Examining Effects of a Repeated Reading Intervention and Predictive Effects of Student Inputs

A Dissertation
SUBMITTED TO THE FACULTY OF
UNIVERSITY OF MINNESOTA
BY

Stephanie M. Snidarich

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

Jennifer J. McComas

December 2015

Acknowledgements

There are two key people who influenced my decision to pursue this degree. The first was my father, Ken Hammerschmidt, who taught me by his example that life is about using your hands and your abilities to serve, and that every goal is attainable through devotion and hard work. The second was my adviser for my masters program, Steve Robinson, who allowed me to accompany him to the Pacific Coast Research Conference. I got my first real taste of academia in that environment, and immediately knew *this* was the work I wanted to be doing. Steve encouraged me and pointed the way from that point forward, and I will be forever grateful for his guidance.

Shortly thereafter, Steve introduced me to Dr. Jennifer McComas, who took me on as her advisee, in which role I have been afforded one opportunity after another to explore and pursue research. Jennifer listened to the the stories I shared from my teaching experiences, and it was gratifying to find that we shared a belief that behavioral interactions between teachers and students around academic tasks might play a key role in the progress of at least *some* students. Our many conversations along these lines have helped me clarify a focus for my own work. On both a personal and professional level, Jennifer has inspired me with her deep understanding and love of behavioral principles, tireless drive, humor, and most of all, her focus on serving *people* through her work. Her example, support, and candor through this process have been invaluable, and I feel fortunate to be the recipient of her knowledge, as well as her friendship.

I would also like to acknowledge several others who have played critical roles in my growth as a scholar. Dr. Kristen McMaster always made time to review my work, answer my many questions, and seriously consider my ideas. I have learned so much from our conversations over the years. Dr. Matt Burns was someone I relied on frequently to bounce around ideas for research and review my analyses. It was also through Matt that I met a number of school psychology students who have since become my colleagues and friends. Dr. Frank Symons influenced me to push my thinking beyond special education and include perspectives and knowledge from other areas, such as philosophy and medicine. Frank is always up for a good, scholarly debate, and our early conversations about dosage helped me to form some of the ideas that underlie this study. Dr. Dana Wagner has been a true friend and colleague; someone who I have relied on for perspective and guidance. It has also been wonderful to conduct research together. Dr. Ted Christ generously allowed me to use FAST measures for my dissertation study and has always made time to provide me with his expert guidance when I had questions about the measures or analyses. Dr. Lori Helman and her work have deepened my understanding of reading instruction, and I had the pleasure of working with her on several PRESS projects.

My progress was buttressed by the support of numerous students who went through this process alongside me. I want to thank the Dead Zone Group: Elizabeth, Quannah, Brooke, Amy, Jessica, Julianne, and honorary member Adele for injecting levity into the graduate school grind, and for your examples as women and scholars. We are finally coming down from the mountain! Additionally, my gratitude for the support I received from Katie Maki during the collection of my dissertation data is heart-felt. You were my right-hand person, and I could not have done this study without your help.

I also want to thank my family—all of you—for being a steadfast presence during the pursuit of this degree, and for believing in and loving me. To my five children, thank you for being endlessly patient and supportive, even through the many “wait a minutes” and “hold on a secs”. To my mom, thank you for providing emotional and logistical supports when I needed them. To my husband Derek, thank you for your belief in me when I did not have energy or confidence left to believe in myself, and for being there whenever I was able to come up for air during the most difficult points along the way.

Finally, I would have never actually made it to this point without my dear friend and colleague, Jessica Simacek. Thank you for making me laugh when I wanted to cry---or simply sleep. Thank you for the many, many, many hilarious conversations and your solidarity throughout these last years. Thank you for drinking piña coladas and gallons of coffee, writing marathons, Pizza Luce, Hawaii, and the writing pants…

Dedication

This thesis is dedicated to my five incredible children, Richard Allen, Nicholas Blue, Jackson Alexander, Josiah Abraham, and Madison Amadea: may you enjoy a lifetime of following your dreams, knowing that I am always with you, cheering you on, just as you have done for me. I love you—you are my greatest joy and treasure.

