Predicting Intervention Effectiveness from Oral Reading Accuracy and Rate Measures through the Learning Hierarchy/Instructional Hierarchy

A DISSERTATION
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Isadora Elisabeth Szadokierski

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

Matthew K. Burns, PhD, Jennifer McComas, PhD

November 2012

Acknowledgements

There are a number of people whose support and encouragement made this dissertation possible. Firstly, I must thank my adviser, Dr. Matt Burns who first sparked my interest in reading and fostered my growth in the area of interventions. He helped me build my character, introduced me to the magic of incremental rehearsal and maintained his faith in me that this day would come. In addition, I am indebted to Dr. Jennifer McComas and her graduate assistants, including notably David Parker, Shawna Petersen-Brown and Katie Haegele who helped collect my data. I literally could not have finished this dissertation without you!

I must also acknowledge my wonderful cohort members, whose friendship, support and teaching have shaped me as a professional and as a person. In particular, Christina Boice-Mallach has been a constant companion on this journey, from meeting on the first day of graduate school to sharing an office and the innumerable hours spent at coffee shops encouraging each other. I look forward to future get-togethers that do not revolve around dissertations!

My incredible parents, Stephen and Beatrice and sister, Victoria have been there for me since the beginning, encouraging me to follow my dreams. Finally, I would never have accomplished this goal without the constant support of my loving husband, Victor. He has been there for me on the good days and the bad, continuing to encourage me, even when I was not always receptive. This doctoral degree is an accomplishment for both of us and I look forward to our future adventures together!

Dedication

This dissertation is dedicated to all of the students and young people whose inquisitiveness and joy of learning inspire me to do my best to make the world a better place for all.

Abstract

The current study used the Learning Hierarchy/Instructional Hierarchy (LH/IH) to predict intervention effectiveness based on the reading skills of students who are developing reading fluency. Pre-intervention reading accuracy and rate were assessed for 49 second and third grade participants who then participated in a brief experimental analysis (BEA) to determine whether each participant responded best to an acquisition or a proficiency intervention package. Analyses indicate significant baseline differences between students who responded to each intervention package. Moreover, accuracy and rate have a positive correlation with proficiency intervention effectiveness and a negative correlation with acquisition intervention effectiveness. Predictive models and potential accuracy and rate cut scores for making intervention decisions were also investigated.

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

Rationale for Study

Reading is one of the most critical skills that students learn in school because it is a precursor to future academic success and a productive adulthood (Gerber, 1997). The opportunity to advance in our society is directly and indirectly related to an individual's reading ability. The results of the 2003 National Assessment of Adult Literacy (Kutner, Greenberg, Jin, Boyle, Hsu & Dunleavy, 2007) showed that in the United States reading proficiency scores were highly correlated with the number of years of schooling completed and household income. Individuals whose literacy was rated at the higher levels were more likely to be employed full time, work in professional occupations and have higher average weekly salaries. Adults with lower levels of literacy, on the other hand, were much more likely to have not completed high school, to work in the service industry, to be out of the labor force or have received public assistance. Furthermore, adults serving time in prison had lower average reading scores than equivalent groups in the general population (Greenberg, Dunleavy, & Kutner, 2007). Overall, higher levels of reading proficiency are related to more positive outcomes in American society.

Reading and School Outcomes. While reading ability level influences an individual's opportunity to succeed professionally in adulthood, the differences between successful readers and struggling readers also impact students during grade school. When a child enters school, factors such as English language development and vocabulary, exposure to literacy concepts, and socioeconomic status predict initial reading achievement (Snow, Burns & Griffin, 1998). Without intervention, students' relative reading achievement often remains stable over the school years (Wadsworth, Fulker & DeFries, 1999). In fact, the gap between struggling readers and their more skilled peers is likely to grow larger as differences in proficiency and motivation affect the number of words read and, thus, the rate of skill growth (Stanovich, 1986; Wagner & Stanovich, 1996).

Poor readers are at greater risk to struggle academically overall as reading becomes an increasingly important element of instruction across content areas (Lerner, 2003). Perhaps as a result of the increasing level of challenge, 23.5% of students diagnosed with a learning disability, the majority of whom have reading problems, drop out of school in comparison to 8% of the general high school population (U.S. Department of Education, National Center for Educational Statistics, 2011).

Behaviorally and emotionally these students are also likely to struggle more than their average achieving peers. Students with lower reading achievement are less engaged during instruction and have higher reported levels of inattention and delinquent behavior (Wasson, Beare & Wasson, 1990; Arnold et al., 2005). These students also typically have poorer social skills, experience greater peer rejection, and are more likely to exhibit antisocial behaviors (Kavale & Forness, 1996; Trzesniewski, Moffitt, Caspi, Taylor & Maughan, 2006). While the link between these negative risk factors and reading achievement is correlational, reading interventions have also been found to have a positive impact on social outcomes (Waznek, Vaughn, Kim & Cavanaugh, 2006; Castillo, Porter, Curtis & Batsche, 2005).

Despite the consistent findings regarding the benefits of strong reading skills, the American public school system has been plagued by low levels of reading achievement and a relatively high prevalence rate of reading disabilities. Results from the 2011 National Assessment of Educational Progress (NAEP) indicated that 67% of fourth graders and 66% of eighth graders fail to achieve grade-level proficiency in reading (Aud, Hussar, Johnson, Kena, Roth et al., 2012). Furthermore, recent statistics show that 5% of youth ages 3 to 21 attending public schools receive special education services under the category of Specific Learning Disabilities (U.S. Department of Education, National Center for Educational Statistics, 2011) and as many as 80% of these students have a primary deficit in the area of reading (Moats & Dakin, 2007). By the time a reading deficit is large enough to warrant special education services, most frequently at age 9, or about third grade (Data Accountability Center, 2012), it is often very difficult to remediate and few of these students exit special education (U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs, 2011).

Students with Reading Difficulties. The causes of reading difficulties have generally been classified into two broad categories: *experiential*, or deficits that are the result of environmental factors, such as lack of exposure to experiences that build school readiness skills, poor reading instruction, and low-socioeconomic status; and, *constitutional*, or deficits that originate within the individual, typically attributed to genetic and/or cognitive differences. In addition, the constitutional category is typically further broken down into individuals who have a specific learning disability in the area of reading and those who have generally low cognitive skills (e.g., intelligence, language abilities) and thus, poorer overall academic achievement.

The low reading performance of this latter group, often called "garden-variety poor readers" or "low achievers", is not unexpected due to their overall low intellectual abilities. However, individuals with a specific difficulty in reading also called dyslexia or learning disability, struggle to learn to read despite average to above average cognitive skills (Stanovich, 1988). Although these distinctions are somewhat artificial given the reciprocal influence of both constitutional and environmental factors on reading achievement (Clay, 1987; Stanovich, 1986) and research indicating few differences between low achievers and those with an identified specific reading deficit (e.g., Ysseldyke, Algozzine, Shinn, & McGue, 1982; Fletcher et al., 1994; Fuchs, Fuchs, Mathes, Lipsey, & Roberts, 2001), they often guide educators' understanding of and response to reading difficulties.

The 2004 federal reauthorization of special education legislation known as the Individuals with Disabilities Education Improvement Act (IDEIA 2004) guarantees students with a specific learning disability (SLD) in reading or other academic areas, the right to receive special education services to address their learning needs. A majority of students with a SLD have been identified by the presence of a significant achievement-ability discrepancy based on scores from standardized achievement and cognitive assessments (Reschly & Hosp, 2004). The presence of this discrepancy is believed to be indicative of a qualitative difference between students with a SLD and garden variety poor readers (Rutter & Yule, 1975). This identification model has been used since the passage of the first comprehensive federal special education legislation, the Education of All Handicapped Children Act of 1975, but it has drawn criticism for being inaccurate, arbitrary, unfair, and for delaying intervention and not improving outcomes for students (for reviews, see Kavale, 2001,Gresham, 2001, and Fuchs, Mock, Morgan, & Young, 2003).

While distinguishing between students who have a SLD and those who struggle for other reasons has been considered essential for determining who receives additional instruction, there is a growing educational movement to provide reading interventions to all struggling readers at earlier ages. In their seminal study, Vellutino and colleagues (1996) found that a majority of students with low initial reading achievement, regardless of cause, moved into or even above the average range when given intensive remediation. This paper helped advance the movement towards early intervention and the response to intervention (RTI) model as a way to both aid a broader range of struggling readers and to identify those students who have a more difficult to remediate learning difficulties (Fuchs & Fuchs, 1998; Gresham et al., 2005).

Response to Intervention (RTI). Response to intervention (RTI; Jimerson, Burns, & VanDerHeyden, 2007; Fuchs, Mock, Morgan, & Young, 2003) is a process of providing increasingly more intensive supplementary instruction to students whose achievement is below that of their peers and systematically monitoring their progress acquiring academic skills. It is a prevention-focused system based on the three tiered public health model where, as a first step, all students are provided with high quality instruction. Those who do not show expected growth with the generally effective standard instruction (“nonresponders”) receive more targeted, evidence-based interventions, typically conceptualized as small group instruction. Students who continue to show little growth in response to the first level of remediation receive a comprehensive, individualized evaluation and more intensive interventions.

RTI offers a number of advantages over traditional practice (for reviews of issues that still need to be addressed regarding RTI, see Vaughn & Fuchs, 2003; Reschly, 2005; Gersten & Dimino, 2006), notably it does not require that the cause of a student’s reading difficulty be determined or diagnosed; instead, any student who struggles is provided early intervention before deficits become so great that they are difficult to remediate. As a result, a large number of students benefit through the early detection and remediation of reading problems. Best results are obtained by taking into account both the research literature on methods that have demonstrated effectiveness and the needs of the individual student in order to promote skill growth.

Critical to the process of selecting an appropriate intervention is consideration of assessment data that indicate the student’s current competencies and deficits. Instructionally relevant data are derived from measures that are direct, idiographic, repeated, and ecologically sensitive (Fuchs & Fuchs, 1986). Two key assessment frameworks that meet these criteria and have become increasingly more prevalent in the schools (Shapiro, Angello, & Eckert, 2004) are curriculum-based measurement (CBM) and curriculum-based assessment (CBA). As the names imply, these strategies involve measuring students’ performance on classroom material in order to monitor their growth over time (CBM; Deno, 1985) and determine the appropriate level of difficulty and objectives area for effective instruction (CBA; Gickling & Thompson, 1985). While these methods allow educators to decide what content to teach and provide a means for evaluating intervention effectiveness over time, they do not aid in determining *how* to teach an individual child, which is a key aspect of the problem-solving process (Griffiths, VanDerHeyden, Parson & Burns, 2006).

Statement of the Problem

Suggestions of common reasons students fail (Daly, Witt, Martens, & Dool, 1997; Duhon, Noell, Witt, Freeland, Dufrene, & Gilbertson 2004) and procedures for testing different interventions components (e.g., brief experimental analysis; Daly, Martens, Dool & Hintze, 1998; Daly, Martens, Hamler, Dool, & Eckert, 1999; Jones & Wickstrom, 2002) can help guide intervention selection, but there is not currently an objective, evidence-based way to determine beforehand the type of intervention to which an individual student is most likely to respond. Being able to predict intervention effectiveness from a student's current performance would allow problem-solving teams to efficiently remediate academic deficits by more accurately targeting interventions to the student's needs. Students could be provided a matched intervention quickly and its effectiveness could be evaluated more closely over time. Skill development could also be monitored to determine when a change in instruction is warranted. In addition, such a guideline would allow a smaller set of relevant interventions to attempt during an individual evaluation such as a brief experimental analysis.

A theory of how reading skills develop helps provide the framework to guide identification of appropriate interventions. The learning hierarchy/instructional hierarchy

(LH/IH) proposed by Haring and Eaton (1978) provides a useful guideline of how skills develop and also which instructional components are matched to effective instruction at each level of skill development. Details will be presented in the literature review chapter, but basically the LH describes skill development as progressing through the four stages that progressively focus on acquisition, proficiency, generalization and adaptation. The matched IH presents instructional components that foster growth at each stage.

Haring and Eaton (1978) suggested that a student remains at the acquisition stage from the time they are first introduced to the skill until he can perform it with between 90% to 100% accuracy; at which point the student moves to the proficiency stage, where the focus shifts to fostering speed, above and beyond accuracy. The authors found this to be true for grade-level fluent reading and nearly all other academic skills (Haring & Eaton, 1978), but they do not provide a specific accuracy cutoff or range below which a student would be in the acquisition stage and above which he would be in the proficiency stage. However, given the above definition of the acquisition stage, it seems possible to use a student's reading accuracy to determine whether he is in the acquisition or the proficiency stage in the area of reading fluency. The student's accurate reading rate (i.e., words read correctly per minute) may provide additional information since rate is the most relevant measure at the proficiency stage (Daly, Lentz, & Boyer, 1996). If the student's placement in the learning hierarchy could be ascertained, then using the instructional hierarchy, it would be possible to link the student's performance to instructional components that foster learning. Thus, it should be possible to predict which types of reading fluency interventions are most likely to be effective based on student's oral reading accuracy and rate.

Purpose of the Study

This study will evaluate the characteristics of oral reading demonstrated by students for whom either acquisition or proficiency reading fluency interventions are found to be most effective through a brief experimental analysis. Based on the IH/LH model, students' reading accuracy (i.e., percentage of words read correctly) and, potentially their rate (i.e., number of words read correctly per minute), should indicate at which stage of the IH a student is performing and thus which types of interventions are most likely to be effective. Thus, the purpose of this study is to determine whether there are differences in baseline accuracy and rate for students for whom the different intervention packages are most effective. If, as expected, differences are found, then a further aim is to determine how accuracy and/or rate scores can be used to best predict intervention effectiveness.

In this study, reading fluency interventions will be examined because grade-level fluency is a key subskill in the development of reading (Chall, 1996) and one for which many evidence-based interventions exist (National Institute of Child Health and Human Development, 2000). One intervention package used in the current study consists of listening passage preview (Daly & Martens, 1994) and phrase drill error correction (O'Shea, Munson, & O'Shea, 1984), which correspond to the modeling and feedback components needed at the acquisition stage. The practice and reinforcement components needed at the proficiency stage will be represented in an intervention package that combines repeated reading (Rashotte & Torgensen, 1985) and contingent reward for rapid reading (Lovitt, Eaton, Kirkwood, & Pelander, 1971).

During the first phase of this study, students' baseline accuracy and rate will be measured by having them orally read grade-level passages unaided. Then a brief experimental analysis will be performed to determine which intervention package has the most positive impact on oral reading rates. Thirdly, the baseline levels of accuracy and rate for students for whom the modeling-error correction (M-EC) intervention was most effective and those for whom the repeated reading-reward (RR-R) intervention was most effective will then be compared. It is expected that accuracy and rate will be significantly different for the two groups and further that it will be possible to predict to which group a student belongs based on these measures.

The major premise of this study is that reading fluency intervention effectiveness can be determined from the easily obtained measures of reading accuracy and rate based on the hypothesized links made through the IH/LH. Specifically, by testing intervention components from the IH and determining which is most effective through BEA, it is possible to determine at which stage of the LH a student is performing. Pre-intervention differences in reading accuracy and rate between the groups of students can then be evaluated to determine whether group membership and thus intervention effectiveness can be predicted based on these measures. Based on Haring and Eaton's (1978) statement that the acquisition stage lasts until the student achieves between 90% to 100% accuracy, at which point she can be seen as transitioning to the proficiency stage, accuracy is expected to be a primary factor in differentiating whether students respond better to the

M-EC intervention or the RR-R intervention. However, accurate reading rate will likely contribute additional information that is helpful for determining intervention effectiveness.

Research Questions

The following questions guide this study:

1. What are the differences in baseline accuracy, rate and normalized rate among students for whom only the modeling-error correction (M-EC) intervention was effective, those for whom only the repeated reading-reward (RR-R) intervention was effective and those for whom neither or both were effective?
2. What is the relationship between intervention effectiveness status and baseline accuracy, rate and normalized rate?
3. How well does students' baseline accuracy and rate or normalized rate predict M-EC effectiveness status?
4. How well does students' baseline accuracy and rate or normalized rate predict RR-R effectiveness status?

In addition, the following exploratory questions will be asked:

What cut score or range of scores for accuracy, rate, or a combination of these measures provides sufficient specificity and sensitivity to predict whether a student will benefit from the M-EC intervention or the RR-R intervention?

A Note about "Fluency"

Before continuing, it is important to know that the word "fluency" is used by educational researchers to describe a variety of different phenomena in learning and reading. Fluency is frequently used to refer to: a) an area of reading development (i.e., reading fluency), b) the speed at which a student accurately executes a skill (e.g., number words read correctly per minute on CBM probes), and, c) the second phase of the LH/IH (i.e., the fluency/proficiency phase). For clarity, in this paper, the area of reading (a) will be referred to as “reading fluency”, the accurate skill speed (b) will be called “rate”, and the stage of the LH (c) will be called “proficiency”. Furthermore, most of the time only the word “rate” will be used, but unless stated otherwise, this refers to “rate of *accurate* reading” (i.e., words read correctly per minute).

Chapter II: Literature Review

Reading is one of the most essential skills students learn at school, making it an indicator of educational attainment and key focus of instruction. When students struggle to learn to read, educators must use evidence-based interventions developed with consideration of current empirical understanding to remediate deficits. The domain of reading development is complex, but a conceptual framework of learning can illuminate principles that can be generalized to foster development across skill areas.

The purpose of the current chapter is to review the current literature on reading development and instruction, focusing on the area of reading fluency. The comprehensive research base on fluency presented, including effective interventions and outcome measures make it a clear choice for investigating the validity of a framework that guides intervention selection for struggling readers. Methods for targeting interventions to individual students' needs will be reviewed, with an emphasis on the brief experimental analysis (BEA) procedure. Finally, Haring and Eaton's (1978) learning hierarchy/instructional hierarchy (LH/IH) is the theoretical scaffold for the current study.

Models of Reading Development

The federal research group, Literacy Information and Communication System defines reading as:

> A complex system of deriving meaning from print that requires all of the following:
>
> - the skills and knowledge to understand how phonemes, or speech sounds, are connected to print;

- the ability to decode unfamiliar words;
- the ability to read fluently;
- sufficient background information and vocabulary to foster reading comprehension;
- the development of appropriate active strategies to construct meaning from print;
- the development and maintenance of a motivation to read. (Literacy Information and Communication Systems, n.d., Research Definitions, para. 2)

Similarly, in their review of the instructional research, the National Reading Panel (National Institute of Child Health and Human Development, 2000) considered reading in terms of the five subskill areas of phonemic awareness, phonics, reading fluency, comprehension and vocabulary. These multifaceted definitions illustrate the varied skills that must interact to produce proficient reading. Each skill is dependent upon different cognitive processes working together to translate the words on the page to a meaningful message understood by the individual (Adams, 2004).

Interactive Models of Reading. Although there is not one comprehensive theory that explains how proficient reading occurs, current models of reading typically describe an interactive cognitive process that involves both bottom-up (i.e., from page to brain) and top-down (i.e., from brain to page) processing of information (Rumelhart, 2004). The key characteristic of such connectionist, or parallel distributed processing models, is the simultaneous processing of different aspects of the input, which is facilitated by reciprocal connections and coordination between the processing components (Rumelhart, 1989). Hypothetical processors of the multiple levels of reading information (e.g., letter, word, meaning, context, connections to background knowledge) consider relevant information from the text and then pass hypotheses about the possible identity of the text to all other processors and through this reciprocal information sharing, the best interpretation is created (Adams, 2004). The efficiency of the connections between the processors is the result of the sum of all previous reading experiences and learning.

Rumelhart (2004) describes a model in which a feature extraction device derives information from the visual image of letters and words on the page and sends it to a pattern synthesizer, which uses orthographic, lexical, syntactic and semantic knowledge to make sense of the text. In this model, bottom-up information about the letters and spelling patterns on the page is simultaneously processed with higher level hypotheses about the likely identity of the word or phrase based on structural and syntactical knowledge of the language and the meaning context in which it is found. The combination of these hypotheses from the multiple levels results in the most probable interpretation of the text for the reader.

Adams (1990; 2004) synthesized the findings of many reading research studies in developing the four processors model of reading, which describes the process used by skilled readers. In this connectionist model, the text serves as input to an orthographic processor, which attempts to recognize letters, spelling patterns, and words. The majority of processing occurs letter by letter, but spelling pattern probabilities, learned from the accumulation of previous reading in the language of concern, facilitate letter identification. As letters are recognized, this information is passed to the meaning processor, which determines which “meaning elements” should be activated. Information about meaning is also automatically passed back to the orthographic processor to aid in word recognition. The reciprocally connected phonological processor further facilitates both the orthographic and the meaning processors by transforming the text into speech sounds and employing the individual’s knowledge of spoken language to hasten word and meaning identification. This process also allows the individual to identify words he has not seen written, but has heard spoken. The redundant activation sent between these three processors promotes automatic word recognition and, thus, increases the speed with which text can be processed. The auditory “image” created by the phonological processor also enhances comprehension because a greater number of words and their temporal order can be retained in auditory memory than by using visual memory alone (Dornbush & Basow, 1970).

While all of this is occurring, the fourth, context processor monitors the overall “context” of the passage (e.g., what the passage is about, where it is being read, activating relevant prior knowledge, etc.). A reciprocal connection with the meaning processor allows for a constant monitoring of the passage context, while also aiding in the selection of the appropriate “meaning elements.” Thus, for a skilled reader, once text begins to be identified by the orthographic processor, activation flows throughout the system as all four processors work together to identify and understand the message of the text. The coordination and collaboration of this system is required for effective and efficient reading and can best be cultivated through reading instruction and experience. While

Adams' and Rumelhart's models describe the process of skilled reading, best practices in reading instruction can best be addressed by understanding how students develop this level of proficiency.

Chall's Model of Reading Development. Chall (1996) presented a well regarded model that divides reading development into six stages. Reading is qualitatively different at each stage as the skills and abilities of the individual, as well as the goals of reading, change over time. In the early stages, the focus is on the medium, or learning the mechanics of accurately and efficiently decoding text. Subsequently, the focus shifts to the message or understanding and gaining knowledge from text. This distinction has also been described as "learning to read" (Stages 0, 1 and 2) versus "reading to learn" (Stages 3, 4 and 5). As a result, texts for students in the earlier stages should contain more high frequency words and simple sentence structure and basic concepts that are likely to be familiar to the child; while texts for students at the later stages become increasingly more complex in writing style and content (Indrisano & Chall, 1995).

Reading development in this model is conceptualized as consisting of stages, but there is overlap between adjacent stages and transitions occur over time. Stage 0, described as "prereading" (birth to about age six), is a time when children gain the foundational knowledge they will need to benefit from formal reading instruction. During this stage, children learn the basic conventions of reading (e.g., that print conveys language and stories, the physical characteristics of text and books, and how to interact with them), that words can be broken down into smaller sounds (i.e., phonemic awareness), that symbols can convey meaning, as well as learning to identify letters by name. In Stage 1, known as “initial reading” (kindergarten through the beginning of second grade), students learn letter-sound correspondences and concentrate on sounding out words. By the end of this stage, children have learned to decode text and can identify some common words by sight. The focus of Stage 2, “confirmation and fluency” (grades two and three), is on “ungluing from print”, or increasing decoding automaticity and prosody when reading. This increased fluency allows the reader to allocate more cognitive resources to comprehension, instead of word identification (LaBerge & Samuels, 1974). During this stage, reading fluency is the focus of development and should be the focus of instruction.

Stage 3, called “reading for learning the new” (grades four through eight) is the first in which the aim of reading shifts to increasing the student’s knowledge. Given students’ level of cognitive development and relative inexperience gaining novel information from print, the texts read during this stage are clear and contain a straightforward presentation of facts, written from a single, authoritative perspective. As students enter high school and Stage 4 (“multiple viewpoints”), the texts become more technically complex and in depth, as varied points of view are presented. Students are expected to critically evaluate information sources and integrate multiple perspectives to achieve a thorough understanding of a topic (Kuhn & Stahl, 2003). The fifth and most advanced stage of reading ability, “construction and reconstruction,” typically develops in college, when the student has developed expertise in a content area and can become a critical reader of texts on the topic. During this stage, when an individual reads about a known subject, he is able to analysis, synthesize and judge detailed material to construct and contribute to his unique understanding of the topic.

