Enhancing Maintenance and Generalization of Incremental Rehearsal Through Theory-Based Modifications

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Shawna M. Petersen-Brown

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Matthew K. Burns

May 2013

Dedication

This thesis is dedicated to my parents for instilling in me the perseverance, courage, and passion to pursue my dreams, and to my husband for relocating his life to facilitate this dream.

Acknowledgements

I could not have achieved this personal and professional goal without the support and encouragement of many.

First, I thank my parents, Gary and Jane Petersen, who have always supported and encouraged me in all of my endeavors. And a big thank you to my husband, Derek Brown, who not only moved 1,000 miles to facilitate my graduate school journey but also patiently endured the many weekend and evening hours spent with my head buried in a computer. And thank you to my unborn first child, who gave me the last bit of motivation I needed to finish my dissertation before his birth. And to the rest of my family members and friends who have offered encouragement and a listening ear over the past five years – thank you!

I am also forever grateful to my advisor, Matt Burns, who from the beginning has been an excellent model of effective scholarship and inquiry. He has also been an integral part of my support system, rom helping me to discover my passions in working with children and conducting research to tirelessly reading and offering edits and suggestions on each chapter of this document, as well as everything in between. I thank him from the bottom of my heart for being an unwavering source of encouragement from the day I was accepted to graduate school! This dissertation and my degree truly would not have been possible without him.

I thank each of my committee members for their support in this research project and throughout my time in graduate school: Jennifer McComas, Kristen McMaster, Sashank Varma, and Jim Ysseldyke. You have each helped me to improve and carry out this dissertation project and have been immensely supportive as I navigated this process.

Abstract

The attainment of basic early literacy skills at an early age is one way to ensure children become proficient readers as adults. Word recognition is an important basic early literacy skill that is related to reading fluency and overall reading competency. Incremental rehearsal (IR) is a flashcard technique that has produced strong outcomes for a variety of outcomes including word recognition. Modifications to IR that are based on theory may enhance maintenance and generalization, outcomes that are frequently overlooked in flashcard intervention research. A number of theoretical frameworks were utilized in this research to enhance maintenance and generalization. The depth of processing framework was used, which suggests that semantic processing of information leads to better maintenance. The theory of common elements and Stokes and Baer's (1977) generalization framework were utilized as frameworks for enhancing generalization. In addition to intervention design, individual factors may also impact intervention efficacy. For example, working memory capacity may influence the amount of verbal information an individual can process. Decoding skill may also lead to enhanced outcomes in word recognition interventions.

The current study examined the impact of theory-based modifications to IR on maintenance and generalization. A within-subjects design was utilized in which all participants were taught seven unknown words in each of three IR variants – IR, IR with vocabulary (IR-V, which leveraged the depth of processing framework), and IR with context (IR-C, which leveraged Stokes and Baer's [1977] generalization framework and the theory of common elements). Auditory working memory and decoding skills were also measured as individual factors and potential moderators of intervention effects.

Primary dependent variables were maintenance and generalization at 1 and 2 weeks after teaching.

Results of the study indicated that maintenance and generalization were high across conditions, suggesting possible ceiling effects. However, 1-week maintenance was significantly greater in IR-V than IR and in IR-C than IR. Additionally, 2-week generalization was significantly greater in IR-V than IR. Effect sizes were small across outcomes in favor of IR-V and IR-C. IR was the most efficient intervention variant in words maintained per minute. A moderating effect of auditory working memory capacity was not observed. However, participants with low decoding skill maintained significantly more words through IR-V than IR at 1 week, while this difference was not observed in participants at higher decoding skill levels. Future research may investigate theory-based modifications applied to different populations (such as students with low decoding skills) or different information types. In addition, future research may investigate theory-based modifications to more efficient modifications of IR (perhaps with fewer opportunities to respond) to enhance the efficiency of these approaches.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

Background

Despite efforts by schools to help all students become proficient in reading, a persistent proportion of the school-aged and adult population lags behind. On the most recent National Assessment of Educational Progress (NAEP), 33% of 4th graders and 24% of 8th graders read at a below basic level (National Center for Education Statistics [NCES], 2011). According to the 2003 National Assessment of Adult Literacy, 12% to 14% of adults (approximately 30 million adults) read at a below basic level (NAAL; Kutner et al., 2007). Additionally, adults who lack basic literacy skills are generally less likely to provide a literacy-rich environment in the home (with 41% of these adults never reading to their young children; Kutner et al., 2007), thus perpetuating the cycle of a lack of sufficient literacy skills.

One way educators can help children become proficient readers as adults is to help them attain basic literacy skills from an early age. Researchers suggest that poor readers become even poorer readers over time while good readers become even better readers over time; this phenomenon is known as the Matthew effect (Stanovich, 1986). Stanovich (1986) explained that good readers read more and therefore have a greater opportunity to learn; these readers become even better at reading through this experience, distancing them even further from their peers who are poor readers. Educators can prevent poor readers from becoming even poorer by helping all students attain levels of basic early literacy skills that will enable them to be successful in consuming and comprehending what they read.

It is important to know which basic literacy skills are important for later reading success. Reading fluency, or the ability to read quickly, accurately, and with appropriate expression (Kuhn, Schwanenflugel, Meisinger, Levy, & Rasinski, 2010; National Institute for Child Health and Human Development [NICHD], 2000) is one critical reading skill. In general, students who are fluent readers can also comprehend what they read (e.g., Fuchs, Fuchs, Hosp, & Jenkins, 2001); therefore, reading fluency is critical for later reading success. Some researchers have emphasized the role of automatic word recognition in fluent reading through theoretical models such as the automaticity model of reading (LaBerge & Samuels, 1974) and the interactive-compensatory model of reading (Stanovich, 1980). Intervention and correlational research also emphasizes the link between word recognition and reading fluency.

Interventions that target word recognition may be important for many struggling readers given the relationship between word recognition and reading fluency. Flashcard interventions are often used to enhance word recognition (Burns & Boice, 2009; Cooke, Guzaukas, Pressley, & Kerr, 1993; Nist & Joseph, 2008; Symonds & Chase, 1992), and they have also been shown to enhance reading fluency and comprehension (Burns, Dean, & Foley, 2004; Tan & Nicholson, 1997). However, maintenance and generalization of outcomes over time has not been a focus of this research. Therefore, maintenance and generalization are critical outcomes in future research on flashcard interventions.

Statement of the Problem

Understanding potential outcomes of flashcard interventions is important given the role of these interventions in enhancing important basic reading skills. Given the nature of most flashcard interventions in which words are practiced in isolation,

generalization is a key outcome. In addition, maintenance is also a key outcome given that students rarely encounter a practiced word immediately after it is taught.

Strategies to enhance maintenance and generalization should be investigated given their importance as outcomes in flashcard intervention research. One potential strategy to enhance maintenance and generalization is to utilize theoretical frameworks to inform intervention modifications. By using frameworks that have been linked to enhanced outcomes in laboratory research, outcomes may be enhanced in applied interventions as well.

Incremental Rehearsal

Incremental rehearsal (IR; Tucker, 1989) is the flashcard intervention that will be used in the current study. In IR, the interventionist introduces unknown items one at a time along with a high proportion of known items (generally, approximately 80% to 90% of words used are known items) to maintain a high rate of success. IR also employs many opportunities to respond (OTRs) as well as OTRs spaced further and further apart to encourage maintenance of items.

IR has shown high retention with a variety of information types, student populations, age group, and with retention intervals ranging from 1 day to 30 days after teaching (e.g., Burns & Boice, 2009; MacQuarrie, Tucker, Burns, & Hartman, 2002; Nist & Joseph, 2008). A recent meta-analysis found a large effect size of IR (Burns, Zaslofsky, Kanive, & Parker, 2012); large effects were consistent across information type, student population, and age group.

Maintenance

Maintenance is defined as correctly performing a response over time without reteaching (Alberto & Troutman, 2003), and it is a key goal of intervention if the intervention is to benefit the child over the long term (Ardoin, 2007). This research leverages theoretical frameworks that have been shown to enhance maintenance in basic research. The depth of processing framework and transfer-appropriate processing were selected for the current study.

Depth of processing. The depth of processing framework is a framework for memory research, which at its inception marked a departure from the way memory was conceptualized at the time. Memory was at that time conceptualized as a series of stores, including short- and long-term memory stores (e.g., Atkinson & Shiffrin, 1968). Craik and Lockhart (1972) advocated a departure from this view, suggesting that memory was procedural. In other words, memory was viewed as a byproduct of cognitive processes (e.g., defining or analyzing) carried out on information rather than a byproduct of a specialized encoding process. This view is supported by neuroimaging research (e.g., McDermott, Petersen, Watson, & Ojemann, 2003). In addition to advocating a procedural view, the depth of processing framework also suggests that the type of procedure carried out on information will impact its durability in memory, where more meaningful analyses lead to greater durability (Craik & Lockhart, 1972). Although depth of processing is probably best considered as a continuum, Craik and Tulving (1975) identified three levels as part of their initial validation of the model; these have persisted in more recent research. These levels, from most superficial to deepest, are (1) structural (relating to the physical properties of a stimulus), (2) phonemic (relating to the sound of the stimulus), and (3) semantic or contextual (relating to the definition or context of a stimulus; Craik &

Tulving, 1975). Craik and Tulving (1975) validated the framework, showing that semantic processing led to superior same-day retention, that time or effort spent processing did not account for results, and that other variations (such as intentional learning) did not affect the pattern of results.

      Transfer-appropriate processing. The transfer-appropriate processing framework emerged out of discussion regarding whether any level of processing was inherently meaningful (or deep) or unmeaningful (or shallow; Morris, Bransford, & Franks, 1977). Morris and colleagues (1977) suggested that the degree to which processing is meaningful depends on learning goals (and the way in which retention is measured). In this initial investigation, retention was best when the type of processing (semantic or phonetic) matched the recognition assessment (standard or phonetic).

Generalization

      Generalization is defined as the application of learning in a way that differs from but has properties similar to original learning with regard to setting, stimuli, or the behavior exhibited. This definition is gathered from the definitions presented by various authors (Alberto & Troutman, 2003; Hull, 1943; Skinner, 1938; Skinner, 1953; Stokes & Baer, 1977). Researchers have lamented the lack of explicit programming of generalization in research and practice (Burns, 2004a; Kendall, 1989; Lyon & Moats, 1997; Stokes & Baer, 1977).

      Transfer is a related term, and it refers to the impact that knowledge from one situation has on learning or performance in a different situation (Mayer & Whitrock, 1996). Transfer is generally utilized in the cognitive literature. Edelstein (1989) recommended the term "transfer" when control is verbally or conceptually mediated.

Because word recognition is generally applied automatically, generalization is considered a more appropriate term in this document. Therefore, while frameworks of generalization and transfer contribute to the current research, the term generalization is used outside of this section.

Generalization frameworks. Various theoretical frameworks of generalization have been proposed (Kirby & Bickel, 1988; Stokes & Baer, 1977; Stokes & Osnes, 1989); Stokes and Baer's framework provided the theory-based modifications for the current study. It was an initial framework for generalization and included nine strategies: (1) train and hope, (2) sequential modification, (3) introduce to natural maintaining contingencies, (4) train sufficient exemplars, (5) train loosely, (6) use indiscriminable contingencies, (7) program common stimuli, (8) mediate generalization, and (9) train to generalize. This framework is defined in greater detail in Chapter 2.

Transfer theory. Thorndike and Woodworth (1901a, 1901b, 1901c) proposed a common elements theory of transfer. In general, they suggested that the more elements (or attributes) a setting shared with a transfer setting, the more transfer will be observed. This theory has been called the "sole viable theory of transfer" (Butterfield & Nelson, 1989, p. 6). Elements have been defined through a variety of lenses, including a behavioral lens (where elements are stimuli and responses) or an information processing lens (where elements are representations, knowledge, and strategies; Butterfield & Nelson, 1989).

Role of Individual Attributes in Intervention Outcomes

Just as the design of an intervention plays a role in intervention outcomes, within-student factors may also play a role in intervention outcomes. Cognitive attributes such as

working memory performance and skill-based factors such as decoding ability are two within-student factors that may play a role in intervention effectiveness.

Working memory. Working memory is defined as a system that both maintains newly acquired information and retrieves stored information so that it can be processed (Becker & Morris, 1999). Deficits in working memory are commonly identified in students with learning disabilities and reading difficulties (Gathercole, Alloway, Willis, & Adams, 2006; Siegel & Ryan, 1989; Swanson, 1994; Swanson, Ashbaker, & Lee, 1996; Wagner et al., 1997). The capacity of working memory is also directly related to reading skills such as reading comprehension and reading fluency (Dixon, LeFevre, & Twilley, 1988; Gathercole et al., 2006; Siegel & Ryan, 1989; Swanson, 1994; Swanson et al., 1996; Swanson & Howell, 2001; Wagner et al., 1997). More specifically, the phonological loop (the component of working memory responsible for processing verbal information) is related to word recognition (Swanson & Howell, 2001).

Skill levels. A student's current skill level in the intervention skill as well as related skills is also likely to play a role in intervention outcomes. Research has shown that students show higher rates of learning when material is correctly matched (Burns, 2002; Burns, 2007a; Gickling & Armstrong, 1978; Shapiro, 1992). LaBarge and Samuels (1974) suggested that accuracy and automaticity in more basic skills allow students to access cognitive resources to apply to higher-level skills. For example, proficiency in decoding skills may facilitate word recognition and successful sight word instruction (Aaron et al., 1999; Samuels, 1988). Taken together, it is possible that the success of a word recognition intervention may depend on a student's level of decoding skills.

<div align="center">

Purpose

</div>

The purpose of the current study is to investigate the effect of adding a vocabulary component to IR on maintenance and generalization of unknown words as well as word knowledge. Additionally, the current study will investigate the effect of adding contextual reading practice to IR on maintenance and generalization of unknown words. Adding a vocabulary component to IR is hypothesized to enhance maintenance because adding semantic information facilitates processing taught words at a greater depth. Adding contextual reading practice to IR is hypothesized to enhance generalization because it increases the common elements between learning and generalization contexts.

The current modifications to IR require processing additional information such as a word's definition, and they may place greater demands on working memory as a result. Therefore, a moderating effect is hypothesized in which the benefit of adding a vocabulary component or contextual reading practice is greater for participants with greater working memory capacity. Additionally, more proficient readers may be more likely to spontaneously decode items regardless of maintenance, so decoding ability will also be examined as a potential moderator. Finally, efficiency is an important consideration when conducting interventions in schools. Therefore, the current study will investigate whether adding components to IR impacts the efficiency of the intervention.

Research Questions

1. What is the effect of adding additional components to IR – a vocabulary component or contextual reading practice – on the maintenance and generalization of target words?

2. What is the relative efficiency of the following IR variations in words learned per minute: IR, IR with vocabulary (IR-V), and IR with contextual reading practice (IR-C)?

3. To what extent does decoding ability moderate the impact of IR with vocabulary or contextual reading practice on maintenance and generalization?

4. To what extent does working memory capacity moderate the impact of IR with vocabulary or contextual reading practice on maintenance and generalization?

5. To what extent does knowledge of word meaning change from before teaching to 1 and 2 weeks after teaching across IR variations?

Hypotheses

Regarding the first research question, the hypothesis was that the addition of a vocabulary component to IR would lead to increased maintenance relative to standard IR as a result of orienting participants toward semantic processing as opposed to phonetic processing. A corresponding increase in generalization would also be expected, since it is reasonable to expect that information that is maintained is more likely to be generalized as well. Regarding the addition of contextual reading practice, the hypothesis was that the addition of contextual reading practice to IR would lead to increased generalization relative to standard IR. Practice in reading multiple sentences with target words embedded is an example of applying the theory of common elements from a cognitive standpoint or Stokes and Baer's (1977) generalization framework (specifically, training sufficient exemplars and programming common stimuli) from a behavioral standpoint.

Regarding the second research question, the hypothesis was that the addition of components to IR would decrease the efficiency of IR-V and IR-C in words maintained per minute of instructional time.

Regarding the third research question, the hypothesis was that more proficient decoders would show a lessened advantage of the two IR modifications (adding vocabulary or contextual reading practice) because they are able to accommodate for a possible lack of maintenance by decoding unfamiliar words.

Regarding the fourth research question, the hypothesis was that individuals with greater working memory capacity would show a greater advantage of the two IR modifications (adding vocabulary or contextual reading practice). The modifications require the participants to process a greater amount of information; this may lead to some difficulty for participants with fewer working memory resources to draw upon.

Regarding the fifth research question, the hypothesis was that knowledge of word meaning would change the most through IR-V. Through exposure to the word used in context, knowledge of word meaning may change to a lesser extent through IR-C. Finally, knowledge of word meaning was hypothesized to change negligibly through IR.

<div align="center">**Definition of Key Terms**</div>

Decoding skill. Skills related to identifying a word by applying letter-sound correspondence rules (Gough & Tunmer, 1986). More primitive levels of this skill are reflected in "sounding out" slowly and with effort, while more advanced levels of this skill are shown through quick, accurate, and silent application.

Depth of processing framework. Also called levels of processing, it is a framework first proposed by Craik and Lockhart (1972). It proposes that the nature of the

processing operations performed on information influences the extent to which it is retained. It emphasizes cognitive processes as opposed to a separate memory operation as the primary mechanism underlying retention.

Generalization. Generalization is defined as the application of learning in a way that differs from but has properties similar to original learning with regard to setting, stimuli, or the behavior exhibited (Alberto & Troutman, 2003; Hull, 1943; Skinner, 1938; Skinner, 1953; Stokes & Baer, 1977).

Incremental rehearsal (IR). IR is a flashcard technique (Tucker, 1989) in which an interventionist introduces unknown items one at a time along with a high proportion of known items (generally, 80% to 90% of words used are known items) to maintain a high rate of success.

Maintenance. Maintenance is defined as correctly performing a response over time without reteaching (Alberto & Troutman, 2003).

Programming for common stimuli. A method for programming generalization proposed by Stokes and Baer (1977); it refers to making the training and generalization settings as similar as possible.

Theory of common elements. A theory of transfer, it suggests that the more attributes a situation shares with a transfer setting, the more transfer that will be observed (Thorndike & Woodworth, 1901a, 1901b, 1901c).

Train sufficient exemplars. A method for programming for generalization proposed by Stokes and Baer (1977); it refers to training exemplars until generalization to the rest of the category occurs.

Semantic processing. Suggested by Craik and Tulving (1975) to be the deepest level of processing, semantic processing orients the learner toward the definition or meaning of a word.

Working memory. Working memory is defined as a system that both maintains newly acquired information and retrieves stored information so that it can be processed (Becker & Morris, 1999).

