of each card contained the intervention word and the ordering label (e.g. card A, card seven, etc.) for use by the administrator when providing error correction. Flash-cards were created each week with participants' designated sight words or sentences according to the condition being implemented for the three days of intervention. Each participant's known, unknown, and semi-known intervention words were randomly distributed across the nine weeks of intervention. For each week of intervention, each participant was assigned a set of nine unknown, seven known, and two semi-known stimulus items.

**Isolation condition stimuli.** Sight words were printed in lowercase 44-point Calibri font in the center of an unlined white 3-in x 5-in index card. On the back side of each card, the sight word and ordering designation were written on the top line for administrator reference.

**Context condition stimuli.** In addition to identifying the sets of known, unknown and semi-known intervention words, the author identified a second set of known words (hereafter referred to as sentence known words) during the screening process. These sentence known words were selected for the purpose of composition of the sentences in which the intervention words were embedded as part of the IRC condition. Sentence creation guidelines dictated that each sentence consist of the intervention word, along with 3–4 words chosen from the sentence known word pool in order to create a logical, grade-appropriate sentence (e.g., "You are early."); "I jump near you.").

**Assessment.** Data were collected utilizing the SR Research Eyelink 1000 eye tracking camera, which has a sampling rate of 100Hz. The Eyelink 1000 allowed for the capturing of participants' pupil location as they read from a 19-in. ViewSonic VG930m monitor. Before the
start of each assessment session, calibration and validation were completed using a nine-point grid. Between each sentence being displayed on the screen, validation was completed using a single point at the start of each sentence. Participants were given a Microsoft Sidewinder Plug and Play game pad in order to inform experimenters when they were finished reading an assessment sentence.

Assessment sentences were created each week based upon each participant's semi-known and unknown intervention words presented during that week’s intervention sessions. These sentences were novel sentences created utilizing each participant's collection of sentence known words, along with the intervention word being assessed (e.g., "It is near you"). Unlike during intervention sessions, intervention words were not underlined within the sentences presented to students during assessment sessions. Participants read the novel sentences from a computer monitor out loud while having their eye movements recorded by the eye tracking camera.

**Dependent Variables**

Multiple dependent variables were measured. First, similar to in previous sight word research, oral reading data were collected regarding the total number of target words read correctly (WRC) during each assessment session. Words were scored as correct if the participant read the word accurately within 3 s of presentation of the sight word. Words were scored as incorrect if the participant did not respond within 3 s or did not pronounce the word correctly.

The second dependent variable measured was participants' total dwell time (TDT) on each intervention word during each assessment session. TDT was collected using the Eyelink 1000 eye tracking camera, indicating the cumulative number of ms during which the eye was fixated on a target stimulus (including the duration of the first fixation, duration of fixations prior to moving to another word, as well as the duration of any fixations upon returning to/rereading the
target word). Eye movement data were summarized as the average TDT across individual sets of intervention words. Thus, TDT was calculated and averaged across the seven unknown and two semi-known intervention words presented during each assessment.

**Design**

A modified alternating-treatment design was used to assess the effectiveness of the two flashcard intervention conditions, IRI and IRC. In order to randomly assign experimental conditions to participants' intervention weeks, researchers flipped a two-sided coin to decide condition order in groups of two. That is, researchers designated "heads" for one condition and "tails" for the second. The first of the two weeks was decided by the coin flip, with the second of the two weeks designated as the alternative condition. This procedure resulted in intervention sequences of IRI, IRC, IRI, IRI, IRC, IRI, IRI for two participants and sequences of IRC, IRI, IRI, IRC, IRI, IRI, IRC and IRC, IRI, IRI, IRC, IRI, IRC, IRC, IRC, IRI, IRC for the two remaining participants. Each phase included three intervention sessions that were conducted three days per week, followed by an assessment session on the fourth day of each week.