In relation to Adams' (2004) four processor model, the alphabetics-based Stage 1 of Chall's (1996) model, may be seen as the time when there is a focus on training the links between the orthographic and phonological processors, with some more limited involvement from the meaning and context processors. Through Stage 2, these processors, particularly the meaning processor, become more interlinked and the entire system becomes more efficient to enable automatic word recognition and fluent reading. Stages 3 and above seem to involve the development of the context processor to acquire knowledge from what it read and to link to the content knowledge areas of the brain. At the same time, associations between all processors continue to strengthen and become more complex as new vocabulary and sentence structures are encountered. Consideration of these two models together provides insight into what happens as a child learns to read and suggests subskills on which to focus instruction and remediation.

Reading Instruction in the Classroom

A brief overview of general trends in reading instruction provides a context in which to understand how knowledge about reading development has been translated into specific teaching strategies. Over the last half century, the field of reading instruction has been dominated by a struggle between two broad, guiding philosophies about what is the best way to teach children to read. Through the 1970's code-based instruction, often called "phonics," was the primary mode of reading instruction implemented in American schools (Chall, 1996). This philosophy supports the use of direct instruction of skills such as phonological awareness, letter-sound correspondences, and blending to foster the decoding of text. The goal of instruction is to help students learn the association between reading stimuli (e.g., letters, words) and their appropriate matched response (e.g., letter sounds, phonological representation of the words). Therefore, the focus of instruction, particularly early in reading development, is on building automaticity of code-based, or bottom-up, processes so that cognitive resources can be freed for comprehension (LaBerge & Samuels, 1974). Key instructional strategies used in the skill-based approach are modeling, demonstration, prompting cueing, feedback, reinforcement, and drill.

In contrast, reading instruction in the 1980s and 1990s was largely dominated by the whole-language philosophy, which suggests that children build on their oral language abilities and learn to read through exposure to a literacy-rich environment (Goodman, 1987). Specifically, the model hypothesizes that readers' desire to construct meaning from text as well as their syntactic and semantic knowledge allowed them to efficiently "guess" words in the text (Goodman, 1967). Thus, in this model, building top-down processes through authentic interactions with text are important to creating skilled reading and little direct instruction of strategies is necessary. The whole-language philosophy falls generally within the constructivist approach to education, which asserts that learners create understanding through their experiences in combination with their previous knowledge (Kamii, 1991).

Over the last 10 years, the pendulum has swung back towards a greater support of code-based instruction (Pressley, 2006). Reviews of the literature indicate that whole language strategies increase students' knowledge of print and reading concepts, which is particularly important in kindergarten, during the prereading stage of development (Stahl & Miller, 1989; Pressley, 2006). However, they lead to smaller increases on key reading indicators for students at-risk for reading difficulty, including those from low socioeconomic status backgrounds (Jeynes & Littell, 2000) and those with lower initial phonological processing skills (Foorman, Francis, Fletcher, Schatschneider, & Mehta 1998; Ball, 1993). Explicit instruction in code-based strategies ensures that students from a variety of backgrounds efficiently learn the fundamental skills of "initial reading," including phonemic awareness, letter-sound correspondences, and word recognition, which are essential for developing skilled reading (Chall, 1996; Stahl & Miller, 1989; Velluntino, 1991; Pressley & Rankin, 1994). Many well known individuals and groups who have researched reading instruction have advocated a balanced literacy approach that incorporates the strengths of both philosophies to meet the needs of the greatest number of students (e.g., Snow, Burns, Griffin, 1998; National Institute of Child Health and Human Development, 2000; Pressley, 2006).

Investigation of evidence-based instructional strategies has also led to a focus on distinguishing between the different components of skilled reading. Most notably, the National Reading Panel (National Institute of Child Health and Human Development, 2000) identified five key subskill areas that have guided research and understanding. The focus of reading development can be seen as progressing through the first four domains of phonemic awareness, phonics, reading fluency and comprehension with vocabulary development occurring throughout. Reading fluency is perhaps the skill area that has received the greatest amount of attention over the past decade (Cassidy & Cassidy, 2009) and is the focus of the current study.

Reading Fluency

Historically, reading fluency has largely been overlooked and has not been the focus of classroom instruction (Dowhower, 1991; National Institute of Child Health and Human Development, 2000). This neglect was likely the result of the notion that reading fluency is merely rapid decoding (Shinn, Good, Knutson, Tilly, & Collins, 1992) and that it is solely a quality of oral reading and, thus, not relevant to the classroom goal of silent reading comprehension (Pikulski & Chard, 2005). However, the theory of automatic information processing in reading (LaBerge & Samuels, 1974) brought increased attention to reading fluency by positing a link between the speed and automaticity of word recognition and the individual's ability to comprehend the text. More recently, the National Reading Panel designated fluency as one of the pillars of reading instruction and concluded that it could be developed through many different instructional procedures (National Institute of Child Health and Human Development, 2000; examples presented below).

Definitions of reading fluency typically incorporate the three observable components of reading accuracy, speed and prosody (Kuhn & Stahl, 2003). Thus, a fluent reader is one who reads words correctly, while using appropriate speed and intonation (National Institute of Child Health and Human Development, 2000). Fluency development is key during Chall's (1996) Stage 2 when children are "ungluing from print," strengthening their decoding skills so they can rapidly recognize words and quickly move through the text. In this way, written text is transformed into a form much like spoken language that can be processed more efficiently (Adams, 1990) and attentional resources are shifted to higher level processes that foster comprehension.

Researchers have consistently demonstrated a positive correlation between reading fluency and comprehension (Pinnell, Pikulski, Wixson, Campbell, Gough, & Beatty, 1995; Deno, Mirkin, & Chiang, 1982; Pikulski & Chard, 2005). Two sets of theories consider different aspects of fluency to explain why this relationship might exist. The first perspective is defined by the theories of automaticity (LaBerge & Samuels, 1974) and the verbal efficiency (Perfetti, 1985). Both models focus on the accuracy and speed aspects of reading fluency and consider the influence of faster word recognition on comprehension. According to these theories, an individual has a limited amount of attention, processing, and memory resources to devote to reading and when decoding is slow and laborious, a majority of these resources are focused on applying symbol-sound correspondences and identifying individual words. This leaves few cognitive resources to focus on the meaning of the text being read. As decoding and word recognition becomes more automatic and efficient, processing resources are freed up to concentrate on constructing meaning.

Another perspective is that reading comprehension is positively affected by the improved expression and phrasing that comes with fluent reading. Prosody is a general linguistic term that describes the rhythmic and tonal features of spoken language (Dowhower, 1991), including stress on the different syllables or words, intonation (i.e., the rise and fall of pitch), and duration of pauses and words (Schreiber, 1991). These suprasegmental features convey information about the phrase structure and focus the listener's attention on key features of the sentence. Written text does not contain such prosodic information and thus, individuals must use morphological, syntactic, and semantic cues in the text to identify phrasal boundaries (Schreiber, 1991; Young & Bowers, 1995). Readers can then use these cues from the text to apply prosodic features to their oral reading of a passage, often called reading with expression by educators, which further enhances their understanding (Dowhower, 1991). These cues also appear to be activated during silent reading to foster comprehension at the word, phrase, and sentence levels (Ashby, 2006). Ascertaining the phrasal organization and syntactic structure provides a greater level of understanding of what is being read, above and beyond identifying the meaning of individual words (Pinnell et al., 1995). Thus, prosodic information helps the student parse the text into larger units and provides clues about meaning that facilitate the process of constructing understanding of the passage.

While both of these perspectives consider the relationship between fluent reading and understanding of the text, neither is clear about whether increased fluency leads to greater comprehension or whether it is a product of improved comprehension. Studies have shown mixed findings (Kuhn & Stahl, 2003) and it is likely that for most of development the relationship is reciprocal, with fluency and comprehension fostering each other (Stecker, Roser, & Martinez, 1998; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003).

Beyond comprehension, fluency is also correlated with proficiency of a range of different sublexical and lexical reading skills (i.e., phonemic awareness, decoding, and sight word recognition), as well as higher level processes (i.e., vocabulary, passage context, and metacognition) indicating its interrelatedness with multiple skills. In fact, research into the theoretical underpinnings of this skill has found that reading fluency is multidimensional and “involves every process and subskill involved in reading” (Wolf & Katzir-Cohen, 2001, p. 220). This conclusion is consistent with connectionist models such as the four processors model (Adams, 2004), which describe reading as a complex process that involves reciprocal input from different systems. Thus, fluency can be seen as both an indicator of the functioning of the overall reading system and an essential, independent feature of skilled reading that acts as an interface between the medium-focused, sublexical skills of reading (e.g., phonemic awareness, alphabetic principle, decoding, sight word recognition) and the message-based skills involved in comprehension (Hudson, Pullen, Lane, & Torgensen, 2009; Kuhn & Stahl, 2003).

Empirical Measurement. Empirical definitions of reading fluency typical focus on reading accuracy and rate (Dowhower, 1991; Schreiber, 1991; National Institute of Child Health and Human Development, 2000), ignoring the third observable component, prosody, as well as the simultaneous development of comprehension (Samuels, 2007). This neglect appears to be the result of the difficulty of quantifying prosody and comprehension quickly, particularly in comparison to words read correctly and rate (see Daane et al., 2005 for a prosody scale used as part of the National Assessment of Educational Progress). Moreover, the proliferation of research on and growing support for reading curriculum-based measurement (CBM; e.g., Shinn, 1989; Fuchs, 2004; Wayman, Wallace, Wiley, Tichá, & Espin, 2007), which uses words read correctly per minute as the key outcome measure of reading ability, has further cemented the link between fluency and the accuracy and rate of oral reading.

Some researchers, including Samuels (2007), believe prosody and simultaneous comprehension are essential features of reading fluency that differentiate truly fluent readers who understand what they read from those who merely bark at print. Certainly all of reading fluency cannot be boiled down to purely word accuracy and rate; however these features are arguably key components of the reading fluency subskill and important indicators of the overall system's efficiency, which includes prosody and comprehension. The current research will consider only reading accuracy and rate, acknowledging that this is a relevant, yet limited perspective of the complex phenomenon of reading fluency. The key outcome variables of this study are percentage of words read accurately (accuracy) and the number of words read correctly per minute. In acknowledgement that prosody and comprehension are not being considered, the latter measure will be called accurate reading rate and not fluency, as is often the case in the research literature.

Reading Fluency Interventions

Given a comprehensive definition of reading fluency, interventions that support a variety of different reading subskills are likely to have a positive impact on fluency (Wolf & Katzir-Cohen, 2001, Hudson, Pullen, Lane, & Torgensen, 2009). However, interventions targeted specifically to promote reading fluency typically foster the accuracy and speed of word recognition in context. What follows is a description of four key types of effective reading fluency interventions that will be used in the current study.

Contingent Reinforcement. Since Edward Throndike and B.F. Skinner, the forefathers of behaviorism first studied it, operant conditioning has been used to modify the behavior of various organisms, including humans. The key principal of this type of conditioning is that a behavior is maintained or eliminated depending on the environmental consequences elicited by the behavior. Reinforcers are consequences that increase the frequency of a behavior, while punishments are consequences that decrease the frequency of the behavior. A reward is an item that is believed to reinforce behavior based on its assumed desirability. It may be a tangible item, such as a toy or piece of candy, a desired activity, such as time playing video games or lunch with the teacher, or removal of an undesirable condition, such as being grounded. The relative desirability and thus motivating quality of a reward can be increased by allowing the individual to choice from a menu of options (Kern et al., 1998).

Typical procedures for using contingent reinforcement (or contingent reward) with reading fluency involve setting a specific goal for performance (e.g., read 70 words correctly per minute, read 10% more words correctly per minute than baseline) and if the student achieves the goal, he is given access to a reward. Contingent reinforcement is not a method of instruction, but the opportunity to access incentives increases student motivation to accurately perform the behavior of concern. The desired behavior may be learning an academic stimulus-response contingency (e.g. being able to correctly read sight words) or increasing the presentation rate of a previously acquired, but infrequently performed skill (e.g., demonstrating a rapid rate of accurate reading). Thus, with respect to increasing reading fluency, contingent reinforcement by itself works best for students who have mastered decoding and have relatively accurate, but slow, automatic word recognition (Haring & Eaton, 1978).

Overwhelmingly, studies have found positive effects for incorporating incentives into the teaching of reading skills (Staats, Minke, Finley, Wolf & Brooks, 1964; Billingsley, 1977) including sight word acquisition (DeVries & Feldman, 1983), reading rate (Lovitt, Eaton, Kirkwood, & Pelander 1971), and reading comprehension (Allyon & Roberts, 1974). Further, focusing a student's attention on increasing the speed of his reading, as occurs by offering a reward for achieving a rate goal, does not have a negative effect on his comprehension of what is read (Roberts & Smith, 1980; Hudson et al., 2009). Adding a reward component to modeling and practice also improves the effectiveness of these instructional methods for increasing reading fluency (Noell et al., 1998). Moreover, generalization of fluency gains from an instructed passage to a novel one is improved by offering a reward (Daly, Bonfiglio, Mattson, Persampieri, & Foreman-Yates, 2005). In sum, contingent reinforcement or reward is a simple, yet powerful instructional component that can be used to improve a student's reading fluency.

Repeated Reading. One of the most widely studied methods for improving reading fluency is repeated reading (RR). The basic procedure involves having the student read the same piece of connected text multiple times to build fluency (Samuels, 1979). Variations include: whether the student reads the passage a fixed number of times (often 3 or 4; Moyer, 1982; Rashotte & Torgensen, 1985) or until a particular rate criteria is obtained (Samuels, 1979); having the student read orally or silently (Rose, 1984);

cueing the student to focus on speed or comprehension (O'Shea, Sindelar, & O'Shea, 1985); and, providing feedback, sometimes through graphing, on reading speed or errors (Meyer & Felton, 1999). The basic RR procedure can also be enhanced by adding instructional components such as pre-reading instruction, error correction, and modeling while the student reads (i.e., choral or prosody reading; Schreiber, 1980) or before (Rose, 1984). This latter type of previewing has evolved into an independent intervention called listening passage preview, which is described below.

The RR procedure was first presented in the educational literature by Samuels (1979) as a way to foster automatic word recognition and comprehension. Through repetition of the passage, the relationship between the written letters and words and their phonological equivalents can be strengthened, resulting in stronger stimulus control. Furthermore, the repetitions produce an increase in the speed of accurate reading and reveal metalinguistic features such as prosody and syntax (Schreiber, 1980; Moyer, 1982). The rapid availability of these features can then work together to foster comprehension of the passage. Thus, when word recognition becomes automatic an individual's limited cognitive resources can be devoted to understanding the text, instead of expending processing effort on decoding (LaBerge & Samuels, 1974).

Reviews of the RR literature demonstrate that this intervention promotes many positive outcomes for normally developing students, as well as those with learning disabilities and adult learners (Therrien, 2004). RR has been shown to produce gains in reading speed and accurate rate on individual passages and promote reading fluency more generally, the primary goals of this method. In his meta-analysis, Therrien (2004) found that RR had a moderate-large impact on reading rate for the practiced text (Cohen's *d* mean effect size = 0.83) and that rate improvements generalized to novel passages to a moderate extent (*d* = 0.50). Generalization of rate gains are a less consistent finding in RR studies and appear more likely when there is extended implementation of RR or when there is a high percentage of content overlap in the practiced and unpracticed passages (Rashotte & Torgensen, 1985; Ardoin, McCall, & Klubnik, 2007).

Consistent with the noted link between fluency and comprehension, studies have generally found an increase in comprehension scores in response to RR (Dowhower, 1987; O'Shea, Sindelar, & O'Shea, 1985). In considering the research base as a whole, Therrien (2004) found that, while smaller than its effect on rate, RR also has a positive impact on reading comprehension. Not surprisingly, the average effect size for practiced passages is larger (*d* = 0.75) than the average generalized effect to novel passages (*d* = 0.25). It is interesting to note that the majority of comprehension gains occur after three readings (Therrien, 2004), so an RR procedure consisting of three repetitions appears to be the most efficient option for promoting understanding.

Many authors note that the RR may increase struggling readers' motivation to read because they are able to have more successful interactions with reading and enjoy the repetitive nature of this technique (Chomsky, 1978; Samuels, 1979; Moyer, 1982; Rashotte & Torgensen, 1985). The improved perception of reading and the increased time spent reading provided by RR may also combat some of the negative outcomes for struggling readers predicted by the Matthew's effect (Stanovich, 1986; Kuhn & Stahl, 2003).

Listening Passage Preview. Often linked with RR, listening passage preview (LPP; Daly & Martens, 1994) provides additional instructional support by having the text modeled to the student before she is asked to read it. Sometimes called assisted reading (Dowhower, 1987) or listening while reading (Rasinski, 1990), during this intervention the student follows along silently, as indicated by sweeping his finger under the text, while listening to a fluent reading of the passage. The student then reads the text orally one or more times to practice her reading. In the literature, fluent models have been provided by audio recordings (e.g., Dowhower, 1987) adult instructors (e.g., Rose, 1984), and peer tutors (e.g., Eldredge, 1990).

When the student reads silently along with the model she is able to connect an accurate reading of the passage with the words on the page. Some have hypothesized that modeling fluent reading provides the student with a guide of the rate the passage should be read (e.g., Cunningham, 1979), but varying the rate of the model does not result in differences in the student's subsequent reading rate, indicating that speed of reading is not the key feature attended to by the student (Skinner, Adamson, Woodward, Jackson, Atchison, & Mims, 1993). Instead, when students listen to reading rates that are only somewhat faster than their current rate (Schreiber, 1980, in Meyer & Felton, 1999), they exhibit improved word accuracy, suggesting that modeling primes the orthographic and phonological processors by providing the child an opportunity to reinforce the connection between the visual form of the words and the correct pronunciation (Skinner, Logan, Robinson, & Robinson, 1997). Further, it seems likely that hearing the words being read cues the child's brain/neurological pathways to more quickly come up with the correct words when they have the opportunity to read the passage on their own (Forbach, Stanners, & Hochhaus, 1974).

LPP has been the subject of less independent research and is often grouped into reviews of RR (e.g., National Institute of Child Health and Human Development, 2000; Therrien, 2004). Comparisons between the two procedures have been mixed. While some studies have found higher accurate reading rates for LPP (Rose, 1984; Rose & Sherry, 1984; Daly & Martens, 1994), others have found no significant differences in the gains (Rasinski, 1990; Dowhower, 1987). This lack of a robust finding in terms of an increased reading rate with LPP may be because this procedure fosters increased accuracy instead of explicitly addressing rate. As a result, any gain in words read correctly per minute is due more to a greater number of accurate words than more words read per minute (Skinner et al., 1997). Moreover, LPP seems to have better results with struggling readers than normally developing students (Dowhower, 1987; Kuhn & Stahl, 2003). Students with reading deficits may benefit more from the emphasis on accuracy provided by LPP than peers who already have well developed decoding strategies and are working on increasing their reading rate.

Students' comprehension for a given passage increases after it is presented using LPP at a level that is consistent with RR (Dowhower, 1987; Kuhn & Stahl, 2003). Some of this improvement may be the result of improvements in reading prosody and the student's ability to parse text into meaningful phrases (Schreiber, 1980; Kuhn & Stahl, 2003). Further, during the passage preview students can concentrate on understanding the text since they are using fewer cognitive resources to decode (Hale, Skinner, Winn, Oliver, & Allin, 2005). There has been relatively little study of LPP's influence on generalized comprehension, but the effect appears to be small (*d* = .10; Therrien, 2004).

Error Correction. In this intervention, the student is given corrective feedback by a tutor about word pronunciation errors or omissions he makes during oral reading. The aim of this method is to improve a student's reading accuracy by providing instruction on the words that he fails to read correctly. Once again, multiple variations of this method have been described in the literature. Some alternatives include: whether the tutor says the correct word (i.e., word supply) or provides cues to help the student figure out the word (i.e., phonic analysis; for comparisons, see Rose, McEntire, & Dowdy, 1982); providing the correct word during reading, immediately after the error, and/or reviewing and practicing the words after the student completes the passage (see Singh, Winton, & Singh, 1985; O'Shea, Munson, & O'Shea, 1984); whether the student re-reads the word, the phrase or the sentence in which the error word is embedded for repeated practice (see Singh, 1990; O'Shea et al., 1984); and the number of times the student must repeat the corrected text (typically 1-3 times; see O'Shea et al., 1984; Singh & Singh, 1986).

Like LPP, the focus of error correction is on improving students' accuracy through the provision of a model of correct pronunciation. Error correction is more individualized and efficient because it identifies those specific words that the student cannot read automatically and helps him learn to correct his errors. Moreover, in drill techniques, such as phrase drill (O'Shea et al., 1984), students must repeat the corrected word or phrase in which it is embedded multiple times, which allows the orthographic and phonological processors to strengthen the relationship between the visual representation of the word and the correct response within the context of the passage. Drill can also be seen as a type of overcorrection that may establish a negative reinforcement contingency that encourages the accurate response the next time the word is encountered because the student finds the drill repetitions somewhat aversive and wants to avoid them (Begeny, Daly, & Valleley, 2006).

Goodman (1973) has expressed concern that word supply, where the tutor provides the correct pronunciation immediately after the student makes the error, has a negative effect on reading comprehension because it focuses that student’s effort on accuracy and the corrections disrupt the student’s attention to the content of what is being read. Allington (1983) also states that this method may cause students to become overly reliant on external correction, preventing them from learning to self-monitor. In fact, error correction has been found to have a positive effect on both accurate rate and comprehension, perhaps because the student is able to focus processing resources on understanding the text instead of decoding (Pany & McCoy, 1988). Also, providing feedback on every error the student makes leads to higher accuracy and better comprehension than correction only for errors that affect meaning, which is recommended by whole word advocates, like Goodman. When students read with higher accuracy they are better able to monitor the content of what they are reading and are more likely to self-correct (Leslie & Osol, 1978; Singh et al., 1985). So, in addition to accuracy, external correction may also have a positive effect on self-monitoring and self-correction; however more research on this topic is needed.

Error correction procedures have been shown to improve the oral reading outcomes of students with learning disabilities (e.g., Pany & McCoy, 1988) and mental retardation (e.g., Singh et al., 1985), as well as other struggling readers (e.g., Jones & Wickstrom, 2000). While a comprehensive review of error correction has not been conducted, one consistent finding is that drill and particularly drill practice within a segment of the text, leads to superior reading accuracy than individual word supply alone (O'Shea et al., 1984; Jenkins, Larson, & Fleisher, 1983). There have been mixed findings regarding the relative effectiveness of phonetic prompting (e.g., providing the first sound) versus word supply techniques and immediate versus delayed feedback (Barbetta, Heward, & Bradley, 1993; Singh & Singh, 1985; Barbetta, Heward, Bradley, & Miller, 1994). The contradictory findings may be a result of differences in the studies' participants' learning histories and disability diagnoses and the context of the error correction. While phonetic prompting may be more useful to students still learning to decode, providing the correct pronunciation of the whole word is a more efficient method for those who struggle more with fluency. Further, the preferable drill method provides immediate feedback when the student makes an error, but also provides delayed practice. Given all these factors, the phrase drill technique, which includes immediate word supply and drill of short segments of the text (O'Shea et al., 1984) is one of the more effective and efficient form of error correction for students building reading fluency.

Targeting Reading Interventions

While many interventions show general effectiveness for improving students' reading skills, they are most successful when they match the individual needs of the student (Daly, Lentz, & Boyer, 1996). Fuchs and Fuchs (1986) proposed that instructionally relevant data are derived from measures that are direct, idiographic, repeated, and ecologically sensitive. Two key assessment frameworks that meet these criteria and are becoming increasingly more prevalent in the schools (Shapiro, Angello, & Eckert, 2004) are curriculum-based measurement (CBM) and curriculum-based assessment (CBA).