Assumptions, Limitations, Delimitations, and Design Controls

Assumptions

Various assumptions were made in designing the current study. First, the sample in the current study is assumed to be representative of the population to which generalizations are drawn. It should be noted that the sample for the current study was a convenience sample from an elementary school in a suburban area in Minnesota, so caution should be exercised in generalizing to a broader student population. However, care was taken to ensure that the students in the current sample represented a variety of achievement levels.

The measures used in this study are assumed to reflect the constructs they purport to measure. A measure of word attack was used as an overall measure of decoding skill. A measure requiring student to recall, rearrange, and restate series of words and numbers was used as a measure of verbal working memory. Prompting students to read taught words at 1 and 2 weeks after teaching was used as a measure of maintenance. Additionally, prompting students to read taught words within sentences was considered a measure of generalization.

Maintenance and generalization are assumed to be important outcomes in intervention research. Both maintenance and generalization have been discussed as important outcomes in previous research (e.g., Ardoin, 2007; Stokes & Baer, 1977), and logically, intervention outcomes would need to be maintained and generalized to classroom-based activities in order to benefit student performance.

Limitations

A sample of convenience was used which included the first participants to return consent forms that also represented an even distribution across quartiles of reading achievement. There may have been systematic differences between students who returned consent forms and those who did not.

Unknown words were located prior to the first session and taught across the first three sessions. It was assumed that previously identified words were still unknown at the time they were taught, in other words that students did not learn the words between initial identification and teaching.

Maintenance and generalization were assessed at 1 and 2 weeks after teaching; longer intervals were considered but not assessed given constraints related to time and access to participants. Additionally, 2-week outcome data may have been impacted by the assessment done 1 week before.

Delimitations

The current research was informed by specific theoretical frameworks that impact maintenance and generalization of intervention outcomes: the depth of processing theory, transfer-appropriate processing theory, Stokes and Baer's (1977) framework for generalization, and the theory of common elements. A variety of other theoretical

frameworks such as the generation effect (Jacoby, 1978) or spacing effect (Bahrick, 1979) could have also provided a basis for theory-based modifications to enhance intervention outcomes.

The current research identified specific within-student attributes that may also affect intervention outcomes. In this study, decoding skill and working memory were included, although a variety of other cognitive or skill-based attributes could have also been included. For example, rapid automatized naming (RAN) speed is one cognitive characteristic that is also associated with reading disabilities and reading difficulties (Denckla & Rudel, 1976; Meyer, Wood, Hart, & Felton, 1998).

The current research utilized IR as the flashcard intervention for this research. Other flashcard interventions such as traditional drill or interspersal have been utilized in prior research (e.g., Nist & Joseph, 2008). However, IR was selected for this study given the strong outcomes associated with IR (Burns et al., 2012).

This study included participants from an elementary school in a suburban area in Minnesota who were not identified as English learners and were in the 2nd and 3rd grade. Generalization to a broader population of students is limited since this sample is not representative of the overall student population.

Design Controls

It is possible that many within-subject factors could impact the results found in the current study; however, a limited number of within-subject factors were measured. A within-subjects design was utilized in which each participant was exposed to each condition to control for the variation caused by these variables.

In addition, potential variations in student performance across weeks due to factors other than the intervention variation itself were controlled by randomly ordering conditions across weeks of the study.

Organization of the Document

This document is organized in five chapters. This first chapter contains background on the issues addressed in this research, a statement of the problem including theoretical underpinnings for this research, and a statement of the purpose and specific research questions. The second chapter includes a detailed review of the literature in the domains addressed in this research, including IR, theories of maintenance and generalization, and within-student factors that may impact intervention outcomes. The third chapter is a review of the methods and procedures implemented in this research. The fourth chapter describes the results of statistical analyses undertaken to answer the research questions previously defined. Finally, the fifth chapter includes a discussion of the results contextualized within previous research; implications for theory, research, and practice; limitations; and conclusions.

CHAPTER 2: LITERATURE REVIEW

The current chapter will summarize relevant literature to contextualize the data within research. The literature review will first describe research regarding reading fluency and word recognition as critical components of reading and as intervention targets. Second, I will present evidence supporting the use of a specific flashcard intervention called incremental rehearsal (IR; Tucker, 1989), which may be used to enhance word recognition. Next, I will explore maintenance and generalization from multiple theoretical perspectives and cognitive- and skill-based factors that may influence maintenance and generalization, particularly as it relates to word recognition, and IR more specifically.

Development of Basic Reading Skills

Despite efforts by educators to help all students become proficient in reading, a persistent proportion of the school-aged and adult population lags behind. On the most recent NAEP, 33% of 4th graders and 24% of 8th graders read at a below basic level (NCES, 2011). According to the 2003 NAAL, 12% to 14% of adults (approximately 30 million adults) read at a below basic level (Kutner et al., 2007). Not surprisingly, a lack of literacy skills limits job possibilities; 57% of adults at a below basic literacy level are not employed, and 35% in this group believe their reading skills limit their job opportunities a lot (Kutner et al., 2007). Adults who lack basic literacy skills are generally less likely to provide a literacy-rich environment for their children (with 41% of these adults never reading to their young children; Kutner et al., 2007), thus perpetuating the cycle of a lack of sufficient literacy skills.

One way educators can ensure children become proficient readers as adults is to help them attain basic literacy skills from an early age. Researchers suggest that poor readers become even poorer readers over time while good readers become even better readers over time; this phenomenon is known as the Matthew effect (Stanovich, 1986). Stanovich (1986) explained that good readers read more and therefore have a greater opportunity to learn; these readers become even better at reading through this experience, distancing them even further from their peers who are poor readers. Educators can prevent poor readers from becoming even poorer by helping all students attain levels of basic early literacy skills that will enable them to be successful in consuming and comprehending what they read.

It is important to know which basic literacy skills are important for later reading success. Reading fluency is one critical reading skill. Reading fluency refers to the ability to read quickly, accurately, and with appropriate expression (Kuhn et al., 2010; NICHD, 2000). In general, students who are fluent readers can also comprehend what they read (e.g., Fuchs et al., 2001), which makes reading fluency critical for later reading success. Research supports a strong relationship between reading fluency and comprehension, with correlations between reading fluency and reading comprehension exceeding .80 (Fuchs et al., 2001; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Pinnell et al., 1995). But how do students become fluent readers? Some researchers have emphasized the role of automatic word recognition (e.g., LaBerge & Samuels, 1974; Stanovich, 1980). LaBerge and Samuels (1974) presented the theory of automatic information processing in reading. According to LaBerge and Samuels, cognitive resources can be allocated to more complex skills (such as reading fluency and comprehension) when

basic reading skills (such as decoding and word recognition) are automatic and do not require significant cognitive resources. Stanovich (1980) presented the interactive-compensatory model of reading. In this model, good readers rely primarily on rapid and automatic word recognition, while poor readers compensate for word recognition deficits by leveraging contextual clues. The labored process of using context clues occupies cognitive resources, making comprehension more difficult. Empirical research has shown strong correlations between word recognition in isolation and reading fluency, with correlations from .50 to greater than .73 across studies (Bowers, 1993; Jenkins et al., 2003). In addition, training word recognition in isolation transfers to enhanced fluency in context (Levy, Abello, & Lysynchuk, 1997). Therefore, it appears that word recognition is a key skill for facilitating reading fluency and ultimately comprehension.

Past theoretical and empirical work supports a facilitative role for word recognition on fluent reading and reading comprehension. Therefore, interventions that target word recognition may be important for many struggling readers. Flashcard interventions have long been used to enhance word recognition (Burns & Boice, 2009; Cooke et al., 1993; Nist & Joseph, 2008; Symonds & Chase, 1992), and they have also been shown to enhance reading fluency and comprehension (Burns et al., 2004; Tan & Nicholson, 1997). However, the outcomes in these studies were assessed immediately. Ultimately, word recognition should be maintained over time and generalized to contextual reading to positively impact reading outcomes in the longer term. Therefore, maintenance and generalization are critical outcomes in research on flashcard interventions.

Incremental Rehearsal

Incremental rehearsal (IR) is a flashcard technique (Tucker, 1989) in which an interventionist introduces unknown items one at a time along with a high proportion of known items (generally, 80% to 90% of words used are known items) to maintain a high rate of success. Additionally, many repetitions or OTRs to unknown items are used, and OTRs are spaced farther and farther apart to encourage the transfer of unknown items into long-term memory.

Researchers have found high retention with IR compared to traditional drill (when known items are not integrated with unknown items) and other interspersal techniques at retention intervals from 1 day to 30 days after teaching (Burns & Boice, 2009; MacQuarrie et al., 2002; Nist & Joseph, 2008). These findings have been shown in a general education population (MacQuarrie et al., 2002), among at-risk students (Nist & Joseph, 2008), and in students with developmental cognitive disabilities or a specific learning disability (SLD; Burns & Boice, 2009). IR has also been used to teach words (Burns, 2007b; Nist & Joseph, 2008; Szakokierski & Burns, 2008), letter sounds (Volpe, Burns, DuBois, & Zaslofsky, 2011), and math facts (Burns, 2004b; Codding, Archer, & Connell, 2010). IR has been applied as a preteaching intervention for students with SLD and was shown to enhance fluency and comprehension (Burns et al., 2004). A meta-analysis of 10 single-case design and 9 group design studies investigating IR yielded a weighted phi coefficient of .77, a large effect size (Burns et al., 2012). Furthermore, the meta-analysis found large effect sizes regardless of type of information taught, whether students had disabilities or not, and regardless of age group.

Efficiency

Although IR has been shown to be effective for enhancing word recognition as compared to other flashcard techniques (e.g., Burns & Boice, 2009; Nist & Joseph, 2008), recent investigations suggest that IR may not be the most efficient flashcard method. Nist and Joseph (2008) showed that while IR resulted in more words learned overall, the rate of words learned per minute (in next-day retention probes) was highest in traditional drill, followed by an interspersal technique in which the distance between presentations of unknown items was not varied. However, another study showed that while traditional drill was more efficient in words learned per minute, the two methods were comparable in words retained (after 1 week) per minute, and IR yielded significantly more words retained overall (Burns & Sterling-Turner, 2010). In a recent meta-analysis of outcomes generated through IR research, 5 studies examined the efficiency of IR compared to other methods (Burns et al., 2012). A small effect was found in favor of comparison methods (weighted phi of -.22) although the confidence interval included 0; this calls into question whether IR is truly less efficient than other methods.

Theory-Based Causal Mechanisms

In investigations aimed at identifying potential causal mechanisms in IR, the number of OTRs impacted retention more than the ratio of known to unknown items (Burns, 2007b; Szadokierski & Burns, 2008). Szadokierski and Burns (2008) systematically varied OTRs and the percentage of known items in a group design. When OTR was held constant and the percentage of known items was varied, effect sizes were small in favor of a greater percentage of known items ($d = 0.03$-0.18). In comparison, effect sizes were large when the number of OTRs was varied, in favor of more OTRs ($d = 1.60$-2.20). This finding is confirmed by a single subject design study (Burns, 2007b) in

which retention in a high OTR condition consistently exceeded a moderate OTR condition (percent of non-overlapping data = 1.0).

Working memory is a system that both maintains newly acquired information and retrieves stored information so that it can be processed (Becker & Morris, 1999), and its capacity may also serve as a potential causal mechanism of IR. A related concept is acquisition rate, which is the amount of information that a child can process without interference (Burns, 2001). A child's working memory capacity, or relatedly their acquisition rate, may also impact retention of words during IR. Haegele and Burns (2010) found that retention of words through IR was maximized when the number of words presented per session aligned with the student's working memory or acquisition rate, as measured through IR. More research on the potential causal mechanisms of IR is needed to understand the role of other aspects of the intervention such as spacing.

This research utilizes theory-based modifications to IR. Theory is critical to both research and practice because it provides a framework for interpreting results and solving practical problems (Tharinger, 2000). It provides insight regarding the reasons certain interventions succeed and the mechanisms underlying their effectiveness (Tharinger, 2000). Hughes (2000) suggests that theory and empirical research are most impactful when they reciprocally influence one another. Burns (2011) identified potential outlets for this reciprocal influence, including identifying causal mechanisms and moderator variables. Burns suggests that identifying causal mechanisms connects interventions to theory but also leads to theory-based modifications of interventions and expansion of existing theories. The theoretical frameworks discussed below show a general lack of application to applied settings and reading intervention specifically. Therefore, empirical

investigation of the application of these theories has potential for improving interventions and expanding the theories themselves.

Future research around IR should continue to investigate the outcomes obtained through IR and how these can be enhanced. Although maintenance has been found to be strong through IR, further improvements may be accomplished by investigating theory-based modifications. Additionally, generalization has not been a focus of flashcard research, including IR research (Burns, 2004a), so explicit programming for generalization presents an opportunity for improvement. Following is a discussion of maintenance and generalization that may inform theory-based modifications to IR.

<div align="center">

Maintenance of Intervention Effects

</div>

Maintenance is defined as correctly performing a response over time without reteaching (Alberto & Troutman, 2003). Maintenance must be a key goal of an intervention if the intervention is to benefit the child over the long term (Ardoin, 2007). However, maintenance is not frequently addressed in intervention research (Burns, 2004a). Theoretical frameworks can suggest attributes of interventions that may lead to enhanced maintenance. The depth of processing framework and related transfer-appropriate processing framework are now discussed.

Depth of Processing

The depth of processing framework is a framework for memory research, which at its inception marked a departure from the way memory was conceptualized at the time. At the time, memory was conceptualized as a series of stores, including short- and long-term memory stores (e.g., Atkinson & Shiffrin, 1968). Memory traces were encoded (through an encoding process such as rehearsal) into these memory stores, and traces

were retrieved from the store when they were needed (Atkinson & Shiffrin, 1968). Craik and Lockhart (1972) advocated a departure from this view, suggesting that memory was procedural. In other words, memory was viewed as a byproduct of cognitive processes (e.g., defining or analyzing) carried out on information rather than a byproduct of a specialized encoding process transferring information into a memory store. Recent support for the procedural hypothesis comes from neuroimaging research. Research has shown that distinct brain regions are activated when phonemic orienting cues are used versus when semantic orienting cues are used for processing words, in addition to some overlapping brain regions (perhaps reflecting areas of the brain for processing verbal information; e.g., McDermott et al., 2003). If the brain had dedicated memory stores, activation patterns would be similar whenever encoding of any type of information occurred (Roediger, Gallo, & Geraci, 2002).

In addition to advocating a procedural view of memory, the depth of processing framework also posits that the type of procedure carried out on information will impact its durability as a memory trace, where more meaningful analyses lead to greater durability (Craik & Lockhart, 1972). Although depth of processing is probably best considered as a continuum, Craik and Tulving (1975) identified three levels as part of their initial validation of the model. These levels, from most superficial to deepest, are (1) structural (relating to the physical properties of a stimulus), (2) phonemic (relating to the sound of a stimulus) and (3) semantic or contextual (relating the definition or context of a stimulus; Craik & Tulving, 1975). Structural processing is marked by fragile memory traces as a result of fewer attentional resources and a lack of meaning attributed to the stimulus. Conversely, when one processes information at a semantic or conceptual level,

he or she attends to the information, analyzes it, and creates meaning and associations to that information (Craik & Lockhart, 1972). In a series of studies, Craik and Tulving (1975) validated the depth of processing framework with groups of undergraduates, showing that semantic processing led to superior same-day retention, that this effect was not due to the amount of time or effort spent processing information, and that the pattern of results did not differ when various modifications were introduced (e.g., cued versus noncued recall, incidental versus intentional learning).

Since the proposal of the depth of processing framework, researchers have investigated why deeper levels of processing generally lead to greater retention. Three potential explanations include (1) encoding elaboration, (2) distinctiveness, and (3) semantic processing supporting an episodic "remembering" as opposed to a general "knowing" or "familiarity". The explanations detailed here may not reflect all potential causal mechanisms previously proposed; however, they reflect potential mechanisms that have been discussed in detail in the literature.

Encoding elaboration. Craik and Tulving (1975) found support for encoding elaboration as a causal mechanism in their initial investigation of the depth of processing framework. Craik and Tulving (1975) carried out an experiment in which participants were shown sentence frames of varying complexity and a focal word. They were then asked to answer "yes" if the word made sense in the sentence and "no" if it did not. After 60 trials, they were asked to recall the focal words without cues (free recall), and then asked to recall the focal words using the sentence frames as cues. Focal words for which the participant answered "yes" were remembered more frequently, with an advantage for more complex sentences. Craik and Tulving (1975) hypothesized that depth of processing

was not enough to explain these results. When the sentence frame allowed the participant to elaborate on what the focal word meant, both through being a complex sentence with more details and by being congruent with the focal word (i.e., a "yes" question), participants remembered the focal word more readily.

Hannon and Craik (2001) used encoding elaboration to explain results of an experiment in which they presented semantically similar, semantically contrasting, and unrelated word pairs, then investigated the likelihood that participants would generate the words when given a different cue and whether participants would recall the rest of the word pair when given one of the original words. They found that similar word pairs led to greatest generation and cued recall, followed by contrasting word pairs, and finally by unrelated. They hypothesized that encoding elaboration was at play given the advantage for similar words. Depth of processing could be used to explain the qualitatively different ways to process information, and encoding elaboration used to describe more extensive processing within a certain depth (Lockhart & Craik, 1978). However, elaboration of encoding may also supplant the depth of processing framework (e.g., Anderson & Reder, 1979). Anderson and Reder (1979) suggest that what has been conceptualized as depth is actually the number and type of elaborations stored; in this case, the varying levels of depth are not needed, only the extent of elaboration must be known.

Distinctiveness. Distinctiveness has also been discussed as a potential mechanism underlying the depth of processing effect. Murdock (1960) defined distinctiveness as the extent to which a stimulus stands out from other stimuli and suggested that distinctiveness is related to later remembering. Distinctiveness is applied to the depth of processing framework because semantic qualities are generally considered to be more

distinctive than other features such as sound and physical features because there is a greater range of possible meanings than sounds or physical features (Gallo, Meadow, Johnson, & Foster, 2008; Moscovitch & Craik, 1976). However, researchers have shown that the depth of processing effect can be eliminated or reduced by reducing distinctiveness (Gallo et al., 2008; Moscovitch & Craik, 1976).