**Procedures**

**Screening.** Before intervention sessions began, an initial screening was conducted to identify participants' pools of known, unknown and semi-known intervention words as well as their sentence known words. Lists of sight words from the kindergarten, first-, second-, third-, and fourth-grade Dolch and Fry word lists were administered to students in order to identify potential intervention words. Sessions were conducted over the course of two days with each participant. An experimenter presented word lists to each participant, one sheet at a time, and

---

1 Due to an error in distribution of conditions for Ted and Dennis, the ordering of these participants' intervention did not follow the specified original pattern, resulting in 5 weeks of IRI intervention and 3 weeks of IRC intervention.
asked the participant to read down the list and to give their best effort to correctly read each word. Words were scored correct if read correctly within 3 s, semi-correct if read correctly within 5 s, and incorrect if read incorrectly or read correctly after 5 s elapsed. On the second day of screening, participants were asked to read the same lists of words in the same manner as on the first day of screening.

Utilizing the two days of screening data, words were designated as known if the participant read them correctly on both occasions and semi-known if they read the word semi-correctly on both occasions or correctly once and semi-correctly once. Words were designated as unknown if they were read incorrectly both times or incorrectly once and semi-correctly once. A second round of screening was performed several weeks into the intervention period in order to identify more intervention words. Unknown words were randomly assigned to each intervention condition as needed in order to maintain the pattern of nine unknown intervention words per participant per week. Similarly, known and semi-known intervention words were assigned to each condition to maintain the pattern of seven known and two semi-known sight words per participant per week. A total of 27 control words also were administered to each participant periodically throughout assessments. Control words were identified as unknown words before the start of intervention and were unseen until presented to participants during assessment. Words utilized for known control assessments consisted of sight words randomly chosen from each student's pool of known intervention words.

Upon completion of screener administrations, the resulting intervention words were sorted into known, unknown, and semi-known groups within a Microsoft Excel document. Items were then randomized within categories and assigned to an intervention week. At this time, a
number of known words were also separated to be used as sentence known words throughout the intervention. Unknown items to be used in control assessments also were set aside for later use.

**Intervention.** Each week students were exposed to either three IRI or IRC intervention sessions. IRI or IRC flash-cards were presented by experimenters using a modified version of the IR procedure employed by MacQuarrie et al. (2002). The IR technique was modified in several substantial ways for the purposes of the current study. First, as opposed to employing three unknown and nine known intervention words per session, two of the known intervention words were replaced with two semi-known sight words. Thus, during each session students practiced three unknown intervention words, two semi-known intervention words and seven known intervention words. Second, in the current study, the first known item was removed after completion of a learning trial, in contrast with MacQuarrie et al.'s procedures, which removed the last known item in the sequence. This modification allowed for more opportunities for students to practice known intervention words before removal from the intervention sequence. Third, in order to reduce sequence effects (i.e., memorization/facilitation due to presenting cards in the same order after the addition of each new card), the cards were shuffled before presentation each time a new known item was added. IRI sessions lasted approximately 15 min and IRC sessions typically lasted around 30 min. If all three intervention sessions could not be conducted within a single week, assessment was not administered for that week's words.

Intervention sessions consisted of three trials across which three unknown, two semi-known and seven known intervention words were presented. Trials began with the experimenter presenting and modeling the first unknown intervention word (U1), then presenting the first known intervention word (K1). Next, the experimenter presented U1, followed by K1, then the second known intervention word (K2). With the addition of each card to the presentation pile,
experimenters shuffled flash-cards in order to present each block in a random order. This process and order was repeated with K3, K4, K5, K6, K7, K8, and K9 until U1 was presented along with all known intervention words. Given this pattern, U1 was presented nine times before the remaining two unknown intervention words were presented. Importantly, K7 and K9 were designated positions for the two semi-known intervention words presented each week. This allowed for the two semi-known intervention words to remain in the participant's stack of stimulus items through the entire three days of intervention.

Upon the completion of trial 1, U1 was moved into the K1 position and K1 was removed from the sequence. Then, in trial 2 the same sequence was repeated beginning with presentation and modeling of U2. After U2 was presented with all known intervention words (K1-K9), U2 was moved into the K1 position, and trial 3 began with the presentation of U3. Trial 3 was administered following the same pattern as the previous trials, and completion of trial 3 ended with U1-U3 in the previous K1-K3 positions, with the remaining K4-K9 in their original positions.