Curriculum-Based Measurement. Perhaps the most popular and well-studied method of instructional decision-making is CBM (Deno, 1985; Fuchs, 2004). The purpose of CBM is to obtain relevant information about a student's academic abilities by systematically measuring her responses on curricular materials. Most often outcome measures are in the form of a rate, or number of correct responses over a given time period, such as the number of words read correctly per minute that is used to evaluate reading fluency. These simple measures have been found to be an indicator of overall ability in the skill area of concern, with a particularly strong research base in the area of reading achievement (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Reschly, Busch, Betts, Deno, & Long, 2009). CBM data are also sensitive to change, allowing for repeated measures and the ability to monitor an individual's skill growth over time (Deno, Marston & Tindal, 1985-1986; Deno, 2002). Moreover, it is possible to determine whether progress is being made in response to instruction or if curricular changes are necessary.

A student's CBM performance can also be compared to a norm group to determine whether she is performing at the expected level. To aid such instructional decision-making, Hasbrouck and Tindal (2006) have published an extensive set of national normative data on students' oral reading fluency rates. Their paper provides the average words read correctly per minute for various percentile ranks at the fall, winter, and spring benchmarks for students in grades 1 through 8. Educators can compare a student's performance to these national norms to evaluate her relative standing compared to grade level peers and determine whether additional intervention is necessary.

CBM is a monitoring tool that can be used to determine whether an individual student needs remedial help and to evaluate the success of a chosen intervention. However, additional strategies are needed to match interventions to an individual student's needs.

Curriculum-Based Assessment. One dimension of instruction that can easily be individualized is the curricular material that is presented. Curriculum-based assessment (CBA; Gickling & Thompson, 1985) involves having the student attempt academic tasks that vary based on their level of difficulty or the instructional objectives they address in order to determine the appropriate content for instruction. Based on Vygotsky's zone of proximal development, Betts (1946) hypothesized that for every skill there is an optimal level of task difficulty that promotes learning known as the instructional level. If the task contains too much new material, it will be at the frustration level and too difficult for the individual to make growth. Conversely, a task at the independent level contains a large percentage of known material and while it is easy for the individual to complete, little new learning occurs.

Gickling and Thompson (1985) hypothesized the instructional level for reading connected text was that on which the student demonstrates 93% to 97% accuracy, while for drill tasks, such as sight word recognition, spelling, or math facts a 70% to 85% rate of accuracy was necessary. Thus, by testing a student's accuracy at different levels of difficulty (e.g., grade levels) it is possible to determine that at which one an optimal level of skill learning will occur. Research has demonstrated that both CBA and CBM result in reliable data that lead to improved student learning (Burns, Tucker, Frame, Foley, & Hauser, 2000; Burns, 2007; Fuchs, Fuchs, Hamlett, & Whinnery, 1991; Marston, 1989).

While these curriculum-based methods allow evaluators to decide what content to teach and provide a means for evaluating intervention effectiveness, other methods must be used to determine *how* to best teach an individual child. The brief experimental analysis (BEA) procedure allows educators to test the effectiveness of different interventions in order to determine how to best teach an individual student.

Brief Experimental Analysis

Empirical evaluation of a student's response to different interventions can be used to individualize instruction. Brief experimental analysis (BEA) is a technique devised to help identify intervention components that increase or improve desired academic behaviors (Daly, Martens, Dool & Hintze, 1998; Daly, Martens, Hamler, Dool, & Eckert, 1999; Jones & Wickstrom, 2002). The basic philosophy guiding BEA is that by using a single-subject, alternating treatment design, one can systematically test a student's performance under different intervention conditions. Comparisons can then be made across conditions to determine which intervention leads to the greatest increase in accurate academic responding over a non-intervention baseline level. This testing of alternatives can occur over a short period of time so an effective intervention can be implemented quickly once a need has been identified.

BEA is derived from the methods of functional analysis, a behavior analysis technique that is used to experimentally test hypotheses about the antecedents and consequences that evoke and maintain certain behaviors (Cone, 1997). McComas, Wacker, Cooper, Asmus, Richman, and Stoner (1996) completed the first study of BEA that examined how different intervention conditions affected academic behaviors. Using a multielement design, the researchers compared the effect of individual and combined instructional strategies on individual students' correct spelling and reading comprehension. For each student, the researchers found one intervention or a combination of interventions that increased accurate academic responding.

Additional research has expanded BEA as a model for addressing academic performance problems (Daly, Witt, Martens & Dool, 1997; Jones & Wickstrom, 2002).While the BEA framework can be used to evaluate interventions across a range of academic areas, the majority of the research has been conducted using oral reading fluency (Daly et al., 1999; Eckert, Ardoin, Daly, & Martens, 2002; Noell, Freeland, Witt & Gansle, 2001). Recent extensions of BEA include application to math instruction (Carson & Eckert, 2003) and using it to select appropriate reading interventions for parent tutoring (Gortmaker, Daly, McCurdy, Persampieri, & Hergenrader, 2007), small groups of students (Bonfiglio, Daly, Persampieri, & Andersen, 2006), and English language learners (Mallory, Gilbertson & Maxfield, 2007).

BEA procedures. A number of procedural variations of BEA are present in the literature (e.g., Daly et al., 1998; Eckert, Ardoin, Daisey, & Scarola, 2000; Gortmaker et al., 2007; Noell et al., 2001). Most vary in the way the tested treatments are combined and ordered (Daly, Andersen, Gortmaker, & Turner, 2006), the number of times an intervention is tested, and whether a no treatment (baseline) condition (e.g., Daly et al., 1999) or the treatment that produced the lowest rate of responding (e.g., Jones & Wickstrom, 2002) is reapplied to confirm the effectiveness of the "best" treatment. However, the essential features of a single-subject, multielement treatment design that includes a reversal remain consistent across them.

The BEA procedure can be broken down into three key steps: 1) collecting pre-intervention baseline data, 2) testing the student's response to different interventions, and 3) confirming the intervention effectiveness (typically through a no intervention-intervention reversal). This entire process can be done in a single session or over multiple sessions, depending on the extensiveness of the BEA. However, typically each session lasts between 15 to 20 minutes (Daly et al., 2002) and a limited number of intervention conditions should be administered during a single session in order to reduce the effect of fatigue on the student's performance (Daly et al., 1999).

When reading fluency is the focus of the BEA, a number of short reading passages from the reading level in which the student is currently being instructed must be selected before the experimental procedure begins. Generic, grade-leveled progress monitoring probes (e.g., those from AIMSweb [EdFormation, 2005] or DIBELS progress monitoring systems [Good & Kaminski, 2002]) are also sometimes used.

The first step in conducting a BEA is collecting pre-intervention baseline data on the outcome of concern. This may be done by giving the student a single probe for efficiency or multiple probes to help control for the impact of measurement error on a single passage.CBM provides an ideal measurement tool that can be used for BEA, particularly oral reading rate and accuracy for evaluating reading fluency. During the first, baseline condition the examiner asks the child to read one or more passages aloud and records his oral reading rates from the first minute (i.e., the number of words read correctly per minute [WCPM] and the number of errors per minute).

In the second step of the experimental process, different treatments are attempted to determine which one leads to the greatest increase in academic responding (i.e., WCPM) over baseline levels. There is no prescribed number or list of interventions that should be attempted, but some frequently included when addressing fluency concerns are those described above, including: contingent reward (Lovitt, Eaton, Kirkwood, & Pelander, 1971), repeated reading (Rashotte & Torgeson, 1985; Rose, 1984), listening passage preview (Daly & Martens, 1994), phrase drill error correction (O'Shea, Munson, & O'Shea, 1984), and easier material (Daly, Martens, Kilmer, & Massie, 1996). The particular treatments selected for testing by a practitioner will vary based on the perceived needs of the student and the school's available resources. While interventions may be tested individually, they can also be combined to create more comprehensive treatment packages that may have stronger effects and often more closely resemble classroom instruction (Daly et al., 2006).

It is typically recommended that the attempted interventions be sequenced hierarchically from least intensive (e.g., providing a contingent reward) to most intensive (e.g., combined treatments or those that include altering the curriculum for the individual student) by adding components to the treatment package in each successive condition (e.g., Daly et al., 1999; 2002; VanAuken, Chafouleas, Bradley, & Martens, 2002). Another option is to start with the most comprehensive treatment package and dismantle it by removing intervention components across treatment conditions (Daly, Persampieri, McCurdy, & Gortmaker, 2005). The goal of ordering interventions hierarchically is to select the most simple intervention that improves academic responding (Daly et al., 1999; Daly et al., 2006; Harding et al., 1994). Practically, it is assumed that interventions that require less adult assistance are preferable because they are easier to implement and, thus, more likely to be carried out in the classroom (Daly et al., 1999). Additionally, children benefit from less intensive interventions that differ only minimally from the desired academic behavior (e.g., independent reading) based on the assumption that the more similar the treatment condition is to the actual environment, the more likely benefits will generalize (Daly et al., 2006).

During the intervention testing phase of BEA, each treatment is applied to a single, novel passage. The effectiveness of the intervention is usually assessed based on the student's reading rate during his final reading of the instructional passage (i.e., after the intervention has been applied). After a treatment has been attempted, the WCPM and errors on the passage are recorded, typically in a graphic format, and compared to baseline performance to examine the impact of the intervention on fluency.

Administration of subsequent interventions continues until one intervention causes a meaningful improvement in reading fluency over baseline levels.

There is no objective rule that states how large the difference in fluency rate between the intervention condition and the baseline condition must be for the treatment to be considered effective. Noell and associates (2001) use the standard of a 20% increase over baseline to establish effectiveness. However, a proportional criterion may not be appropriate for all students. In particular, a student who is only slightly below the norm may not make up this difference, particularly in a single performance. A raw number criterion such as an improvement of 20 WCPM may also not be useful because a student's baseline reading level affects the amount of growth that can be expected under optimal conditions. For example, while a 20 WCPM gain is possible for children reading at a relatively high level (e.g., 75 WCPM), it would be an unreasonable expectation for children beginning at a low level (e.g., 25 WCPM; Daly et al., 1999). To empirically address this issue, Burns and Wagner (2008) used meta-analytic procedures to investigate the percent of growth and number of WCPM over baseline present for effective intervention conditions within published BEA studies. They found an average growth of 73% and an average increase in fluency of 30.19 WCPM and suggested that these values might serve as criteria for determining an effective intervention. However, these values are averages that may not be appropriate for every individual student as a result of the concerns discussed above. Thus, to avoid these issues, Daly and colleagues (1998) defined an effective intervention as one that caused "a clearly visible difference in performance relative to baseline and other treatment conditions" (p. 211). While somewhat vague and subjective, this rule may be sufficient for making judgments of efficacy until more objective criteria, perhaps dependent upon the student's baseline fluency rates, are devised.

Once an apparently effective treatment condition has been identified based on achieving the selected criteria, the final phase of the BEA is conducted to finalize the intervention selection. A mini-reversal or retesting of the intervention's impact confirms that the observed improvement is due to the intervention and not to the relative difficulty of the passage or measurement error, conditions which are more likely when a single passage is used to judge efficacy. The mini-reversal consists of administering a baseline/no intervention condition followed by a re-administration of the intervention that led to the largest gain on novel passages. A replication of results (i.e., a return to a lower accurate rate during the baseline condition and a similar level of increase when the intervention is re-administered) demonstrates that the intervention reliably improves performance. If a reversal does not replicate the initial findings (i.e., the intervention does not result in a replicable increase in rate), then other interventions are administered until another demonstrates a significant increase in academic responding, at which point another mini-reversal is conducted. The treatment condition that causes a notable increase in academic responding during both the initial administration and the mini-reversal is considered effective for the individual student (Jones & Wickstrom, 2002). This intervention is then selected for long-term application to address the student's learning needs, while being continually monitored to ensure effectiveness (Wilber & Cushman, 2006).

In sum, through the BEA procedure an effective, individualized intervention strategy is identified by determining which treatment condition leads to the greatest, replicable increase in reading fluency over baseline rate. BEA has been the subject of much recent research within the field of educational psychology and the following section describes key conclusions drawn from this research.

Empirical findings. The research literature consistently supports the use of BEA as an effective and efficient method for identifying appropriate fluency interventions for individual students. Multiple studies have found that through BEA it was possible to identify at least one intervention that improved reading fluency to such an extent that it was considered effective for all students tested (Daly et al., 1998, 1999; Eckert et al., 2000, 2002; Fontanni-Axelrod, 2006; Noell et al., 2001; Jones & Wickstrom, 2002; Wagner, McComas, Bollman & Holton, 2006; Wilber & Cushman, 2006). While such results are promising, it may not always be possible to determine the optimal treatment for all individuals using BEA (Daly et al., 1999). In such cases, BEA can help rule out ineffective treatments (Daly et al., 1999), but an extended analysis of potential treatments may be warranted to help discriminate the most effective intervention for the student (Daly et al., 2002).

Another consistent finding is that intervention effectiveness varies between individuals with the same apparent fluency problem (Daly et al., 1998, 1999; Eckert et al., 2000, 2002; Fontanni-Axelrod, 2006; Noell et al., 2001; Jones & Wickstrom, 2002). In other words, different interventions resulted in the greatest improvement in fluency for different children dependent on their individual instructional needs. Furthermore, it is not always the case that more intensive interventions lead to greater increases in fluency or even the maintenance of gains seen during less intensive conditions.

Extended analyses. In order to determine how well the findings of brief analyses correspond to those of more typical, long-term experimental procedures, their results have been compared to those of extended analyses. Findings generally confirm the intervention selected through BEA. Despite the use of only a few data points in decision making, the effects of the best treatments selected using BEA are stable over time as indicated by treatment fluency averages consistently higher than baseline averages in the extended analysis (Daly et al., 2002; Fontanni-Axelrod, 2006; Jones & Wickstrom, 2002). Moreover, extended analysis confirms the classification of interventions as either effective or ineffective 83% of the time (Noell et al., 2001). However, the difference in reading fluency scores between the most and least effective intervention conditions diminished over an extended implementation (VanAuken et al., 2002), indicating that while BEA may be a good predictor of initial relative effectiveness, it is less predictive over time. As a student's skill level improves due to the implementation of the effective intervention, it would be expected that their instructional needs would change and the selected intervention would no longer improve academic responding. Thus, BEA may need to be re-administered after a period of time to re-establish the most appropriate intervention (VanAuken et al., 2002). Overall, the findings of BEA tend to match those of more extended analyses and offer treatment utility in that they contribute to beneficial treatment outcomes (Hayes, Nelson, & Jarrett, 1987).

Learning Hierarchy/Instructional Hierarchy

The learning hierarchy/instructional hierarchy (LH/IH; Haring & Eaton, 1978) framework presents a four stage learning sequence and matched interventions that foster progression through each stage. Initially named the learning hierarchy by the original authors and called an instructional hierarchy by later researchers (e.g., Daly, Lentz, & Boyer, 1996) I will use the term "LH/IH" in order to acknowledge the two components of this framework. BEA researchers often use the LH/IH as a basis to select interventions to be tested (Daly et al., 1997, 1998; Jones & Wickstrom, 2002; Wilber & Cushman, 2006) and the assumed link between these two hierarchies is the focus of the current study.

The LH/IH falls broadly within the behaviorist perspective of learning and guides instructional efforts with the goal of promoting stimulus control (e.g., linking letters/written words with their phonological representation) and skill generalization (Ardoin & Daly, 2007). This framework proposes that the learning of any skill progresses through four increasingly complex stages of development (*learning hierarchy*) and that different instructional methods foster development at each stage (*instructional hierarchy*).

The learning hierarchy was described by Haring and Eaton (1978) after considering how students learn different academic skills. The first stage, *acquisition*, lasts from the time the individual first attempts the skill until he learns how to perform it with relatively high accuracy. It was hypothesized that 90% to 100% accuracy was necessary before the student began transitioning into the *fluency building/proficiency* stage. During this second stage, referred to in this paper as the "proficiency stage," the individual practices the behavior until he can accurately perform it at the speed needed to effectively use it. Finally, in the *generalization* and *application/adaptation* phases the individual learns to produce the skill in different contexts and modify it to solve problems in novel situations, respectively.

As depicted in Table 1, matched to the four stages of the learning hierarchy are different instructional methods which are hypothesized to provide the type of practice necessary to support the advancement of learning at each stage; this heuristic is known as the instructional hierarchy (IH; Daly, Lentz, & Boyer, 1996). For example, students in the acquisition stage are learning how to correctly perform a skill and thus benefit most from explicit instructional techniques, such as demonstration and modeling. Thus, listening passage preview (Daly & Martens, 1994), which involves demonstration of fluent reading and error correction techniques such as phrase drill (O'Shea et al., 1984) are interventions that are beneficial to students at the acquisition stage in demonstrating reading fluency.

Students at the proficiency stage, on the other hand, can perform the skill with adequate accuracy, but need repeated practice and reinforcement to be able to obtain a rate of fluency that allows them to use the skill in a meaningful way. Multiple exposures inherent within instructional techniques such as repeated reading (Rashotte & Torgensen, 1985) and reinforcement provided by reward for rapid reading (Lovitt, Eaton, Kirkwood, & Pelander, 1971) or performance feedback (Eckert, Dunn & Ardoin, 2006) help foster growth within the proficiency stage.

Students in the generalization and adaptation phases are best served by being asked to use the skill in diverse contexts so they can learn when to apply the skill and how to use it more broadly. Generalization and adaptation can be supported by giving students multiple opportunities to read passages with a high level of content or word overlap (Ardoin, Eckert & Cole, 2008).

Table 1

Learning Hierarchy (Figure 2-1 from Haring and Eaton, 1978)

Level	Emphasis	Strategies
Acquisition	accuracy of response	1. demonstration
		2. models
		3. cues
		4. routine drill
Fluency or	speed	1. repeated novel drills
Proficiency		2. reinforcement
Generalization	novel stimulus	1. discrimination training
		2. differentiation training
Adaption	adapted response	1. problem solving
		2. simulations

Daly, Lentz, and Boyer (1996) consider the types of academic responding that are most relevant for measuring progress at the different stages of the IH/LH. Accuracy (i.e., the percentage of responses that are correct) is the key outcome at the acquisition stage, while accurate rate (i.e., the number of accurate responses performed during a given time period) is important at the proficiency stage. Generalization or "displaying a recently acquired behavior either in multiple settings or in the appropriate context in which the individual is expected to demonstrate the behavior" (Daly et al., 1996, p. 375) is important in the third stage of the IH. Adaption of the learned skill in novel situations is the desired outcome during the final stage, but it is more difficult to measure and has not been widely studied.

Daly and colleagues (1996) also demonstrate that the treatment components provided in the IH show the greatest effectiveness for improving reading skills when the appropriate form of academic responding is used as an outcome measure. For example, self-monitoring, which includes a modeling component suggested for the acquisition stage, improved sight word accuracy, but the addition of contingent reinforcement (a proficiency stage strategy) provided no additional benefit to accuracy (Lalli & Shapiro, 1990). However, reinforcement increases oral reading rates (Lovitt et al., 1971), which is consistent with what would be predicted by the LH/IH model since reinforcement is a proficiency intervention component and rate is the matched form of academic responding during this stage. Thus, in addition to providing a guide of skill development and suggesting treatment components that foster growth at each of the stages, the LH/IH implies the relevant forms of academic responding that should be used to evaluate intervention effectiveness at each stage.

The LH/IH model provides a useful framework for identifying effective intervention components based on individual student needs and implies relevant forms of academic responding for each stage of the learning hierarchy. Students' pre-intervention level of accuracy and accurate rate indicate the type of intervention that is most likely to be successful. Consistent with contentions of the LH/IH, individuals with lower accuracy and rate scores demonstrate improvement in response to more intensive, acquisition interventions, while those with high accuracy, but low rate scores respond best to proficiency interventions (Chafouleas, Martens, Dobson, Weinstein, & Gardner, 2004; Burns, Codding, Boice, & Lukito, 2010). Further, once students' reading accuracy was improved using an acquisition stage intervention, they demonstrated a greater increase in reading rate in response to a proficiency intervention (Parker & Burns, in press).These applications of the LH/IH to predict intervention effectiveness were found in the areas of reading (Chafouleas et al., 2004; Parker & Burns, in press) and math (Burns et al., 2010), supporting the validity of a skill-by-treatment interaction.

There have been some efforts to integrate the LH/IH into an instructional decision making approach (Daly, Chafouleas, & Skinner, 2005; Burns, Christ, Boice, & Szadokierski, 2009), but there is currently limited empirical evidence regarding identification of the LH stage at which a student is currently performing. Haring and Eaton (1978) suggest that students remain at the acquisition stage until they perform a skill with 90% or greater accuracy, at which time the primary focus of learning shifts to the proficiency stage. The lower bound of the instructional level for reading (i.e., 93%; Gickling & Thompson, 1985) has also been suggested as an accuracy cut score for differentiating performance at the acquisition or proficiency stages (Burns et al., 2009; Parker & Burns, in press). Alternatively, Daly, Chafouleas and Skinner (2005) have proposed subjectively evaluating whether a student tends to struggle more with accurate performance or rate of accurate performance as part of their instructional decision-making framework.

Present Study

While BEA is an effective method for selecting interventions that improve academic responding, it would be useful for practitioners to have a guideline that would predict intervention effectiveness without requiring testing. Evidence of a skill-by-treatment effect (Burns, Codding, Boice, & Lukito, 2010; Chafouleas, Martens, Dobson, Weinstein, & Gardner, 2004) indicates that the LH/IH could provide a useful framework selecting a matched intervention based on a student's current skill performance.

By taking advantage of the assumed links between instructional strategies and the stages of the LH, it is possible to use BEA techniques to deduce where a student falls in the learning sequence for the particular skill. By testing interventions that include strategies on the IH and determining which is most effective, the LH stage at which a student is currently performing can be inferred. Consideration of baseline characteristics of groups of students that are identified as performing at the acquisition or proficiency stages of development based on their intervention response can be used to evaluate the validity of the LH as a description of reading fluency growth. The links between the LH and IH would also be confirmed by developmental differences between the two groups of students. Moreover, indicators of stage membership could be used to provide guidance on intervention selection for struggling students.

The goal of the current research is to consider how students' accuracy and accurate rate when reading grade-level passages relates to whether they are in the acquisition or the proficiency stage for oral reading fluency. Accuracy and rate are two of the key features of fluent readers and these measures are used to effectively evaluate performance with in the CBA and CBM frameworks and through the BEA procedure. Moreover, these forms of academic responding are linked to the acquisition and proficiency stages of the LH and are likely to be indicators of a student's place within the LH/IH. However, given the focus of this study on oral reading accuracy and rate, the implications are limited to these two observable features of reading fluency and not necessarily to other components such as prosody and comprehension.

In this study, a student's baseline levels of reading accuracy and accurate rate are obtained, followed by implementation of the BEA procedure to evaluate his response to two intervention packages. One package combines listening passage preview and phrase drill, which include modeling and error correction (M-EC), respectively, instructional strategies linked with the acquisition stage of the learning hierarchy. The other intervention package is composed of repeated reading and reward (RR-R), which include the repeated drill practice and reinforcement strategies that are best suited for students in the proficiency stage of the instructional hierarchy.

Students are categorized based on which of the two intervention packages was found to be most effective and average baseline accuracy and rate measures of each of these groups of students are compared. In addition, because students are from different grade levels and tested at different times of the year, rate data will be turned into a normalized rate measure. In this case, students' baseline rate was transformed into z-scores that represented how a student's performance compared to the normative expectations for grade level and time of year (Hasbrouck & Tindal, 2006).