Words processed at a deep, as opposed to shallow, level show consistently lower false recognition ($d = 3.09$; Gallo et al., 2008), which may offer support for distinctiveness as an underlying causal mechanism for depth of processing. Other studies have manipulated distinctiveness and studied its effect on recall. In one study, participants were shown words and asked a question that oriented participants toward phonemic, categorical, or semantic processing (Moscovitch & Craik, 1976). Participants were given the question again and asked to recall the word. To manipulate distinctiveness, some words shared the same cue. The authors reported that when cues were shared, recall decreased across levels, particularly at the semantic level (to the extent that the depth of processing effect was dampened or eliminated); recall was enhanced when cues were unique, particularly for the semantic level of processing. Another study found that false recall was greatest with semantically related foils after semantic processing, as compared to visually related foils after visual processing (Thapar & McDermott, 2001). Subsequent research found that false recall increased only when the method of processing (semantic or phonological) matched the way in which distinctiveness was reduced on the recognition test (with semantically or phonologically related foils; Chan, McDermott, Watson, & Gallo, 2005).

Know versus remember. A final potential underlying mechanism discussed here is the propensity for semantic processing to engender remembering, or specific, episodic recollection, while shallower levels of processing are more likely to support knowing or familiarity, which is considered to be a vague recollection in the absence of details (Sheridan & Reingold, 2012; Tulving, 1985; Yonelinas, 2002). Words processed at a deep or semantic level were more likely to be remembered, while words processed at a shallow level were more likely to be recalled (Sheridan & Reingold, 2012; Yonelinas, 2002). Additionally, words that were remembered were more likely to be recalled correctly than known or familiar words (Sheridan & Reingold, 2012). While a compelling explanation for the depth of processing effect, research merging the remember/know or recollection/familiarity paradigms with levels of processing is minimal (Gardiner & Richardson-Klavehn, 2000).

Remaining Questions Regarding the Depth of Processing Framework

Although the depth of processing framework has been researched extensively over the past 40 years, issues and challenges still remain. The following issues are discussed here: circularity, the lack of an independent index of depth, challenges to the memory stores model, and the impact of the recall situation on the depth of processing effect.

Circularity. A common criticism of the depth of processing framework is that it is unfalsifiable and circular (Eysenck, 1978; Lockhart, 2002; Lockhart & Craik, 1978; Lockhart & Craik, 1990). Lockhart (2002) describes the perception that because there is not an independent index of depth, deeper processing is simply defined by the better remembering that results. It is true that better remembering cannot be predicted unless

retrieval conditions are known since better remembering depends somewhat on the retrieval situation (e.g., the match between encoding and retrieval). However, the depth of processing framework is still able to make predictions about the probability of better remembering with the retrieval task known (Lockhart, 2002).

Lack of an independent index of depth. The lack of a systematic framework for defining depth is related to other issues with the depth of processing framework, namely its potential circularity. Relatedly, is depth defined qualitatively or quantitatively, and how does this impact how we describe depth? Depth is not synonymous with strength, particularly given the interaction with the retrieval cue, and is therefore not a quantitative construct (Lockhart & Craik, 1978). Moreover, a unidimensional, numerical definition of depth is not likely to be useful (Lockhart & Craik, 1990), and a numerical definition of depth would depend on the observed degree of remembering, which returns to the issue of circularity in defining depth. Depth of processing can and has been independently defined at a basic level: structural, phonemic, and semantic processing is a common, though crude, depth configuration used in the research, and defining depth at a more granular level (i.e., defining depth within the category of semantic processing) would be difficult (Lockhart & Craik, 1990).

Challenges to the memory stores model. A common misconception of Craik and Lockhart's (1972) initial introduction of the depth of processing framework was that it was meant to eliminate the distinction between short- and long-term memory stores (Lockhart & Craik, 1990). In reality, they intended to replace the idea that items are perceived and transferred to structural memory stores through specific encoding processes with the idea that memory is couched within other cognitive processes such as

comprehension (Craik & Lockhart, 1972; Lockhart & Craik, 1990). As evidence of memory processes distinct from perception and encoding, Tulving (2001) presented examples where perception is markedly different than information that is later recalled. For example, some people report memories of events that did not occur, these events were never perceived but still remembered.

   The influence of the recall situation. In his critique of the depth of processing framework, Eysenck (1978) concluded that more emphasis on retrieval processes was warranted. Research supports the idea that the impact of processing depth is moderated by retrieval conditions (Fisher & Craik, 1977; Jacoby & Dallas, 1981; Morris et al., 1977; Roediger et al., 2001; Stein, 1978), which is reflected in the related transfer-appropriate processing and encoding specificity frameworks; the transfer-appropriate processing framework is discussed below. Generally, both frameworks emphasize the match between the processing orientation and later retrieval cues. The depth of processing framework may be most useful when learning objectives are semantic or conceptual.

Transfer-Appropriate Processing

   The transfer-appropriate processing framework arose out of questions regarding the inherent meaning of meaningful (deep) versus unmeaningful (shallow) processing within the depth of processing framework (Morris, et al., 1977). Morris and colleagues (1977) proposed that no method of processing is inherently meaningful or not meaningful, rather, the degree to which processing is meaningful depends on learning goals. In their initial investigation, the experimenters oriented participants to process words semantically or phonetically, and they provided both standard and phonetic recognition assessments. To encourage semantic processing, the experimenter read a sentence (e.g.,

"The train has a silver engine."), and the participant responded "yes" or "no" based on whether the sentence made sense. To encourage phonetic processing, the experimenter read a sentence (e.g., "Train rhymes with rain."), and the participant responded "yes" or "no" based on whether the sentence was true. During the standard recognition assessment, the experimenter read a list of words, and the participant responded whether each word had been previously studied. During the phonetic recognition assessment, the experimenter read a list of words, and the participant responded whether a rhyming word had been studied.

Morris and colleagues (1977) showed that semantic processing was superior when a standard recognition assessment was used, but when a phonetic recognition task was used, performance improved through phonetic processing. Notably, the advantage of semantic processing with a standard recognition test was greater than the advantage of phonetic processing with a phonetic recognition test. This suggests that there is an overall advantage to semantic processing. Encoding elaboration appears to have played a role as well; answering "no" to a semantic processing prompt led to decreased recognition and possibly lower encoding elaboration. This impact was even greater in phonetic processing.

Other researchers have also found support for the transfer-appropriate processing framework (e.g., Blaxton, 1989; Franks, Bilbrey, Lien, & McNamara, 2000; Martin-Chang & Levy, 2006; Park & Rugg, 2008; Schendan & Kutas, 2007). Blaxton (1989) showed that semantic study yielded better performance on semantic tests, whereas study without context yielded better performance on data-driven tests (such as word fragment completion). Martin-Chang and Levy (2006) taught third grade good and poor readers words both in context and isolation. They tested retention by asking participants to read

words in isolation and found an advantage for words taught in isolation. Unfortunately, the authors did not include a retention test in context, so only partial support for the transfer-appropriate processing framework was established in this study. Park and Rugg (2008) found support for the transfer appropriate processing framework through neuroimaging. Specifically, they found distinct regions were activated for picture versus word processing. Additionally, they found regions of the brain activated only when study and test modalities were congruent.

It is important to note that not all studies have found support for the transfer-appropriate processing framework. For instance, one study found that using two study conditions, studying intact words and solving anagrams, solving anagrams led to better recall regardless of the testing conditions (solving anagrams versus intact words; Mulligan & Lozito, 2006). However, solving anagrams at test also led to greater false alarms (saying that a word had previously been studied when it had not; Mulligan & Lozito, 2006), which may reflect the overall superiority of more effortful processing, such as solving an anagram, for later recognition, or reflect a stronger episodic memory of the act of solving the anagram.

Effectiveness in Reading Intervention

The depth of processing and transfer-appropriate processing frameworks are both well researched frameworks which may be leveraged to enhance learning in applied settings. Research on the depth of processing framework has most recently focused on encoding elaboration, the role of distinctiveness of stimuli in the depth of processing effect (even showing that the levels of processing effect can be reduced or eliminated by manipulating distinctiveness; Chan et al., 2005), and the large effect of level of

processing on remembering versus knowing (e.g., Sheridan & Reingold, 2012). Research

on the transfer appropriate processing framework has focused on establishing the benefits

of congruent learning and testing situations (Blaxton et al., 1989; Park & Rugg, 2008).

Research regarding the application of the depth of processing framework has been

largely conducted in laboratory settings with adults and fairly short retention intervals.

An important next step is to test these frameworks in applied settings where remembering

is a key outcome. To the extent that these investigations involve research in different

populations, with different types of information, and over different spans of time, support

for these frameworks will be more robust. To that end, the depth of processing and

transfer-appropriate processing frameworks will be utilized in the current investigation to

potentially enhance maintenance outcomes associated with IR. This use of the depth of

processing framework within IR has been supported in previous research (Petersen-

Brown & Burns, 2011), but more research is needed to investigate the utility of these

frameworks in a school setting. Just as theoretical frameworks may be utilized to enhance

maintenance, they may also be used to enhance generalization. Following is a discussion

of relevant frameworks for enhancing generalization.

Generalization and Transfer of Intervention Effects

In the literature overall, generalization refers to a behavioral construct where

transfer refers to a cognitive construct. Both refer to the application of learned content in

situations that are different in some way(s) from the situation in which content was

learned. Transfer, generalization, and the theories underlying each are reviewed. Outside

of this section, the term generalization is used. Edelstein (1989) recommended the term

transfer when control is verbally or conceptually mediated. Because word recognition is

typically and optimally applied automatically and without thought (e.g., LaBerge & Samuels, 1974; Stanovich, 1980), generalization is considered a more appropriate term for the current review. Importantly, however, theories of both generalization and transfer contribute to the current investigation and are discussed below.

Generalization

Generalization has been defined in various ways by many researchers over time. Skinner (1938, 1953) and Hull (1943) referred to generalization of both stimuli and responses. Generalization may occur when a strengthening of a particular response leads to a strengthening of other similar responses to the same stimuli. Alternatively, generalization may occur when a certain response is controlled by a particular stimulus and other stimuli that are similar. Others (e.g., Alberto & Troutman, 2003; Stokes & Baer, 1977) emphasized setting elements and defined generalization as the application of a behavior under different situations or conditions with less support than in the learning environment. One might consolidate these definitions by saying that generalization is the application of learning in a way that differs from but has properties similar to original learning with regard to setting, stimulus, or the behavior exhibited.

Over the years, researchers have lamented the lack of explicit programming of generalization in research and practice (Burns, 2004a; Kendall, 1989; Lyon & Moats, 1997; Stokes & Baer, 1977). To this end, a variety of frameworks have been presented for enhancing generalization in practice and providing a systematic way to program for generalization in research. The frameworks proposed by Stokes and Baer (1977), Kirby and Bickel (1988), and Stokes and Osnes (1989) are discussed below.

Theory. Stokes and Baer (1977) formulated the initial framework for generalization programming. The "technology of generalization" they defined has nine parts:

(1) train and hope: most commonly used method, it involves no explicit programming of generalization;

(2) sequential modification: in this method, the participant is trained and generalization is assessed; if generalization is not present, then the participant is retrained in all nongeneralized settings;

(3) introduce to natural maintaining contingencies: when natural contingencies are present, the behavior generalizes naturally because reinforcement is present in the environment and supports the behavior;

(4) train sufficient exemplars: this method involves training exemplars until generalization to the rest of the category occurs;

(5) train loosely: minimal control is exercised over the stimuli used and responses allowed;

(6) use indiscriminable contingencies: under this method, reinforcement across generalization settings is unpredictable so that the participant is more likely to demonstrate the behavior across generalization settings in an effort to obtain reinforcement;

(7) program common stimuli: this method refers to making the training and generalization settings as similar as possible;

(8) mediate generalization: this method is in use when a subject-controlled

mediator, such as self-reinforcement or a set of written instructions, is used to

enhance generalization; and

(9) train "to generalize": this method is used by reinforcing generalization, though

examples are also presented of training the participant to see the training and

generalization stimuli or responses as the same thing, this requires cognitive

effort by the participant.

Stokes and Baer's framework was clearly influential as it was the first attempt to

explicitly define a technology of generalization and appears to have influenced the

frameworks that followed (Kirby & Bickel, 1988; Stokes & Osnes, 1989). However,

neither Stokes and Baer nor any others that followed succeeded in influencing researchers

to consistently adopt these frameworks or to consider generalization a goal of researchers.

Kirby and Bickel. Kirby and Bickel (1988) offered a framework that reorganized

Stokes and Baer's (1977) framework with regard to how aspects of training and

generalization settings are controlled. Broadly, they defined three tactics for stimulus

control: controlling the stimuli that initiate the behavior, treatment of extraneous setting-

based stimuli, and controlling reinforcement schedules. The authors presented two ways

of controlling stimuli: presenting naturally occurring or not naturally occurring, but

supplemental, members of the stimulus class. These methods seem most similar to Stokes

and Baer's (1977) suggestion to program common stimuli, train sufficient exemplars, or

to mediate generalization.

To control extraneous stimuli, one could vary these stimuli. This reduces the

correlation that these extraneous, noncritical elements have with critical elements and

help ensure that extraneous stimuli will not control responding. This method is most similar to Stokes and Baer's (1977) strategy of training loosely. One could also control extraneous stimuli by maximizing common extraneous features between the training and generalization environments. This is most similar to Stokes and Baer's (1977) suggestion to train common stimuli and is also similar to the transfer theory of common elements, described below.

Finally, to control reinforcement schedules, one could alter the reinforcement schedule (from continuous to variable). This is similar to Stokes and Baer's (1977) suggestion to use indiscriminable contingencies to ensure that participants will increase the desired response across settings because reinforcement is unpredictable. To control reinforcement, one could also arrange for reinforcement to occur in generalization settings. This method is most similar to introducing to natural maintaining contingencies (if the reinforcement in the generalization setting is naturally occurring).

Taken as a whole, Kirby and Bickel's (1988) framework refers to ways to enhance generalization across stimulus and setting elements; they do not address ways to enhance generalization to similar behaviors. For instance, Kirby and Bickel refer to varying stimulus elements when training loosely, whereas Stokes and Baer refer to variations in both stimulus and the responses that receive reinforcement. Kirby and Bickel also refer to several programming methods as nonmethods: train and hope, introduce to natural maintaining contingencies, and train to generalize. While train and hope truly does not involve any intentional generalization programming, selecting behaviors that will be naturally reinforced surely requires planning and intent. Additionally, training to generalize, while it may refer to simply reinforcing instances of

generalization, may also require instruction in abstract principles that help an individual

see a class of examples as "the same thing" – this should surely be considered a method

of generalization and is most similar to cognitive views of transfer, described below.

  Stokes and Osnes. Stokes and Osnes (1989) also redefined and refocused Stokes

and Baer's (1977) original generalization framework. Their framework is a simplified

version of Stokes and Baer's original framework and focuses on three strategy categories:

exploit current functional contingencies, train diversely, and incorporate functional

mediators.

  To exploit current functional contingencies, one could focus on naturally

occurring consequences of behavior. Relatedly, one could introduce mechanisms in the

environment that lead to reinforcement of generalized behavior (such as mechanisms for

recording performance that lead to praise or teaching participants to ask someone how

they are doing to initiate reinforcement). Stokes and Osnes also emphasized modification

of consequences that may decrease desired behavior or increase undesired behavior, a

concept that was not explicitly addressed in Stokes and Baer's (1977) original framework.

Finally, one can reinforce instances of generalization when it occurs, a method that may

be especially important if reinforcement does not occur naturally.

  To train diversely, one accepts and/or employs a diverse range of stimuli,

responses, and settings. One may train multiple related stimuli for which the same

behavior is reinforced. Concurrently, one may also teach or at least reinforce multiple

similar responses. One may also vary aspects of the setting or other antecedents.

Collectively, these methods are reminiscent of Stokes and Baer's (1977) suggestions to

train loosely and train multiple exemplars. Relatedly, one may provide variable

reinforcement, a method that is equivalent to providing indiscriminable contingencies (Stokes & Baer, 1977).

To incorporate functional mediators, one ensures that discriminative stimuli exist in both training and generalization environments. These stimuli may be transported across settings (possibly by the participant himself) or stimuli common to the generalization environment may be incorporated into the training setting. A self-mediated stimulus may include written instructions, a self-management device, or another device that serves as a reminder to demonstrate the desired behavior.

Challenges for generalization. One lamentation common to each group of researchers who proposed a generalization framework is the need for increased focus on generalization, and not just as something that becomes a focus late in training (Kirby & Bickel, 1988; Stokes & Baer, 1977; Stokes & Osnes, 1989). Despite the emphasis on defining generalization and proposing frameworks for its enhancement, a "train and hope" approach still dominates 35 years after Stokes and Baer's (1977) original call to action (when it is a consideration at all). A greater emphasis on generalization is still important after all of these years, especially because many researchers agree that generalization does not happen automatically (e.g., Kendall, 1989; Stokes & Baer, 1977).

Effectiveness in reading intervention. Generalization is a key consideration in reading intervention research, particularly when skills (such as letter sounds or word recognition) are not taught in context. In order to fluently read printed words in text, students need to be able to generalize word recognition skills to contextual settings such as books. Thus, students may be taught to recognize words in one setting (e.g., on a

flashcard) but then be required to recognize the same word in a different setting (e.g., in connected text).

The frameworks described above may be employed to improve generalization of taught words. Most generally, training loosely may be an effective way to train participants to generalize words they have learned in a flashcard setting (Stokes & Osnes, 1989). Training loosely would most likely involve varying extraneous stimuli, although different variations on the same word (with different prefixes and suffixes) could also be used (Kirby & Bickel, 1988). Within Stokes and Baer's (1977) original framework, training multiple exemplars and programming common stimuli are most relevant.

Examples of utilizing generalization frameworks exist in the academic intervention literature, although these examples are few. The multiple exemplar approach for programming generalization has been compared to a repeated reading intervention (Ardoin, Eckert, & Cole, 2008; Ardoin, McCall, Klubnik, 2007; Silber & Martens, 2010). Ardoin and colleagues (2007) compared a multiple exemplars approach in which students read two high content overlap passages twice to a repeated reading intervention in which participants read the same passage four times. In this instance, most students showed greater gains when reading a third high content overlap passages with repeated reading than the multiple exemplars approach. Subsequent research found that a multiple exemplars approach in which students read three different passages with the same words in a different order led to greater fluency on medium word overlap passages (Ardoin et al., 2008), and repeatedly reading a subset of sentences that sampled the range of stimulus and response variations showed greater gains on generalization passages than simply repeatedly reading a passage and required less time to complete (Silber & Martens, 2010).

This line of research is one example of explicitly programming for generalization using the generalization framework proposed by Stokes and Baer (1977).

Daly and colleagues (2005) used the common stimuli (Stokes & Baer, 1977) method of generalization programming by implementing a repeated reading intervention and found enhanced reading fluency to high content overlap passages that were also an instructional match for the student as compared to no practice or passages that were not an instructional match for the student. Relatedly, an investigation of word recognition programmed common stimuli by introducing word families (Mesmer et al., 2010). The common stimuli procedure produced an improvement in performance on trained words and a slight improvement in generalization to nontrained words.