Utilizing this framework, participants were introduced to three previously unknown words, two semi-known words and seven previously known sight words during each intervention session. Upon completion of each week's three intervention sessions all seven known and two semi-known intervention words were replaced with previously unknown intervention words in the order outlined above. That is, over the course of each intervention session, three known intervention words were removed from the stack and three previously unknown intervention words were incorporated into the stack of stimulus items. This procedure resulted in each unknown intervention word being presented a total of nine times before being sorted into the stack of stimulus items as a known item. Upon conclusion of the third intervention session each
week, all nine originally unknown intervention words had been worked into the stack as known intervention words.

During both intervention conditions (IRI and IRC), experimenters followed strict protocol upon presenting flash-cards to students. With the first presentation of an unknown intervention word, experimenters showed the card, modeled the word or sentence ("The word is---, say the word") and provided corrective feedback until the participant read the word correctly. For both known intervention words and unknown intervention words, the flash-card was presented and participants were prompted by the examiner if they did not immediately respond ("say the word", "say the sentence"). If participants pronounced a word incorrectly, the experimenter provided immediate error correction and asked the participant to repeat the word ("no, the word is---, say the word"). If a participant did not respond to a word within 3 s of presentation, the experimenter counted the word as incorrect and provided error correction to the participant ("the word is---, say the word").

**Incremental rehearsal in isolation.** Known (K) and Unknown (U) intervention words were systematically interspersed at a 9:1 ratio during each trial. After an unknown intervention word was presented with all nine known intervention words, the unknown intervention word (U1) was treated as a known intervention word (K1) and a new unknown intervention word (U2) was introduced.

**Incremental rehearsal in context.** Known (K) and Unknown (U) intervention words were systematically interspersed across trials at a 9:1 ratio within simple 3-5 word sentences. After an unknown intervention word was presented with all nine known intervention words, the unknown intervention word (U1) became treated as a known intervention word (K1) and a new unknown intervention word (U2) was introduced.
**Assessment.** An assessment session was conducted once per week following three days of intervention. All words assessed were trained over the previous three intervention sessions preceding the assessment session. On the assessment day each week, participants were presented with a list of 11 short novel sentences of 3-5 words, with each sentence containing one of the nine unknown intervention words or two semi-known intervention words trained throughout the week. Participants read these sentences aloud from a computer monitor while their eye movements were recorded by the eye tracking camera. Participants were administered the assessment sentences utilizing an Eyelink Experiment Builder program which presented each sentence in black text in the middle of a white screen. Before each sentence was presented, a one-point drift correction check at the screen position of the start of the sentence was used to ensure that the eye tracking camera would accurately capture each participant's eye movements while reading. Participants then read the sentence orally, pressing a button on a game pad to progress to the next sentence.

Assessment sentences were created using the same parameters utilized in creating intervention sentences in the context condition. However, intervention words were not underlined in the assessment sentences. The intervention word was used to create a novel 3-5 word sentence consisting of the intervention word and known sentence words.

**Interobserver agreement and treatment integrity.** All intervention sessions were audiotaped and reviewed to assess interobserver agreement and treatment integrity. Interobserver agreement and treatment integrity were calculated for one-third of each participant's total number of trials over the course of the intervention period, with equal numbers of IRI and IRC sessions being reviewed. To assess treatment integrity, an independent observer indicated whether or not the experimenter in the recording implemented each step correctly,
including word presentation order, modeling, prompting, error correction, and pre-session preparations. While listening to the recording, the independent observer also assessed interobserver agreement by scoring responses as correct/incorrect and then comparing his/her score sheet to that of the experimenter. Agreement was defined as the experimenter and the independent observer both marking a word as being read correctly or incorrectly by the participant in the audio recording. Interobserver agreement was calculated by dividing the total number of agreements by the total number of possible agreements and multiplying by 100. The overall mean agreement was 99% (range=94% to 100% across all sessions sampled). Treatment integrity was calculated by dividing the total number of correctly implemented steps by the total number of possible steps and multiplying by 100. The overall mean percentage of steps implemented correctly across sessions sampled, averaged across all students, was 98% (range=67% to 100%). Procedural integrity was low (67%) for one session due to the magnification of a small number of missed steps, given the relatively small number of total steps included in the calculation of procedural integrity.