The research questions guiding this study are:

1. What are the differences in baseline accuracy, rate and normalized rate among students for whom only the modeling-error correction (M-EC) intervention was effective, those for whom only the repeated reading-reward (RR-R) intervention was effective and those for whom neither or both were effective?
2. What is the relationship between intervention effectiveness status and baseline accuracy, rate and normalized rate?
3. How well does students' baseline accuracy and rate or normalized rate predict M-EC effectiveness status?
4. How well does students' baseline accuracy and rate or normalized rate predict RR-R effectiveness status?

In addition, the following exploratory questions will be asked:

What cut score or range of scores for accuracy, rate, or a combination of these measures provides sufficient specificity and sensitivity to predict whether a student will benefit from the M-EC intervention or the RR-R intervention?

Based on the theoretical foundations of the LH/IH and the relevant empirical findings presented above, the following results are expected for each of the research questions:

1. It is hypothesized that baseline accuracy percentage, accurate rate and normalized rate averages will be significantly different between the two groups.

Specifically, all will be higher for students who respond better to RR-R intervention because they are in the more advanced proficiency stage. Further, there will likely be a greater difference in the accuracy percentage than the rate measures based on Haring and Eaton's emphasis on accuracy as an indicator of progress from the acquisition to the proficiency stage.

In addition it is hypothesized that students who do not show a differential response to either intervention will fall into one of two categories. They may show a similar response to both interventions and have baseline accuracy and rate scores between the other groups indicating that they are transitioning from the acquisition stage to the proficiency stage. Alternatively, they may show little response to either intervention and baseline scores that are well below those of the other two groups indicating that they are still developing more basic reading skills such as phonics.

2. It is hypothesized that as baseline accuracy, rate and normalized rate increase, the likelihood of responding better to the RR-R intervention will increase because students are more likely to be further along on the learning hierarchy. It is unclear whether both factors together or one individually will best account for individual variability in intervention effectiveness, although the theoretical framework of the LH/IH indicates that accuracy should be the best predictor since it is the measure aligned with progress in the acquisition stage. On the other hand, accurate rate may provide a better measure since it includes information both about accuracy and speed of performance, important factors at both the acquisition and proficiency stages.

The normalized rate data are predicted to have a stronger relationship with effectiveness data than the raw rate data because a student's intervention needs are hypothesized to be based on how discrepant they are from grade-level expectations. Furthermore, this normalized data controls for the students being tested from different grade levels and at different times of the year, factors which likely impact accurate reading rate.

3. It is hypothesized that accuracy alone will be a better predictor of intervention effectiveness for students for whom the M-EC intervention was effective than rate or normalized rate. Daly and colleagues (1996) state that accuracy is the form of academic responding best matched to the acquisition stage, in which modeling and error correction interventions are believed to be the most effective.

4. Similarly, it is hypothesized that accurate rate will be a better predictor of intervention effectiveness for students for whom the RR-R intervention was effective based on it being the form of academic responding matched to progress during the proficiency stage.

Once again, it is expected that normalized rate may be a superior predictor of the RR-R intervention effectiveness than raw rate data because it takes into account measurement factors that influence reading rate, but may not indicate a student's learning process.

Furthermore, it is hypothesized that students whose accuracy is above or near 90% will be more likely to respond to RR-R than M-EC based on Haring and Eaton's original description of the LH/IH (1978). The accuracy cut score may also be closer to

93% needed to achieve the instructional level in reading (Gickling & Thompson, 1985) as noted by Parker & Burns (in press). It is unclear whether accurate rate will add unique information to the prediction of which type of intervention students will respond to best, but it may, based on its link to understanding proficiency stage development (Daly et al., 1996) .

Chapter III: Methods

Participants

The participants in this study were 23 second and 32 third grade students (32 male and 23 female) who were nominated by their teachers as struggling readers. The students' reading skills were confirmed by assessing oral reading performance on the first five passages used in the study. Students who had a median baseline rate above the 50th percentile for their grade level and time of year (based on the expectations reported by Hasbrouck & Tindal, 2006) were eliminated from the study. Further, students who read below 11 WRC/min (10th percentile, Fall second grade; Hasbrouck & Tindal, 2006) on three of the first five grade level passages were also eliminated from the study because they were likely developing a more foundational skill (phonemic awareness, phonics, etc.). Additional qualifications for participation were that the students were native English speakers and were not receiving special education services for moderate-severe developmental cognitive delay, autism, or a visual disability due to the possible impact of these conditions on oral reading fluency. Four second grade students (two male and two female) and two third grade students (both male) were removed from the participant pool because their baseline reading scores were above the 50th percentile and no students were eliminated due to scores below the 10th percentile. In total 49 students participated in the full study and only their data were included in analyses.

The participants were enrolled in two public school districts in the urban and suburban areas of a large mid-western city. Student participants came from three elementary schools and two elementary summer school programs. It should be noted that two of the second graders participated during the summer before they started second grade (although one of these was eliminated from analyses) and three participated during the summer after their second grade year. Also, four of the 3rd graders participated in the summer after their third grade year.

Materials

Reading passages. A total of 10 similar second and 10 similar third grade reading passages were needed for the study. Grade level passages were selected because there were expected to be challenging for the students and leave room for growth in both accuracy and rate. At the initial stage of passage selection, 13 passages were randomly drawn from the each of the grade level sets of AIMSweb (EdFormation, 2005) oral reading fluency monitoring probes. To reduce the variability in reading results due passage difficulty, the 10 passages with the most similar readability scores were kept. Specifically, the lexile scores (Stenner, Burdick, Sanford, & Burdick, 2007) for each passage in the grade level sets were converted into normalized z-scores and those three passages with the most discrepant scores (i.e., the three passages with the highest absolute z-scores) were eliminated. Therefore, 10 second and 10 third grade passages with similar readability levels of were used in this study.

Measures

Two key variables measured in this study were *rate*, the number of words read correctly per minute during oral reading fluency assessments, and *accuracy*, the percentage of words read correctly of those attempted during oral reading. Rate and accuracy data were collected during the baseline phase of this study and students' median rate and accuracy on the 10 baseline passages were considered the baseline levels of these variables used in the group analyses examining the research questions. Baseline median scores rather than means were used because median values are traditionally used to evaluate CBM oral reading fluency data given that they limit the impact of extreme values on the average (Shinn, 1989).

Rate and accuracy were recorded as students read passages aloud. The examiner followed along on a separate copy of the passage, using a stopwatch to track the duration of reading, and marked down errors and the last word read at the end of one minute. Following Shinn’s (1989) recommendations for curriculum-based measurement (CBM), a word was considered correct if it was pronounced accurately within 3 seconds. Mispronunciations, omissions, and substitutions were considered errors. Rate was calculated by determining the total number of words read correctly (WRC) in one minute, while accuracy was calculated by dividing the number of correct words by the total number of words attempted in one minute. While interrater reliability was not collected in the current study, previous reviews of this method indicated high agreement (above .80) between different scorers of rate and accuracy during oral reading (Burns, Tucker, Frame, Foley, & Hauser, 2000; Goffreda & DiPerna, 2010).

Baseline data were collected from second and third grade students at different times of the school year. Due to the typical growth in accurate reading rate over these years, the expectations for students’ rate performance varied depending on the grade level and time of year the data were collected. For example, while 51 WRC is at the 50th percentile for fall of second grade, is it just above the 10th percentile in spring of third grade

(Hasbrouck and Tindal, 2006).

In order to account for the rate differences over time, group analyses were also completed using normalized rate data, where students' baseline median rates were transformed to z-scores describing their discrepancy from grade level and time of year expectations. The median and standard deviation scores presented by Hasbrouck and Tindal (2006) were used to evaluate students' relative standing. Specifically, the median words read correct for the matched grade and time of year was subtracted from the student's rate score and the result was divided by the corresponding standard deviation. Z-score computations are typically computed using mean values, but in the present analysis large sample medians (i.e., 50th percentile) were used. Evaluation of the reported rates at different percentile values indicated that the samples presented by Hasbrouck and Tindal (2006) were close to normally distributed and thus the means would be similar to the median values presented.

It should be noted that average median and standard deviation scores were used to calculate z-scores for students tested between times of year provided (e.g., a winter-spring mean was used for students tested in March). Also, students whose data were collected in the summer after they completed the grade were compared to spring 2nd or 3rd grade norms since there is typically little growth over summer vacations (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996). Finally, the one student tested in the summer before she began 2nd grade was compared to fall 2nd grade norms. Because baseline accuracy was a percentage of words read correctly, it was not impacted by time of year in the same way and thus was not normalized for analyses.

General Procedures

Participants were solicited by asking teachers to send home informational letters and consent forms (Appendix A) with students that met the qualifications. Those students whose parents provided consent were considered for the study, but they were required to read and sign assent forms (Appendix B) before they were enrolled in the study.

The study's design was composed of two parts; first, the single-subject evaluation of intervention package effectiveness, followed by an analysis of differences between students who responded best to one or the other intervention package. The single-subject evaluation was further broken down into the baseline data collection phase, followed by the brief experimental analysis (BEA) phase. In other words, participants initially read all 10 grade level passages to establish baseline reading rate and accuracy. Then a BEA was completed with each student to determine to which intervention they responded best and thus into which group they should be categorized. Finally, group analyses of the students' baseline data were completed.

Students met with an experimenter five times over a 2 to 3 week span. The first two sessions (baseline phase) typically occurred on different days during one week and the last three sessions (BEA phase) occurred on different days of the following week. Occasionally, due to availability restrictions of the students or the experimenter, some sessions occurred on the same day (with a break in between) or there was a more extended time between sessions. However, the more intensive BEA intervention sessions never occurred on the same day and there were never fewer than 2 days between the baseline and BEA phases.

Data were collected by the primary investigator and school psychology graduate students conducting a reading intervention program at one of the elementary schools. All interactions between the experimenters and participating students occurred at a table in a quite area within the school, typically an empty room, the school's library, or in a quiet hallway. Sessions lasted between 5 and 35 minutes and were scheduled at a time that reduced disruption to classroom instruction. The first two baseline sessions typically lasted about 10 minutes each, while the last three BEA sessions usually lasted between 5 and 30 minutes, depending on whether it was an intervention session or a shorter reversal session.

During the first two baseline sessions of the study, students were administered the passages from their respective grade level, which had been randomly ordered and divided into two sets of five. The student was asked to read each passage aloud for 1 minute while the experimenter followed along on a separate copy of the passage, marking down errors and the last word read after 1 minute in order to calculate rate and accuracy. Because learning was not the goal during this phase, the experimenter did not correct errors or omissions, but told the student to "go on" when he hesitated for 3 seconds. More detailed instructions for baseline sessions can be found in Appendix C.

The eight passages on which the student's reading rates were the most similar were used in the BEA phase of the study. Similarity was evaluated by transforming each individual student's baseline rate data into standardized z-scores and eliminating the two passages with the largest absolute z-scores. Further, the eight selected passages were rank ordered by rate and assigned in an alternating fashion to the two instructional conditions (three passages each) and the reversal condition (two passages). Rank ordering the passages helped control for and minimized the effect of initial passage difficulty for an individual student on outcomes during the BEA phase of the study. In addition, because all passages used during the BEA phase of the study had been read during the baseline phase of the study, the repetition was believed to have a similar impact on reading rate and accuracy across passages.

Intervention Packages

The two intervention packages implemented during the BEA phase of this study were created using the Learning Hierarchy/Instructional Hierarchy (LH/IH; Haring & Eaton, 1978) to identify instructional components matched to the acquisition and proficiency stages of learning reading fluency. Differences in responsiveness to the two packages were assumed to indicate at which stage a student was functioning.

Apart from the different intervention components, the packages were designed to be as similar as possible to eliminate the effect of other factors on the outcome. Specifically, in each condition, the student was exposed to the passage three times (either modeled by the examiner or read by the student), with rate data measured during the third reading. Additionally, the first two exposures were longer (i.e., the entire passage read by the examiner or the student allowed to read for 90 seconds) and included more words that typically read during the first minute of the final reading in order to present the student with words she might encounter if she read the passage faster on the final reading. This limited the impact of unknown or novel words on the final outcome rate and accuracy data.

Modeling-error correction intervention package. The modeling-error correction (M-EC) intervention package was composed of listening passage preview (Daly & Martens, 1994) and the phrase drill error correction procedure (O'Shea, Munson, & O'Shea, 1984). The modeling provided during the passage preview and the immediate feedback included in the error correction procedure aimed to improve the student's reading accuracy, which is the focus of the acquisition stage of the LH.

The M-EC intervention procedure began with the experimenter modeling fluent reading of the entire passage by reading it aloud while the student followed along on another copy. The student was then asked to read the passage aloud and when she incorrectly read or omitted a word or hesitated for 3 seconds, she was corrected. The examiner corrected the student by immediately telling her the correct pronunciation of the word, having her repeat it once and then continue reading. When the student had read the passage for 90 seconds she was asked to stop. The experimenter then pointed out the error words and the three to five word phrases in which they were embedded and had the student repeat each once and three times, respectively.

The student was then asked to read the passage a final time, while the examiner recorded rate. Errors were not corrected during the post-intervention final reading of the passage and instead the student was told to "go on." More detailed instructions for this intervention package can be found in Appendix D.

Repeated reading-reward intervention package. The second intervention package was composed of repeated reading (Rashotte & Torgensen, 1985) and contingent reward for rapid reading (Lovitt, Eaton, Kirkwood, & Pelander, 1971). These interventions included the instructional components of practice (through multiple readings) and reinforcement that aim to increase reading rate (without explicitly addressing accuracy), which is important during the proficiency stage of learning.

The repeated reading-reward (RR-R) intervention package began by having the student read the passage aloud twice for 90 seconds. Once the time criterion was met the first time, the examiner stopped the student and instructed him to start at the beginning again. Contrary to what is often the case in the repeated reading procedure, no feedback on accuracy was given, so if the student misread a word she was not corrected and the examiner responded to hesitations of 3 seconds by telling the student to “go on.” Error correction was not in order to focus the intervention on increasing rate, the goal of the proficiency stage, and not building accuracy.

After the second 90 second reading, the examiner showed the student three tangible items (i.e., a pencil, a sticker, and a packet of fruit snacks) and asked him to pick which one he liked best and would like to try to earn. The examiner then told the student that if he read the passage faster than the first time he had read it that day (i.e., met the examiner’s goal of reading 20% more words correctly in one minute than read in the first minute of the first 90 second reading), he would earn the reward. The examiner recorded the student’s rate in the first minute of the student’s third reading of the passage. If the student successfully met the 20% goal, then he was given the selected tangible reward. More detailed instructions for this intervention package can be found in Appendix E.

Brief Experimental Analysis Phase

A brief experimental analysis (BEA) that tested the two fluency intervention packages was conducted with each student over three sessions. During the first two sessions, each intervention was tested on three different reading passages. The interventions were administered in an alternating fashion (i.e., session 1: ABA; session 2: BAB) and the intervention that was attempted first was randomly determined across participants. During each intervention trial, the student's rate was recorded for the first minute of the student's last reading of the passage (i.e., after the intervention had been implemented).

After these two sessions, the effectiveness of the intervention packages was evaluated based on the average gain in rate over the median baseline level. The intervention that produced the greatest median rate gain was tested again during the subsequent reversal session to confirm the effectiveness of this intervention. During the third BEA session, a baseline, no-intervention passage condition was given where the student was asked to read a passage a single time for 1 minute with no error correction and correct rate was recorded. This was followed by an administration of the selected, most effective intervention package to the final passage.

Data Analyses

Group classification. Students were classified into one of three intervention effectiveness groups (i.e., those for whom the M-EC intervention package had the greatest impact on rate, those for whom the RR-R intervention package had the greatest impact on rate, and those for whom neither was differentially effective) based on the results of the BEA. The effectiveness of the intervention packages was evaluated using the single-subject analysis technique called percentage of points exceeding the median

(PEM; Ma, 2006). Similar to percentage of nonoverlapping data (PND; Scruggs, Mastropieri, & Casto, 1987), PEM indicated the percentage of rate data points obtained using the “better” intervention package (i.e., that with the higher median rate during the BEA) that were greater than the median rate obtained with the other package. This method is preferable to PND because the impact of outlying data points was moderated by using the median as an indicator of the less effective intervention’s effectiveness instead of its highest outcome data point.

If the interventions were equally effective, then the odds that a data point would be above the median would be at a chance level; thus, we would expect a PEM of 50% if there was no difference in effectiveness between the two intervention packages. If more than half of the “better” intervention package’s data points fell above the median of the other intervention, then that intervention was determined to be the most effective and the student was classified into the corresponding effectiveness group. This intervention was considered the one that best met the student’s oral reading fluency needs and suggested at which phase of the learning hierarchy he was functioning. However, if 50% or fewer data points were above the median, indicating a high amount of overlap, then the student was classified into the third, Neither Intervention Effectiveness group (see Figures 1, 2, and 3 in the Results chapter).

Group comparisons. The main focus of the data analyses was to examine differences in the baseline median accuracy and rate measures, as well as the normalized rate measures for students categorized into the three intervention effectiveness groups. To address the first research question considering these differences, three one-way analyses of variance (ANOVAs) were conducted to evaluate the mean differences in baseline rate, accuracy, and normalized rate for students in the three intervention effectiveness groups. Due to the multiple comparisons, a Bonferroni corrected alpha ($\alpha = .05/3 = .017$) was used to evaluate the significance of the results of the three ANOVAs.

Additionally, the effect size of the baseline measures on predicting intervention effectiveness group membership was calculated using eta-squared (η^2). This simple measure indicates the percentage of variance in the outcome data (i.e., group membership) accounted for by the predictor variable (i.e., baseline measures) by dividing the treatment sum of squares by the total sum of squares. It should be noted that eta-squared is considered a biased estimator, tending to inflate the effect size of a sample (Howell, 2002). However, it may be used to consider the relative effect of different predictors on the same set of data. Finally, Bonferroni post hoc tests also using a .017 alpha level were used to evaluate the three pair-wise comparisons of means within each significant ANOVA.

The second research question regarding the strength of the relationship between baseline rate, accuracy, and normalized rate and the effectiveness status of each intervention group was addressed by finding point bi-serial correlations. Specifically, each intervention effectiveness group's data were converted into a dichotomous variables where 1 indicated belonging to the group (i.e., the intervention was effective or neither intervention was effective for the third group) and 0 indicated that the individual did not belong to the group. The dichotomous data for each of the three groups were correlated with the baselines measures yielding a total of 9 correlations (i.e., three intervention effectiveness groups by three baseline measures).

The third and fourth research questions regarding how well intervention effectiveness can be predicted by baseline accuracy and rate were addressed by fitting logistical regression models to the data. Logistical regression models were selected instead of linear models since the resulting effectiveness data were dichotomous. In considering the third question, accuracy and rate independently and then together were used to predict the success or failure of the M-EC intervention for all students. In addition, normalized rate data alone and in combination with accuracy were used to predict the effectiveness of the M-EC intervention. Accuracy, rate, and normalized rate were also used to predict the success or failure of the RR-R intervention to address the fourth research question.

In linear regression models, the R^2 statistic can be used to indicate the percentage of variability in an outcome measure accounted for by the independent variable(s) in a model. Logistic regression does not yield an equivalent statistic for evaluating variability accounted for by the model, but pseudo R^2 statistics have been developed by Cox and Snell (1989) and Nagelkerke (1991). Hosmer and Lemeshow (2000) stated that such pseudo R^2 values are difficult to interpret within a logistic regression model, but can be used to compare predictive improvement in models that include the same data. Similarly, an omnibus chi-square (χ^2) test of the model indicates improvement in its predictive value over a null, constant-only model.

The model's goodness of fit to the data can best be evaluated by comparing observed outcomes to predicted outcomes. As a basic indicator, the overall percentage of the cases correctly classified by the model when a .5 probability cut point is used was calculated. A .5 probability cut score was selected because this is the point at which the intervention is more likely to be effective than ineffective, based on the model.

In addition, Hosmer and Lemeshow (1980) proposed a test that compares the observed results to those predicted by the model when the data are divided into deciles. In this test the null hypothesis is that the model matches the data, so $p > .05$ indicates that the model fits well. Therefore, overall percentage of cases correctly classified by the models and results of the Hosmer and Lemeshow test were evaluated, in addition to the pseudo R^2 values and chi-square statistics, in order to compare the logistical regression models and determine how to best predict intervention effectiveness.

For each variable included within a logistical regression model, an odds ratio derived from the model coefficient was computed in order to understand how changes in the predictor variable impacted the odds that an individual responds best to the given intervention. Odds ratio values greater than 1 indicated that the chances than an intervention was effective increased with gains in the predictor value. Odds ratios less than 1 indicated that the odds declined with increases in the predictor variable.

The single variable logistic regression models were also considered in more detail in order to address the exploratory question about the identifying baseline data cut scores for predicting intervention effectiveness. A cut score is useful in practice for predicting which intervention a student is more likely to respond to without conducting a more intensive evaluation, such as a BEA. Specifically, the rate, accuracy, or normalized rate cut scores were derived by finding the point on the model at which the chances that an intervention was effective was greater than the chances that is it ineffective (i.e., the point at which the probability is .5).

Receiver operating characteristic (ROC) curves were also used to evaluate the quality of prediction obtained by the single-variable models. ROC curves graph the relationship between sensitivity (rate of true positive) and specificity (rate of true negative) at different points in the model. Sensitivity is the proportion of students who responded to the intervention that were correctly identified by the model as likely to respond. Conversely, specificity is the proportion of students who did not respond to the intervention who were correctly identified as not likely to respond.

The area under each of the curves is also an indicator of overall accuracy of the model and the further it is from .5 (chance level) and the closer it is to 1, the better it differentiates students for whom the intervention is effective from those for whom it is not. It has been suggested that ROC curves provide more meaningful results when calculated using at least 100 observations, with 50 or more in each of the effective and non-effective groups (Metz, 1978). There were only 49 data points in the current model, so the results presented are tentative and should be considered cautiously.

Because ROC curves graph a model's sensitivity and specificity when predicting intervention effectiveness at different baseline values, it is possible to find the point in the model that optimizes these characteristic. Thus, an alternative cut score can that maximizes sensitivity and specificity in determining intervention effectiveness was also established.

Chapter IV: Results

Research Questions

The research questions that guided the current study were as follows:

1. What are the differences in baseline accuracy, rate and normalized rate among students for whom only the modeling-error correction (M-EC) intervention was effective, those for whom only the repeated reading-reward (RR-R) intervention was effective and those for whom neither or both were effective?
2. What is the relationship between intervention effectiveness status and baseline accuracy, rate and normalized rate?
3. How well does students' baseline accuracy and rate or normalized rate predict M-EC effectiveness status?
4. How well does students' baseline accuracy and rate or normalized rate predict RR-R effectiveness status?

In addition, the following exploratory questions will be asked:

What cut score or range of scores for accuracy, rate, or a combination of these measures provides sufficient specificity and sensitivity to predict whether a student will benefit from the M-EC intervention or the RR-R intervention?

Preliminary Analyses

The median baseline measures of rate (words read correctly per minute) and accuracy (percentage of words read correctly) across the 10 baseline passages for each of the participants were used as the key predictor variables in the data analyses. The median rate for the 49 participants across the grade levels ranged from 9.5 to 99 words read correctly per minute (WRC), with a mean of 45.3 words (SD = 29.6). Median accuracy scores ranged from 50% to 99%, with a mean of 79.9% (SD = 14.8%). The normalized rate z-scores ranged from -1.6 to -0.13, with a mean of -0.8 (SD = 0.41).

Statistical analyses generally assume that the data samples are normally distributed, which can be assessed by examining the skewness and kurtosis statistics. A data set is considered normally distributed if the skewness and kurtosis have an absolute value less than twice the standard error for the statistic (Howell, 2002). The standard error for skewness in these baseline samples was .34 and the skewness for median rate (0.54), median accuracy (-0.409), and normalized rate (-0.20) fell within the normal range. The standard error for kurtosis was .67 and kurtosis scores for all three baseline data sets fell within two standard errors (-1.13, -1.11, and -.89, respectively). These results indicate that the three baseline data sets, rate, accuracy, and normalized rate can be considered normally distributed.