The studies reviewed here do not represent a comprehensive view of the literature explicitly programming generalization using the frameworks presented above. However, they are a sampling of the application of generalization frameworks to reading interventions, using the methods utilized in the current study. Overall, findings suggest that there are performance benefits when generalization is explicitly programmed.

Transfer

Theory. Transfer, or the impact that knowledge from one situation has on learning or performance in a different situation (Mayer & Whitrock, 1996), has been actively debated over the past century. Thorndike and Woodworth (1901a, 1901b, & 1901c) investigated whether improvement in one mental function generates improvement in another mental function. They concluded that in general, improvements in one mental function don't necessarily transfer to another function. In fact, they suggested that it might even be the opposite in some cases. Thorndike and Woodward also defined a

theory of transfer, the common elements theory (also called identical elements theory). At that time, they defined element rather loosely: in general, this meant that the more attributes a situation shared with a transfer setting, the more transfer will be observed (whether it is desired or not). Over the past century, researchers have defined the common elements theory more precisely, and it has been called the "sole viable theory of transfer" (Butterfield & Nelson, 1989, p. 6).

In defining elements, a variety of viewpoints have been expressed (Butterfield & Nelson, 1989). Elements have been defined behaviorally, as stimuli and responses. In this case, when stimuli are similar and desired responses are also similar, positive transfer is likely. When stimuli are similar but desired responses are dissimilar, negative (or undesired) transfer is likely. Finally, when stimuli are dissimilar but desired responses are similar, no transfer is likely. Elements have also been defined through an information processing lens, where elements are defined as representations, knowledge, and strategies. In general, the way an individual represents a problem impacts whether he or she sees a potential benefit of transfer and whether he or she will transfer previously learned knowledge and strategies to the current problem. This ties in with the concept of encoding specificity (and, relatedly, transfer appropriate processing), or that retrieval is influenced in part by the encoding operations conducted on the original stimuli (Tulving & Thomson, 1973).

Similar to defining elements, explaining how transfer actually happens can be framed differently through different theoretical lenses (Butterfield & Nelson, 1989). Thorndike and Woodworth's (1901a) original definition of transfer did not hypothesize a mechanism. The behaviorist view emphasizes the strength of the stimulus-response

relationship facilitating transfer. From a cognitive view, a variety of cognitive processes impact transfer, including how information is encoded as well as the match between the current problem and existing mental models.

In defining transfer, researchers have also defined two types of transfer: low-road and high-road transfer (Perkins & Salomon, 1989). Low-road transfer occurs automatically and is encouraged by varied practice in a variety of contexts. High-road transfer is more intentional and is encouraged by making concepts more general and abstract and imposing categories.

Effectiveness in reading intervention. The literature appears to lack empirical research examining the application of the theory of common elements to applied educational settings or to reading intervention more specifically. However, Butterfield and Nelson (1989) suggested potential education implications of the theory of common elements. They suggested an emphasis on common critical elements when initially teaching a concept, then gradually building in variability of noncritical stimuli. They also encouraged instruction on all critical elements to ensure a complete and accurate mental model of a concept so that positive transfer is more likely. Butterfield and Nelson's (1989) suggestions apply to the transfer of word recognition to a variety of contexts because students must learn the critical elements of the stimuli (e.g., the letters that make up a word). After learning these elements, students may begin to be exposed to noncritical elements, such as prefixes, suffixes, and a variety of contexts, but these should be varied so that correlation with critical elements does not cause a lack of transfer.

Role of Cognitive Attributes in Maintenance and Generalization

While the design of an intervention certainly plays a role in intervention outcomes,

within-student factors may also play a role in outcomes such as maintenance and generalization. Working memory is one such factor that may play a role. Working memory has been defined as a system that both maintains newly acquired information and retrieves stored information so that it can be processed (Becker & Morris, 1999). Deficits in working memory are commonly identified among students with learning disabilities and reading difficulties in general (Gathercole et al., 2006; Siegel & Ryan, 1989; Swanson, 1994; Swanson et al., 1996; Wagner et al., 1997). Relatedly, a lack of response to intervention is another hallmark feature of learning disabilities and reading difficulties. The evidence highlighted below suggests that there is a potential relationship between working memory and the maintenance and generalization of intervention effects.

Miller's (1956) discussion the limited capacity of working memory is now widely accepted, that the amount of information that we are able to process and later remember is limited. The capacity of working memory is shown to be related to reading skills such as reading comprehension and reading fluency (Dixon et al., 1988; Gathercole et al., 2006; Siegel & Ryan, 1989; Swanson, 1994; Swanson et al., 1996; Swanson & Howell, 2001; Wagner et al., 1997). The relationship of working memory to reading skill is another reason that it may be an important factor in the maintenance and generalization of intervention effects.

Phonological Loop

Within Baddeley and Hitch's (1974) model of working memory, the phonological loop is the store responsible for processing verbal information. Working memory for verbal information has been directly linked to the development of reading skills. Researchers have shown, for example, that word span in kindergarten is significantly and

moderately related to reading skill in first grade (Mann, 1984; Mann & Liberman, 1984).

Other researchers have found a moderate relationship between verbal working memory

(in addition to visual working memory) and word recognition (Swanson & Howell, 2001).

Wagner and Torgeson (1987) suggested that the efficiency of coding verbal information

in working memory is the origin of memory span differences between good and poor

readers.

Role of Skill Levels in Intervention Outcomes

While intervention design plays a key role in intervention outcomes including

maintenance and generalization, factors within student, including working memory

(above) and skill level, are likely to influence intervention outcomes as well. For

example, research has shown that students learning material that is correctly matched for

difficulty show higher rates of learning (Burns, 2002; Burns, 2007a; Gickling &

Armstrong, 1978; Shapiro, 1992). This has led to the designation of an instructional level

for connected text reading (Gickling & Armstrong, 1978) and drill tasks (e.g., Burns,

2004a; Gickling & Thompson, 1985).

While material that is the correct difficulty level may impact intervention

outcomes, proficiency in related skills may also impact intervention outcomes. For

instance, LaBerge and Samuels (1974) suggested that accuracy and automaticity in more

basic skills allow students to access cognitive resources to apply to higher-level skills.

LaBerge and Samuels (1974) applied this concept in detail to the facilitation of reading

comprehension. However, this concept can also be applied to other reading skills because

proficiency in decoding skills may facilitate word recognition (Samuels, 1988) and

success in word recognition interventions (Aaron et al., 1999). Decoding and sight word

reading are related and decoding skill likely facilitates successful sight word instruction (Aaron et al., 1999). Additionally, phonics interventions have facilitated word recognition (Joseph, 2002) and reading fluency when combined with repeated practice (Devault & Joseph, 2004). Therefore, the success of an intervention (such as a flashcard intervention) meant to enhance word recognition may depend partially on a student's level of decoding skill.

Conclusion

A substantial proportion of students read at a below basic level (NCES, 2011), and reading difficulties that begin in school often carry over into adulthood and have an impact on one's life outcomes (Kutner et al., 2007). Basic literacy skills must be developed at an early age to prevent difficulties later in life (e.g., Stanovich, 1986). Reading fluency is one critical basic reading skill; reading fluency itself depends on other basic skills including automatic word recognition (e.g., LaBerge & Samuels, 1974). While other basic reading skills are certainly critical intervention targets, word recognition is an important basic reading skill that will be targeted in this research.

IR is a flashcard technique (Tucker, 1989) and is the intervention used in the current study. Using a large number of OTRs as well as a high proportion of known items, IR has shown favorable outcomes for a large number of outcomes, including letter sounds (Volpe et al., 2011), math facts (e.g., Burns, 2004b), word recognition (Burns & Boice, 2009; Nist & Joseph, 2008), and reading fluency (Burns, 2002, 2007b). While a strong research base supports the use of IR, this study aimed to enhance outcomes obtained through IR by integrating theories of maintenance and generalization.

The body of research integrating theory into existing interventions is minimal, nonexistent in some cases. Moreover, many theories have limited or no research supporting its application in any applied setting, let alone in schools among children. This presents an opportunity to improve practice in addition to improving empirical support for theoretical constructs. Although many theories exist which may influence maintenance and generalization, this study focuses on three: the depth of processing framework for enhancing maintenance and Stokes and Baer's generalization framework and the theory of common elements for enhancing generalization. Each of these theories has a rich history in the literature, although their application to applied settings varies considerably.

The depth of processing framework is a procedural view of memory; in other words, it suggests that memory is a byproduct of the cognitive processes performed on a given piece of information (Craik & Lockhart, 1972). The particular cognitive process carried out influences the durability of a memory trace; more meaningful analyses will lead to greater durability (Craik & Lockhart, 2002). This study will accomplish a more meaningful analysis of the words taught through IR by integrating semantic information when teaching words. The transfer appropriate processing framework has brought about questions as to whether any type of processing is inherently meaningful (Morris et al., 1977). However, when traditional recognition or recognition within a meaningful context is desired, semantic processing is likely to create a meaningful and durable memory (Morris et al., 1977). The depth of processing is one framework that has not been researched in applied settings; in fact, a recent investigation by the author (Petersen-

Brown & Burns, 2011) marks an initial attempt to investigate the depth of processing framework within a school setting.

Stokes and Baer's (1977) generalization framework for generalization was an attempt to define a technology of generalization to bring together the variety of ways in which generalization had been previously programmed. This initial theoretical definition was accompanied by a lamentation that the majority of investigations do not program generalization intentionally, but instead use a "train and hope" approach. Kirby and Bickel (1988) and Stokes and Osnes (1989) have reorganized Stokes and Baer's framework, but the original remains the dominant framework used to date. Stokes and Baer's framework has been applied in research, but its explicit application to reading intervention research is primarily found in its application to the repeated reading intervention (e.g., Ardoin et al., 2007, 2008). In the current investigation, two methods, programming common stimuli and training multiple exemplars, are utilized to program generalization.

A separate theoretical construct from the cognitive literature was considered and suggested similar modifications to IR: the theory of common elements (Thorndike & Woodworth, 1901a, 1901b, 1901c). The theory of common elements is the oldest of the three theories utilized in this study, first suggested by Thorndike and Woodworth in 1901. Generally, the theory of common elements implies that transfer of learning from one situation to another is more likely if the individual perceives critical commonalities between a concept and the problem with which he or she is currently faced (Butterfield & Nelson, 1989). In this study, the theory of common elements is used to increase the common elements between training and the generalization setting.

While intervention design is undoubtedly an important consideration, individual characteristics, both cognitive and skill-based, likely play a role in intervention outcomes. The current study will examine one cognitive attribute, working memory, and one skill-based attributed, decoding skill. Working memory deficits are consistently identified among students with reading difficulties (e.g., Gathercole et al., 2006), and it is a key consideration in IR given the amount of information presented in a typical IR session. Decoding skills may facilitate word recognition (Samuels, 1988), and potentially success in word recognition interventions.

The purpose of the current study as outlined in the previous chapter is to investigate the effect of integrating theory with IR on intervention outcomes (specifically, integration of the depth of processing framework for enhancing maintenance and word knowledge and the integration of Stokes and Baer's generalization framework and the theory of common elements for enhancing generalization). This study also investigates efficiency of the modified versions of IR as well as the role of auditory working memory and decoding skill on intervention outcomes. To review, the current study will investigate whether adding components to IR impacts the efficiency of the intervention. In sum, the following research questions will be examined:

1. What is the effect of adding additional components to IR – a vocabulary component or contextual reading practice – on the maintenance and generalization of target words?

2. What is the relative efficiency of the following IR variations in words learned per minute: IR, IR with vocabulary, and IR with contextual reading practice?

3. To what extent does decoding ability moderate the impact of IR with vocabulary or contextual reading practice on maintenance and generalization?

4. To what extent does working memory capacity moderate the impact of IR with vocabulary or contextual reading practice on maintenance and generalization?

5. To what extent does knowledge of word meaning change from before teaching to 1 and 2 weeks after teaching across IR variations?

CHAPTER 3: METHOD

Participants and Setting

Assuming a moderate effect size ($f = .25$), a moderate correlation among repeated measures ($r = .50$), and 3 repeated measures, a total sample size of 28 was required to achieve power of .80. A total sample of 40 was desired to account for possible attrition. Forty-one second and third grade students participated in the current study. Thus, all 41 students participated through the entire study and no attrition was observed (although 1-week maintenance and generalization data was inadvertently not collected for one student for the IR-V condition and a different student for the IR-C condition). Students were selected with a clustered approach in which participants were recruited from four second-grade and three third-grade classrooms from one elementary school. Teachers distributed consent forms to students and their parents and collected signed consent forms. Potential student participants were also grouped according to their reading achievement quartile based on the Fall administration of the Measures of Academic Progress (MAP; Northwest Evaluation Association, 2011). The first 10 to 11 students from each quartile that returned their signed informed consent form were included in the study so that each quartile was equally represented. As consent forms were returned, teachers referred to a data management system and indicated each student's MAP quartile to the investigator. The investigator included the first 10 to 11 students to return consent forms in each MAP quartile. English language learners were excluded, but students receiving special education services were included.

A total of 10 participants from the first quartile of the Fall MAP administration, 10 from the second quartile, 11 from the third quartile, and 10 from the fourth quartile were included in the study for a total of 41. Thirty-two participants (78.0%) were in

second grade and 9 (22.0%) participants were in third grade. Moreover, 26 (63.4%) females and 15 (36.6%) males participated, and 30 participants (73.2%) were Caucasian, 5 (12.2%) were African American, 5 (12.2%) were Asian, and 1 (2.4%) was Hispanic. Although the selected participants included more Caucasian participants than would be expected based on the school's population, these differences were not significant (χ^2[3] = 3.19, p = .36).

The students were recruited from a suburban elementary and middle school in the Upper Midwest. The school had approximately 600 students in kindergarten through eighth grade; 38.5% received free or reduced lunch, 58.2% were Caucasian, 15.2% Asian, 20.5% African American, 5.2% Hispanic, and 0.9% from other ethnicities.

Six interventionists conducted intervention and assessment sessions. Five interventionists were graduate students in educational psychology at a local university. Each of these interventionists had prior experience delivering interventions to students in schools and implementing IR. One interventionist was an advanced undergraduate studying psychology at a local university who did not have prior experience delivering interventions to students or delivering IR.

Of the five graduate students, two interventionists were in the fourth year of a doctoral program in educational psychology; each of these interventionists had approximately four years of experience delivering interventionists to students in schools. One interventionist was in the second year of the educational psychology doctoral program and had approximately two years of experience delivering interventions to students in schools. One interventionist was in the first year of the same doctoral program and had one year of experience as a special education teacher and one year of experience

delivering interventions to students in schools. One interventionist was in the first year of an educational specialist program in educational psychology and had approximately one year of experience delivering interventions to students in schools.

Each interventionist was trained in the specific procedures of the current study (see training description below). Additionally, each interventionist delivered approximately the same number of intervention sessions across IR variants to control for teacher effects.

Independent Variables

Each participant participated in three variants of IR in three weekly sessions. The variants included IR, IR with vocabulary (IR-V), and IR with contextual reading practice (IR-C). Each IR variant used eight known words, and interventionists taught seven unknown words in each of the three IR sessions.

IR. The interventionist began by placing one unknown word in front of the eight known words. The interventionist presented the unknown word by saying, "This word is _____. What word is this?" When the participant replied by saying the word correctly, the interventionist replied, "Good. Can you use it in a sentence?" The participant then provided a sentence that included the unknown word (used semantically or syntactically correctly or incorrectly). Then, the participant read the unknown word and the first known word in the stack of flashcards. Next, the participant read the unknown word and the first two known words, followed by the unknown word and the first three known words. This procedure continued in this way until the participant read the unknown word and all eight known words. The last known word was then dropped from the stack, and the second unknown word was added to the top of the stack. The interventionist

presented the newly added unknown word, and the procedure was repeated. Seven

unknown words were taught to each participant. If the participant made an error in

reading a word at any time during the procedure, the interventionist corrected the

participant by saying, "This word is ____. What word is this?" When the participant said

the word correctly, the participant said, "Good," and continued with the procedure.

 IR-V. Interventionists conducted the IR-V procedure as described above under

the "IR" heading with the following modifications. The interventionist presented each

unknown word by saying, "This word is ____. _____ means ____. What is this word,

and what does it mean?" The author supplied the definitions written on the back of each

flashcard. When the participant correctly identified the unknown word and supplied its

meaning, the interventionist said, "Good. Can you use it in a sentence?" If the unknown

word was used in a sentence but in a syntactically or semantically incorrect way or if the

participant could not provide a sentence, the interventionist provided the sentence written

on the back of the flashcard. Whenever the participant saw the unknown word, he or she

read the word and repeated the definition. If the participant made an error in reading the

unknown word or in providing the definition, the interventionist corrected the participant

by saying, "This word is ____. _____ means ____. What is this word, and what does it

mean?" After a word was taught and was no longer at the front of the stack, the

interventionist did not ask the participant to supply the definition of the word any longer,

and the correction procedure did not include presentation of the word meaning. Therefore,

for each unknown word, the participant repeated the definition a minimum of eight times.

 IR-C. Interventionists conducted the IR-C procedure as described above under

the "IR" heading, with the following modifications. The author wrote the target word on

the front of each card, along with three sentences with a single clause containing the target word. Using "similar" as an example, the interventionist presented the unknown word by saying, "This word is 'similar.' What is this word?" When the participant replied by saying the word correctly, the interventionist replied, "Good. Now read 'similar' in the sentences on the card." Whenever the participant saw the unknown word at the front of the stack, he or she read the word and one of the three sentences on the card. After a word was taught and was no longer at the front of the stack, the interventionist did not ask the participant to read the sentences any longer. Therefore, for each unknown word, the participant read the word in context 10 times (three times when initially introduced and one time each in the remaining seven exposures).

Measures

   Dependent variables. Dependent variables included maintenance and generalization of taught words at one and two weeks, efficiency in words learned per minute, and knowledge of word meanings.

   Maintenance of taught words. Interventionists measured maintenance of taught words by showing participants the words they learned both 1 and 2 weeks prior on the same flashcards used during teaching. For words used in the IR-C condition, interventionists covered the sentences to avoid giving clues. If the participant was able to read the word correctly within 2s, the word was maintained. The word was not maintained if the participant required more than 2s to read the word correctly, said the word incorrectly, or did not say the word. The data for this variable consisted of the number of words maintained after 1 and 2 weeks.