Abstract

The ability to read fluently is associated with positive outcomes in school and adulthood. Low reading achievement is thus a critical issue—a remarkably pervasive problem among students in certain demographic groups, and one that persists in spite of an ever-expanding knowledge base of effective instructional approaches and interventions.

One possible factor may be that struggling readers are not doing or receiving "enough of what works"—specifically, that the dose of effective strategies has been insufficient to develop reading proficiency. Research supports that quantity and quality of practice are important to developing fluency in a practiced skill such as reading (Bryan & Harter, 1897; Chase & Simon, 1973; Ericsson, Krampe, Tesch-Romer, 1993; Geary, 1995; Williams & Hodges, 2005). Related to this notion, two additional factors, (1) time spent away from school (e.g., summer break) when students may not have access to literacy activities and (2) student responses and behaviors during reading instruction or intervention, may influence the development and maintenance of reading fluency

This study examined the effects on oral reading fluency of a repeated reading intervention implemented during a short (four-week) summer program with students whose reading was accurate but slow. Also examined was the degree to which student input variables related to treatment implementation (i.e., accuracy, minutes of intervention attended, number of 1 min readings completed, number of words read, and student engagement) predicted changes in oral reading fluency.

Participants included 79 students in second and third grades who were at or below the 50% percentile for reading rate according to grade level norms, but able to read passages with at least 93% accuracy. Students were randomly assigned to an intervention group that received core literacy instruction delivered by their summer school teacher and a supplementary repeated reading intervention implemented four times per week, or a control group that received core literacy instruction only.

Overall, the repeated reading intervention increased oral reading fluency more than core instruction alone. Post hoc analysis also indicated that the intervention was more effective for relatively high-level readers (26-50th percentile) than for low-level readers (0-25th percentile). Additionally, the cumulative number of words read correctly across all intervention sessions attended was the only significant predictor of posttest oral reading fluency.

Results of this study were contextualized within existing research on reading fluency intervention and treatment implementation. Implications for practice were discussed along with limitations of the study and directions for future research.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

INTRODUCTION

Low reading achievement has arguably reached epidemic proportions in our country. In 2011 alone, 66% of a representative sample of the 3.6+ million fourth grade students in the U.S. (NCES, 2014) scored below proficient levels in reading, with even higher percentages reported for certain demographic groups. For instance, 82% of students in the sample of below-proficient readers were eligible for free and reduced-price lunch. Similarly, disproportionate percentages of struggling readers who scored below proficient levels were children of color (e.g., Black: 84%, American Indian/Alaskan Native: 82%, Hispanic: 82%), compared to 57% of White children (NAEP, 2013). These statistics translate into millions of children who fail to read adequately and on time. Unfortunately, the inability to read proficiently by grade three has been associated with sobering outcomes such as failure to graduate on time (Hernandez, 2011), lower educational attainment (Rumberger & Lamb, 2003), increased unemployment (Bureau of Labor Statistics for 2012-13; Sum, Khatiwada, & McLaughlin, 2009), and higher rates of arrest or incarceration (Sum et al., 2009). Conversely, children who are successful readers early in their school careers are more likely to experience positive academic success, including high school and college graduation, and post-high school employment (Reder, 2010). This achievement gap in reading and its associated social concerns has been referred to as "the civil rights issue of our day" (Rodriguez, as cited by Marty, 2013).

The ability to read fluently is an essential aspect of proficient reading and one of the five core areas of literacy instruction (Kuhn & Stahl, 2003; NRP, NICHD, 2000; Samuels & Farstrup, 2006). Reading fluency is closely related to reading comprehension, which is important because the fundamental purpose of reading is to comprehend text (Pinnell et al., 1995). There is no agreed-upon definition of reading fluency in the research literature, however many agree that basic components include speed, accuracy, and expression (Kuhn & Stahl, 2003; NRP, NICHD, 2000; Samuels & Farstrup, 2006). Fluent reading has also been described as "intrinsically elegant in both form and cadence…we certainly know it when we see it, and are quick to celebrate it, along with the trajectory of success it portends." (Kame'enui & Simmons, 2001, p. 203). Given the connection between proficient reading and positive student outcomes in school and adulthood, it is essential to have instructional programs and interventions that are effective for developing and maintaining students' reading fluency.