The correlation of accuracy with rate and normalized rate across the 49 participants was significant, $r = .86$ and $r = .75$ ($p \leq .05$), respectively. These high correlations were not surprising because rate was the numerator in calculating the accuracy percentage. It also indicates that there is a high degree of redundancy in these variables, which impacted their ability to independently predict variability in intervention effectiveness as noted in the following analyses.

Frequency. The points exceeding the median (PEM) technique indicated that the modeling-error correction intervention (M-EC) was most effective for 21 students, the repeated reading-reward intervention (RR-R) was most effective for 24 students, and

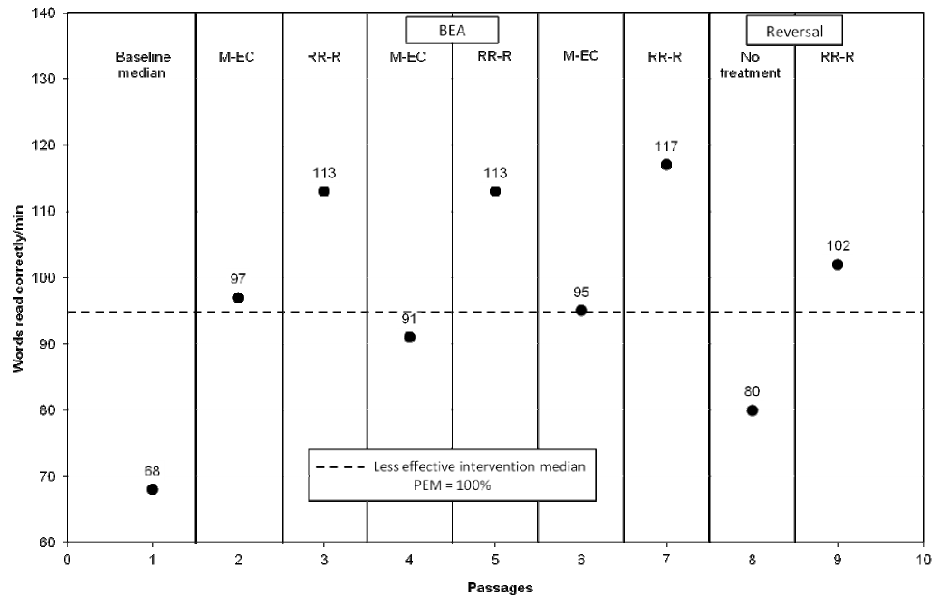

Figure 1. Points exceeding the median (PEM) graph of brief experimental analysis (BEA) data for participant 1. RR-R = repeated reading-reward, M-EC = modeling-error correction.

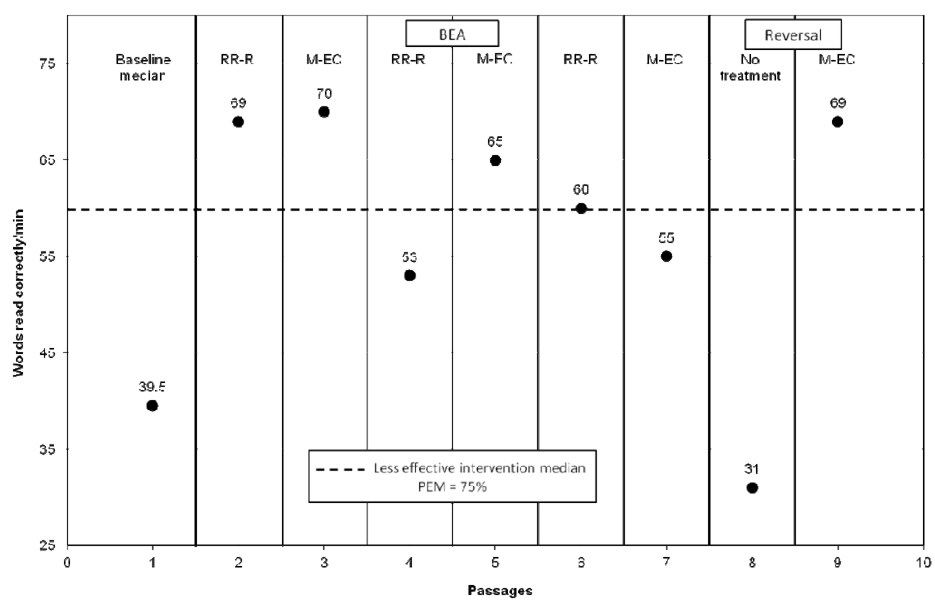

Figure 2. Points exceeding the median (PEM) graph of brief experimental analysis (BEA) data for participant 59. RR-R = repeated reading-reward, M-EC = modeling-error correction.

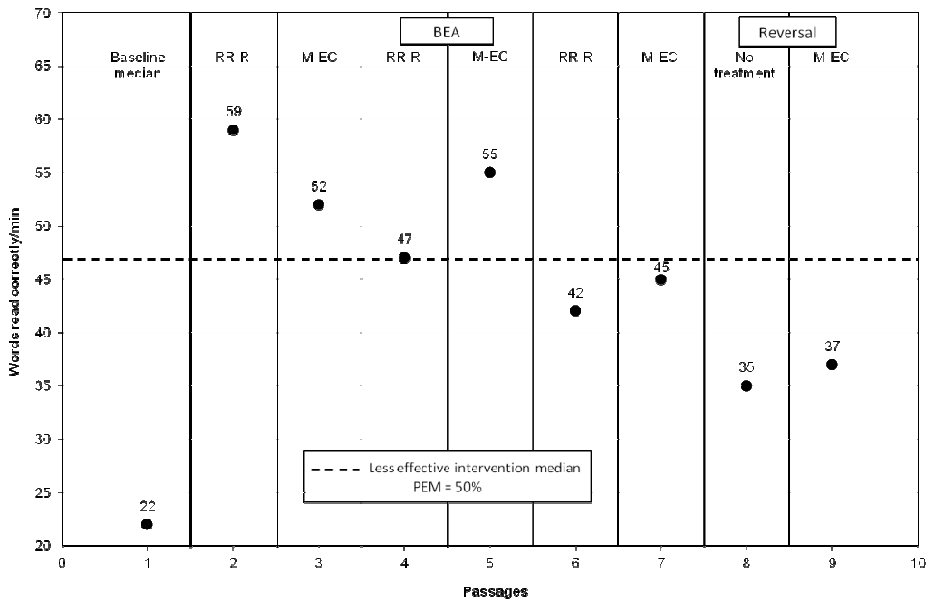

Figure 3. Points exceeding the median (PEM) graph of brief experimental analysis (BEA) data for participant 16. RR-R = repeated reading-reward, M-EC = modeling-error correction.

neither intervention was differentially effective for four students. Figure 1 presents a graph for a participant for whom 100% of the effective intervention's (RR-R) data points fell above the above the median of the less effective intervention, while Figure 2 graphs a participant for whom 75% of the effective intervention's (M-EC) data points fell above the other's median. Finally, Figure 3 presents a graph for a participant who was assigned to the Neither Intervention effectiveness group because only 50% of the more effective intervention's data points fell above the other's median. The characteristics of the three effectiveness groups including their average levels of baseline median rate and accuracy and normalized rate are presented in Table 2.

Table 2

Effectiveness Group Characteristics

Group	*n*	Gender (% female)	Grade (% second)	Median rate *M* (*SD*)[a]	Median accuracy *M* (*SD*)[b]	Normalized rate *M* (*SD*)[c]
M-EC	21	52.4	57.1	23.4 (13.8)	69 (11.1)	-1.04 (0.32)
RR-R	24	33.3	16.7	63.3 (27.1)	89.1 (11.7)	-0.61 (0.38)
Neither	4	50	50	53.1 (33.6)	81.6 (11.3)	-0.57 (0.45)
Total	49	42.9	38.8	45.3 (29.6)	79.9 (14.8)	-0.79 (0.41)

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward

[a] Words read correctly per minute. [b] % words read correctly. [c] Z-score.

Neither Intervention effectiveness group. Four students demonstrated similar levels of rate improvement in response to both intervention packages and based on the PEM analyses were categorized in the Neither Intervention effectiveness group. It was noted that all students in this group demonstrated improvement in their reading rates over baseline in response to the interventions, but neither intervention was found to be differentially effective.

While the average rate and accuracy measures for the Neither Intervention effectiveness group fell between those of the M-EC and RR-R effectiveness groups, an examination of the individual student data indicated diversity within this group. Two of the students had median baseline rates of 22 and 27 words read correctly per minute

(WRC) while the other two had rates of 75 and 88.5 WRC. Additionally, the students with the lower rates had accuracy scores of 70% and 79.4%, while the two higher readers had accuracies of 96.5% and 97.2%. Looking more closely at the BEA results, the M-EC intervention was selected as the better intervention implemented during the reversal for three of the students, but for all four students the reversal did not confirm the effectiveness of the selected intervention. Finally, the students were from both grade levels (three in second and one in third grade), indicating that grade level did not predict membership in this group.

Research Question 1: Group Differences

Differences in baseline rate, accuracy, and normalized rate between students who fell into the three intervention effectiveness groups were evaluated using one-way analyses of variance (ANOVAs) and Bonferroni post-hoc tests of all pair-wise comparisons. One-way ANOVAs found significant differences across the effectiveness groups for baseline rate ($F[2, 46] = 17.3$, $p \leq .017$), accuracy ($F[2, 46] = 17.47$, $p \leq .017$), and normalized rate ($F[2, 46] = 8.9$, $p \leq .017$). The effect sizes for the rate ($\eta^2 = 0.429$) and accuracy ($\eta^2 = 0.432$) measures fell in the moderate range, suggesting that about 43% of the variance in effectiveness group membership was accounted for by variance in these baseline measures. In contrast, there was a smaller relative effect ($\eta^2 = 0.279$) for the variance in normalized rate predicting group membership. It should be reiterated that this measure of effect size should be considered cautiously as it tends to be biased, but does allow comparisons of relative effects across the same data set.

Given the significant results, Bonferroni post-hoc tests were used to evaluate the

Table 3

Mean Difference Comparisons using Bonferroni Post-Hoc Tests and Associated p-values

	Baseline rate[a]	Baseline accuracy[b]	Normalized rate[c]
RR-R – M-EC	39.9 ($p \leq .001$)*	20.1 ($p \leq .001$)*	0.43 ($p \leq .001$)*
RR-R – Neither	10.1 (1.00)	7.5 (.692)	-0.45 (1.00)
M-EC – Neither	-29.8 (.064)	-12.6 (.144)	-0.48 (.057)

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward

[a] Words read correctly per minute. [b] % words read correctly. [c] Z-score.

* $p < .017$ (Bonferroni corrected alpha)

three pair-wise comparisons of means for each ANOVA. For all three sets of baseline data, there were significant differences ($\alpha = .017$) between the means of the M-EC and RR-R effectiveness groups. Specifically, rate, accuracy and normalized rate were significantly higher for students in the RR-R group than the M-EC group. However, the mean differences between the Neither Intervention effectiveness group and both the M-EC and RR-R intervention groups were not significant (see Table 3). The large *p*-values noted in some of the Neither Intervention effectiveness comparisons are due to the small number of participants ($n = 4$) in this group and the relatively large spread of the data within this group.

Research Question 2: Relationship Between Baseline Measures and Intervention Effectiveness

Point bi-serial correlations between baseline measures and the dichotomous

Table 4

Point Bi-serial Correlations Between Intervention Effectiveness Status[a] *and Median Baseline Measures Rate (N = 49)*

Effectiveness group	Baseline rate	Baseline accuracy	Normalized rate
M-EC	-.65 **	-.64 **	-.53 **
RR-R	.6 **	.6 **	.43*
Neither	.08	.04	.16

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward

[a] Where each participant was classified with a 1 if they were in the particular effectiveness group and a 0 if they were not in the effectiveness group.

* $p < .05$, ** $p \leq .001$

effectiveness data were calculated to consider the relationship between these variables. The results presented in Table 4 indicate that there were significant relationships between baseline rate, accuracy, and normalized rate and effectiveness of the two intervention packages. Specifically, as raw rate, accuracy and normalized rate increased, the chances that the RR-R intervention was effective increased, while the chances that the M-EC intervention was effective decreased. There was not a significant correlation between baseline measures and membership in the Neither Intervention Effectiveness group.

Research Question 3 and 4: Predicting Intervention Effectiveness from Baseline Measures

Single Variable Models. Logistic regression was used to consider how well raw baseline rate and normalized rate data and accuracy predicted the effectiveness of each of the intervention packages. As a first step, single variable models were tested that included rate, accuracy and normalized rate independently to predict whether or not each of the intervention packages was effective. Table 5 presents the pseudo R^2 values and omnibus chi-square test values for these models predicting M-EC and RR-R effectiveness.

Table 5

Pseudo R^2 and Chi Square (χ^2) Values for the Single Variable Logistic Regression Models Predicting Intervention Effectiveness (N = 49)

	Cox and Snell pseudo R^2	Nagelkerke pseudo R^2	Model χ^2
Models			
M-EC			
Rate only	.42	.56	26.57 **
Accuracy only	.38	.51	23.20 **
Normalized rate only	.27	.36	15.19 **
RR-R			
Rate only	.34	.45	20.12 **
Accuracy only	.36	.48	21.75 **
Normalized rate only	.18	.24	9.85 *

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward

* $p < .05$, ** $p \leq .001$

Table 6

Percentage of Cases Correctly Classified and Hosmer and Lemeshow Test Chi Square Results with p-value for the Single Variable Logistic Regression Models Predicting Intervention Effectiveness (N = 49)

	Percentage of cases correctly classified	Hosmer and Lemeshow test $\chi^2(p)^a$
Models		
M-EC		
Rate only	83.7	3.91 (.79)
Accuracy only	81.6	6.89 (.55)
Normalized rate only	77.6	15.1 (.06)
RR-R		
Rate only	75.5	7.74 (.46)
Accuracy only	83.7	12.36 (.14)
Normalized rate only	73.5	11.2 (.13)

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward

[a] Non-significant *p*-values indicate the model fits the data well.

* $p < .05$, ** $p \leq .001$

As noted in the Methods chapter, the pseudo R^2 values cannot be interpreted in exactly the same way as the R^2 statistic in a linear regression, but they do provide a general comparative indicator of model fit. These values also cannot be submitted to a hypothesis test to determine the likelihood of these results. Overall, the pseudo R^2 statistics and chi-square values for all six logistical regression models indicate that baseline measures increased the ability to predict whether the M-EC and RR-R intervention packages would be successful over constant-only, null models. The raw rate and accuracy models for both interventions demonstrated stronger relative predictability than the models based on normalized rate.

The percentage of cases correctly predicted by each single variable model and the results of the Hosmer and Lemeshow tests are presented in Table 6. All models correctly classified the large majority of observed cases indicating that baseline data provided useful information for predicting both M-EC and RR-R intervention effectiveness. Once again, in both cases, the rate and accuracy models were somewhat better than the normalized rate models because they predicted a greater number of cases. In addition, it was noted that the rate model was slightly better at predicting M-EC effectiveness, while the accuracy model predicted a great number of RR-R effectiveness cases.

The results of the Hosmer and Lemeshow tests also indicated that the predicted results of all six models generally matched the observed results. For both interventions, the rate and accuracy models provided a better fit to the data than the normalized rate models. In particular, when predicting M-EC intervention effectiveness with the normalized rate model, the Hosmer and Lemeshow test neared significance ($p = .06$) indicating that this model was a less good fit to the data than the other models.

Table 7 presents the odds ratios and 95% confidence intervals for the predictor variables within each of the single variable models. In all cases, the odds ratios were significant, although the *p*-value for the normalized rate model predicting RR-R effectiveness was larger than for the other models. Given that a majority of normalized data fall within six standard deviation units and that the data used in this analysis fell within two standard deviations, the odds ratio may not provide a good description of how changes in normalized rate impact the chances that either intervention is effective. This also accounts for the relatively large confidence intervals for these models' odds ratios.

Table 7

Predictor Variable Odds Ratios and 95% Confidence Intervals from the Single Variable Logistic Regression Models Predicting Intervention Effectiveness (N = 49)

	Odds ratios	95% CI
Models		
M-EC		
Rate only	0.92 **	0.88 – 0.97
Accuracy only	0.89 **	0.83 – 0.94
Normalized rate only	0.04 **	0.005 – 0.28
RR-R		
Rate only	1.06 **	1.03 – 1.09
Accuracy only	1.12 **	1.06 – 1.2
Normalized rate only	11.63 *	2.09 – 64.79

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward

* $p < .05$, ** $p \leq .001$

The odds ratios for each of the models predicting M-EC effectiveness was less than one, which indicated that as rate, accuracy, or normalized rate increased, the chances that the M-EC intervention was effective decreased. Specifically, for each one word increase in words read correctly (rate), the odds that the M-EC intervention was effective were multiplied by 0.92. For the accuracy model, the odds that M-EC was effective were multiplied by 0.89 for each 1% increase in accuracy. Finally, with each one standard deviation unit increase in normalized rate, the odds that the M-EC intervention was effective were multiplied by 0.04.

The odds ratios for all of the RR-R models were greater than one, indicating that as baseline data values increased, the chances that this intervention was effective also increased. For each 1 words increase in rate, the odds that the RR-R intervention was effective were multiplied by 1.06. In the accuracy model, a 1% increase in accuracy led to a 1.12 times increase in the odds. Finally, for each one standard deviation unit increase in normalized rate, the odds that the RR-R intervention was effective were multiplied by 11.63.

Multivariable Models. Given that all of the single variable models yielded significant results for the intervention groups, combined models were tested to determine whether the prediction of intervention effectiveness could be improved by including both rate and accuracy data. One model for each intervention effectiveness group included baseline rate and accuracy as predictor variables, while the other included normalized rate and baseline accuracy scores. The pseudo R^2 values and omnibus chi-square test values for these multivariable models predicting M-EC and RR-R effectiveness are presented in Table 8. The results indicate that the two variable models contributed significant predictive power over a null, constant-only model.

Table 8

Pseudo R^2 and Chi Square (χ^2) Values from the Multivariable Logistic Regression Models Predicting Intervention Effectiveness (N = 49)

	Cox and Snell pseudo R^2	Nagelkerke pseudo R^2	Model χ^2
Models			
M-EC			
Rate and Accuracy	.43	.57	27.07 **
Normalized Rate and Accuracy	.38	.52	23.73 **
RR-R			
Rate and Accuracy	.37	.49	22. 63**
Normalized Rate and Accuracy	.36	.48	21.83 **

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward

* $p < .05$, ** $p \leq .001$

It was noted that many of the pseudo R^2 and omnibus chi-square test values for the multivariable models were similar to those from the single variable models, indicating that they perhaps did not contribute additional predictive value. Specifically, for the models predicting M-EC effectiveness, the rate and accuracy multivariable model was similar to the rate only model, while the normalized rate and accuracy model has results similar to the accuracy only model. When considering the RR-R effectiveness models, the normalized rate and accuracy model had pseudo R^2 and omnibus chi-square test values like those in accuracy only model.

Table 9

Percentage of Cases Correctly Classified and Hosmer and Lemeshow Test Chi Square Results with p-value from the Multivariable Logistic Regression Models Predicting Intervention Effectiveness (N = 49)

	Percentage of cases correctly classified	Hosmer and Lemeshow test $\chi^2(p)^a$
Models		
M-EC		
Rate and Accuracy	83.7	5.13 (.74)
Normalized Rate and Accuracy	81.6	5.6 (.69)
RR-R		
Rate and Accuracy	83.7	7.97 (.44)
Normalized Rate and Accuracy	83.7	8.59 (.38)

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward

[a] Non-significant *p*-values indicate the model fits the data well.

* $p < .05$, ** $p \leq .001$

The goodness of fit between the models and the actual data as measured by percentage of cases correctly classified and the results of the Hosmer and Lemeshow test are presented in Table 9. The non-significant results of the Hosmer and Lemeshow tests indicate that there is substantial overlap between the results obtained using the models and the actual data. In addition, a high percentage of cases were correctly classified by all the multivariable models. It was noted that the percentage correctly classified data points was the same as that of the highest single variable model; the rate only and accuracy only models, respectively, for M-EC effectiveness and the accuracy only model for the RR-R effectiveness models.

Table 10 presents the odds ratios and their 95% confidence intervals for each of the individual variables in the multivariable models. Odds ratios in these multivariable models indicate the impact of a one unit increase in the variable on the chances that an intervention is effective, all other variables being equal. While the variable trends appear similar to the single variable models (i.e., increases in the baseline measures generally decreased the odds that M-EC would be effective and increased the odds that RR-R would be effective), the odds ratios are mostly non-significant. This result was likely due to the high correlation between accuracy and the two rate variables and indicated that each variable did not contribute additional predictive value to the model in the presence of the other. Thus, the single variable models independently appear to provide a better predictor of intervention effectiveness.

Table 10

Predictor Variable Odds Ratios and 95% Confidence Intervals from the Multivariable Logistic Regression Models Predicting Intervention Effectiveness (N = 49)

Models	Variable	Odds ratios	95% CI
M-EC			
Rate and Accuracy	Rate	0.942	0.880 – 1.008
	Accuracy	0.964	0.871 – 1.067
Normalized Rate and Accuracy	Normalized Rate	0.396	0.033 – 4.743
	Accuracy	0.901 *	0.835 – 0.972
RR-R			
Rate and Accuracy	Rate	1.023	0.975 – 1.073
	Accuracy	1.081	0.977 – 1.196
Normalized Rate and Accuracy	Normalized Rate	0.696	0.060 – 8.043
	Accuracy	1.133 *	1.045 – 1.227

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward

* $p < .05$, ** $p \leq .001$

In addition, the majority of the test statistics for the single and multi-variable models yielded stronger results for the baseline rate data over the normalized rate data based indicating a greater predictive value of the raw data. This finding indicates that students' actual rate level provides a better indicator of their level of reading fluency development and thus which intervention is most likely to be successful than a relative comparison to grade-level, time of year expectations.

Exploratory Cut Score Analyses

Further analysis of the single variable models allows for the delineation of cut scores that can be used in practice to predict to which intervention a student is most likely to respond. One way to establish intervention effectiveness cut scores is to determine the point at which the likelihood that an intervention is effective is greater than the likelihood that it is ineffective; this change occurs at the point that probability of intervention effectiveness is .5. Table 11 presents the three logistic regression models for each intervention and the .5 probability cut scores for each model.

The interpretation of the cut scores is somewhat different based on the intervention models. In particular, because the models for the effectiveness of the M-EC intervention have a negative variable coefficient, this indicates that the intervention is more likely to be effective for rate, accuracy, and normalized rate values below the presented cut scores. On the other hand, the RR-R intervention is more likely to be effective for measures above the cut scores presented.

These cut scores provide an indicator of which intervention is more likely to be successful based on the student’s baseline reading data. However, for each predictor, there is a range for which neither intervention is more likely to be effective based on the single variable models. Specifically, students with baseline reading rates between 34 and 44 WRC/min (or normalized rates between .91 and .77 standard deviation units below grade-level expectations) or baseline accuracy of between 76.6% and 81.6% may respond

Table 11

Logistic Models and Cut Scores when Probability is .5 for the Single Variable Models of Intervention Effectiveness

Intervention	Variable (x)	Logistic Model	p = .5 Cut Score
M-EC	Rate	$\ln(\frac{p}{1-p})$ = -0.08x + 2.738	34.23 WRC/min
	Accuracy	$\ln(\frac{p}{1-p})$ = -0.121x + 9.271	76.62%
	Normalized Rate	$\ln(\frac{p}{1-p})$ = -3.296x – 2.994	-0.91 SD units
RR-R	Rate	$\ln(\frac{p}{1-p})$ = 0.055x – 2.43	44.18 WRC/min
	Accuracy	$\ln(\frac{p}{1-p})$ = 0.117x – 9.545	81.58%
	Normalized Rate	$\ln(\frac{p}{1-p})$ = 2.453x + 1.886	-0.77 SD units

Note. M-EC = modeling-error correction; RR-R = repeated reading-reward; WRC/min = words read correctly per minute; SD units = standard deviation units

best to either the M-EC or RR-R interventions. As a result, alternate probability cut scores determined by considering sensitivity and specificity may provide a better way of differentiating students who are likely to respond to one intervention or the other.