Generalization of taught words. Interventionists measured generalization of taught words by showing participants the words they learned both 1 and 2 weeks prior within sentences (sentences were different from those used in the IR-C condition), a method used in previous research to measure generalization (Nist & Joseph, 2008). If the participant read the target word correctly within 2s, it was generalized. If the participant required more than 2s to read the word correctly, read the word incorrectly, or did not read the word, it was not generalized. The data for this variable consisted of the number of words generalized after 1 and 2 weeks.

Efficiency. Efficiency in each condition was measured in terms of words maintained at 1 week by total session time. During each IR session, the interventionist began timing the session as he or she presented the first unknown word and stopped timing the session immediately after completing the IR procedure. Efficiency was calculated by dividing the total words maintained at 1 week by the total number of minutes to deliver the session, which resulted in words maintained per minute.

Knowledge of word meanings. Interventionists recorded knowledge of word meanings prior to the first session. Additionally, at both 1 and 2 weeks after learning a set of unknown words, interventionists asked participants, "What does this word mean?" as interventionists were conducting the maintenance assessment. Interventionists recorded the participants' responses verbatim, and if participants defined a homonym, interventionists asked, "What else could this word mean?"

There were five levels of word knowledge defined based on frameworks suggested by Durso and Shore (1991). Unknown words received a score of 0 and were rated as such if they were either defined incorrectly or not defined (i.e., participant said,

"I don't know"). Partially known words received a score of 1 and were rated as such if the response reflected only the correct part of speech but did not reflect any critical misunderstandings (i.e., using "museum" as an example, the participant might have said, "It's an interesting thing," which reflects that it is a noun without reflecting critical misunderstandings). There were three levels of known words, (a) known with context, (b) known but lacking key details, and (c) known with a definition. Known words defined in context received a score of 2 and were rated as such when the participant stated an example or used the word in a meaningful sentence that reflected more detail than just the part of speech (i.e., using "museum" as an example, the participant might have said, "I go to the museum," which reflects that it is not simply a noun but a place as well; the participant could have also said, "an art museum," which is an example). Known words whose definitions lacked key details received a score of 3 and were rated as such when the participant gave a definition that lacked a key feature and was either very vague or included only ancillary features (i.e., using "museum" as an example, the participant might have said, "a place where you see stuff," which is vague and lacking a statement of what one might see at a museum). Known words were defined independent of a context with key features included, meaning that participants provided a correct definition beyond using it in a sentence or stating an example (i.e., using museum as an example, the participant might have said, "a place with lots of artwork"); these definitions received a score of 4. The data were the number of words receiving a given rating within a certain condition, across participants, at three separate time points (prior to teaching, 1 week after teaching, and 2 weeks after teaching).

Potential moderators. Based on the possibility that working memory and decoding skills may impact maintenance and generalization of unknown words, the following measures were administered and investigated as potential moderators in the study.

Working memory subtest. Interventionists administered the Auditory Working Memory (AWM) subtest of the Woodcock-Johnson III Tests of Cognitive Abilities (WJ III – COG; Woodcock, McGrew, & Mather, 2001) as a measure of auditory working memory. Only interventionists who had been trained in the administration and interpretation of the WJ III-COG battery administered this assessment and collected interobserver agreement (IOA) information. In the AWM subtest, participants hear a sequence of common nouns and numbers (e.g., bread 9 tub 1 5). The administrator instructs them to say the words first in the order they heard them followed by the numbers in the order they heard them (e.g., bread tub 9 1 5). Median internal consistency reliability of the auditory working memory is .87 overall, and internal consistency is .90, .90, and .89 for seven-, eight-, and nine-year-olds, respectively (McGrew & Woodcock, 2001). The data consisted of the standard score ($M = 100$, $SD = 15$) based on grade level.

Decoding subtest. Investigators administered the Word Attack (WA) subtest of the Woodcock-Johnson III Tests of Achievement (WJ III – ACH; Woodcock et al., 2001) as a measure of decoding skill. Once again, only interventionists who had been trained in the administration and interpretation of the WJ III – ACH battery administered this assessment and collected IOA information. In the WA subtest, participants read a series of nonsense words with a variety of phonetic features. Median internal consistency

reliability is .87 overall, and internal consistency is .92, .92, and .89 for seven-, eight-,

and nine-year olds, respectively (McGrew & Woodcock, 2001). The data were the

standard score ($M = 100$, $SD = 15$) based on grade level.

Procedure

      Training interventionists. Prior to the beginning of the study, the primary

investigator conducted a 90-minute training session with the data collectors in which she

reviewed the procedures for IR, the IR variants used in the current study, and other study

procedures (e.g., assessing maintenance and generalization and timing sessions). Data

collectors were blind to the hypotheses or theoretical bases for IR variants, and were told

that different variations on IR were being investigated. The training period included

opportunities for data collectors to practice implementing IR procedures. Following the

training session, the data collectors implemented each IR variant while the primary

investigator observed and feedback provided as needed until each data collector correctly

implemented 100% of steps listed on the fidelity checklists (described below).

      Locating words. The primary investigator selected 25 words as potential known

words and 180 different words of varying difficulty as potential unknown words. The

investigator selected all words from the 1,000 most common English words and

elementary content words in math, social studies, geography, and science (Fry & Kress,

2006). These lists contain words that participants are likely to encounter when reading a

book or content area materials, and they are therefore considered to be high utility.

      A total of 25 potential known words were randomly selected from the first 100

words in the common word list (Fry & Kress, 2006) because they are the most common

words used in written material. Thus, it was assumed that a majority of second and third

graders would be able to recognize these words quickly and easily. Only 8 known words were needed to conduct intervention sessions, but the investigator chose additional words (25 total, up to 17 more words than needed) to ensure all participants would know at least 8 words.

The investigator randomly selected the 180 potentially unknown words from the 501st word through the 1,000[th] word, and from the elementary content area lists. First, 20 words from each set of 100 words from the 501st to the 1000th in the common word list were randomly selected. In addition, 20 words each from the Elementary Geography, Elementary Math, Elementary Science, and Elementary Social Studies word lists were randomly selected as potential unknown words. Words were included if they could be succinctly defined, generally in under 10 words. After selecting the unknown words, they were piloted with a high-achieving third grade student; the student knew all but 4 of the 180 potentially unknown words. Based on this, 24 randomly selected words from the Intermediate Geography, Intermediate Math, Intermediate Science, and Intermediate Social Studies word lists were added to ensure that interventionists would be able to identify enough unknown words for high-achieving participants.

Word length has previously been connected to decoding difficulty, particularly for struggling readers (Manis, 1985). Additionally, measures of word length are often associated with word difficulty in readability indices, measured either by the number of letters (Coleman & Liau, 1975; Smith & Senter, 1967;) or syllables (Fry, 1968; Gunning, 1952; Kincaid, Fishburne, Rogers, & Chissom, 1975). The list of words selected for the current study had a mean word length of 6.80 letters (range = 3-13) and a mean syllable length of 2.27 (range = 1-5). Because correlation between number of letters and syllables

was quite high ($r = .87$), the number of letters was used as a measure of difficulty due to the greater variation. Words for the current study were sorted into the following 6 groups: 3- and 4-letter words (e.g., son, meet), 5-letter words (e.g., equal), 6-letter words (e.g., voyage), 7-letter words (e.g., polygon), 8-letter words (e.g., boundary), and 9- to 13-letter words (e.g., tradition, evaporation).

Materials. The investigator created three sets of flashcards using the 204 potential unknown words. For the IR condition, the investigator typed unknown words on 3 X 5 cards using 40-point, sans-serif font. For the IR-V condition, the investigator typed unknown words on 3 X 5 cards using 40-point, sans-serif font. On the back of each unknown word card for the IR-V condition, the primary investigator typed a short definition of the word as well as a sentence including the target word for use in teaching. Definitions were short dictionary definitions or synonyms that were 10 words or fewer in length. Sentences were also written on the back of each card to account for the possibility that the participant may be unable to create a sentence correctly using the word. Sentences were written to imply the word's definition, for example, "When the movie *commences*, it is time to be quiet."

For the IR-C condition, the investigator typed unknown words on 4 X 6 cards using 40-point, sans-serif font. The investigator typed three short sentences containing the target word below the word using 24-point, sans-serif font. Including multiple sentences with the target word is an example of training multiple exemplars from Stokes and Baer's (1977) framework. A previous investigation showed that three exemplars was effective in facilitating generalization and was not significantly less effective than five exemplars (Hupp, 1986). Additionally, recent research that used multiple exemplars to

enhance generalization used three exemplars and was successful in increasing

generalization (Ardoin et al., 2008). It is also an example of programming common

stimuli or common elements, because seeing the word in context is an element or stimuli

common to the teaching and generalization settings. The investigator did not write

definitions and sentences on the back of the cards used for the IR-C condition.

Determining session order. Session order and words to be taught during each

session were randomly determined. Word list forms were used with the condition order at

the top and space for 21 unknown words and participant-provided definitions below.

Condition order was randomly selected on each of the 41 forms. Next to each unknown

word space, the investigator randomly selected the condition in which each word would

be taught so that there were seven randomly selected unknown words for each condition.

First session. Prior to beginning the first session, the student assent form was

read to each student and each verbally provided assent and signed the form. Next,

interventionists completed the WA subtest. Third, interventionists selected unknown

words; this began by asking the participant to read words beginning with the list of six-

letter words. Words were identified as unknown each time a participant read a word

incorrectly or took longer than 3s to read the word. If the participant read six words in a

row correctly, the interventionist progressed to the next list of longer words. If the

participant read six words in a row incorrectly, the intervention progressed to the next list

of shorter words. This process continued until the interventionist identified 21 unknown

words. Next, the interventionist read each word to the participant and asked him or her,

"What does this word mean?" If the participant defined a homonym of the word, the

interventionist asked, "What else could the word mean?" Next, the interventionist located

the IR variant to be used in the first session at the top of the form and located the flashcards that were randomly assigned to that condition. Finally, the interventionist conducted the IR variant randomly selected to occur during the first session.

Second session. During the second session, the interventionist presented the participant with a set of seven sentences; each sentence included one of the seven words taught the previous week. The participant read the sentences. If participants experienced any hesitation reading any non-target words, the interventionist supplied the word. If participants had trouble reading the target word, they were asked to try their best and then move to the next word. Next, the interventionist presented the flashcards used the previous week and asked the participant to read the word as a measure of maintenance. After showing each flashcard, the interventionist asked, "What does this word mean?" If the participant defined a homonym of the word, the interventionist asked, "What else could the word mean?" Next, the interventionist conducted the condition that was randomly selected to occur during the second session.

Third session. During the third session, the interventionist first administered the AWM subtest. Then, the interventionist presented the participant with a set of 14 sentences; each sentence included one of the seven words taught the first week or one of the seven words taught the second week. The participant read the sentences and the interventionist proceeded as described above in the Measures section. Next, the interventionist presented the 14 flashcards taught during the first and second sessions and asked the participant to read the word as a measure of maintenance. After showing each flashcard, the interventionist asked, "What does this word mean?" If the participant defined a homonym of the word, the interventionist asked, "What else could the word

mean?" Next, the interventionist conducted the condition that was randomly selected to occur during the third session.

Fourth session. During the fourth session, the interventionist presented the participant with a set of 14 sentences; each sentence included one of the seven words taught the second week or one of the seven words taught the third week. The participant read the sentences and the interventionist proceeded as described above in the Measures section. Next, the interventionist presented the 14 flashcards taught during the second and third sessions and asked the participant to read the word as a measure of maintenance. After showing each flashcard, the interventionist asked, "What does this word mean?" If the participant defined a homonym of the word, the interventionist asked, "What else could the word mean?"

Fifth session. During the fifth session, the interventionist presented the participant with a set of seven sentences; each sentence included one of the seven words taught the third week. The participant read the sentences and the interventionist proceeded as described above in the Measures section. Next, the interventionist presented the seven flashcards taught during the third session and asked the participant to read the word as a measure of maintenance. After showing each flashcard, the interventionist asked, "What does this word mean?" If the participant defined a homonym of the word, the interventionist asked, "What else could the word mean?"

Research Design

The study used a within-subjects group design. There were three repeated measures in this study, corresponding to the three IR variants used in the current study. Dependent variables included maintenance and generalization of taught words. Each

participant participated in each IR variant in a randomly selected order; these were the independent variables. In addition, two potential moderators (decoding skills and auditory working memory) were identified and measured for each participant as described above.

Implementation Fidelity and Interobserver Agreement

Information on implementation fidelity was collected during 26.8% of sessions across conditions, or 11 out of 41 sessions for each IR variant. Between 21.1% and 50.0% of sessions were observed for each interventionist. A 17-item checklist was used that detailed both general IR procedures (e.g., last known word is removed when a new unknown word is added) and procedures specific to the current study (e.g., asking the participant to read the word in a sentence during IR-C, starting a timer once the interventionist started teaching the first word). Observers included interventionists and one other student in the same educational psychology doctoral program who was trained on the study's procedures. Mean implementation fidelity was 99.2% (range = 94.1% to 100%). The only step interventionists missed during observed sessions was starting a timer when the interventionist began teaching the first word.

IOA was collected on maintenance and generalization at one and two weeks and on AWM and WA subtests. Overall, observers collected IOA on 28.9% of maintenance and generalization sessions, 30.1% of one-week maintenance and generalization sessions and 27.6% of two-week maintenance and generalization sessions. Average IOA was 97.5% (range = 71.4% to 100%) across all maintenance and generalization sessions; IOA for maintenance (98.7%) was slightly higher than IOA for generalization (96.3%). When reasons for disagreements were cited, it was based on whether the participant said the word correctly within 2s and slight variations in pronunciation.

Observers trained in the administration and interpretation of the WJ III-COG subtest collected IOA on 24.3% of AWM subtests. Average agreement was 96.6% (range = 79.2% to 100%). Observers trained in the administration and interpretation of the WJ-ACH collected IOA on 26.8% of WA subtests. Average agreement was 95.0% (range = 87.5% to 100%).

Interrater reliability was also collected on 25% of the word definitions. The primary investigator trained a second rater, who conducted intervention sessions in the study, on the definition rating scheme. The primary investigator and the second rater established reliability by first going through ratings together; then they rated a set of definitions independently and discussing the respective ratings. The primary investigator then randomly selected 25% of the word definitions from each of the three time points (pretest, 1 week, and 2 weeks) for interrater reliability purposes. The primary investigator calculated a kappa coefficient to account for chance agreement, $k = .72$. Values between .60 and .80 are considered good (Landis & Koch, 1977).

Data Analyses

To answer the first research question regarding the effect of adding contextual reading practice or vocabulary on maintenance and generalization, the primary investigator used a one-way repeated measures analysis of variance (ANOVA), with three repeated measures (one for each condition) and one group. Planned analyses were used to identify significant differences between conditions.

The second research question, which addressed efficiency of the three different IR variations, was answered by calculating efficiency as described above to compute a words-maintained-per-instructional-minute score. Significant differences in efficiency

were examined with a one-way repeated measures ANOVA. Finally, a planned analysis was conducted to identify significant differences in efficiency between conditions.

A one-way ANOVA was used to answer the third research question regarding decoding skill as a moderator of intervention effects. First, WA standard scores were categorized as low (standard score 99 and under), moderate (standard score between 100 and 114), and high (standard score 115 and higher). The difference between words maintained and generalized between IR-V and IR and IR-C and IR was calculated for each participant. A one-way ANOVA was used to determine whether overall differences between conditions were significantly different based on the level of decoding skill.

A one-way ANOVA was used to answer the fourth research question regarding auditory working memory capacity as a moderator of intervention effects. First, AWM standard scores were categorized as low (standard score 99 and under), moderate (standard score between 100 and 114), and high (standard score 115 and higher). The difference between words maintained and generalized between IR-V and IR and IR-C and IR was calculated for each participant. A one-way ANOVA was used to determine whether overall differences between conditions were significantly different based on the capacity of auditory working memory.

Data for the fifth research question, regarding increases in word knowledge across IR variants, were analyzed using chi-squared analyses. Word knowledge was rated into five categories, with 0 indicating an unknown word and 4 indicating the provision of a correct definition of the word, as discussed above. The number of definitions rated 0, 1, 2, 3, and 4 were tabulated prior to teaching, at 1 week after teaching, and at 2 weeks after teaching within each IR variant. Chi-squared analyses were used to compare the number

of definitions receiving each rating across conditions within time points. Chi-squared

analyses were not used to compare the number of definitions receiving each rating across

time points within a particular condition. These data did not meet the independence

assumption because the same words are represented at each time point. Thus, word

knowledge was dichotomized (with a rating of 0 indicating no word knowledge and a

rating of 1, 2, 3, or 4 indicating some word knowledge) and compared across points using

Cochran's Q tests (one test was used for each condition).

CHAPTER 4: RESULTS

Descriptive statistics for the primary variables are included in Table 1. Overall, maintenance and generalization across all conditions were relatively high. Additionally, maintenance and generalization were generally negatively skewed and leptokurtic, but not all values exceeded an absolute value of 2. In terms of session length, IR was generally the shortest, followed by IR-V and IR-C, which were nearly equal (see Table 1). Session length was positively skewed and leptokurtic across all three conditions, indicating that session length did not vary considerably, but a few sessions took considerably longer than the mean. Descriptive statistics for AWM and WA are shown in Table 2. AWM standard scores ranged from 91 to 134 and averaged 109, showing that the group was at the highest end of the average range as compared to same-grade peers. WA standard scores ranged from 83 to 132 and averaged 103. AWM and WA standard scores were nearly normally distributed as indicated by skewness and kurtosis values near 0.

Bivariate relationships between the variables were next examined using Pearson correlation coefficients. Correlations between the various measures of maintenance and generalization across conditions were generally significant and moderate to strong (ranging from .62 to .81). Correlations between the covariates and maintenance and generalization are shown in Table 3. The relationships between AWM standard scores and maintenance and generalization were positive, but weak and nonsignificant. The relationships between WA standard scores and maintenance and generalization were positive and generally moderate and significant. Session length was negatively and moderately correlated with WA standard scores, but it was not significantly correlated

with AWM standard scores. Finally, the correlation between WA and AWM standard scores was positive and moderate.