Audio recordings of assessment sessions were reviewed by two independent observers in order to establish whether participants correctly read words out loud while reading sentences during eye tracking assessments. Observers were provided with audio recordings of each assessment session, along with lists of assessment sentences that were read. Observers listened to each assessment sentence and marked whether the intervention word was pronounced correctly or incorrectly, then listed any other errors committed by the participant while reading each sentence. Interobserver agreement for assessment sessions was determined for each intervention word included in an assessment session. Agreement was calculated in the same manner as during intervention sessions, with the total number of agreements divided by the total
number of words assessed then multiplied by 100. Interobserver agreement of assessment sessions was 100% across all participants.
CHAPTER 3
DATA ANALYSIS PROCEDURES

Oral Reading Data

Descriptive statistics were calculated to assess the overall level and trend of the WRC data across assessment sessions. Statistics of interest included: (a) the percentage of total unknown and semi-known intervention words read correctly within and across assessment sessions and (b) the percentage of total intervention words read correctly between IRI and IRC conditions. The same procedures were used in evaluating students’ performance on assessments of known and unknown control words.

Eye Movement Data

Visual analysis procedures were used to assess the level and variability of each participant's TDT across assessment sessions. Control assessments were also used to compare average levels of TDT during control word assessments versus average levels of TDT during intervention word assessments.
CHAPTER 4

RESULTS

Oral Reading Data

**Context versus isolation.** Figure 1 illustrates each participant's total number of words read correctly (WRC) on assessment sessions following intervention in IRC and IRI conditions. Students' reading accuracy on unknown intervention words did not differ as a function of intervention condition (IRI vs. IRC) (Table 1). Visual analysis of the oral reading data indicated substantial overlap of data points across conditions for all participants (Figure 1). One participant, Dennis (Figure 1, top right), had greater levels of accuracy for the IRI condition (76%) than the IRC condition (59%); however, this was likely a function of the increasing trend in Dennis' performance and the fact that the latter two sessions were IRI conditions.

Figure 2 illustrates each participant's performance measured by words read correctly (WRC) across unknown intervention words, known control words and unknown control words. Although visual analyses suggest no differences in students' accuracy between intervention conditions, these data do indicate that the interventions resulted in gains, as participants' accuracy on intervention words exceeded their accuracy on unknown control words. All participants had greater percentages of WRC in each IRI and IRC assessment compared to their number of WRC in both unknown control assessments (Table 2 and Figure 2). Visual analysis of these data indicate there were no overlapping data points between participants' IRC and IRI sessions and unknown control assessments. One participant, Ted (Figure 2, bottom right), was administered two known control assessments and only one unknown control assessment;
however, he also read more words correctly in IRI and IRC assessments than in the unknown control assessment.

**Semi-known versus unknown.** Differences in performance were noted in students’ reading accuracy as a function of whether words were initially identified as unknown versus semi-known intervention words. All students performed more accurately on weekly assessments when reading semi-known intervention words compared to when reading unknown intervention words (Table 3).