ROC curves. The single variable models were compared by calculating receiver operator characteristic (ROC) curves, which graph the relationship between sensitivity (true positive) and specificity (true negative) for different points in the model. The ROC curves using rate, accuracy, and normalized rate to predict intervention effectiveness are presented in Figures 4 (M-EC) and 5 (RR-R).

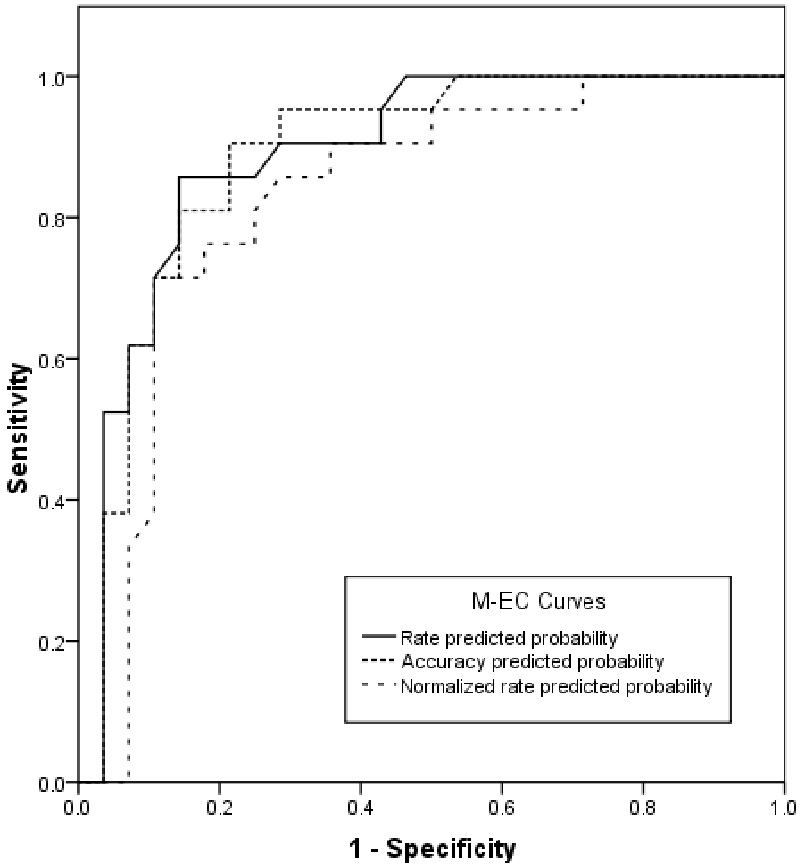

Figure 4. ROC graph of single variable models predicting M-EC intervention effectiveness. Diagonal segments are produced by ties.

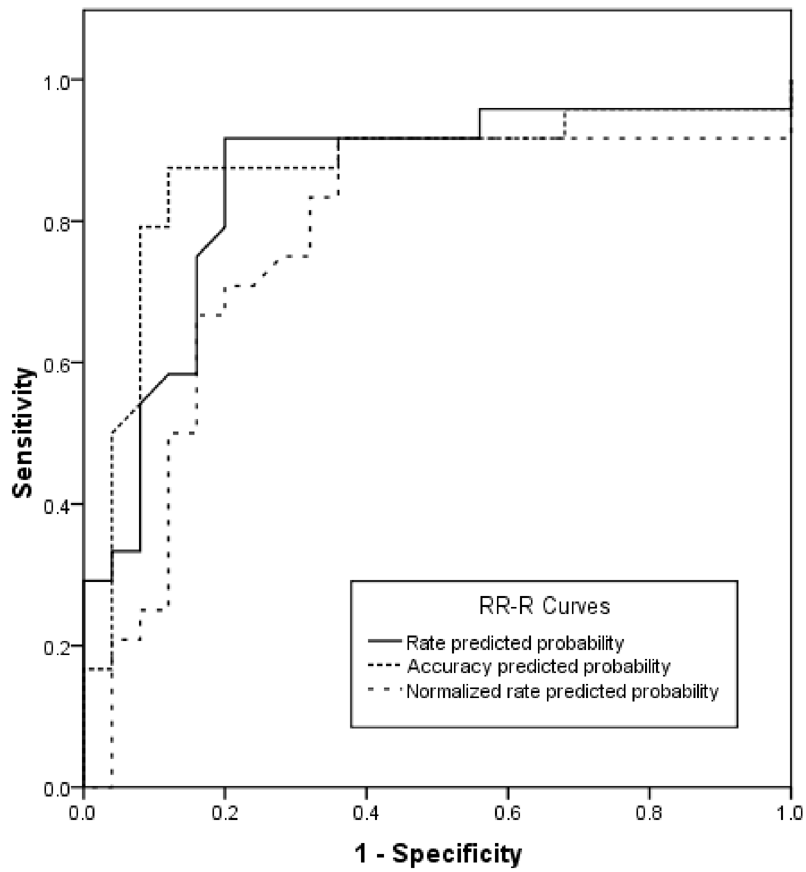

Figure 5. ROC graph of single variable models predicting RR-R intervention effectiveness. Diagonal segments are produced by ties.

The closer the area under the ROC curve is to 1, the better it differentiates students for whom the intervention is effective from those for whom it is not. When predicting M-EC effectiveness, the area under the curve was .89 for the model using baseline rate as the predictor, .89 for the model using baseline accuracy, and .83 for the model using normalized rate. Considering the RR-R effectiveness models, the area under the curve was .85 when rate was the predictor, .87 when accuracy was the predictor, and .77 when normalized rate was the predictor. In all cases, the area was significantly greater ($p \leq .001$) than the .5 chance level. Also, it was noted that for both interventions, the rate and accuracy models demonstrated a higher level of precision in predicting the effectiveness than the normalized rate model.

Thus, the results of the ROC analyses indicate that baseline rate, accuracy and normalized rate were able to effectively differentiate students who responded to the interventions from those who did not respond. For both intervention packages, the rate and accuracy data was a better predictor than the normalized rate data, again confirming the superiority of these measures.

Further cut score analyses. The closer that a point on the ROC curve falls to the top left part of the graph, the higher the sensitivity and specificity of the cut score when identifying students who are predicted to respond to an intervention. Using the ROC curves, it is possible to find the cut point in each model that optimizes sensitivity and specificity (Table 12). These cut scores do the best job of both correctly identifying students for whom the intervention is likely to be effective and those for whom the intervention is not likely to be effective. In contrast to the potential cut scores presented in Table 10, the probability of intervention effectiveness associated with this cut score may be higher or lower than .5 depending on where sensitivity and specificity was found to be optimized.

Results indicated that the optimal cut scores for the M-EC intervention were a rate of 30.6 WRC/min, an accuracy of 84.6% or a normalized rate of -0.94 SD units. Students who scored below these cut points were more likely to respond to M-EC intervention,

Table 12

Cut Scores, Associated Optimal Sensitivity, Specificity and Probability Values for the Single Variable Models of Intervention Effectiveness

Intervention	Variable (x)	Cut Score	Sensitivity	Specificity	*p*
M-EC	Rate	30.6 WRC/min	.857	.857	.572
	Accuracy	84.63%	.905	.786	.275
	Normalized Rate	-0.939 SD units	.714	.893	.525
RR-R	Rate	32.6 WRC/min.	.917	.8	.346
	Accuracy	85.01%	.875	.88	.599
	Normalized Rate	-0.939 SD units	.917	.64	.397

M-EC = modeling-error correction, RR-R = repeated reading-reward

WRC/min = words read correctly per minute, SD units = standard deviation units

while those who score higher were less likely to respond to this intervention. Similarly, the cut scores above which students were more likely to respond to the RR-R intervention were 32.6 WRC/min, 85% and -0.94 SD units. The similarity of these cut scores may indicate a natural transition point in reading development or may be a spurious result of the nearly dichotomous nature of the data.

Results Synthesis

Overall, a significant difference in baseline rates, accuracy and normalized rates were found between students who responded best to the M-EC intervention package and those who responded to the RR-R intervention package. As predicted by the LH/IH, these baseline measures were significantly lower for students who responded best to the M-EC intervention which addressed the more basic skill of reading accuracy. A significant predicted relationship between the baseline measures and intervention outcomes was found, although the single variable models and raw rate and accuracy variable were found to be stronger predictors of intervention outcomes. Further analysis of the data revealed cut scores that optimized the sensitivity and specificity of the models and offer a decision point that can be used in practice.

Chapter V: Discussion

Results of the current study indicated that, as expected, there were significant differences in participants' baseline measures of accuracy and rate based on whether they responded best to a reading fluency intervention package matched to the acquisition or the proficiency stages of the LH/IH. Moreover, the relationship between baseline measures and the outcomes was strong enough that intervention effectiveness for both the modeling-error correction (M-EC) and the repeated reading-reward (RR-R) packages could be predicted from accuracy and rate data. Further analysis of the logistical regression models yielded provisional accuracy and rate cut scores that can be used by educators to quickly make intervention decisions. A close examination of these findings in the context of the research questions that guided this study are presented below.

Research Question 1: Group Differences

Consistent with predictions in the literature review, the results of this study indicated that there were differences in the baseline accurate reading rate and reading accuracy percentages for students who responded best to either the M-EC or the RR-R intervention packages. Students who responded best to the M-EC intervention demonstrated significantly lower average rate (23 vs. 63 words read correctly) and accuracy (69% vs. 89%) and a higher percentage were in second grade than students in the RR-R effectiveness group, indicating that they were at a more basic level of development. Higher averages were found for students who responded best to the interventions recommended for the more advanced, proficiency stage of the LH. Furthermore, the average baseline accuracy for the RR-R group was 89.1%, which is close to the 90% accuracy that Haring and Eaton (1978) suggested was the point at which students transition from the acquisition stage to the proficiency stage of learning. This finding is also consistent with Parker and Burns (in press) who found that students responded best to a proficiency intervention package only once their accuracy percentage was reliably 93% or higher.

Students who responded best to each of the intervention packages demonstrated differences in baseline characteristics that corresponded to the predictions of the LH/IH (Haring & Eaton, 1978). The M-EC intervention contained instructional components (i.e., modeling and feedback) linked to the first, acquisition stage of reading fluency development, while the RR-R intervention package contained instructional components (i.e., repetition and contingent reinforcement) linked to the second, proficiency stage. Given the desire to isolate the specific instructional components relevant to each stage, intervention packages were designed to only foster the relevant skill. So, for example, the RR-R intervention packages did not involve feedback on accuracy, although actual implementation of such intervention would likely include error correction. As a result, BEA findings that one of these intervention packages worked best can be seen as a proxy for student functioning at the matched stage of the LH.

These results provide support to the theoretical underpinnings of the LH/IH framework and in particular how it predicts performance regarding oral reading accuracy and rate, two key measures of reading fluency. Specifically, the significant baseline differences indicated that students who respond best to the two interventions fell into unique groups that demonstrated different levels of competence corresponding to IH/LH stages. Furthermore, this finding confirmed the links between the intervention strategies that address relevant needs at the two basic levels of the learning hierarchy. The group differences also indicated that further analysis including consideration of the predictive relationship between baseline rate and accuracy, proposed in the additional research questions, was warranted.

Neither Intervention effectiveness group. Only four students were categorized in the Neither Intervention effectiveness group, so solid conclusions about students who did not demonstrate differential responsiveness to either intervention cannot be made. While the average baseline rate and accuracy scores for the students in this group fell between those of the two intervention groups, the data do not indicate that these students were a unique group transitioning from the acquisition to the proficiency stage. The four students demonstrated a fairly large range of baseline characteristics which were not significantly different from students who fell in the two intervention effectiveness groups.

In all cases, both interventions led to an increase in rate scores over baseline, implying that that this group was not made up of students whose reading development was focused on a more basic skill. In addition, these students likely would have demonstrated improvement in response to either intervention package. The variability in baseline data along with the small number of students in this group made it difficult to predict membership based on the current data.

Research Question 2: Relationship between Baseline Measures and Intervention Effectiveness

Point bi-serial correlations were calculated to further examine the relationships between baseline measures and intervention effectiveness. Moderate, significant coefficients were found when describing the relationship between baseline levels of rate, accuracy and normalized rate and intervention effectiveness. Specifically, as rate, accuracy and normalized rate increased, it was more likely that the RR-R intervention package was effective and less likely that the M-EC intervention package was effective. Baseline rate and accuracy measures did not demonstrate a significant correlation with membership in the Neither Intervention effectiveness group.

At a basic level, the correlation between reading accuracy and rate and the outcomes of fluency interventions is expected based on the empirical definition of reading fluency that includes these qualities (Dowhower, 1991; Schreiber, 1991; National Institute of Child Health and Human Development, 2000). In the context of the LH/IH, Haring and Eaton (1978) suggested that accuracy was the key measure that indicated a student's advancement from the acquisition to the proficiency stage. Current data suggest that reading rate can also be used to predict reading fluency development, which is consistent with the use of this outcome measure for evaluating growth within the proficiency stage (Daly et al, 1996). The directions of the correlations between M-EC and RR-R and accuracy and rate also correspond with the underlying premise of the LH/IH described by Haring and Eaton (1978) that as students gain competence with a skill their intervention needs change.

The correlations between the baseline measures and M-EC effectiveness were somewhat stronger than those for RR-R effectiveness, suggesting that baseline data provided a somewhat better indicator of effectiveness for the acquisition interventions. This conclusion was also reflected in the following logistic regression analyses.

Research Questions 3 and 4: Predicting Intervention Effectiveness from Baseline Measures

Given that baseline measures did have a significant relationship with intervention effectiveness, the next step was to consider the predictive value of the data. Single and multivariable logistical regression models using baseline accuracy, rate and normalized rate to predict the effectiveness of the M-EC and RR-R intervention packages were evaluated. Multiple statistics indicated that the regression models were strong and matched the observed data well. Overall results indicated that when a student's reading accuracy and rate was low, the M-EC intervention was more likely to be effective, whereas when baseline accuracy and rate were high, the RR-R intervention was more likely to be effective. Moreover, consistent across analyses were the findings that raw rate was a better predictor than normalized rate and that the single variable models were as good as or better at predicting intervention effectiveness than multivariable models.

Rate versus normalized rate. Analyses conducted with normalized rate data provided a way to consider whether intervention effectiveness was better described using a student's performance relative to rate expectations. The high percentage of cases correctly predicted by all of the models and the insignificant results of the Hosmer and Lemeshow tests indicated that there was a good fit between the observed data and the models. The proportion of variability in intervention effectiveness predicted by baseline data cannot be delineated in logistical regression models. However, comparisons of pseudo R^2 values indicated that the rate only and accuracy only models were better predictors of both interventions' effectiveness than the normalized rate only model. In addition, receiver operator characteristic (ROC) curve analyses also supported the conclusion that rate and accuracy only models better differentiated students than the normalized rate models. The results indicate that normalizing the data based on grade level and time of year did not lead to more accurate models for predicting intervention effectiveness.

The superiority of raw rate data over the normalized rate data supports the use of a student's accurate reading speed, without necessarily considering how discrepant he is from his peers, to select an appropriate intervention. Thus, it appears that reading rate on grade level material, regardless of how it compares to normative expectations, is an indicator of underlying reading fluency skill. This implies that a student's progress through Chall's (1996) Stage 2 of reading development, the overall efficiency of the reading system (Hudson, Pullen, Lane, & Torgensen, 2009; Kuhn & Stahl, 2003) and, as a result, type of intervention that is more likely to be effective, can be evaluated through the simple measurement of rate on grade level material for second and third grade students.

Multivariable models. When accuracy and rate or normalized rate were combined in the multivariable models, they did not add substantial predictive value above and beyond the single variable models. In addition, the odds ratios for the multivariable models were not significant indicating that when combined the variables do not inform intervention effectiveness. The high correlation between accuracy and rate measures likely accounted for the fact that the two significant predictor variables were not significant when combined. These findings along with a desire for a simpler model indicate a preference for single variable models using rate or accuracy only to predict the effectiveness of both intervention packages.

Single variable models. The odds that the M-EC intervention was effective decreased by about 8% for every one word increase in baseline reading rate and decreased by about 11% for every one percent increase in reading accuracy. In contrast, the odds of RR-R intervention effectiveness increased by 6% for every one word increase in rate and increased by about 12% for every one percent increase in accuracy. The reciprocal relationship between these predictive models discovered in the current study is not surprising since the majority of participants were found to respond best to one of the two intervention packages.

The rate only and accuracy only models for both interventions exhibited comparable percentages of cases correctly classified and pseudo R^2 values, indicating that both models similarly predicted effectiveness. The high correlation (r = .86) between these two types of baseline data likely accounts for the similarity in model statistics. However, it appears that rate was a somewhat stronger predictor of M-EC effectiveness, while accuracy was a somewhat stronger predictor of RR-R effectiveness.

This result is the opposite of what was hypothesized in the literature review because it contradicts the contention of Daly, Lentz, and Boyer (1996) that accuracy is the form of academic responding matched to measuring growth in the acquisition stage and rate is matched to proficiency stage. However, upon further consideration, this predictive relationship is not inconsistent with this model. It appears that when the matched form of academic responding is the necessary target of intervention, it is more variable, while the other baseline measure is more constant. Thus, at the acquisition stage, accuracy is the focus of improvement, but rate is a better predictor because it is more consistently low for students at this stage and higher for students who have moved on to the proficiency stage. Similarly, students at the proficiency stage demonstrate consistently high accuracy, while rate is the target of growth.

The relatively small number of students in the data set and the reciprocal relationship of the data may contribute to the finding that rate is a better predictor of M-EC effectiveness and accuracy is a better predictor of RR-R effectiveness. Further analysis with a larger data set is necessary to evaluate the accuracy of the stability of the predictive form of measurement while the matched form of academic responding grows.

Exploratory Analyses

The strength of the accuracy and rate models in predicting the effectiveness of both interventions led to further analysis of these single variable models. Specifically, two types of cut scores for making instructional decisions were investigated. First, cut scores were calculated by determining the point at which the probability that an intervention was effective was greater than the probability that it was not effective (i.e., $p = .5$). Students with reading rates of 34 or fewer words correct per minute or accuracy below 77% were more likely to respond to the M-EC intervention. The RR-R intervention was more likely to be effective with students with rates above 44 words read correctly or accuracy above 82%. However, these cut scores could be improved by finding the point at which optimized the prediction of intervention effectiveness.

Using the ROC models it was possible to find cut scores for each model that provided the best sensitivity (true positives) and specificity (true negatives) for predicting intervention effectiveness based on the current data. Based on this method, the rate cut score for the M-EC intervention was 31 words read correctly, while the accuracy cut score was 85%. Similarly, for the RR-R intervention the rate cut score was 33 words read correctly, while the accuracy cut score was also 85%. Once again, the similarity of these cut score may be due to the reciprocal nature of the data, but it may also describe a natural transition point between the acquisition and proficiency stages in the area of reading fluency.

An accuracy cut score of 85% is fairly consistent with Haring and Eaton's (1978) suggestion that it was only once a student achieved 90% accuracy that the focus of learning shifted to increasing the rate of skill performance. This cut score is also lower than the 93% accuracy that Parker and Burns (in press) found to predict greater reading growth in response to proficiency interventions. Rather than being consistent with the 93% to 97% instructional level for reading tasks (Gickling & Thompson, 1985), an 85% cut score is more consistent with the upper end of the instructional level for drill tasks, including word recognition (i.e., 70% to 85%). Perhaps a student does not move to the proficiency stage of the LH for reading fluency until he is at the independent level regarding word recognition. At this point, the focus of skill development can shift to other facets of the reading, including possibly addressing the rate component of fluency.

Based on the data collected in this study, it was possible to ascertain possible rate and accuracy cut scores that could be used to identify interventions for students developing reading fluency. However, further research is needed to confirm these cut score values, their ability to differentiate students' intervention responsiveness, and how they relate to other instructional decision-making frameworks (e.g., curriculum-based assessment).

Potential Theoretical Implications

The findings of the current study lend support to the LH/IH (Haring & Eaton, 1978) as a description of how learning progresses for the oral accuracy and rate components of reading fluency. As would be predicted by this model, students who responded best to the acquisition intervention package demonstrated significantly lower average reading rate and accuracy on grade level material. Students who responded best to the proficiency interventions demonstrated higher average baseline skills, as would be expected at this more advanced stage of development.

The results also reinforce the underlying tenant of the LH/IH that there is a match between the stage of skill development and the type of intervention that facilitates growth. Specifically, students who demonstrated lower reading accuracy and rate demonstrated greater change in response to the more intensive modeling and error correction instructional strategies, while those with more advanced skills responded better to the repetition and reward components.

Furthermore, the stage of the LH at which the individual student is performing can be identified based on their current level of accuracy and rate. However, the baseline measure that best predicted intervention effectiveness for each stage was not the subskill that was the emphasis of development at that stage. While accuracy represents the key outcome for measuring a student's progress within the acquisition stage (Daly et al., 1996), it is a low accurate reading rate that best predicts performance at the acquisition stage. In contrast, a high reading accuracy is an indicator of performance at the proficiency stage while rate is the measure of growth within this stage.

Support for the LH/IH as an explanation for how reading fluency develops also informs broader models of reading development. LaBerge and Samuels (1974) proposed that reading decoding must be fast and automatic before a greater proportion of cognitive resources can be devoted to comprehension of text. The results of this study support the notion that achieving fluent reading skills can be broken down into the steps of first obtaining accurate decoding followed by increasing the rate of word reading. Comprehension skills are likely developing simultaneously and once an appropriately level of accuracy and rate is achieved, the focus of reading development shifts to comprehension. This progression is also consistent with Chall's (1996) model of reading development which begins with a focus on decoding during the initial reading stage and then on increasing automaticity during the confirmation and fluency stage. The subsequent "reading to learn" stages place the emphasis on increasingly more complex levels of comprehension.

Practical Implications

Currently, instructional decisions for struggling readers are often made based on analyzing their response to a selected intervention (Deno, 1985, Gickling & Thompson, 1985). Methods such as curriculum-based measurement (CBM) or brief experimental analysis (BEA) require time for the intervention to be implemented and the response measured before the data can be analyzed to determine intervention effectiveness.

The current study used the tenets of the LH/IH (Haring & Eaton, 1978) framework in combination with BEA to identify pre-intervention characteristics of students who responded best to two reading fluency intervention packages. Students at the acquisition and proficiency stages demonstrated different baseline accuracy and fluency scores which could be used to predict the results of the BEA. Thus, an appropriate intervention for improving accurate reading rate could be identified more efficiently based on a student's accuracy and rate reading grade level passages. Furthermore, the student's on-going response to the intervention could continue to be evaluated using CBM data and changes in baseline rate and accuracy could be used to determine when a change in interventions is warranted.

This model of intervention selection based on a student's current needs fits well within the response to intervention (RTI) framework. When students initially demonstrate a need for additional support, decisions regarding the type of tier 2 intervention (typically, small group instruction) could be made quickly based on a student's baseline reading performance. In addition, if a student does not demonstrate growth in response to the intervention matched to their needs, there is greater justification that a more intensive, tier 3 intervention is needed. The results of the current study suggest preliminary cut scores that can be used for selecting the appropriate intervention for second and third grade students using easy to collect rate and accuracy data. A possible decision model could state that if a student performing below the 50^{th} percentile both reads above 32 words read correctly and has an accuracy score of 85% of greater, then a proficiency intervention should be attempted to boost reading rate. In contrast, if the student is below these cut scores on one or both of these baseline measures, then an acquisition intervention should be administered to address accuracy until the student increases their rate and accuracy.