Maintenance and generalization at 1 and 2 weeks

The first research question inquired about maintenance and generalization at 1 and 2 weeks, which was analyzed with a repeated measures ANOVA, with condition as the within-subjects factor, to compare 1- and 2-week maintenance and generalization across the three IR variants. For 1-week maintenance, Mauchly's Test of Sphericity was not significant, $\chi^2(2) = 1.53$, $p = .47$. There was a significant difference between the three IR variants on 1-week maintenance, $F(2, 76) = 3.54$, $p = .03$. Partial eta-squared was moderate, $\eta^2 = .09$. Post-hoc pairwise comparisons using a Bonferroni correction ($p = .017$) suggested that 1-week maintenance was higher in IR-V and IR-C than IR, though not significantly using the adjusted p-value ($p = .02$ and $p = .03$, respectively). Cohen's d effect sizes were small, as shown in Table 1. There was also no significant difference between IR-V and IR-C. Because 1-week maintenance did not meet the normality assumption (particularly as a result of kurtosis values far above 2.0), Friedman's test of ranks was also used to validate the results of the ANOVA. Although results were nonsignificant, $\chi^2(2) = 3.76$, $p = .15$, the order of effects was the same – mean rank for IR was 1.79, 2.14 for IR-V, and 2.06 for IR-C.

For 1-week generalization and 2-week maintenance, the repeated measures ANOVA was not significant. For 1-week generalization, Mauchly's Test of Sphericity was not significant, $\chi^2(2) = 0.75$, $p = .69$. The difference between the three IR variants was not significant, $F(2, 76) = 2.55$, $p = .09$. Effect sizes showing the difference between IR-V and IR and IR-C and IR were small (see Table 1). Again, results were verified using

Friedman's test of ranks as a result of kurtosis values greater than an absolute value of 2.0. Results were nonsignificant using this test, $\chi^2(2) = 5.59$, $p = .06$, and the direction of effects was the same as the parametric test – mean rank for IR was 1.76, 2.21 for IR, and 2.04 for IR-C. For 2-week maintenance, Mauchly's Test of Sphericity was not significant, $\chi^2(2) = 0.28$, $p = .87$. The difference between the three IR variants was not significant, $F(2, 80) = 1.27$, $p = .29$. Effect sizes comparing IR-V to IR and IR-C to IR were near zero. Results were verified using Friedman's test of ranks as a result of skewness and kurtosis values exceeding an absolute value of 2.0. Results were nonsignificant using this test, $\chi^2(2) = 3.47$, $p = .18$. The direction of effects was the same as with the ANOVA, with the mean rank for IR being 2.05, 2.12 for IR-V, and 1.83 for IR-C.

For 2-week generalization, Mauchly's Test of Sphericity was not significant, $\chi^2(2) = 1.11$, $p = .58$. There was a significant difference between the three IR variants, $F(2, 80) = 3.63$, $p = .03$. Partial eta-squared was moderate, $\eta^2 = .08$. Post-hoc pairwise comparisons, using a Bonferroni-adjusted significance level of $p = .017$, suggested that 2-week generalization was significantly higher in IR-V than IR ($t(40) = 2.72$, $p = .01$), but there was no significant difference between IR-C and IR-V or IR-C and IR. Results were verified using Friedman's test of ranks as a result of skewness and kurtosis values exceeding an absolute value of 2.0. Results were significant using this test, $\chi^2(2) = 6.78$, $p = .03$. The effect sizes comparing IR-V and IR and IR-C and IR were both small as shown in Table 1. The direction of effects was the same, with the mean rank for IR being 1.76, 2.21 for IR-V, and 2.04 for IR-C.

Words learned per minute of instructional time

The second research question addressed efficiency of the different IR approaches. Efficiency was defined as the number of words maintained at 1 week after instruction per minute of instructional time. Words maintained per minute of instructional time were greater in IR ($M = 0.55$, $SD = 0.17$) than in IR-V ($M = 0.41$, $SD = 0.11$), $d = 1.02$, and IR-C ($M = 0.41$, $SD = 0.13$), $d = 0.94$. Differences in efficiency were compared using a repeated-measures ANOVA with condition as the within-subjects factor. Mauchly's test of sphericity was significant, $\chi^2(2) = 12.69$, $p = .002$. $F(1.54, 57.05) = 31.08$, $p = <.001$. Because the test of sphericity was significant, a non-parametric analysis was used. The Friedman's test of ranks was significant, $\chi^2(2) = 31.16$, $p < .001$ with IR being more efficient than the other two variants. The mean rank for IR was 2.74, 1.58 for IR-V, and 1.68 for IR-C.

Role of Working Memory and Decoding Skills

The third research question inquired about the role of working memory and decoding skills in the two outcome variables, for which the repeated-measures ANOVA was significant: 1-week maintenance and 2-week generalization.

Auditory working memory. Participants were grouped by level of AWM capacity: low (standard score up to 99, $n = 6$), medium (standard score from 100 to 114, $n = 23$), and high (standard score of 115 or above, $n = 12$). The difference between 1-week maintenance in IR-V versus IR and IR-C versus IR was calculated for each participant. Average differences across participants are shown in Figure 1. Overall differences were compared across the three AWM groups. Differences across AWM groups were not significant overall for IR-V versus IR, 1-week maintenance ($F[2, 37] = 0.10$, $p = .90$), and this effect was small (Cohen, 1988), $\eta^2 = .01$. Differences across AWM groups were

not significant overall for IR-C versus IR, 1-week maintenance ($F[2, 37] = 0.91$, $p = .41$). This effect was small to moderate, $\eta^2 = .05$.

Next, the difference in 2-week generalization in IR-V versus IR and IR-C versus IR was calculated for each participant. Average differences across participants are shown in Figure 1. Overall differences were compared across the three AWM groups using an ANOVA. Differences across AWM groups were not significant overall for IR-V versus IR, 2-week generalization ($F[2, 38] = 1.06$, $p = .36$), and this effect was small to moderate, $\eta^2 = .05$. Differences across AWM groups were not significant overall for IR-C versus IR, 2-week maintenance ($F[2, 38] = 0.62$, $p = .54$), and this effect was small, $\eta^2 = .03$.

Decoding skills. Similar to AWM, participants were grouped by level of WA skill: low (standard score up to 99; $n = 12$), medium (standard score from 100 to 114; $n = 20$), and high (standard score of 115 or above; $n = 8$). The difference between 1-week maintenance in IR-V versus IR and IR-C versus IR was calculated for each participant. Differences are shown in Figure 2. Overall differences were compared across the three WA groups using an ANOVA. Differences across WA groups were not significant overall for IR-V versus IR, 1-week maintenance ($F[2, 37] = 2.26$, $p = .12$). The effect size was moderate to large, $\eta^2 = .11$. Post-hoc comparisons suggest that the difference between 1-week maintenance through IR-V and IR was significant for participants with low decoding skills (mean difference = 0.97, $p = .049$), but not for participants with medium or high decoding skills. Differences across WA groups were not significant overall for IR-C versus IR, 1-week maintenance ($F[2, 37] = 0.41$, $p = .86$). In addition, the effect was small, $\eta^2 = .01$.

Next, the difference in 2-week generalization in IR-V versus IR and IR-C versus IR was calculated for each participant. Differences are shown in Figure 2. Overall differences were compared across the three WA groups using an ANOVA. Differences across WA groups were not significant overall for IR-V versus IR, 2-week generalization (F[2, 38] = 0.31, p = .74), and this effect was small, η^2 = .02. Differences across WA groups were not significant overall for IR-C versus IR, 2-week maintenance (F[2, 38] = 0.67, p = .52), and this effect was small to moderate, η^2 = .03.

Word Knowledge

Word knowledge was rated into five categories, with 0 indicating an unknown word and 4 indicating the provision of a correct definition of the word, as discussed in Chapter 3. The number of definitions rated 0, 1, 2, 3, and 4 were tabulated prior to teaching, at 1 week after teaching, and at 2 weeks after teaching. Information on word knowledge ratings is shown in Table 4. Chi-squared analyses were used to compare the proportion of each rating across conditions within time points because the data were ordinal. Cochran's Q analyses were used to investigate changes in word knowledge across time points within conditions because the data were dependent across time points.

Work knowledge differences were compared across IR variants at each time point using chi-squared analyses. Prior to teaching, word knowledge did not differ across conditions, χ^2 (8) = 6.43, p = .60. At 1 week after teaching, word knowledge did differ across conditions, χ^2 (8) = 60.23, p < .001. Finally, at 2 weeks after teaching, word knowledge differed across conditions, χ^2 (8) = 56.99, p < .001. Cochran's Q tests were used to compare word knowledge across time points for each IR variant. To conduct these analyses, word knowledge was dichotomized as described in the Methods chapter.

Across time points within IR, word knowledge significantly changed, $Q(2) = 13.32$, p

$= .001$. Across time points within IR-V, work knowledge significantly changed, $Q(2) =$

82.49, $p < .001$. Finally, across time points within IR-C, word knowledge significantly

changed, $Q(2) = 41.03$, $p < .001$.

CHAPTER 5: DISCUSSION

Maintenance at 1 week was found to be greater in IR-V and IR-C than in IR ($d =$ 0.37 and 0.27, respectively). Further, generalization at 2 weeks was found to be greater in IR-V than in IR ($d = 0.31$). Despite some nonsignificant results, effects were generally small in favor of IR-V and IR-C across outcomes. However, IR was found to be more efficient than IR-V ($d = 1.02$) and IR-C ($d = 0.94$) in words maintained at 1 week per minute of instructional time. This should be balanced with the finding that word knowledge increased through IR-V and IR-C. Finally, some evidence of a moderating effect of decoding skills was found related to 1-week maintenance in which participants with low decoding skills were found to maintain significantly more words at 1 week through IR-V than IR while this benefit was not as pronounced among participants with average and high decoding skills. However, these moderating effects were not consistent across other outcomes.

Effect of Additional Components to IR

Both maintenance and generalization were relatively high across conditions, with median values at the maximum value of seven words in many cases. This suggests that IR was an effective intervention for enhancing maintenance and generalization of taught words, regardless of modification. Although maintenance and generalization were high across conditions, both outcomes were generally higher in the IR-V and IR-C conditions, with small effect sizes across outcome variables. These effects were statistically significant in the case of 1-week maintenance and 2-week generalization. One-week maintenance was significantly higher in IR-V and IR-C than IR ($d = 0.37$ and 0.27, respectively), although these differences were not significant once a Bonferroni

correction was applied to account for the three comparisons. Two-week generalization was significantly higher in IR-V than IR ($d = 0.31$), although the difference between IR-C and IR was not significant ($d = 0.21$).

In general, the depth of processing framework (Craik & Lockhart, 1972), the theory of identical elements (Thorndike & Woodworth, 1901a, 1901b, 1901c), and Stokes and Baer's (1977) generalization framework were all supported in this study, although outcomes were not statistically significant in all cases and effect sizes were small. Specifically, Craik and Lockhart's (1972) suggestion that deeper processing leads to greater memory durability was supported. Additionally, Craik and Tulving's (1975) identification of three levels of processing with semantic or contextual processing being deeper than phonemic processing was also supported. Causal mechanisms were not tested in this study, although encoding elaboration (e.g., Hannon & Craik, 2001), distinctiveness (e.g., Moscovitch & Craik, 1976), and knowing versus remembering (e.g., Sheridan & Reingold, 2012) are all plausible mechanisms underlying these results. In addition, Stokes and Baer's (1977) suggestions to program common stimuli and train multiple exemplars were also supported, although effect sizes were small. In addition, the research that was conducted with the repeated reading intervention and training multiple exemplars (Ardoin et al., 2008; Ardoin et al., 2007; Silber & Martens, 2010) was also supported.

IR-V was meant to enhance maintenance in particular (although some resulting increase in generalization was hypothesized) and IR-C meant to enhance generalization in particular. However, this differentiation was not observed. IR-V showed a similar impact on maintenance and generalization; likewise, IR-C showed a similar impact on

maintenance and generalization. One possibility is that IR-V resulted in enhanced

maintenance through the depth of processing theory, and a corresponding increase in

generalization resulted as well as hypothesized. IR-C may have also oriented participants

toward a deeper level of processing (the deepest level has been defined as semantic or

contextual; Craik & Tulving, 1975) and thus resulted in similar effects. Maintenance and

generalization were nearly equal across conditions in this case, and it is not likely that

generalization would ever be higher than maintenance. Therefore, any differential effect

of IR-C on generalization would have been difficult to detect in the current investigation.

Ceiling effects may have also played a role in detecting any differential effects of the IR

modifications on maintenance versus generalization.

Future research in this area might investigate the role of the depth of processing

framework with other types of information previously taught through IR (such as math

facts or letter sounds), with other populations (such as students struggling with word

recognition), or in other interventions. In addition, this study could be implemented so

that the potential for ceiling effects is reduced (possibly through longer retention intervals

or by presenting more words) to provide greater potential for detecting effects of

intervention modifications.

Relative Efficiency of IR Variants

IR-V and IR-C took considerably longer and had slightly better results (small

effect sizes, generally). Effect sizes comparing words learned per minute were large for

IR-V and IR-C, with standard IR showing greater words learned per minute ($d = 1.02$ and

0.94, respectively). However, it is worth noting that IR-V had the added advantage of

significantly increasing word knowledge. It is difficult to quantify this benefit, but it certainly is an added benefit of the IR-V modification.

Previous research identified IR as the least efficient in words retained 1 day later per minute of instructional time (Nist & Joseph, 2008), while another study found that IR was equal in efficiency to traditional drill when considering words retained 1 week later (Burns & Sterling-Turner, 2010). Based on this, it is likely that IR with additional components would be less efficient than other flashcard methods such as traditional drill since it is less efficient than IR. However, given the results of Burns and Sterling-Turner (2010), it is possible that these findings may change over longer retention intervals. Nist and Joseph (2008) suggested that IR may be the best procedure when students do not respond to other flashcard methods; likewise, IR-V and IR-C may be useful when students have not responded to other flashcard interventions.

Future research may also investigate whether IR-V can be reconfigured to be more efficient (for example, by using fewer known words and therefore fewer OTRs) while still yielding results that are favorable in an absolute and relative sense. Although outcomes would be expected to decrease given that OTRs have been previously identified as a potential causal mechanism in IR (Burns & Szadokierski, 2008), this effect may be eliminated or reduced by using the depth of processing framework through IR-V. In this way, each OTR within IR-V could be said to be "higher quality" because it includes semantic information that according to the depth of processing theory leads to a more durable memory trace.

Decoding and Working Memory as Moderators for Effects of IR

Auditory working memory was not related to outcomes, but word attack skills showed a moderate and significant relationship to outcomes. Perhaps the small relationship between maintenance or generalization and working memory could be attributed to all participants having average to above average working memory skills on the selected assessment. Students with poor working memory may have had greater difficulties processing and later recalling target words. This hypothesis is in keeping with previous research showing a moderate relationship between working memory and word recognition (Swanson & Howell, 2001) as well as research finding that working memory capacity is related to reading skill (e.g., Swanson, 1994).

The moderate relationship between decoding skills and outcomes highlights the importance of decoding skills for word recognition. Perhaps decoding skills facilitate processing of target words during teaching, or perhaps decoding skills enable students to spontaneously decode a word that was actually not recalled as hypothesized. Perhaps decoding skills play a role on both ends, facilitating processing and recall as well as compensating for a lack of recall. This is consistent with LaBarge and Samuels' (1974) theory of automatic information processing, which suggests that automaticity in more basic skills (in this case, decoding) facilitates the application of cognitive resources to more complex skills (in this case, word recognition). It is also consistent with research suggesting that decoding skills facilitate word recognition (e.g., Aaron et al., 1999; Samuels, 1988).

When words maintained and generalized were compared across IR variants for participants with low, average, and high working memory and decoding skills, a potential moderating effect of decoding skills was found on the relationship between IR variant

and 1-week maintenance. Specifically, participants with low decoding skills benefitted more from IR-V as compared to IR than participants with medium or high decoding skills. Students with low decoding skills may not have been able to spontaneously decode words as readily as those with average or high decoding skills; thus, deeper processing may have facilitated enhanced retention for these students.

Future research could examine IR modifications with students who have low levels of decoding skills to confirm whether these students in particular benefit from theory-based modifications such as IR-V. Care should be taken to ensure these results are maintained over time, since poor reading skills have been previously associated with poor working memory capacity as discussed above (e.g., Swanson, 1994).

Effects of IR Modifications on Word Knowledge

Word knowledge was not significantly different across conditions prior to teaching, although it did significantly differ at 1 and 2 weeks after teaching, with IR-C and IR-V showing higher levels of word knowledge. Although statistical significance was obtained for all three IR variants when comparing word knowledge across time, word knowledge appears to have changed most in IR-V and IR-C, as shown in Table 4. Although IR-V and IR-C were less efficient than IR in words learned per minute of instructional time, the addition of word knowledge through IR-V and IR-C adds value that should be considered when evaluating the efficiency and effectiveness of IR-V as stated above.

Although it is perhaps not surprising that word knowledge increased through IR-V especially, this effect has not been investigated through previous research. This may be a result of the predominance of research with adults in controlled settings in which

participants know the words to be remembered (e.g., Craik & Tulving, 1975). Additional research would be needed to see if IR-V results in word knowledge that transfers to generalized settings such as comprehension of text. Another possible area for additional research is investigating IR-V as a complement to vocabulary instruction.

Additional Exploratory Findings

Two-week outcomes were found to be greater than 1-week outcomes. In other words, a brief OTR 1 week after teaching prevented reductions in (and seems to have increased) maintenance over time, which is consistent with previous IR research (MacQuarrie et al., 2002) and with the concept of a testing effect (e.g., McDaniel, Anderson, Derbish, & Morrisette, 2007; McDaniel & Masson, 1985). However, another interpretation speaks to the impact of distributed practice in the sense that many OTRs conducted in one session resulted in lower outcomes one week later than one "reminder" OTR. An additional interpretation of this finding is that practice in recalling information is a powerful aide to future recall, possibly even more powerful than additional practice (e.g., Karpicke & Roediger, 2007; Roediger & Karpicke, 2006).

Implications for Research

The current study has potential implications for future research. First, subsequent research could investigate the impact of the depth of processing framework on outcomes associated with IR for other types of information such as math facts and letter sounds, with other interventions targeting word recognition, and with other populations such as students with disabilities or students with low decoding skills. Likewise, researchers could investigate the influence of the theory of common elements and Stokes and Baer's (1977) generalization framework on outcomes associated with IR for other types of

information, with other interventions targeting word recognition, and with other populations. Particularly in the case of the depth of processing framework and the theory of common elements, little to no research exists to support its use in an applied setting. Future research should also investigate increases in word knowledge that result from integration of semantic properties when learning target words. In other words, does this word knowledge translate into enhanced understanding of words in context, or does it simply yield declarative knowledge of word definitions?

Second, the efficiency of the methods used in the current investigation was called into question by the small and generally non-significant effect sizes in the current investigation in addition to significantly greater time spent to implement the modifications. However, these challenges may also present an opportunity for future research. Past research has linked the outcomes associated with IR to the high number of OTRs (Szadokierski & Burns, 2008), although the number of OTRs increases the time required to implement the intervention. It is possible that a theory-based modification may provide "higher quality" OTRs. Thus, future research could investigate the use of theory-based modifications paired with fewer OTRs (perhaps by reducing the number of known words used in IR) to understand the impact on outcomes and efficiency compare to conventional IR.