Statement of the Problem

The reading achievement gap has prompted research that assists teachers in their efforts to ensure that all students develop proficient reading skills, including reading fluency, in a timely manner. The result is an expanding knowledge base (e.g., Kamil, Pearson, Moje, Afferbach, 2011) with multiple demonstrations of instructional approaches and interventions that have been effective for increasing the reading skills of many students, including students from lower-income backgrounds (Marcuso, & Rodman, 2011; Stevens et al., 2008; Yurick, Robinson, Cartledge, Lo, & Evans, 2006) and various racial and ethnic groups (Jeynes, 2008; Vadasy & Sanders, 2008, 2010; Ross & Begeny, 2011; Yurick et al., 2006). In spite of this knowledge, statistics on reading and post-secondary outcomes are grim evidence that the gap persists, disparities in reading proficiency remain, and we are left asking why some students benefit from instructional approaches and intervention while other students do not.

Addressing reading disparities is complicated. The multidimensional nature of the reading achievement gap suggests that numerous school-related factors contribute to its existence (Boyken & Noguera, 2011). For instance, ineffectual schools were historically thought to be a factor, although that belief has been brought into question by multiple research studies indicating that the instruction students receive during the regular school year can, on average, effectively decrease inequality in learning rates (see studies cited in Alexander, Entwistle, & Olson, 2001; Downey, von Hippel, & Broh, 2004; Entwisle & Alexander, 1992, 1994). Other factors related to the school environment include complex issues such as cultural incongruity between school and home (Banks, 2001), issues surrounding race (Ladson-Billings, 2006), and students' academic efficacy (Uwah, McMahon, & Furlow, 2008), to name a few. These issues are important to address because they represent the possibility that meaningful differences in students' school experiences may impact their reading achievement. In addition to the potential factors mentioned, another factor may warrant consideration: for some students, the dose of effective and relevant instructional programming may be insufficient to develop their reading proficiency. Put simply, some struggling readers may not be doing or receiving enough of what works.

Enough of What Works: A Different Approach to the Reading Achievement Gap

The concept that underlies "enough of what works" is simple: students must engage in/receive enough of an instructional program or intervention that is targeted to their needs in order to make sufficient gains to close their personal reading achievement gap. There is some support for this paradigm in research findings that confirm the relation between the quantity and quality of practice and resultant proficiency in a practiced skill (Bryan & Harter, 1897; Chase & Simon, 1973; Ericsson, Krampe, Tesch-Romer, 1993; Williams & Hodges, 2005). Specifically in the area of reading, there is research that substantiates wide differences in the amount of text students read both in and out of school (Allington, 1984; Heyns, 1987), and confirms that students' reading achievement is related to the quantity and quality of their reading (Shany & Biemiller, 1995; Topping, Samuels, & Paul, 2007). Research also suggests that practice may be particularly important for the acquisition and maintenance of procedural skills such as fluent reading (Geary, 1995). Related to this notion are two additional factors that potentially influence whether students do or receive enough of what works: time spent away from school (e.g., during the summer months) and dimensions of implementation.

Time spent away from school. Reading fluency can degrade over periods of non-use (Geary, 1995). Thus, when students do not engage in reading for extended periods of time—such as during the summer months—they miss out on necessary reading practice, which can contribute to decreased oral reading fluency. This phenomenon is commonly referred to as "summer slide" or summer learning loss (Heyns, 1987), but can happen during any significant school break. Other potential contributors to learning loss are access to books and engagement in literacy activities during school breaks, which may differ for students depending, in part, on their family income status (Allington et al., 2010; Heyns, 1987). Some of the most detrimental effects of learning loss are on the reading skills of students who come from low-income families (Alexander et al., 2001; Downey, et al., 2004; Hayes, & Grether, 1969, 1983; Heyns, 1987), many of whom are also students of color. Research has found that while the reading skills of students in higher-income families often increase over school breaks, the reading skills of students in lower-income families often stay the same or decrease, as evidenced by grade-level equivalent reading scores following summer break that are lower than corresponding previous spring scores, relative to national norms (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996; see also Alexander et al., 2001; Entwistle & Alexander, 1992, 1994, Heyns, 1987). The cumulative result of repeated regression in reading ability can be an achievement gap as wide as months or years for some students. Between first and sixth grade for example, the loss in reading development during the summer months alone can compound to as much as 1.5 years (Cooper et al., 1996). Instructional programs and intervention that take place over school breaks may help attenuate these effects.