**Eye Movement Data.**

**Context versus isolation.** Figure 3 illustrates each participant's mean TDT in ms on assessments across semi-known intervention words and unknown intervention words. Results of eye tracking assessments similarly indicated no consistent difference in results obtained as a function of the IRI and IRC conditions. Visual analysis of TDT data indicated substantial overlap of data points across conditions for all participants, indicating that one condition did not result in greater intervention effects than did the other. Consistent with the oral reading accuracy data, visual analyses of differences in students' TDT on intervention and unknown control words suggest that the interventions did improve students' reading of the intervention words. Measures of participants' mean TDT on known control and unknown control assessments substantially differed for three students (Colin, Dennis, Paula) (Figure 3), with lower average TDT on known control assessments than unknown control assessments. Analysis of Ted's control assessments showed lower rates of TDT on unknown control words than known control words (Figure 3, bottom right). However, Ted was administered only two known control assessments and one unknown control assessment, thus limiting the number of words on which these data were collected.
**Semi-known versus unknown.** For two out of four participants (Ted and Colin), results of eye tracking assessments indicated substantially lower levels of TDT for semi-known intervention words than for unknown intervention words (Table 4). For Ted, all data points for TDT on semi-known words were lower than those for TDT on unknown words (Figure 4, bottom right). For Colin, eight of nine semi-known TDT data points were lower than unknown TDT data points (Figure 4, top left). Results were inconclusive for Paula and Dennis, who showed similar rates of TDT for both semi-known and unknown intervention words.
CHAPTER 5
DISCUSSION

In this study, the author compared the learning effects of teaching sight words to young readers with or without accompanying context. Utilizing eye movement measures and oral reading data made it possible to transcend surface-level measures of reading accuracy and further explore reading behaviors underlying sight word reading. Intervention utilizing IR methodology was adapted to deliver sight word instruction to participants in context (IRC) or isolation (IRI) conditions for three sessions per week, and assessment of sight word learning was completed using both oral reading data collection and eye tracking technology upon completion of intervention each week. Participants learned to read previously unknown sight words correctly at similar levels across both the IRC and IRI conditions. They did not benefit from the presence of contextual information as predicted. Given the established effectiveness of IR as a method of sight word intervention, the growth of students across both conditions is consistent with previous research (MacQuarrie et al., 2002). However, the perceived lack of improvement with the addition of contextual information is not consistent with previous research predicting that such information would provide added benefit over single-word stimulus presentations (López et al., 2011). Given that students received IRC intervention on sight words embedded within one set of stimuli but were assessed using those words within novel sentences, these findings lend further support to the idea that students often struggle to generalize sight word learning across multiple contexts (Tan & Nicholson, 1997).
Comparing reading performance on semi-known and unknown intervention words, students consistently read semi-known intervention words correctly more often than unknown intervention words. Further, students displayed consistently lower TDT on semi-known intervention words than on unknown intervention words, which is consistent with previous research indicating that greater familiarity with stimulus items leads to lower TDT (Ardoin et al., 2013). However, one factor that makes these conclusions difficult to interpret is that students performed consistently more accurately on intervention words during intervention sessions than they did when reading the same words presented during assessment sessions. Previous literature regarding the instructional hierarchy may account for this discrepancy in results between intervention and assessment sessions (Daly, Lentz, & Boyer, 1996). Specifically, the instructional hierarchy suggests that students must first acquire and become fluent with stimuli before being able to generalize that information in novel contexts. Given that assessment of sight word learning in the current study occurred shortly after completion of intervention each week, it is likely that students did not receive an adequate number of learning trials to truly acquire and generalize sight words to novel assessment contexts. Students also performed better on known control assessments than intervention word assessments, making it difficult to determine how familiar a student must be with a word before it becomes automatically decodable and can be read accurately across contexts.

**Limitations**

A number of limitations in the present study also impede our ability to adequately differentiate expected participant performance on unknown and semi-known intervention words. One such limitation is that the total number of semi-known intervention words administered each week was substantially fewer than the number of administered unknown intervention words.
Given the procedures utilized to designate semi-known words, along with the fewer overall number of semi-known items, there was significantly more error surrounding assessment of semi-known intervention words.

**Implications and Future Directions**

Findings from the present study indicate several implications for future practice. First, greater familiarity with stimulus materials appears to directly affect the ease with which students may learn new material. Given this finding, students may require a greater number of response opportunities and learning trials to become familiar with materials so that they may generalize and adapt information to new learning contexts. Secondly, students may perform better on academic assessments when the assessment items are presented in a similar context to that in which the content was originally taught. Students should either be taught new material through a variety of methods and visual contexts, or teachers should make an effort to develop assessments that resemble instruction. Instructors should also assess learning using both proximal measures (i.e., assessment materials that look similar to original learning materials) and distal measures (i.e., assessment materials that assess learning in novel formats and contexts).