Future research may also further investigate the role of cognitive and skill-based factors in intervention effectiveness. For instance, what is the nature of the relationship between decoding skills and outcomes of interventions targeting word recognition, including IR? Future research may aim to define this relationship as well as potentially identifying a level of decoding skill that may be desired before attempting word

recognition interventions. Additionally, the role of working memory capacity in intervention outcomes was not supported in this study although past research has shown a relationship between working memory and performance in word recognition interventions (Swanson & Howell, 2001). Future research may further investigate the impact of working memory capacity on IR and other interventions that rely on working memory capacity.

Implications for Theory

Much research has been devoted to building the evidence base for a wide variety of educational interventions, documented through resources such as the What Works Clearinghouse (Institute of Education Sciences, 2013). However, much of intervention research focuses on measuring focal outcomes and comparing gains between intervention conditions. While research comparing interventions is important, research that focuses on thoughtful modifications to a single intervention may also be helpful in improving practice. Utilizing theory is one way in which thoughtful modifications may be made. However, many theories have a strong research base in controlled settings but have not been applied in educational settings. This study is an attempt to use a research-based intervention (IR) as a foundation for applying cognitive and behavioral theory to enhance intervention outcomes while at the same time building the research for theories implemented in an applied setting.

Considerable evidence has established the effectiveness of IR for multiple purposes (including teaching math facts and letter sounds; e.g., Burns, 2004b; Volpe et al., 2011) and across multiple populations (students in general education, students with learning disabilities, and students with developmental cognitive disabilities; e.g., Burns &

Boice, 2009). Multiple investigations have established the utility of IR for enhancing word recognition (e.g., Burns & Boice, 2009; Nist & Joseph, 2008). IR has produced favorable outcomes when compared to other flashcard interventions, though it has been shown to be potentially less efficient than other flashcard methods (Nist & Joseph, 2008).

Through this study, the depth of processing framework was integrated into IR with the intent of positively impacting maintenance of taught words. This framework has been well established in laboratory research with adults (e.g., Craik & Tulving, 1975; Gallo et al., 2008; Moscovitch & Craik, 1976) but has not been applied to an educational setting prior to an initial investigation by the author (Petersen-Brown & Burns, 2011). The depth of processing theory was integrated by orienting processing toward the semantic qualities of target words; this was accomplished by pairing the word meaning with the word as it was practiced.

The theory of common elements (Thorndike & Woodworth, 1901a, 1901b, 1901c) and Stokes and Baer's (1977) generalization framework were also leveraged as a basis for drawing attention to critical elements and varying other attributes of training in order to influence the generalization of taught words. For example, the current data support Stokes and Baer's (1977) framework, specifically the strategies of training sufficient exemplars and programming common stimuli. Participants were trained to read target words within three different sentences in the IR-C condition. Reading words in context then becomes a common stimulus between the training task and the generalization assessment. In addition, including three different sentences in training is an example of training multiple exemplars. Training multiple exemplars (three in particular) has previously been shown to enhance generalization (Ardoin et al., 2008;

Hupp, 1986). Stokes and Baer's framework has been used to systematically program for generalization in applied settings including schools and clinics to improve intervention outcomes for adults and children alike. The framework has been applied to interventions focused on communication (Durand & Carr, 1991), social skills (Chandler, Lubeck, & Fowler, 1992), as well as academic (Ardoin et al., 2008) outcomes. However, most investigations do not explicitly program for generalization; even when it is measured a more passive "train and hope" approach is most common (Stokes & Baer, 1977).

Outside of the theories explicitly investigated here, the results may have implications for other theoretical frameworks. For instance, the finding that greater levels of decoding skill were generally associated with improved intervention outcomes may support theories of reading development that emphasize the role of decoding skills in enhancing word recognition. For example, the theory of automatic information processing in reading suggests that automaticity in more basic reading skills (such as decoding) facilitate the allocation of cognitive resources to higher-level skills (such as word recognition; LaBerge & Samuels, 1974). This theory was supported through the current research, specifically the finding that decoding skills were moderately related to outcomes.

While this research was based on theoretical frameworks, it also contributes to the research base supporting and refining these theoretical frameworks. While theory provides frameworks for interpretation and problem solving in empirical research (Tharinger, 2000), research should also influence theory (Hughes, 2000). In this case, theories of maintenance and generalization, which had not been applied either at all or with consistency to applied settings, gained empirical support through the current study.

In addition, the theory of automatic information processing (LaBerge & Samuels, 1974) was supported through the relationship found between decoding skill and outcomes. Further, the current research may have other implications for LaBerge and Samuels' (1974) theoretical framework; specifically through the finding that a low level of basic skills might be compensated for by the use of theory-based modifications.

Limitations

The current investigation builds upon the existing evidence base for IR and is an investigation of the influence of theory-based modifications on intervention outcomes. While this investigation suggests possible implications for practice, research, and theory, it has significant limitations that should be considered. First, the lack of consistent statistical significance of outcome variables may suggest a greater sample size is needed in future investigations to see significant effects. The lack of consistent statistical significance and generally small effect sizes were also potentially due to ceiling effects; at the very least, the lack of variability in outcomes may have obscured any effects of IR variants. Future investigations may pursue various paths to avoid this issue in the future: increasing the number of unknown words, increasing retention intervals, and/or working with struggling students who are less likely to remember a substantial number of words.

Related to study participants, a convenience sample was used in which participants were second and third graders from a specific school in a suburban area who were among the first to return consent forms. There may be systematic differences between those who returned consent forms and those who did not. In addition, this sample may not be representative of broader populations, so these results may not be generalizable to other populations. While the current study recruited participants across

achievement levels, participants were generally high performing, particularly on an auditory working memory assessment. Future investigations should ensure that participants not only represent lower levels of achievement based on standardized achievement tests but also on covariates. Future investigations may also focus on students with lower achievement levels because these are the students who are most likely to participate in a one-on-one intervention such as IR in a school setting.

With regard to study materials, unknown words used in the current study may not reflect the broader sample of words students are expected to recognize. Many words that are important for children to automatically recognize may not lend themselves to being succinctly defined and therefore may not be presented through IR-V; this limitation may limit broader applicability of IR-V specifically. In addition, unknown words were located prior to the first session and taught across the first three sessions; it is possible that students may have learned the words previously identified as unknown prior to teaching. Although randomly selecting words for each week and randomly sequencing IR variants should control for any systematic effects this may have had, it is a potential limitation.

Finally, maintenance and generalization were assessed at 1 and 2 weeks after teaching. The 1-week assessment may have influenced results of the 2-week assessment. In addition, longer retention intervals are important and may provide greater utility to the results since students may not encounter taught words in a classroom context until a point further in the future.

Conclusions

This research marks not only an addition to the growing literature supporting IR as an intervention that can support a wide variety of outcomes in the schools. It also

marks an attempt to apply educational theory to the classroom; application of theoretical frameworks to applied settings has been inconsistent at best and nonexistent in many cases. The current investigation suggests that both cognitive and behavioral theory can be successfully integrated in interventions utilized in educational settings, and further that these modifications may lead to improved outcomes. Although the theories leveraged in this study have been researched and discussed or decades, there is a dearth of support from applied settings, particularly in the case of the depth of processing framework and the theory of common elements. The current research adds support to these theories as well as potentially enhancing their relevance in application. Finally, this research suggests new methods for securing effective interventions in schools: when improved outcomes from intervention are desired, one may seek a different intervention but also may integrate thoughtful, theory-based modifications to an existing intervention.

Table 1

Descriptive Statistics for Primary Outcome Variables

	n	*Mean*	*SD*	Skewness	Kurtosis	*d*
Incremental Rehearsal (IR)						
1-week maintenance	41	5.78	1.49	-1.79	3.58	--
1-week generalization	41	5.29	1.42	-0.94	0.76	--
2-week maintenance	41	6.41	1.12	-2.60	7.14	--
2-week generalization	41	5.83	1.39	-1.78	3.49	--
Session length	41	11.08	2.77	3.01	12.18	--
IR with Vocabulary						
1-week maintenance	40	6.25	1.06	-1.91	5.24	0.37
1-week generalization	40	5.78	1.39	-1.46	2.46	0.35
2-week maintenance	41	6.49	0.93	-2.15	4.82	0.08
2-week generalization	41	6.22	1.17	-1.92	3.95	0.31
Session length	41	16.14	3.77	1.78	4.96	1.55
IR with Contextual Reading Practice						
1-week maintenance	40	6.15	1.27	-3.13	13.56	0.27
1-week generalization	40	5.60	1.52	-1.60	3.51	0.21
2-week maintenance	41	6.27	1.34	-3.46	13.64	-0.12
2-week generalization	41	6.10	1.28	-2.14	5.67	0.21
Session length	41	15.91	3.88	2.34	7.68	1.45

Note. *d* is compared to Incremental Rehearsal

Table 2

Descriptive Statistics for Potential Moderators

	N	Min	Max	*M*	*SD*	Skew	Kurtosis
Auditory Working Memory	41	91	134	109.49	10.50	.49	-.35
Word Attack	41	83	132	105.98	11.19	.25	.13

Note. Data are grade-based standard scores

Table 3

Correlations Between Covariates and Primary Outcome Variables

	1-Week Maintenance			1-Week Generalization			2-Week Maintenance			2-Week Generalization			Session Length			
	IR	IR-V	IR-C	IR	IR-V	IR-C	IR	IR-V	IR-C	IR	IR-V	IR-C	IR	IR-V	IR-C	WA
AWM	.06	.15	.19	.20	.12	.32*	.23	.12	.11	.22	.09	.09	-.14	-.08	-.08	.31*
WA	.56*	.41*	.58*	.55*	.47*	.65*	.59*	.51*	.41*	.59*	.59*	.51*	-.29	-.53*	-.53*	--

Note. AWM = Auditory Working Memory, WA = Word Attack, IR = Incremental rehearsal, IR-V = Incremental rehearsal with vocabulary, IR-C = Incremental rehearsal with contextual reading practice

*p<.05

Table 4

Word Knowledge Across Conditions and Time Points

	Level of Word Knowledge				
	0	1	2	3	4
Prior to teaching					
Incremental Rehearsal (IR)	192	14	27	11	43
IR with Vocabulary	197	12	20	9	49
IR with Contextual Reading Practice	177	19	32	9	49
1 week after teaching					
Incremental Rehearsal (IR)	179	13	23	15	56
IR with Vocabulary	133	6	27	11	102
IR with Contextual Reading Practice	132	23	46	27	51
2 weeks after teaching					
Incremental Rehearsal (IR)	167	17	23	23	57
IR with Vocabulary	124	8	24	9	121
IR with Contextual Reading Practice	136	15	46	9	67

Note. 0 = unknown, 1 = partially known, 2 = known with context, 3 = definition with

ancillary feature, 4 = definition with key features

Figure 1. Benefit of IR-V Over IR and IR-C Over IR by Level of Auditory Working Memory Capacity
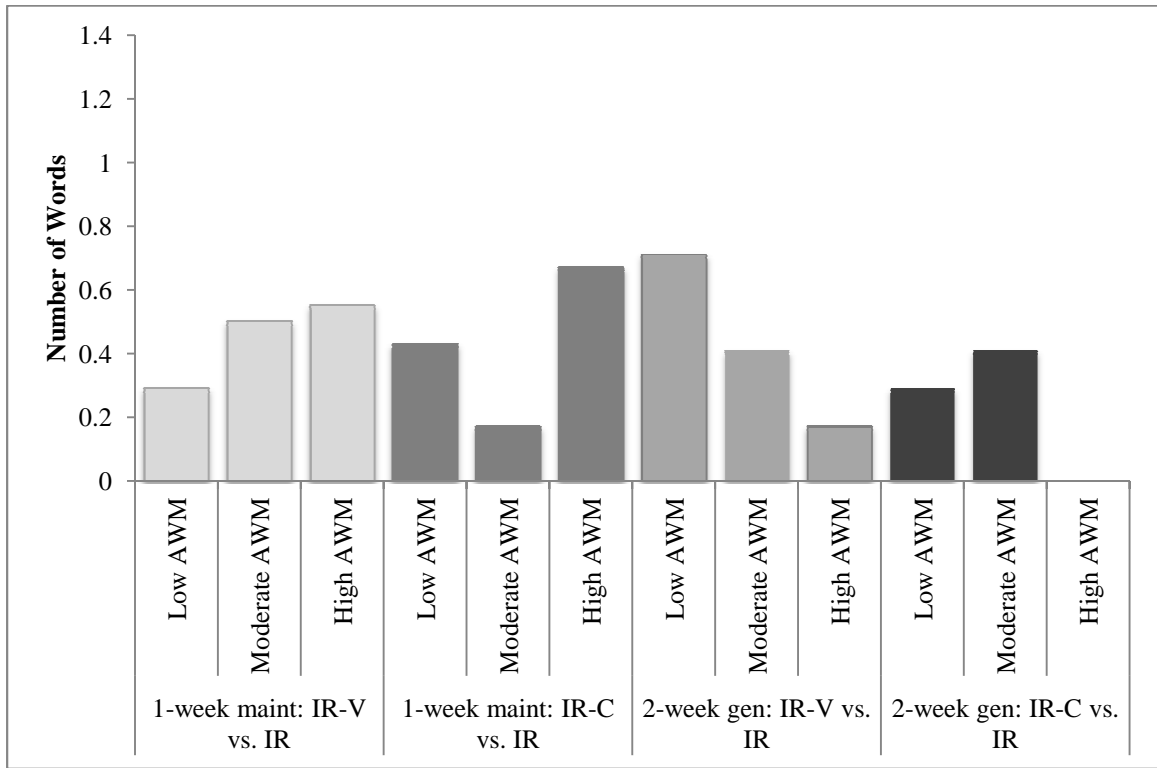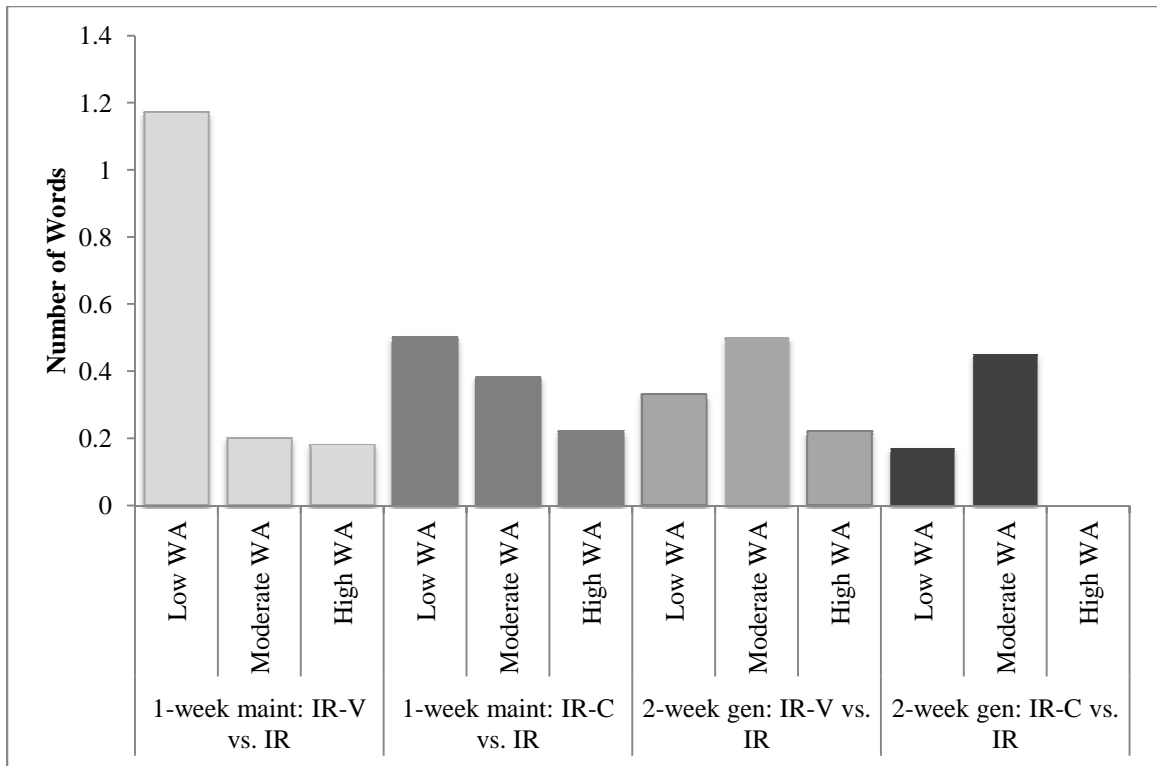
Figure 2. Benefit of IR-V Over IR and IR-C Over IR by Level of Decoding Skill

Bibliography

Aaron, P. G., Joshi, R. M. Ayotollah, M., Ellsberry, A., Henderson, J., & Lindsey, K. (1999). Decoding and sight-word naming: Are they two independent components of word recognition skill? *Reading and Writing: An Interdisciplinary Journal*, *11*, 89-127.

Alberto, P. A. & Troutman, A. C. (2003). *Applied behavioral analysis for teachers* (6th ed.). Upper Saddle River, NJ: Merrill Prentice Hall.

Anderson, J. R. & Reder, L. M. (1979). An elaborative processing explanation of depth of processing. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ardoin, S. P. (2007). The response in response to intervention: Evaluating the utility of assessing maintenance of intervention effects. *Psychology in the Schools*, *43*, 713-725.

Ardoin, S. P., Eckert, T. L., & Cole, C. A. S. (2008). Promoting generalization of reading: A comparison of two fluency-based interventions for improving general education student's oral reading rate. *Journal of Behavioral Education*, *17*, 237-252.

Ardoin, S. P., McCall, M., Klubnik, C. (2007). Promoting generalization of oral reading fluency: Providing drill versus practice opportunities. *Journal of Behavioral Education*, *16*, 55-70.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation* (pp. 89-195). New York: Academic press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-89). San Diego, CA: Academic Press.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *108*. 296-308.

Becker, J. T., & Morris, R. G. (1999). Working memory(s). *Brain and cognition*, *41*, 1-8.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 657-668.

Bowers, P. G. (1993). Text reading and rereading: Determinants of fluency beyond word recognition. *Journal of Reading Behavior*, *25*, 133-153.

Burns, M. K. (2001). Measuring sight-word acquisition and retention rates with curriculum-based assessment. *Journal of Psychoeducational Assessment*, *19*, 148-157.