Dimensions of implementation. Any instructional program or intervention that is implemented must first be matched to the instructional needs of the student. The instructional hierarchy framework by Haring and colleagues (1978) is a useful heuristic for matching components of interventions to the type of assistance that is most likely to benefit a given student. For instance, if a student is working to achieve accuracy, modeling and error correction may be most supportive, while practice opportunities may be most supportive for a student who is working to achieve fluency (Haring, Lovett, Eaton, & Hansen, 1978; Daly, Martens, Dool, & Hintze, 1998). Procedures must also be appropriately challenging. If the activity is too difficult or too easy, the student may be frustrated or bored and demonstrate off-task behavior (Treptow, Burns, & McComas, 2007). Implementing a generic instructional program or intervention that is not appropriate for a student may be ineffective and waste precious instructional time and resources.

After identifying an instructional approach or intervention that is properly matched to student needs, it may be important to examine *dimensions of implementation* that are potentially related to *doing enough of what works,* and may therefore impact a procedure's effectiveness. Common aspects of treatment delivery include the amount of intervention (i.e., minutes, sessions) or how closely an interventionist follows procedures.

Strong treatment delivery is critical for an instructional approach or intervention to work as intended, and also facilitates accurate evaluation of its effectiveness (Bellg et al., 2004; Lichstein, Riedel, & Grieve, 1994). Effectiveness has been defined as "the ability of an intervention to have the desired beneficial effect in actual use" (Dorland, p. 531 as cited in Saunders, 1994). In within-subject single case research designs (SCDs) where the student serves as his or her own control, effectiveness is demonstrated by replicated changes in level, trend, and/or variability in a response (dependent variable), relative to baseline or a comparison condition that coincides with the implementation of instruction or intervention (independent variable). In experimental group research, effectiveness is often demonstrated by a significant positive change in the mean level of a response for a treatment group, relative to a control or comparison group. However, effectiveness that has been demonstrated for one individual in an SCD study may not generalize to another individual. Similarly, a group mean that indicates effectiveness *on average* can (and often does) include scores which show that the intervention was successful for some students, and less so for others (Gall, Gall, & Borg, 1999; Gast & Ledford, 2010; Shadish, Cook, & Campbell, 2002). It is possible that dimensions of implementation may account for some of the differences in effectiveness across students. Understanding *how* these dimensions are related to differential intervention effectiveness for individual students may be critical to addressing the reading achievement gap.

Dimensions of implementation often associated with strong treatment delivery include (1) a high level of *procedural fidelity* (Noell & Gansle, 2014), commonly measured as accurate and consistent *adherence* to an established protocol (e.g., Dane & Schneider, 1998; Gresham, 2014; Sanetti, Gritter, & Dobey, 2011) and (2) a suitable level of *treatment* [intervention] *intensity*, commonly measured as the amount of a specified intervention topography, which is increasingly referred to as *dosage* (Dane & Schneider, 1998; Mellard, McKnight, & Jordan, 2010; Warren, Fey, & Yoder, 2007). Simply put, the important components of a given activity, sometimes referred to as *active ingredients*, must be delivered as specified, and the student must have access to enough of these active ingredients to make desired levels of progress.