Further analysis with a larger and more diverse set of students is needed to confirm and clarify these decision points. However, this intervention selection model would make intervention decisions within the RTI framework more efficient and linked to features of the individual student’s pre-intervention level of performance. Moreover, such a cut score model provides a simple method for making intervention decisions that could be used in conjunction with large-scale triannual CBM screenings or as part of an individual assessment of needs.

Potential Implications for Research. The current study is a unique exploration of the relationship between baseline skill level and intervention responsiveness that has the potential of many practical applications. Firstly, the methods used in the current study were unique in that they combined single-subject and group methodologies to draw conclusions about an underlying theory of learning. Specifically, BEA was used to measure qualities of students’ reading functioning and categorize them into groups which could then be further analyzed. While time consuming, this methodology could be used in future research to consider other factors that impact and predict individual differences in response to intervention.

For example, these procedures could be extended to other reading skills including letter/sound recognition, decoding and comprehension or to other academic skill areas such as math fact fluency or spelling. While the current study's focus was on the acquisition and proficiency stages because they are most likely to be the target of intervention for struggling readers, consideration should be given to how the methodology and results could be applied to the generalization and adaptation stages of the LH. Such a comprehensive model could help teachers determine how to focus adjust their instruction based on students' needs across a skill's learning hierarchy.

Future research efforts should also focus on confirming and extending the conclusions of this study regarding reading accuracy and rate for second and third graders and addressing the limitations. Replications would help verify the findings of baseline differences in second and third grade students who respond differentially to fluency interventions that correspond to the predictions of the LH/IH and would allow for more precise cut scores to be developed based on larger data sets. Students from additional grade levels could be evaluated to determine whether the cut scores are consistent regardless of grade level, or if they vary at times when reading fluency is not the focus of instruction or skill development. Furthermore, studies of normally developing students would also indicate whether the findings were consistent whether reading development was progressing as expected or delayed.

Predictions of intervention effectiveness could be evaluated through extended analyses of response to intervention to determine whether the expected gains in fluency were obtained. Additionally, students who were judged to be in the acquisition stage could be followed until their reading rates and accuracy surpass the cut scores and then a BEA could be completed to determine whether they demonstrated a stronger response to the proficiency intervention package, as would be predicted.

Finally, the cut score model presented in this study could be incorporated into larger reading intervention decision frameworks such as those presented by Daly, Chafouleas and Skinner (2004) or Burns and colleagues (2010). These models include simple assessments that can be completed to determine which reading skill should be the focus for an individual student and provide intervention suggestions. Typically decision points in these models are based on theoretical explanations of reading development, but the current methodology has the advantage of being empirically based. Thus, there are a number of opportunities to extend the current study in ways that have important theoretical and practical applications.

Limitations

The current study should be considered within the context of limitations which may have impacted the data collected. There was variability in the grade levels and times of year that participants were tested. Analyses that attempted to control for these two factors did not improve the quality of the models evaluated, indicating that they had a limited impact on the results. However, future studies should work to control for or take into account these variables in order to obtain more accurate cut scores.

In addition, student participants were drawn from different school settings and different evaluators completed the assessments. Inter-rater reliability was not collected to ensure fidelity of procedure implementation, although graduate student evaluators were familiar with the CBM and BEA procedures and were trained and given detailed instruction sheets (Appendixes C, D, and E) by the primary investigator.

The number of students in each of the intervention effectiveness groups (M-EC: 21, RR-R: 24) was also lower than the 30 per group initially desired when the study began. While more definitive results may have been attained with a greater *n*, the strengths of the relationships obtained in the current study despite the lower number of participants indicate that real differences between groups were found.

Additionally, the use BEA to differentiate student responsiveness to the acquisition and proficiency intervention packages may have impacted the results. First of all, the BEA procedure only allowed the examiner to identify the best of the two intervention packages rather than the one that might be deemed effective by allowing the student to meet a desired rate criterion (i.e., reading at the 40^{th} percentile). Secondly, the effectiveness of the selected intervention was assumed based on studies supporting a match between the results of a BEA and more extended analyses (Noell, Freeland, Witt & Gansle, 2001Jones & Wickstrom, 2002). However, the actual impact of the selected intervention package on students' reading fluency skills was not formally evaluated as part of the current study. Different methods of determining the effectiveness of the two intervention packages may have yielded different conclusions.

Finally, it should be noted that the results of this study apply to a limited subset of the skill of reading fluency. The focus of reading accuracy and rate neglected other important features, including prosody and simultaneous comprehension (Samuels, 2007). Additionally, as noted in the comprehensive models of reading development, reading fluency is impacted by competence in a number of other reading related skills (Wolf & Katzir-Cohen, 2001; Hudson et al., 2009). As such, it is difficult to isolate fluency as the single skill measured and addressed with interventions in this study. However, measurement and intervention were based on the reading fluency literature and all students demonstrated improvement in reading rates in response to the intervention packages. Thus, to the extent that it was possible, it appears that important components of reading fluency were appropriately targeted in the current study.

Conclusions

Haring and Eaton's (1978) Learning Hierarchy/Instructional Hierarchy (LH/IH) presents a framework of how skills develop and what forms of instruction best facilitate growth at each stage. The current study used the tenets of the LH/IH to evaluate the baseline characteristics of students who were developing the reading fluency skill of accurate reading rate. The results confirmed the hypothesis that students at the acquisition and proficiency stages for this skill demonstrated different reading rates and accuracy and further that intervention effectiveness could be predicted based on these measures. While additional research is needed to confirm and extend these findings, the current study supports the LH/IH as a useful theoretical model for guiding decision-making within the response to intervention framework.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Adams, M. J. (2004). Modeling the connection between word recognition and reading. In N. J. Unrau & R. B. Ruddell, (Eds), *Theoretical models and processes of reading* (5th Ed., pp. 1202-1227). Newark, DE: International Reading Association.

Allington, R. L. (1983). The reading instruction provided readers of differing reading abilities. *The Elementary School Journal, 83*, 548-559. doi: 10.1086/461333

Allyon, T., & Roberts, M. D. (1974). Eliminating discipline problems by strengthening academic performance. *Journal of Applied Behavior Analysis, 7*, 71-76. doi: 10.1901/jaba.1974.7-71

Ardoin, S. P., & Daly, E. J. III. (2007). Introduction to the special series: Close encounters of the instructional kind-How the Instructional Hierarchy is shaping instructional research 30 years later. *Journal of Behavioral Education, 16*, 1-6. doi: 10.1007/s10864-006-9027-5

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. doi: 10.1007/s10864-008-9066-1

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. doi: 10.1007/s10864-9020-z

Arnold, E. M., Goldston, D. B., Walsh, A. K., Reboussin, B. A., Daniel, S. S., Hickman, E., & Wood, F. B. (2005). Severity of emotional and behavioral problems among poor and typical readers. *Journal of Abnormal Child Psychology, 33*, 205-217. doi: 10.1007/s10802-005-1828-9

Ashby, J. (2006). Prosody in skilled silent reading: Evidence from eye movements. *Journal of Research in Reading, 29*, 318-33. doi: 10.1111/j.1467-9817.2006.00311.x

Aud, S., Hussar, W., Johnson, F., Kena, G., Roth, E., Manning, E., Wang, X., & Zhang, J. (2012). *The Condition of Education 2012* (NCES 2012-045). U.S. Department of Education, National Center for Educational Statistics. Washington, DC. Retrieved from http://nces.ed.gov/pubsearch

Ball, E. W. (1993). Phonological awareness: What's important and to whom? *Reading and Writing: An Interdisciplinary Journal, 5*, 141-159.

Barbetta, P. M., Heward, W. L., & Bradley, D. M. C. (1993). Relative effects of whole-word and phonetic-prompt error correction on the acquisition and maintenance of sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 26*, 99-110. doi: 10.1901/jaba.1993.26-99

Barbetta, P. M., Heward, W. L., Bradley, D. M., & Miller, A.D. (1994). Effects of immediate and delayed error correction on the acquisition and maintenance of basic sight words by students with developmental disabilities. *Journal of Applied Behavior Analysis, 27*, 177-178. doi: 10.1901/jaba.1994.27-177

Begeny, J. C., Daly, E. J. III, & Valleley, R. J. (2006). Improving oral reading fluency through response opportunities: A comparison of phrase drill error correction with repeated readings. *Journal of Behavioral Education, 15*, 229-235. doi: 10.1007/s10864-006-9028-4

Betts, E. A. (1946). *Foundations of reading instruction*. New York: American Book.

Billingsley, F. F. (1977). The effects of self- and externally-imposed schedules of reinforcement on oral reading performance. *Journal of Learning Disabilities, 10*, 20-30. doi: 10.1177/002221947701000904

Bonfiglio, C. M., Daly, E. J. III., Persampieri, M., & Andersen, M. (2006). An experimental analysis of the effects of reading interventions in a small group reading instructional context. *Journal of Behavioral Education, 15*, 93-109. doi: 10.1007/s10864-006-9009-7

Burns, M. K. (2007). Reading at the instructional level with children identified as learning disabled: Potential implications for response–to-intervention. *School Psychology Quarterly*, 22, 297-313. doi: 10.1037/1045-3830.22.3.297

Burns, M. K., Christ, T. J., Boice, C. H., & Szadokierski, I. (2009). Special education in an RTI Model: Addressing unique learning needs. In T. A. Glover and S. Vaughn (Eds.), *The promise of response to intervention* (pp. 267-285). New York: Guilford Press.

Burns, M. K., Codding, R. S., Boice, C. H., & Lukito, G. (2010). Meta-analysis of acquisition and fluency math interventions with instructional and frustration level skills: Evidence for a skill-by-treatment interaction. *School Psychology Review, 39*, 69-83.

Burns, M. K., Tucker, J. A., Frame, J., Foley, S., & Hauser, A. (2000). Interscorer, alternative-form, internal consistency, and test-retest reliability of Gickling's model of curriculum-based assessment for reading. *Journal of Psychoeducational Assessment, 18*, 353-360. doi: 10.1177/073428290001800405

Burns, M. K. & Wagner, D. (2008) Determining an effective intervention within a brief experimental analysis for reading: A meta-analytic review. *School Psychology Review, 37*, 126-136.

Carson, P. M., & Eckert, T. L. (2003). An experimental analysis of mathematics instructional components: Examining the effects of student-selected versus empirically-selected interventions. *Journal of Behavioral Education, 12*(1), 35-54. doi: 10.1023/A:1022370305486

Cassidy, J. & Cassidy, D. (2009). What's hot for 2009. *Reading Today, 26*(4), 1, 8, 9.

Castillo, J. M., Porter, L. J., Curtis, M., & Batsche, G. (2005). The effects of evidence-based reading intervention on socially important outcomes. *NASP Communique, 34*(2), 15-16.

Chafouleas, S. M., Martens, B. K., Dobson, R. L., Weinstein, K. S., & Gardner, K. B. (2004). Fluent reading as the improvement of stimulus control: Additive effects of performance-based interventions to repeated reading on students' reading and error rates. *Journal of Behavioral Education, 13*, 67-81. doi: 10.1023/B:JOBE.0000023656.45233.6f

Chall, J. S. (1983). *Stages of reading development.* New York: McGraw-Hill.

Chall, J. S. (1996). *Learning to read: The great debate* (1967). New York: McGraw Hill.
Chomsky, C. (1978). When you still can't read in third grade: After decoding, what? In S. J. Samuels (Ed.), *What Research has to say about reading instruction* (pp. 13-30). Newark, DE: International Reading Association.
Clay, M. (1987). Learning to be learning disabled. *New Zealand Journal of Educational Studies*, *22*, 155-173.
Cone, J. D. (1997). Issues in functional analysis in behavioral assessment. *Behaviour Research and Therapy, 35*(3), 259-275. doi: 10.1016/S0005-7967(96)00101-5
Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research, 66*, 227-268. doi: 10.3102/00346543066003227
Cox, D. R., & Snell, E. J. (1989). *Analysis of binary data* (2^{nd} ed.). Chapman & Hall/CRC.
Daane, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005). *Fourth-grade students reading aloud: NAEP 2002 Special Study of Oral Reading*. (NCES 2006-469). U.S. Department of Education. Institute of Education Sciences, National Center for Educational Statistics. Washington, DC: U.S. Government Printing Office.
Daly, E. J., III, 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 oral reading performance. *Journal of Applied Behavior Analysis*, *38*, 485-497. doi: 10.1901/jaba.2005.113-04
Daly, E. J., III, Bonfiglio, C. M., Mattson, T., Persampieri, M., & Foreman-Yates, K. (2006). Refining the experimental analysis of academic skills deficits: Part II. Use of brief experimental analysis to evaluate reading fluency treatments. *Journal of Applied Behavior Analysis, 39*(3), 323-331. doi: 10.1901/jaba.2006.13-05
Daly, E. J. III, Chafouleas, S., & Skinner, C. H. (2005). *Interventions for reading problems: Designing and evaluating effective strategies*. New York: The Guilford Press.
Daly, E. J. III, Lentz, F. E., & Boyer, J. (1996). The instructional hierarchy: A conceptual model for understanding the effective components of reading interventions. *School Psychology Quarterly, 11*, 369-386.
Daly, E. J. III, & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27*, 459-469. doi: 10.1901/jaba.1994.27-459
Daly, E. J. III, Martens, B. K., Dool, E. J., & Hintze, J. M. (1998). Using brief functional analysis to select interventions for oral reading. *Journal of Behavioral Education, 8*, 203-218. doi: 10.1023/A:1022835607985
Daly, E. J. III, Martens, B. K., Hamler, K. R., Dool, E. J., & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve

oral reading fluency. *Journal of Applied Behavior Analysis, 32,* 83-94. doi: 10.1901/jaba.1999.32-83

Daly, E. J., III, Martens, B. K., Kilmer, A., & Massie, D. (1996). The effects of instructional match and content overlap on generalized reading performance. *Journal of Applied Behavior Analysis, 29,*507-518. doi: 10.1901/jaba.1996.29-507

Daly, E. J., III, Murdoch, A., Lillenstein, L., Webber, L., & Lentz, F. E. (2002). An examination of methods for testing treatments: Conducting brief experimental analyses of the effects of instructional components on oral reading fluency. *Education & Treatment of Children, 25*(3), 288-316.

Daly, E. J., III, Persampieri, M., McCurdy, M., & Gortmaker, V. (2005) Generating reading interventions through experimental analysis of academic skills: Demonstration and empirical evaluation. *School Psychology Review, 34,* 395-414.

Daly, E. J. III, Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic and performance problems. *School Psychology Review, 26,* 554-574.

Data Accountability Center. (2012). *Part B Child Count, Report Year: Fall 2012, 50 States and D.C.* [Data table created using IDEA Data Explorer tool]. Retrieved from https://www.ideadata.org/DACAnalyticTool/Intro_2.asp

Deno, S. L. (1985) Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Deno, S. L. (2002). Problem-solving as "best practice." In A. Thomas and J. Grimes (Eds.) *Best practices in school psychology IV* (Vol. 1, pp. 37-55). Bethesda, MD: National Association of School Psychologists.

Deno, S. L., Marston, D., & Tindal, G. (1985-1986). Direct and frequent curriculum-based measurement: An alternative for educational decision making. *Special Services in the Schools, 2,* 5-27. doi: 10.1300/J008v02n02_02

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36-47.

DeVries, M., & Feldman, D. (1983, February). *The effect of a token reinforcement program on the sight word acquisition rate of learning disabled students in a rural school program.* Paper presented at the Annual Convention of the Association for Children and Adults with Learning Disabilities, Washington, DC.

Dornbush, R. L., & Baslow, S. (1970). The relationship between auditory and visual short-term memory and reading achievement. *Child Development, 41*, 1033-1044. doi: 10.1111/1467-8624.ep10403872

Dowhower, S. L. (1987). Effects of repeated reading on second-grade transitional readers' fluency and comprehension. *Reading Research Quarterly, 22,* 389-406. doi: 10.2307/747699

Dowhower, S. L. (1991). Speaking of prosody: Fluency's unattended bedfellow. *Theory into Practice, 30,* 165-175. doi: 10.2307/1476878

Duhon, G. J., Noell, G. H., Witt, J., Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N. (2004). Identifying academic skill and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33,* 429-443.

Eckert, T. L., Ardoin, S. P., Daisey, D. M., & Scarola, M. D. (2000). Empirically evaluating the effectiveness of reading interventions: The use of brief experimental analysis and single case designs. *Psychology in the Schools, 37*(5), 463-473. doi: 10.1002/1520-6807(200009)

Eckert, T. L., Ardoin, S. P., Daly, E. J., III, & Martens, B. K. (2002). Improving oral reading fluency: A brief experimental analysis of combining an antecedent intervention with consequences. *Journal of Applied Behavior Analysis, 35*(3), 271-281. doi: 10.1901/jaba.2002.35-271

Eckert, T. L., Dunn, E. K., & Ardoin, S. P. (2006). The effects of alternate forms of performance feedback on elementary-aged students' oral reading fluency. *Journal of Behavioral Education, 15*, 149-162. doi: 10.1007/s10864-006-9018-6

Eldredge, J. L. (1990). Increasing reading performance of poor readers in the third grade by using a group assisted strategy. *Journal of Educational Research, 84*, 69-77.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., ... Shaywitz, B. A. (1994). Cognitive profiles of reading disability: Comparisons of discrepancy and low achievement definitions. *Journal of Educational Psychology, 86*, 6-23. doi: 10.1037/0022-0663.86.1.6

Fontanni-Axelrod, A. M. (2006). Linking assessment to intervention: Utilizing experimental analysis of reading problems to develop individualized interventions. Dissertation Abstracts International Section A: Humanities and Social Sciences, 66 (12-A), 4296.

Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read: Preventing reading failure in at-risk children. *Journal of Educational Psychology, 90*, 37-55.

Forbach, G. B., Stanners, R. F., & Hochhaus, L. (1974). Repetition and practice effects In a lexical decision task. *Memory and Cognition, 2*, 337-339. doi: 10.3758/BF03209005

Fuchs, L. S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*, 188-192.

Fuchs, L. S., & Fuchs, D. (1986). Linking assessment to instructional intervention: An overview. *School Psychology Review, 3*, 318-323.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Whinnery, K. (1991). Effects of goal line feedback on level, slope, and stability of performance within curriculum-based measurement. *Learning Disabilities Research & Practice, 6*, 66-74.

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. doi: 10.1207/S1532799XSSR503_3

Fuchs, D., Fuchs, L. S., Mathes, P. G., Lipsey, M. W., & Roberts, P. H. (2001, August). *Is "learning disabilities" just a fancy term for low achievement?: A meta-analysis of reading differences between low achievers with and without the label.* Paper presented at the Learning Disabilities Summit: Building a Foundation for the Future, Washington, DC.

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2003). Responsiveness-to-intervention: Definitions, evidence, and implications for the learning disability construct. *Learning Disabilities Research & Practice, 18,* 157-171. doi: 10.1111/1540-5826.00072

Gerber, P. J. (1997). Employment of people with learning disabilities: The big picture. In P. Gerber & D. Brown (Eds.), *Learning disabilities and employment* (pp. 3-18). Austin, TX: PRO-ED.

Gersten, R., & Dimino, J. A., (2006). RTI (Response to intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly, 41,* 99-108. doi: 10.1598/RRQ.41.1.5

Gickling, E., & Thompson, V. (1985). A personal view of curriculum-based assessment. *Exceptional Children,* 52, 205-218.

Goffreda, C. T., & DiPerna, J. C. (2010). An empirical review of psychometric evidence for the dynamic indicators of basic early literacy skills. *School Psychology Review, 39,* 463-483.

Good, R. H., III, & Kaminski, R. A. (Eds.) (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6^{th} ed.). Eugene, OR: Institute for the Development of Educational Achievement.

Goodman, K. (1967). Reading: A psychlinguistic guessing game. *Journal of the Reading Specialist, 64,* 126-135.

Goodman, K. (1973). Miscues: Windows on the reading process. In F. Gollasch (Ed.) *Language and literacy: The selected writings of Kenneth Goodman* (Vol. 1, pp. 123-134). Boston: Routledge & Kegan Paul.

Goodman, K. S. (1987). Acquiring literacy is natural: Who skilled Cock Robin? *Theory into Practice,26,* 368-373.

Gortmaker, V. J., Daly, E. J. III, McCurdy, M., Persampieri, M. J., & Hergenrader, M. (2007). Improving reading outcomes for children with learning disabilities: Using brief experimental analysis to develop parent-tutoring interventions. *Journal of Applied Behavior Analysis, 40,* 203-221. doi: 10.1901/jaba.2007.105-05.

Greenberg, E., Dunleavy, E., & Kutner, M. (2007). Literacy *Behind Bars: Results from the 2003 National Assessment of Adult Literacy* (NCES 2007-473). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Gresham, F. (2001, August). *Responsiveness to intervention: An alternative approach to the identification of learning disabilities.* Paper presented at the Learning Disabilities Summit: Building a Foundation for the Future, Washington, DC.

Gresham, F., Reschly, D. J., Tilly, W. D., Fletcher, J., Burns, M., Christ, T., … Shinn, M. (2005). Comprehensive evaluation of learning disabilities: A response to intervention perspective. *The School Psychologist, 59,* 26-30.

Griffiths, A., VanDerHeyden, A. M., Parson, L. B, & Burns, M. K. (2006). Practical applications of response-to-intervention research. Assessment for Effective Intervention, 32, 50-57. doi: 10.1177/15345084060320010701

Hale, A. D., Skinner, C. H., Winn, B. D., Oliver, R., & Allin, J. D. (2005). An investigation of listening and listening-while-reading accommodations on reading comprehension levels and rates in students with emotional disorders. *Psychology in the Schools, 42,*39-51. doi: 10.1002/pits20027.

Harding, J., Wacker, D. P., Cooper, L. J., Millard, T., & Jensen-Kovalan, P. (1994). Brief hierarchical assessment of potential treatment components with children in an outpatient clinic. *Journal of Applied Behavior Analysis*, *27,* 291–300. doi: 10.1901/jaba.1994.27-291

Haring, N. G., & Eaton, M. D. (1978). Systematic instructional technology: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The fourth R: Research in the classroom*. Columbus, OH: Merrill.

Hasbrouck, J. & Tindal, G. A. (2006) Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59,* 636-644. doi: 10.1598/RT.59.7.3

Hayes, S. C., Nelson, R. O., & Jarrett, R. B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist,* 963-974. Doi: 10.1037/0003-066X.42.11.963

Hosmer, D. W., & Lemeshow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics-Theory and Methods, 9*, 1043-1069. doi: 10.1080/03610928008827941

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgensen, J. K. (2009). The complex nature of reading fluency: A multidimensional view, *Reading & Writing Quarterly, 25*, 4-32. Doi: 10.1080/10573560802491208

Individuals with Disabilities Education Improvement Act of 2004, Public Law 108-446 (2004).

Indrisano, R., & Chall, J. S. (1995). Literacy development. *Journal of Education, 177*, 63-83.

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. doi: 10.1037/0022-0663.95.4.719

Jenkins, J. R., Larson, K., & Fleisher, L. (1983). Effects of error correction of word recognition and reading comprehension. *Learning Disability Quarterly, 6*, 139-145.

Jeynes, W. H., & Littell, S. W. (2000). A meta-analysis of studies examining the effect of whole language instruction on the literacy of low-SES students. *The Elementary School Journal, 101*, 21-33.

Jimerson, S., Burns, M. K., & VanDerHeyden, A. M. (2007). *Handbook of Response to Intervention: The science and practice of assessment and intervention.* New York: Springer.

Jones K. M., & Wickstrom, K. F. (2002). Done in sixty seconds: Further analysis of the brief assessment model for academic problems. *School Psychology Review, 31*, 554-568.

Kamii, C. (1991). *Early literacy: A constructivist foundation for whole language.* Washington, DC: National Education Association.

Kavale, K. A. (2001). *Discrepancy Models in the Identification of Learning Disability*. Paper presented at the Learning Disabilities Summit: Building a Foundation for the Future, Washington, DC.