Burns, M. K. (2002). Utilizing a comprehensive system of assessment to intervention using curriculum-based assessments. *Intervention in School and Clinic*, *38*, 8-13.

Burns, M. K. (2004a). Empirical analysis of drill ratio research: Refining the instructional level for drill tasks. *Remedial and Special Education*, *25*, 167-173.

Burns, M. K. (2004b). Using incremental rehearsal to increase fluency of single-digit multiplication facts with children identified as learning disabled in mathematics computation. *Education and Treatment of Children*, *28*, 237-249.

Burns, M. K. (2007a). Reading at the instructional level with children identified as learning disabled: Potential implications for response-to-intervention. *School*

Psychology Quarterly, *22*, 297-313.

Burns, M. K. (2007b). Comparison of drill ratio and opportunities to respond when rehearsing sight words with a child with mental retardation. *School Psychology Quarterly*, *22*, 250-263.

Burns, M. K. (2011). School psychology research: Combining ecological theory and prevention science. *School Psychology Review*, *40*, 132-139.

Burns, M. K. & Boice, C. H. (2009). Comparison of the relationship between words retained and intelligence for three instructional strategies among students with below-average IQ. *School Psychology Review*, *38*, 284-292.

Burns, M. K., Dean, V. J., & Foley, S. (2004). Preteaching unknown key words with incremental rehearsal to improve reading fluency and comprehension with children identified as reading disabled. *Journal of School Psychology*, *42*, 303-314.

Burns, M. K., & Sterling-Turner, H. (2010). Comparison of efficiency measures for academic interventions based on acquisition and maintenance. *Psychology in the Schools*, *47*, 126-134.

Burns, M. K., Zaslofsky, A. F., Kanive, R., & Parker, D. C. (2012). Meta-analysis of incremental rehearsal using phi coefficients to compare single-case and group designs. *Journal of Behavioral Education*, *21*, 185-202.

Butterfield, E. C., & Nelson, G. D. (1989). ERIC annual review paper: Theory and practice of teaching for transfer. *Educational Technology Research and Development*, *37*, 5-38.

Chan, J. C., McDermott, K. B., Watson, J. M., & Gallo, D. A. (2005). The importance of

    material-processing interactions in inducing false memories. *Memory &*

    Cognition, *33*, 389-395.

Chandler, L. K., Lubeck, R. C., & Fowler, S. A. (1992). Generalization and maintenance

    of preschool children's social skills: A critical review and analysis. *Journal of*

    Applied Behavior Analysis, *25*, 415-428.

Codding, R. S., Archer, J., & Connell, J. (2010). A systematic replication and extension

    of using incremental rehearsal to improve multiplication skills: An investigation

    of generalization. *Journal of Behavioral Education*, *19*, 93-105.

Coleman, M., & Liau, T. L. (1975). A computer readability formula designed for

    machine scoring. *Journal of Applied Psychology*, *60*, 283-284.

Cooke, N. L., Guzaukas, R., Pressley, J. S., & Kerr, K. (1993). Effects of using a ratio of

    new items to review items during drill and practice: Three experiments. *Education*

    and Treatment of Children, *16*, 213-234.

Craik, F. I. M. & Lockhart, R. S. (1972). Levels of processing: A framework for memory

    research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671-684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in

    episodic memory. *Journal of Experimental Psychology: General*, *104*, 268-294.

Daly III, E. J., Bonfiglio, C. M., Mattson, T., Persampieri, M., & Foreman-Yates, K.

    (2005). Refining the experimental analysis of academic skills deficits: Part I. An

    investigation of variables that affect generalized reading performance. *Journal of*

    Applied Behavior Analysis, *38*, 485-497.

Denckla, M. B., & Rudel, R. G. (1976). Rapid 'automatized' naming (R.A.N.): Dyslexia

differentiated from other learning disabilities. *Neuropsychologia*, *14*, 471-479.

DeVault, R., & Joseph, L. M. (2004). Repeated readings combined with word boxes phonics technique increases fluency levels of high school students with severe reading delays. *Preventing School Failure*, *49*, 22-27.

Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology*, *80*, 465-472.

Durand, V. M., & Carr, E. G. (1991). Functional communication training to reduce challenging behavior: Maintenance and application in new settings. *Journal of Applied Behavior Analysis*, *24*, 251-264.

Durso, F. T., & Shore, W. J. (1991). Partial knowledge of word meanings. *Journal of Educational Psychology: General*, *120*, 190-202.

Edelstein, B. A. (1989). Generalization: Terminological, methodological and conceptual issues. *Behavior Therapy*, *20*, 311-324.

Eysenck, M. W. (1978). Levels of processing: A critique. *British Journal of Psychology*, *69*, 157-169.

Fisher, R. P., & Craik, F. I. (1977). Interaction between encoding and retrieval operations in cued recall. *Journal of Experimental Psychology: Human Learning and Memory*, *3*, 701-711.

Franks, J. J., Bilbrey, C. W., Lien, K. G., & McNamara, T. P. (2000). Transfer-appropriate processing (TAP). *Memory & Cognition*, *28*, 1140-1151.

Fry, E. (1968). A readability formula that saves time. *Journal of Reading*, *11*, 513-516.

Frye, E. B., & Kress, J. E. (2006). *The reading teacher's book of lists* (5th ed.). San Francisco: Jossey-Bass.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an

    indicator of reading competence: A theoretical, empirical, and historical analysis.

    Scientific Studies of Reading, *5*, 239-256.

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of

    processing elicit a distinctiveness heuristic: Evidence from the criterial

    recollection task. *Journal of Memory and Language*, *58*, 1095-1111.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. E.

    Tulving, & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229-244).

    New York: Oxford University Press.

Gathercole, S. E., Alloway, T. P., Willis, C., & Adams, A. (2006). Working memory in

    children with reading disabilities. *Journal of Experimental Child Psychology*, *93*,

    265-281.

Gicking, E. E., & Armstrong, D. L. (1978). Levels of instructional difficulty as related to

    on-task behavior, task completion, and comprehension. *Journal of Learning*

    Disabilities, *11*, 32-39.

Gickling, E. E., & Thompson, V. P. (1985). A personal view of curriculum-based

    assessment. *Exceptional Children*, *52*, 205-218.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability.

    Remedial and Special Education, *7*, 6-10.

Gunning, R. (1968). *The technique of clear writing*. McGraw-Hill.

Haegele, K. M., & Burns, M. K. (2010). The effect of modifying instruction set size

    based on the acquisition rate among students identified with a learning disability.

    Manuscript submitted for publication.

Hannon, B., & Craik, F. I. M. (2001). Encoding specificity revisited: The role of

    semantics. *Canadian Journal of Experimental Psychology*, *55*, 231-243.

Hughes, J. N. (2000). The essential role of theory in the science of treating children:

    Beyond empirically supported treatments. *Journal of School Psychology*, *38*, 301-

    330.

Hull, C. L. (1943). *Principles of behavior.* New York: Appleton-Century-Crofts.

Hupp, S. C. (1986). Use of multiple exemplars in object concept training: How many are

    sufficient? *Analysis and Intervention in Developmental Disabilities*, *6*, 305-317.

Institute of Education Sciences, National Center for Education Evaluation and Regional

    Assistance, What Works Clearinghouse (2013). What Works Clearinghouse.

    Retrieved from http://ies.ed.gov/ncee/wwc.

Jacoby, L. L. (1978). On interpreting the effects of repetition: Solving a problem versus

    remembering a solution. *Journal of Verbal Learning and Verbal Behavior*, *17*,

    649-667.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical

    memory and perceptual learning. *Journal of Experimental Psychology: General*,

    110(3), 306-340.

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources

    of individual differences in reading comprehension and reading fluency. *Journal*

    of Educational Psychology, *95*, 719-729.

Joseph, L. M. (2002). Facilitating word recognition and spelling using word boxes and

    word sort phonic procedures. *School Psychology Review*, *31*, 122-129.

Karpicke, J. D., & Roediger III, H. L. (2007). Repeated retrieval during learning is the

key to long-term retention. *Journal of Memory and Language*, *57*, 151-162.

Kendall, P. C. (1989). The generalization and maintenance of behavior change: Comments, considerations, and the "no-cure" criticism. *Behavior Therapy*, *20*, 357-364.

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975) Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesh Reading Ease Formula) for Navy Enlisted Personnel. *Research Branch Report 8* (75).

Kirby, K. C., & Bickel, W. K. (1988). Toward an explicit analysis of generalization: A stimulus control interpretation. *The Behavior Analyst*, *11*, 115-129.

Kuhn, M. R., Schwanenflugel, P. J., Meisinger, E. B., Levy, B. A., & Rasinski, T. V. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, *45*, 230-251.

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., and Dunleavy, E. (2007). Literacy in Everyday Life: Results From the 2003 National Assessment of Adult Literacy (NCES 2007–480). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology*, *6*, 293-323.

Landis, R. J., & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174.

Levy, B. A., Abello, B., & Lysynchuk, L. (1997). Transfer from word training to reading in context: Gains in reading fluency and comprehension. *Learning Disability*

Quarterly, *20*, 173-188.

Lockhart, R. S. (2002). Levels of processing, transfer appropriate processing, and the concept of robust encoding. *Memory*, *10*, 397-403.

Lockhart, R. S., & Craik, F. I. M. (1978). Levels of processing: A reply to Eysenck. *British Journal of Psychology*, *69*, 171-175.

Lockhart, R. S., & Craik, F. I. M. (1990). Levels of processing: A retrospective commentary on a framework for memory research. *Canadian Journal of Psychology*, *44*, 87-112.

Lyon, G. R. & Moats, L. C. (1997). Critical conceptual and methodological considerations in reading intervention research. *Journal of Learning Disabilities*, *30*, 578-588.

Manis, F. R. (1985). Acquisition of word identification skills in non-disabled readers. *Journal of Educational Psychology*, *77*, 78-90.

Mann, V. A. (1984). Longitudinal prediction and prevention of early reading difficulty. *Annals of Dyslexia*, *34*, 117-136.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities*, *17*, 592-599.

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

Mayer, R. E., & Whitrock, M. C. (1996). Problem-solving transfer. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational technology* (pp. 47-62). New York: Simon & Schuster.

MacQuarrie, L. L., Tucker, J. A., Burns, M. K., & Hartman, B. (2002). Comparison of

retention rates using traditional drill sandwich, and incremental rehearsal flash
card methods. *School Psychology Review*, *31*, 584 – 595.

McDaniel, M. A., Anderson, J. L., Dorbish, M. H., & Morisette, N. (2007). Testing the
testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*, 4-
5.

McDaniel, M. A., & Masson, M. E. J. (1985). Altering memory representations through
retrieval. *Journal of Experimental Psychology: Learning, Memory, and
Cognitition*, *11*, 371-385.

McDermott, K. B., Petersen, S. E., Watson, J. M., & Ojemann, J. G. (2003). Procedure
for identifying regions preferentially activated by attention to semantic and
phonological relations using functional magnetic resonance imaging.
Neuropsychologia, *41*, 293-303.

McGrew, K. S., & Woodcock, R. W. (2001). *Woodcock-Johnson III technical manual*.
Itasca, IL: Riverside.

Mesmer, E. M., Duhon, G. J., Hogan, K., Newry, B., Hommema, S., Fletcher, C., & Boso,
M. (2010). Generalization of sight word accuracy using a common stimulus
procedure: A preliminary investigation. *Journal of Behavioral Education*, *19*, 47-
61.

Meyer, M. S., Wood, F. B., Hart, L. A., & Felton, R. H. (1998). Selective predictive
value of rapid automatized naming in poor readers. *Journal of Learning
Disabilities*, *31*, 106-117.

Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus
transfer appropriate processing. *Journal of verbal learning and verbal behavior*,
16, 519-533.

Moscovitch, M., & Craik, F. I. M. (1976). Depth of processing, retrieval cues, and
uniqueness of encoding as factors in recall. *Journal of Verbal Learning and
Verbal Behavior*, *15*, 447-458.

Mulligan, N. W., & Lozito, J. P. (2006). An Asymmetry Between Memory Encoding and
Retrieval Revelation, Generation, and Transfer-Appropriate Processing.
Psychological science, *17*, 7-11.

Murdock, Jr., B. B. (1960). The distinctiveness of stimuli. *Psychological Review*, *67*, 16-
31.

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

National Institute of Child Heath and Human Development (2000). *Report of the
National Reading Panel. Teaching children to read: an evidence-based
assessment of the scientific research literature on reading and its implications for
reading instruction: Reports of the subgroups* (NIH Publication No. 00-4754).
Washington, DC: U.S. Government Printing Office.

Nist, L. & Joseph, L. M. (2008). Effectiveness and efficiency of flashcard drill
instructional methods on urban first-graders' word recognition, acquisition,
maintenance, and generalization. *School Psychology Review*, *37*, 294-308.

Northwest Evaluation Association (2011). Measures of academic progress. Retrieved
    from http://www.nwea.org/

Park, H., & Rugg, M. D. (2008). Neural correlates of successful encoding of semantically
    and phonologically mediated inter-item associations. *Neuroimage*, *43*(1), 165-172.

Perkins, D. N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational
    Researcher*, *18*, 16-25.

Petersen-Brown, S., & Burns, M. K. (2011). Adding a vocabulary component to
    incremental rehearsal to enhance retention and generalization. *School Psychology
    Quarterly*, *26*(3), 245-255.

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatey, A.
    S. (1995). *Listening to children read aloud: Data from NAEP's integrated reading
    performance record (IRPR) at grade four*. Washington, DC: National Center for
    Education Statistics.

Roediger III, H. L., Gallo, D. A., & Geraci, L. (2002). Processing approaches to
    cognition: The impetus from the levels-of-processing framework. *Memory*, *10*,
    319-332.

Roediger III, H. L., & Karpicke, J. D. (2006). Test enhanced learning: Taking memory
    tests improves long-term retention. *Psychological Science*, *17*, 249-255.

Samuels, S. J. (1988). Decoding and automaticity: Helping poor readers become
    automatic at word recognition. *The Reading Teacher*, *41*, 756-760.

Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations
    supporting visual object identification and memory. *Journal of Cognitive
    Neuroscience*, *15*, 111-135.

Shapiro, E. S. (1992). Use of Gickling's model of curriculum-based assessment to

    improve reading in elementary age students. *School Psychology Review*, *21*, 168-

    176.

Sheridan, H. & Reingold, E. M. (2012). Levels of processing influences both recollection

    and familiarity: Evidence from a modified remember-know paradigm.

    Consciousness and Cognition, *21*, 438-443.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally

    achieving and subtypes of learning disabled children. *Child Development*, *60*,

    973-980.

Silber, J. & Martens, B. K. (2010). Programming for the generalization of oral reading

    fluency: Repeated readings of entire text versus multiple exemplars. *Journal of*

    Behavioral Education, *19*, 30-46.

Skinner, B. F. (1938). *The behavior of organisms.* New York: Appleton-Century.

Skinner, B. F. (1953). *Science and human behavior.* New York: Macmillan.

Smith, E. A., & Senter, R. J. (1967). *Automated readability index.* AMRL-TR-66-22.

    Wright-Patterson AFB, OH: Aerospace Medical Division.

Stanovich, K. E., (1980). Toward an interactive-compensatory model of individual

    differences in the development of reading fluency. *Reading Research Quarterly*,

    16, 32-71.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual

    differences in the development of reading fluency. *Reading Research Quarterly*,

    16, 32-71.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual

    differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360-

    407.

Stein, B. S. (1978). Depth of processing reexamined: The effects of the precision of

    encoding and test appropriateness. *Journal of Verbal Learning and Verbal*

    Behavior, *17*, 165-174.

Stokes, T. F. & Baer, D. M. (1977). An implicit technology of generalization *Journal of*

    Applied Behavior Analysis, *10*, 349-367.

Stokes, T. F., & Osnes, P. G. (1989). An operant pursuit of generalization. *Behavior*

    Therapy, *20*, 337-355.

Swanson, H. L. (1994). Short-term memory and working memory: Do both contribute to

    our understanding of academic achievement in children and adults with learning

    disabilities? *Journal of Learning Disabilities*, *27*, 34-50.

Swanson, H. L., Ashbaker, M. H., & Lee, C. (1996). Learning-disabled readers' working

    memory as a function of processing demands. *Journal of Experimental Child*

    Psychology, *61*, 242-275.

Swanson, H. L., & Howell, M. (2001). Working memory, short-term memory, and

    speech rate as predictors of children's reading performance at different ages.

    Journal of Educational Psychology, *93*, 720-734.

Symonds, P. M. & Chase, D. H. (1992). Practice vs. motivation. *Journal of Educational*

    Psychology, *84*, 282-289.

Szadokierski, I., & Burns, M. K. (2008). Analogue evaluation of the effects of

    opportunities to respond and ratios of known items within drill rehearsal of

Esperanto words. *Journal of School Psychology*, *46*, 593-609.

Tan, A., & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology*, *89*, 276-288.

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition*, *29*, 424-432.

Tharinger, D. (2000). The complexity of development and change: The need for the integration of theory and research findings in psychological practice with children. *Journal of School Psychology*, *38*, 383-388.

Thorndike (1906)

Thorndike, E. L., & Woodworth, R. S. (1901a). The influence of improvement in one mental function upon the efficiency of other functions (I). *Psychological Review*, *8*, 247-261.

Thorndike, E. L., & Woodworth, R. S. (1901b). The influence of improvement in one mental function upon the efficiency of other functions (II): The estimation of magnitudes. *Psychological Review*, *8*, 384-395.

Thorndike, E. L., & Woodworth, R. S. (1901b). The influence of improvement in one mental function upon the efficiency of other functions (III): Functions involving attention, observation, and discrimation. *Psychological Review*, *8*, 553-564.

Tucker, J. A. (1989). *Basic flashcard technique when vocabulary is the goal*. Unpublished teaching materials, University of Tennessee at Chattanooga, TN: Author.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, *26*, 1-12.

Tulving, E. (2001). Does memory encoding exist? In M. Naveh-Benjamin, M. Moscovitch, & H. L. Roediger III (Eds.), *Perspectives on human memory and cognitive aging* (pp. 6-27). New York: Psychology Press.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352-373.

Volpe, R. J., Burns, M. K., DuBois, M., & Zaslofsky, A. F. (2011). Computer-assisted tutoring: Teaching letter sounds to kindergarten students using incremental rehearsal. *Psychology in the Schools*, *48*, 332-342.

Wagner, R. K., & Torgeson, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *The Psychological Bulletin*, *101*, 192-212.

Wagner, R. K., Torgeson, J. K., Roshotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R.,… 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.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Itasca, IL: Riverside.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, *46*, 441-517.