Other dimensions of implementation that may be important to consider are the responses and behaviors of the student, also referred to as student inputs or *client acts* (Baker, 2012). Such responses and behaviors have been conceptualized as either *treatment receipt* or *treatment enactment*. Treatment receipt has to do with the student's level of understanding, knowledge, and ability to use the treatment. Treatment enactment has to do with the student's application of the treatment (Lichstein et al., 1994). Other conceptualizations of the construct include Jones and colleagues' *participant adherence* (essential components of the intervention implemented by the student, as planned) and *participant dosage received*, (Jones, Clarke, & Power, 2008), as well as Dane and Schneider's *participant responsiveness*, described as the student's level of enthusiasm or participation (Dane & Schneider, 1998). Gresham (2014) extended Dane and Schneider's conceptualization by adding that participant responsiveness also includes the student's level of engagement in the treatment and/or the degree to which he or she finds it relevant (Gresham, 2014). However, a limited number of studies have included, mentioned, or assessed student inputs as a dimension of implementation (see review by Barnett et al., 2014). Instead, the focus typically centers on dimensions that pertain to treatment delivery (e.g., Barnett et al., 2014; Warren et al., 2007). Therefore, there is a need to expand the study of treatment implementation to include student inputs, and the degree to which they moderate the effectiveness of instructional programs and interventions.

The underlying rationale for considering student inputs is that it is not uncommon for a student to behave in a manner that impacts his or her progress. In the course of reading fluency instruction or intervention, a student may complete all steps or skip steps, read quickly or slowly, perform flawlessly or make mistakes, or be on- or off-task. Such responses are distinct from the behavior of the interventionist. For example, even as an interventionist follows the steps of a protocol and directs a student to read a passage, the student may demonstrate task avoidance by leaving to use the restroom or reading the passage with low accuracy. In this way, students' inputs can influence treatment delivery and impact its effectiveness. It is thus possible that even under conditions where a teacher or interventionist consistently completes all the activity components for which he or she is responsible (high adherence), and the student is provided with a sufficient amount of access to the intervention's active ingredients (dosage), variance in student inputs may lead to differences in student progress.

In sum, the social significance of the reading achievement gap underscores the need for research to identify ways to reduce and permanently eliminate disparities in reading achievement for all students. It is critical that all students learn to read proficiently and in a timely manner. Given the vast numbers of students who currently lag behind their more skilled peers, there is a need to improve the effectiveness of instructional approaches and interventions so they are potent enough to *accelerate* progress and close reading achievement gaps for individual students, as well as provide a buffer for the learning loss that often occurs during school breaks (e.g., summer).

Purpose of the Study and Research Questions

The purpose of this study was to explore the notion of "enough of what works" by evaluating the effects of a reading fluency intervention implemented with second and third grade struggling readers during a school break, along with student input factors that predict and potentially moderate those effects. The study comprised two foci. One focus was to evaluate a well-known reading fluency practice intervention, *repeated reading* (RR) and its effects on a particular form of reading fluency, *oral reading fluency*, that can serve as an indicator of overall reading competence (Fuchs, Fuchs, Hosp, & Jenkins, 2001, p. 239). Repeated reading (RR) has been shown to be effective for increasing the oral reading fluency of students in various demographic groups (Therrien, 2004; NRP NICHD, 2000), and it was hypothesized that RR would attenuate reading loss by increasing and/or maintaining students' reading fluency over four weeks of a summer program. Additionally, there is little evidence to support RR's effectiveness during school break programming when intervention time may be limited because of short program duration, low attendance, or motivational issues. A second focus was to examine student inputs as a dimension of implementation related to treatment intensity and procedural fidelity. Specifically, it was posited that variance in responses and behavior during intervention could alter the quantity and quality of practice, and potentially contribute to differences in intervention effectiveness for individual students.

The following research questions guided the study:

1. What are the effects on oral reading fluency of a repeated reading (RR) intervention implemented during four weeks of summer school with students in 2nd and 3rd grade whose reading is accurate but slow?
2. What dimensions of RR implementation that include student inputs predict and/or correlate with posttest oral reading fluency scores, controlling for students' oral reading fluency at pretest?

I hypothesized that students receiving intervention would increase their oral reading fluency more than students in the control group. I also hypothesized that the best predictor of progress would be the amount of reading fluency practice students engaged in, as measured by the total number of words read across all the intervention sessions attended.