Kavale, K. A., & Forness, S. R. (1996). Social skill deficits and learning disabilities: A meta-analysis. *Journal of Learning Disabilities*, *29*, 226-237. doi: 10.1177/002221949602900301

Kern, L., Vorndran, C. M., Hilt, A., Ringdahl, J. E., Adleman, B. E., & Dunlap, G. (1998). Choice as an intervention to improve behavior: A review of the literature. *Journal of Behavioral Education,8*, 151-169. doi: 10.1023/A:1022831507077

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of Educational Psychology, 95*, 3-21. doi: 10.1037/0022-0663.95.1.3

Kutner, M., Greenberg, E., Jin, Y., Boyle, B., Hsu, Y., & 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. doi: 10.1016/0010-0285(74)90015-2.

Lalli, E. P., & Shapiro, E. S. (1990). The effects of self-monitoring and contingent reward on sight word acquisition. *Education and Treatment of Children*, *13*, 129-141.

Lerner, J. (2003). *Learning disabilities: Theories, diagnosis, and teaching practices* (9th Edition). Boston: Houghton Mifflin Company.

Leslie L., & Osol, P. (1978). Changes in oral reading strategies as a function of quantities of miscues. *Journal of Reading Behavior, 10*, 442-445.

Literacy Information and Communication System. (n.d.). *Research Definitions*. Retrieved from http://lincs.ed.gov/research/researchdef.html

Lovitt, T. C., Eaton, M., Kirkwood, M. E., & Pelander, A. (1971). Effects of various reinforcement contingencies on oral reading rate. In E. Ramp & B. L. Hopkins (Eds.) *A new direction for education: Behavior analysis* (pp. 54-71). Lawrence: University of Kansas Press.

Ma, H. –H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification*, *30*, 598-617. doi: 10.1177/0145445504272974

Mallory, K. J., Gilbertson, D., & Maxfield, J. (2007). Use of brief experimental analysis for selecting reading interventions for English language learners. *School Psychology Review*, *36*, 291-310.

Marston, D. B. (1989). Curriculum-based measurement: What it is and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York: Guilford Press.

McComas, J. J., Wacker, D. P., Cooper, L. J., Asmus, J. M., Richman, D., & Stoner, B. (1996). Brief experimental analysis of stimulus prompts for accurate responding on academic tasks in an outpatient clinic. *Journal of Applied Behavior Analysis, 29*(3), 397-401. doi: 10.1901/jaba.1996.29-397.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8,* 283-298.

Meyer, M. S., & Felton, R. H. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of Dyslexia, 49*, 283-306. Doi: 10.1007/s11881-999-0027-8

Moats, L. C. & Dakin, K. E. (2007). *Basic Facts about Dyslexia and Other Reading Problems.* Baltimore, MD: International Dyslexia Association.

Moyer, S. B. (1982). Repeated reading. *Journal of Learning Disabilities, 15,* 619-623.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika, 78,* 691-692. doi: 10.1093/biomet/78.3.691

National Institute of Child Health 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* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

Noell, G. H., Gansle, K. A., Witt, J. C., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., Gilbertson, D. N., & Northup, J. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31*, 659-663. doi: 10.1901/jaba.1998.31-659

Noell, G. H., Freeland, J. T., Witt, J. C., & Gansle, K. A. (2001). Using brief assessments to identify effective interventions for individual students. *Journal of School Psychology, 39*, 335-355. doi: 10.1016/S0022-4405(01)00072-3

O'Shea, L. J., Munson, S. M., & O'Shea, D. J. (1984). Error correction in oral reading: Evaluating the effectiveness of three procedures. *Education and Treatment of Children, 7,* 203-214.

O'Shea, L., Sindelar, P., & O'Shea, D. (1985). The effects of repeated readings and attentional cues on reading fluency and comprehension. *Journal of Reading Behavior, 17*, 129-142. doi: 10.1080/10862968509547535

Pany, D., & McCoy, K. M. (1988). Effects of corrective feedback on word accuracy and reading comprehension of readers with learning disabilities. *Journal of Learning Disabilities, 21*, 546-550. doi: 10.1177/002221948802100905

Parker, D. C., & Burns, M. K. (in press). Using the instructional level as a criterion to determine student phase of learning for reading fluency: Evidence for the learning hierarchy. *Reading and Writing Quarterly.*

Perfetti, C. A. (1985). *Reading ability*. New York: Oxford University Press.

Pikulski J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *The Reading Teacher*, *58*, 510-519. doi: 10.1598/RT.58.6.2

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). *Listening to children read aloud: Data from NAEP's integrated reading performance record (IRPR) at Grade 4* (Report No. 23-FR-04). Retrieved from the ERIC website: http://www.eric.ed.gov/PDFS/ED378550.pdf.

Pressley, M. (2006). *Reading instruction that works: The case for balanced teaching.* New York: Guilford Press.

Pressley, M., & Rankin, J. (1994). More about whole language methods of reading instruction for students at risk for early reading failure. *Learning Disabilities Research & Practice, 9*, 157-168.
Rashotte, C. A., & Torgesen, J. K. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly*, 20, 180–188.
Rasinski, T. V. (1990). Effects of repeated reading and listening-while-reading on reading fluency. *Journal of Educational Research, 83*, 147-150.
Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlation evidence. *Journal of School Psychology, 47*, 427-469. doi: 10.1016/j.jsp.2009.07.001
Reschly, D. J. (2005). Learning disabilities identification: Primary intervention, secondary intervention, and then what? *Journal of Learning Disabilities, 38*, 510-515. doi: 10.1177/00222194050380060601
Reschly, D. J., & Hosp, J. L. (2004). State SLD identification policies and practices. *Learning Disability Quarterly, 27*, 197-213. doi: 10.2307/1593673
Roberts, M., & Smith, D. D. (1980). The relationship among correct and error oral reading rates and comprehension. *Learning Disability Quarterly, 3*, 54-64. doi: 10.2307/1510426
Rose, T. L. (1984). The effects of two prepractice procedures on oral reading. *Journal of Learning Disabilities, 17*, 544-548. doi: 10.1177/002221948401700709
Rose, T. L., McEntire, E., & Dowdy, C. (1982). Effects of two error-correction procedures on oral reading. *Learning Disability Quarterly, 5*, 100-105. doi: 10.2307/1510570
Rose, T. L., & Sherry, L. (1984). Relative effects of two previewing procedures on LD adolescents' oral reading performance. *Learning Disability Quarterly, 7*, 39-44. doi: 10.2307/1510259
Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach. In M. Posner (ed.), *Foundations of cognitive science* (pp. 133-159). Cambridge, MA: MIT Press.
Rumelhart, D. E. (2004). Toward an interactive model of reading. In N. J. Unrau & R. B. Ruddell, (Eds), *Theoretical models and processes of reading* (5th Ed., pp. 1149-1179). Newark, DE: International Reading Association.
Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16*, 181-197. doi: 10.1111/j.1469-7610.1975.tb01269.x
Samuels, S. J. (1979). The method of repeated readings. *The Reading Teacher, 32*, 403-408.
Samuels, S. J. (2007). Commentary: The DIBELS test is speed of barking at print: What we mean by reading fluency. *Reading Research Quarterly, 42*, 563-566.
Schreiber, P. A. (1980). On the acquisition of reading fluency. *Journal of Reading Behavior, 12*, 177-186.
Schreiber, P. A. (1991). Understanding prosody's role in reading acquisition. *Theory into Practice, 30*, 158-164. doi: 10.1080/00405849109543496

Scruggs, T. E., Mastropieri, M. A., & Casto, G. (1987). The quantitative synthesis of single-subject research: Methodology and validation. *Remedial and Special Education, 8*, 24-33. doi: 10.1177/074193258700800206

Shapiro, E. S., Angello, L. M., & Eckert, T. L. (2004). Has curriculum-based assessment become of staple of school psychology practice? An update and extension of knowledge, use, and attitudes from 1990 to 2000. *School Psychology Review, 33,* 249-257.

Shinn, M. R. (1989). *Curriculum-based measurement: Assessing special children.* New York: Guilford Press.

Shinn, M. R., Good, R. H. III, Knutson, N., Tilly, W. D. III, & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459-479.

Singh, J., & Singh, N. N. (1985). Comparison of word-supply and word-analysis error-correction procedures on oral reading by mentally retarded children. *American Journal of Mental Deficiency, 90*, 64-70.

Singh, N. N. (1990). Effects of two error-correction procedures on oral reading errors: Word supply versus sentence repeat. *Behavior Modification, 14*, 188-199.

Singh, N. N. & Singh, J. (1986). Increasing oral reading proficiency: A comparative analysis of drill and positive practice overcorrection procedures. *Behavior Modification, 10*, 115-130. doi: 10.1177/01454455860101007

Singh, N. N., Winton, A. S. W., & Singh, J. (1985). Effects of delayed versus immediate attention to oral reading on the reading proficiency of mentally retarded children. *Applied Research in Mental Retardation, 6*, 283-293. doi: 10.1016/0270-3092(85)90002-5

Skinner, C. H., Adamson, K. L., Woodward, J. R., Jackson, Jr., R. R., Atchison, L. A., & Mims, J. W. (1993). A comparison of fast-rate, slow-rate, and silent previewing interventions on reading performance. *Journal of Learning Disabilities, 26*, 674-681. doi: 10.1177/002221949302601005

Skinner, C. H., Logan, P., Robinson, S. L., & Robinson, D. H. (1997). Demonstration as a reading intervention for exceptional learners. *School Psychology Review, 26*, 437-447.

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

Staats, A. W., Minke, K. A., Finley, J. R., Wolf. M., & Brooks, L. O. (1964). A reinforcer system and experimental procedure for the laboratory study of reading acquisition. *Child Development, 35*, 209-231.

Stahl, S. A., & Miller, P. D. (1989). Whole language and language experience approaches for beginning reading: A Quantitative research synthesis. *Review of Educational Research, 59*, 87-116. doi: 10.3102/00346543059001087

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*, 360-407. doi: 10.1598/RRQ.21.4.1

Stanovich, K. E. (1988). The right and wrong places to look for the cognitive locus of reading disability. *Annals of Dyslexia, 38*, 154-177. doi: 10.1007/BF02648254

Stecker, S. K., Roser, N. L., & Martinez, M. G. (1998). Understanding oral reading fluency. In T. Shanahan & F. V. Rodriguez-Brown (Eds.), *47th yearbook of the National Reading Conference* (pp. 295-310). Chicago: National Reading Conference.

Stenner, A. J., Burdick, H., Sanford, E.E., & Burdick, D. S. (2007). *The Lexile Framework for Reading Technical Report*. Durham, NC: MetaMetrics, Inc.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education, 25*, 252-261. doi: 10.1177/07419325040250040801

Trzesniewski, K. H., Moffitt, T. E., Caspi, A., Taylor A., & Maughan, B. (2006). Revisiting the association between reading achievement and antisocial behavior: New evidence of an environmental explanation from a twin study. *Child Development*, *77*, 72-88. doi: 10.1111/j.1467-8624.2006.00857.x

U.S. Department of Education, Office of Special Education and Rehabilitative Services, Office of Special Education Programs (2011). *30th Annual Report to Congress on the Implementation of the Individuals with Disabilities Education Act, 2008*, Washington, D.C.

U.S. Department of Education, National Center for Educational Statistics. (2011). *Digest of Educational Statistics, 2010* (NCES 2011-015). Retrieved from National Center for Educational Statistics website: http://nces.ed.gov/pubs2011/2011015.pdf

VanAuken, T. L., Chafouleas, S. M., Bradley, T. A., & Martens, B. K. (2002). Using brief experimental analysis to select oral reading interventions: An investigation of treatment utility. *Journal of Behavioral Education, 11*, 163-179.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research and Practice, 18*, 137-146. doi: 10.1111/1540-5826.00070

Velluntino, F. R. (1991). Introduction to three studies on reading acquisition: Convergent findings on theoretical foundations of code-oriented versus whole-language approaches to reading instruction. *Journal of Educational Psychology, 83*, 437-443. doi: 10.1037/0022-0663.83.4.437

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult to remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 4*, 601-638. doi: 10.1037/0022-0663.88.4.601

Wadsworth, S. J., Fulker D. W., & DeFries, J. C. (1999). Stability of genetic and environmental influences on reading performance at 7 and 12 years of age in the Colorado Adoption Project. *International Journal of Behavioral Development, 23*, 319-332. doi: 10.1080/016502599383847

Wagner, D., McComas, J. J., Bollman, K., & Holton, E. (2006). The use of functional reading analysis to identify effective reading interventions. *Assessment for Effective Intervention, 32*, 40-49. doi: 10.1177/15345084060320010601

Wagner, R. K., & Stanovich, K. E. (1996). Expertise in reading. In K. A. Ericsson (Ed.), *The road to excellence: The acquisition of expert performance in the arts and sciences, sports, and games* (pp. 189-225). Hillsdale, NJ: Erlbaum.

Wasson, B. B., Beare, P. L., & Wasson, J. B. (1990). Classroom behavior of good and poor readers. *Journal of Educational Research, 83*, 162-165.

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education, 41*, 85-120. doi: 10.1177/00224669070410020401

Waznek, J., Vaughn, S., Kim, A. H., & Cavanaugh, C. L. (2006). The effects of reading interventions on social outcomes for elementary students with reading difficulties: A synthesis. *Reading and Writing Quarterly, 22*, 121-138. doi: 10.1080/10573560500242192

Wilber, A., & Cushman, T. P. (2006). Selecting effective academic interventions: An example using brief experimental analysis for oral reading. *Psychology in the Schools, 43,* 79-84. doi: 10.1002/pits.20131

Wolf, M., & Katzir-Cohen, T. (2001). Reading fluency and its intervention. *Scientific Studies of Reading, 5*, 211-239. doi: 10.1207/S1532799XSSR0503_2

Young, A., & Bowers, P. G. (1995). Individual difference and text difficulty determinants of reading fluency and expressiveness. *Journal of Experimental Child Psychology*, *60*, 428-454. doi: 10.1006/jecp.1995.1048

Ysseldyke, J. E., Algozzine, B., Shinn, M., & McGue, M. (1982). Similarities and differences between students labeled underachievers and learning disabled. *Journal of Special Education, 16*, 73-85. doi: 10.1177/002246698201600108

Appendices

Appendix A

CONSENT FORM

Predicting Intervention Effectiveness from Oral Reading Accuracy and Rate Measures

Your child is being invited to be in a research study of how to predict which type of reading instruction will work best for individual students. Your child was selected as a possible participant because he or she was identified by his/her teacher as being in the process of learning to read at an appropriate rate. We ask that you read this form and ask any questions you may have before agreeing to allow your child to be in the study.

This study is being conducted by: Isadora Szadokierski, M.A., a graduate student in the Educational Psychology Department at the University of Minnesota, under the guidance of Matthew Burns, Ph.D., an Associate Professor of Educational Psychology at the University of Minnesota.

Background Information

The purpose of this study is: To learn the reading characteristics of students who will benefit most from to two types of instruction. Students at different stages of the learning process need different types of instruction to improve their reading ability. Early on, students must learn to read words correctly and then the focus shifts to practicing reading with appropriate speed. If a teacher could tell at which stage a student was reading, then she could predict how to best teach the student.

Procedures:

If you decide to allow your child to be in this study, he or she will be asked to read 10 grade-level passages out loud for 1 minute while the researcher records his/her reading rate (number of words read correctly in 1 minute) and accuracy (percentage of words read correctly). First, your child will read all passages with no help. Then he or she will read the passages again, but this time the researcher will provide brief instruction using two different interventions that help students at different stages of learning. The researchers will then determine if it possible to predict which intervention has the best results using the original levels of accuracy and rate.

The study will last over a two-week period. Your child will meet with the researcher for two 15-20 minute sessions during the first week and three 10-15 minute sessions during the second week. The following procedure will be used:

1) Each student will accompany the researcher to a quiet area within the school.

2) The study will be explained to the student and he or she will be asked if they wish to participate. Those that do not wish to participate will be allowed to return to the classroom without any negative consequences.

3) During each of the first two sessions, the student will read five passages out loud for 1 minute each with no instruction from the researcher.

4) During the next three sessions, the researcher will alternate between two interventions to pre-teach the passages before measuring the student's reading. In one intervention, the researcher will read the passage once to the student and then correct any errors he or she makes in the second reading the passage. In the other, the student will practice reading the passage aloud twice and will be offered a reward for fast reading on the third time.

Risks and Benefits of being in the Study
The study has a few potential risks: First, your child will be asked to read aloud while the researcher monitors performance, which may be uncomfortable for some children. Also, we will attempt to conduct the study during non-instructional times, but some class time will likely be missed.

The benefits to participation are: There are few direct benefits for your child. Information about the type of reading instruction that lead to the best outcomes for your child will be provided to the teacher and may or may not be used in future instruction. However, your child's participation will help us learn how to predict the best reading instruction for individual students, which will help future children learn to read.

Compensation:
Your child will receive a small reward (a sticker, pencil, or fruit snacks) as part of one of the interventions.

Confidentiality:
The records of this study will be kept private. In any sort of report we might publish, we will not include any information that will make it possible to identify a participant. Research records will be stored securely and only researchers will have access to the records.

Voluntary Nature of the Study:
Participation in this study is voluntary. Your decision whether or not to let your child participate will not affect you or your child's current or future relations with the University of Minnesota or their school. If you decide to let your child participate, **he or she is free to withdraw at any time without explanation**.

Contacts and Questions:
The researchers conducting this study are: Isadora Szadokierski, under the guidance of Dr. Matthew Burns. If you have any questions, please contact Ms. Szadokierski at 651-295-3032 or szado001@umn.edu or Dr. Burns at 341 EdSciBldg, 56 E. River Rd., Minneapolis, MN 55455, 612-626-7324, or burns258@umn.edu.

If you have any questions or concerns regarding this study and would like to talk to someone other than the researcher(s), **you are encouraged** to contact the Research Subjects' Advocate Line, D528 Mayo, 420 Delaware St. Southeast, Minneapolis, Minnesota 55455; (612) 625-1650.

You will be given a copy of this information to keep for your records.

Statement of Consent:
I have read the above information. I have asked questions and have received answers. I consent for my child to participate in the study.

Name of Child:____

Signature of parent or guardian:_________________________ Date: ___________

Signature of Investigator:______________________________ Date: ___________

Appendix B

Assent Form

Predicting Intervention Effectiveness from Oral Reading Accuracy and Rate Measures

You are invited to participate in a research project. We are hoping to learn how to best teach kids to read. You can be in this project because you are in second or third grade and are learning to read. Your parent(s) have given permission for you to take part in this study. If you have any questions, please ask at any time.

If you agree to be in the study, you will go with the researcher and practice reading five times over the next two weeks. Each session will last between 10 and 25 minutes. The researcher will measure how fast you read out loud and how many words you read correctly. Also, sometimes the researcher will help you with your reading and sometimes she won't.

You might not like reading out loud while the researcher follows along, but being in this study will help us learn how to do a better job teaching other kids to read.

The results will be used for research only. No one, not even your teacher or parents, will know how you did. Being in this study is totally up to you, and no one will be mad at you if you don't want to do it.

Signing here means that you have read this paper or had it read to you and that you are willing to be in this study. If you don't want to be in this study, don't sign. Remember, being in this study is up to you, and no one will be mad at you if you don't sign this or even if you change your mind later.

Signature of participant_______________________________________

Signature of person explaining study_________________________________

Date______________________

Appendix C

Predicting Intervention Effectiveness from Oral Reading Accuracy and Rate Measures Baseline Instructions

Baseline includes 10 passages, read over 2 sessions. The spoken directions are mainly for the first session, but the student should be reminded during the second session.

<u>*Passage 1*</u>

1. Say "*I have some stories for you to read today*."

2. Place the passage in front of the student.

3. Say "*When I say 'begin' start reading aloud at the top of the page. Read across the page* (demonstrate) *and try to read each word. If you come to a word you don't know, try to sound it out, but if you still can't figure it out, skip it. Be sure to do your best reading. Do you have any questions?*"

4. Say "*Begin*" and start stopwatch when student says first word.

5. Follow along, putting a slash through words read incorrectly.

6. If the student hesitates for 3 seconds, tell them to "*go on*" and mark word incorrect.

7. At 1 minute, say "*Stop*" and put a] at the last word read.

<u>*Passage 2-5*</u>

1. Say "*Here is the next story*" and place passage in front of student.

2. Say "*When I say 'begin', start reading aloud at the top of this page. (Remember to do you best reading)*".

3. Repeat monitoring steps above.

<u>*End*</u>

1. Say "*You did great work today! Thanks for your help*."

Appendix D

BEA Session Instructions

There are 3 BEA sessions. During the first two, the following interventions are tested 3 times. The third session consists of a baseline passage (no intervention), followed by the "best" intervention (i.e., that which lead to the greatest increase in rate over baseline) applied to a single passage.

1. Tell the student "*Today we are going to do things a little differently. I am going to have you read the stories again, but now I'm going to help you in different ways.*"

Modeling-Error Correction Intervention

2. "*This time I am going to read the story to you once and then have you read it back to me. Please follow along while I read*" Read the entire passage at a fluent speed. (Record time). Although it is not required, encourage the student to follow along with his finger and try to redirect attention to place in passage if overly distracted.

3. After done reading passage, say "*Now you read the story back to me. When you get to a word you don't know, I will tell it to you and I want you to say it back to me and keep reading. Then at the end we'll practice those words again.*"

4. Have student read for 90 seconds, <u>providing missed and incorrectly read words and having student repeat them back</u>. Mark errors & place a] after last word read after 60 and 90 seconds. Place a small "1" above errors made on this reading and at the]s. Also, place a circled "60/90" and the 60/90-second], respectively. Then say "*Stop.*"

5. Practice all missed words and phrases. Say "*This word* (point) *is* ____. *Say it back to me. Good. Now read this phrase out loud <u>three times</u>,* __________ (point). Phrase is generally the clause in which the word is embedded. If multiple words missed in the same sentence, practice each word and then have student read whole sentence 3 times.

6. "*Now, when I say begin start reading at the beginning of the story. Be sure to do your best reading.*" (Do not provide correction, just say "go on" is hesitating for 3 secs). Mark errors with a "2" above them and place a circled "60" next to] after last word read after 60 seconds. Say "*Stop*" after 1 min.

Appendix E

BEA Session Instructions

There are 3 BEA sessions. During the first two, the following interventions are tested 3 times. The third session consists of a baseline passage (no intervention), followed by the "best" intervention (i.e., that which lead to the greatest increase in rate over baseline) applied to a single passage.

1. Tell the student "*Today we are going to do things a little differently. I am going to have you read the stories again, but now I'm going to help you in different ways.*"

Repeated Reading-Reward Intervention

1. "*This time I want you to read the story out loud two times. When I tell you to stop the first time, I want you to start at the beginning again. If you come to a word you don't know, try your best, but if you still don't know it, skip it. After you've read the story twice we'll do something different. Ready?*"

2. Let the student read the passage for 90 seconds. Tell the student to "*Go* on" if they hesitate for 3 secs or more. Mark errors and]s at 60 and 90 seconds with a "1". After 90 seconds tell student "*Stop. Start at the top again. Begin.*"

3. Let the student read for 90 seconds. Mark errors and]s with a "2". While student reads, try to determine the student's rate for the first reading (above). Then determine the "reading goal" by adding 20% to the accurate 1st reading rate. After 90 seconds tell student "*Stop.*"
- To calculate goal, round accurate rate/1st 60 secs. down to nearest 0/5; take 10% and multiply x 2; add to accurate rate. Student must read at least at least this many words correctly on 3rd reading to earn prize.

4. Take out prize box and say "*I brought my prize box with some fun things inside. I have pencils* (take out 2), *stickers* (take out 2), *and fruit snacks. Which of these would you most like to earn?*" Place child's selection to the side (still visible) and put others & prize box away.

5. "*To earn this prize you have to read the story faster than the first time you read it today. I'll keep track and tell you at the end whether you met the goal. Ready?*" Record students correct rate for the 1st minute. Reward prize if earned.

- Students often "speed read," which for this is ok, but it can be difficult to track errors.