Results of this study suggest areas of research not yet adequately explored. First, knowing the effect of context on students' reading accuracy would be beneficial in pairing appropriate reading interventions with specific students in the future. Secondly, individual components that vary between the initial learning trial and assessment of new material (i.e., computer assessment versus paper-and-pencil assessment) could benefit instructors in determining the most effective format in which to test students' learning. Greater knowledge of the reasons for the lack of generalization shown by young readers may allow more consistent assessment of content mastery.
Teaching sight words in context does not appear to benefit young readers more than the isolated presentation of sight words. However, students appear to benefit from some familiarity with stimulus items before learning trials take place. Thus, it may be important for instructors to preview future learning materials with students in order to more effectively learn those materials when the material is eventually taught in the classroom.
REFERENCES


Table 1

*Mean Percentage of Words Read Correctly During Assessment Sessions in IRC and IRI Conditions*

<table>
<thead>
<tr>
<th>Participant</th>
<th>IRC (Mean, Range)</th>
<th>IRI (Mean, Range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colin</td>
<td>53(44-89)</td>
<td>62(33-78)</td>
</tr>
<tr>
<td>Dennis</td>
<td>59(56-67)</td>
<td>76(67-89)</td>
</tr>
<tr>
<td>Paula</td>
<td>81(67-100)</td>
<td>85(67-100)</td>
</tr>
<tr>
<td>Ted</td>
<td>89(78-100)</td>
<td>91(56-100)</td>
</tr>
</tbody>
</table>
Table 2

*Mean Percentage of Words Read Correctly Across Unknown Intervention, Unknown Control, and Known Control Target Words*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Unknown Intervention</th>
<th>Unknown Control</th>
<th>Known Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colin</td>
<td>58 (33-89)</td>
<td>33(33)</td>
<td>78</td>
</tr>
<tr>
<td>Dennis</td>
<td>69 (56-89)</td>
<td>39(33-44)</td>
<td>100</td>
</tr>
<tr>
<td>Paula</td>
<td>83 (67-100)</td>
<td>28(22-33)</td>
<td>100</td>
</tr>
<tr>
<td>Ted*</td>
<td>90 (56-100)</td>
<td>56</td>
<td>100</td>
</tr>
</tbody>
</table>

*Ted was administered two Known Control programs and one Unknown Control program.*
Table 3

*Mean Percentage of Words Read Correctly Across Semi-Known and Unknown Intervention Words*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Semi-Known</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colin</td>
<td>94(50-100)</td>
<td>58(33-89)</td>
</tr>
<tr>
<td>Dennis</td>
<td>94(50-100)</td>
<td>69(56-89)</td>
</tr>
<tr>
<td>Paula</td>
<td>93(50-100)</td>
<td>83(67-100)</td>
</tr>
<tr>
<td>Ted</td>
<td>94(50-100)</td>
<td>90(56-100)</td>
</tr>
</tbody>
</table>
### Table 4

*Mean Total Dwell Time Across Semi-Known and Unknown Intervention Target Words*

<table>
<thead>
<tr>
<th>Participant</th>
<th>Semi-Known</th>
<th>Unknown</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colin</td>
<td>1972</td>
<td>4725</td>
</tr>
<tr>
<td>Dennis</td>
<td>2049</td>
<td>2752</td>
</tr>
<tr>
<td>Paula</td>
<td>2073</td>
<td>2130</td>
</tr>
<tr>
<td>Ted</td>
<td>730</td>
<td>1575</td>
</tr>
</tbody>
</table>
Figure 1. Number of words read correctly within each assessment session, comparing performance in context (IRC) and isolation (IRI) conditions.
Figure 2. Percentage of words read correctly by participants on isolation (IRI), context (IRC), known control and unknown control assessments.
Figure 3. Mean total dwell time across intervention words within context and isolation assessments.
Figure 4. Mean total dwell time across target words within intervention, known control, and unknown control assessments. "Intervention" results refer to previously unknown intervention words whereas "semi-known" refer to semi-known intervention words.