Delimitations

The following limitations were placed on the study:

1. Participants were limited to a convenience sample of second and third grade struggling readers (at or below the 50th percentile for ORF rate, according to grade-level norms) who had been invited to attend a summer school program, held in either of two participating schools within an urban school district in the Midwestern United States.
2. Interventions were implemented for the relatively short duration of the summer program (four weeks).
3. The intervention targeted oral reading fluency only.
4. Reading outcome measures were limited to ORF pre- and posttests.

Organization of the Dissertation

This dissertation is comprised of four additional chapters. Chapter 2 provides a review of peer-review published literature in the areas of (a) oral reading fluency, (b) repeated reading, and (c) dimensions of treatment implementation. Chapter 3 describes the methodology used in the current study and (a) characteristics of the study participants, interventionists, and setting, (b) measures utilized to for screening and to assess ORF, (c) dimensions of implementation potentially related to posttest ORF along with measurement procedures, (d) intervention procedures, and (e) experimental design and data analysis. Chapter 4 reports the results for each research question, including two post hoc research questions that developed in the course of data analysis. Chapter 5 discusses the results of the study in the context of existing research and conceptualizations of treatment implementation. Limitations of the study are then identified, followed by proposed implications for practice and directions for future research. The chapter then closes with a final conclusion.

CHAPTER TWO

LITERATURE REVIEW

Organization of the Chapter

This chapter reviews existing peer-review published literature in the areas of (a) oral reading fluency (ORF), (b) repeated reading, and (c) dimensions of treatment implementation. First, definitions and views on the construct of reading fluency are reviewed, along with its importance to reading development. Second, literature is presented pertaining to the reading fluency intervention, repeated reading (RR). Third, a discussion of intervention intensity and procedural fidelity ensues in order to provide a framework of theoretical support for the examination of dimensions of repeated reading implementation that may be critical to the effectiveness of the intervention in general, and with individual students. A particular focus will be on the influence of variables that comprise responses of the students themselves. The chapter will conclude with a summary, followed by a review of the study purpose and research questions.

Reading Fluency

Reading fluency has been defined by many as the ability to read "quickly, accurately, and with proper expression" (NRP, NICHD, 2000, Ch. 3, p. 6), and is widely accepted as fundamental to reading success (Kuhn & Stahl, 2003; NRP, NICHD, 2000; Samuels & Farstrup, 2006). The speed and accuracy aspects of this definition are grounded in LaBerge and Samuels' theory of automaticity, which is the process by which an individual first perceives, and then makes meaning of, written word stimuli through a series of information processing stages (Laberge & Samuels, 1974). This process occurs almost instantaneously for fluent readers; words are quickly recognized or decoded and their meanings interpreted in a manner that is highly accurate and seemingly effortless. As posited in the theory and supported by other research on the topic (e.g., Perfetti, 1985; Stanovich, 1980), information processing that reaches a high level of automaticity reduces attentional demands at the visual or phonological levels so it can remain focused at the semantic level, dedicated to making meaning (Laberge & Samuels, 1974). Put simply, the less attention a reader has to expend on the decoding task, the more attention he or she has available for comprehension. The argument continues that it is therefore critical to develop decoding ability to the point of instantaneous word recognition. Only then can the vast majority of the reader's attention be focused on understanding the text.

The automaticity theory has additional support in Stanovich's interactive-compensatory model (1980) where he posits that when readers cannot recognize words automatically, they utilize information from multiple contextual sources (orthographic, phonological, semantic, syntactic) in an attempt to construct meaning. Like LaBerge and Samuels, Stanovich would argue that as word recognition skills become more automatic, additional cognitive resources are then available for comprehension. Additionally, there is support for the assertions of the three authors in the form of multiple studies that have reported strong positive relations between reading fluency and reading comprehension, and that growth in reading comprehension corresponded with growth in oral reading fluency (O'Connor, White, & Swanson, 2007; Pinnell et al., 1995, Therrien, 2004).

In spite of its elegant simplicity, the automaticity model does not represent other complex semantic processes that might factor into fluent reading (Logan, 1977).

Alternate views have considered reading fluency to be a predominantly oral reading task that involves parsing words into meaningful sequences (Schreiber, 1980) and reading with prosody (Schwanenflugel, Hamilton, Wisenbaker, & Stahl, 2004). Others have considered reading fluency to be the understanding and comprehension that results *from* reading with speed, accuracy, and expression (Thurlow & van den Broek, 1997). These views represent additional aspects of fluency that should certainly be examined further. Nevertheless, the literature currently provides sufficient evidence to support a definition of reading fluency that includes speed, accuracy, and prosody—with the understanding that the fundamental purpose of reading is to comprehend text. All subsequent references to reading fluency in this study will assume this conceptualization unless otherwise noted.

Reading fluency is important to reading development for multiple reasons. One is the possible influence of the "Matthew Effects" phenomenon, also referred to as "rich-getting-richer" effects. The term is a reference to the biblical verse, "For unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath" (Matthew, XXV:29, Stanovich, 1986, p. 381). As the reasoning goes, good readers read more and become even better readers while poor readers read less and fall further behind (Stanovich, 1986). The Matthew Effects phenomenon aligns with behavioral principles. Good readers read easily which can make the task pleasant, even reinforcing, so that good readers often engage in *more* reading, and become even *better* readers. In contrast, poor readers struggle, which can make reading increasingly aversive and lead to task avoidant behavior that enables the poor reader to escape the unpleasant task (Burke, Hagan-Burke, & Sugai, 2003; McComas, Hoch, Paone, & El-Roy, 2000; Stanovich, 1986). For example, students may engage in behavior that results in delays or avoidance of reading, such as frequent trips to the bathroom during independent reading, paging through a book but not actually reading it, negotiating or arguing with the adult in charge to reduce the amount of reading, or taking a long time to choose a book. Students may also engage in more challenging behavior such as aggression or property destruction. Another reality to consider alongside task avoidance is that some students simply read more slowly than others so that within a given amount of time (e.g., independent reading), a student with a slower reading rate is unable to read the same amount as a student who is able to read more quickly, even if he or she is on task the whole time. Students who read more slowly may thus engage in *less* reading. Over time, such students may perform at increasingly lower levels relative to proficient peers. Beyond its common sense appeal, there is also some research support for Matthew Effects. For instance, results of a study conducted by Allington (1984) indicated that there were significant discrepancies in the amount of reading that was done by good and poor readers. In a week's time, poor readers in the first grade read an average of 386 words (range = 16-739 words) compared to good readers who read an average of 1,121 words (range = 181-1933). A similar pattern was also found for students in third and fifth grades (Allington, 1984).

Reading fluency is also important because of its relation to comprehension. Literacy instruction in the earlier grades focuses on *learning to read*, but that focus shifts to *reading to learn* in later grades (Durkin, 1978). Students who struggle to read fluently grade-level texts need to devote a great deal of their attention to figuring out the words (e.g., decoding, guessing based on contextual clues). Under these conditions, less attention is left for processing the meaning of the text, which could compromise their comprehension (LaBerge & Samuels, 1974). In this way, low reading fluency may initially inhibit only a student's progress in reading, but eventually inhibit progress in other subjects as well—those that rely on the ability to read and process written text, such as science, history, or literature studies.

Struggling readers are not likely to spontaneously improve their skills without intervention (Torgesen, 1998), thus the importance of building reading fluency early in a child's school career is paramount. Fortunately, reading fluency is an ability that can be developed through instruction and improved with practice. This contention is supported by the results of a large body of research on reading practice procedures that have improved the reading fluency of many students with different characteristics and abilities. Multiple studies have shown positive effects for students in general education, students with disabilities, students learning English as a second language, students from low-income families, and students of various race and ethnicities (O'Connor et al., 2007; Pinnell et al., 1995; Samuels, 1979; Therrien, 2004).

In sum, reading fluency is the ability to read quickly, accurately, and with prosody. It is important for students to develop reading fluency in a timely manner to increase the likelihood that reading becomes a pleasant—even reinforcing—activity that a student engages in willingly and often. Furthermore, fluent reading is important for comprehension, which is the fundamental purpose of reading. There is currently a number of reading practice procedures that research has shown to be effective for improving reading fluency. Repeated reading is one such well-known example.

Repeated Reading

It has been said, "the most effective device that can be applied to learning is to increase the amount of drill or practice" (Symonds & Chase, 1992, p. 289). A common reading procedure that provides practice opportunities is called *repeated reading* (RR), first introduced in the late 1970's (Samuels, 1979). RR is a fluency-oriented approach that may work primarily because it increases the amount of reading that students do. RR provides students with a structured opportunity to engage in relatively high amounts of reading practice via repeated reading of a segment of text. Such practice is likely RR's active ingredient, responsible for increasing students' reading fluency.

In one of the simplest forms of RR, a student repeatedly reads a short passage (fiction or non-fiction) until a criterion ORF level is achieved, typically measured as the number of words read correctly per minute (WRCM). After reaching the criterion ORF level, the student is provided with another passage and the procedure is repeated. In early trials of RR, students with intellectual disabilities were pre-tested on a passage and then read the passage repeatedly on their own (ostensibly, silently) and without error correction in a process that was termed *unassisted repeated reading* (Meyer & Felton, 1999). When students were re-tested on the same passage, the results showed that reading speed and accuracy had increased on the practice passage. Other results included that students increased their speed on the first reading of a novel passage, and decreased the number of readings necessary to reach criterion on a passage. These findings were promising because they suggested that in addition to becoming more proficient with practice text, students were generalizing the training to new, unpracticed text (Samuels, 1979).

In early RR studies where students read alone and often silently (unassisted RR), reading behavior was not verified through direct observation (e.g., Samuels, 1979), which threatened the internal validity of those studies, and inhibited the ability to establish a functional relation between RR and student gains in reading. Later RR protocols began to require students to read aloud, a change that helped resolve the internal validity threat previously mentioned, and likely enhanced the intervention's effectiveness as well. First, reading aloud may encourage higher quality engagement because it makes reading visible. Protocols that require oral reading enable teachers to verify that students are on-task, and to document the amount and accuracy of reading, so they can then take necessary action—such as making adjustments to passage difficulty, or providing prompts, praise, or corrective feedback. Second, reading aloud provides a more salient model than silent reading. A student can see his or her teacher, tutor, or other students reading aloud; he or she can observe fluent and accurate reading and on-task behavior. Studies have shown that such observation of a verbal model can increase modeled behaviors in observing students, even under conditions where the students' behavior was unreinforced (Brody, Lahey, & Combs, 1978; Lahey, 1971).

Protocols that include oral reading have taken a number of forms. In contrast to unassisted RR, another early example was *assisted* RR, where a student read a passage orally with a fluent reader who served as a model (Eldredge, 1990; Young, Bowers, &

MacKinnon, 1996). Another example was *prosodic* RR where the student's focus was directed to the syntactic or rhythmic cues in the passage by listening to or reading along with a fluent reader (Schreiber, 1980). In still another example, students were directed to read the passage aloud while a tutor listened and marked the number of WRCM and errors (Sindelar, Monda, & OShea, 1990). At present, repeated oral reading of practice passages to another individual (e.g., teacher, paraprofessional, or peer) is often part of RR protocols in the literature (Therrien, 2004).

Research has been conducted to determine the effectiveness of RR for developing the reading fluency of students at various levels of reading ability, with and without disabilities (Therrien, 2004). Less is known about the conditions under which RR is an appropriate intervention to select, although matching an intervention to a student's level of need can have important implications for its effectiveness. Using the instructional hierarchy as a guide (Haring et al., 1978), RR may be a good choice for students whose reading is accurate but slow. The heavy dose of practice opportunities is supportive when the primary intervention focus is to improve reading speed. Meanwhile, readers who are inaccurate may benefit from RR-like protocols that include re-reading of text, but also immediate error correction and/or modeling. Other times, a different intervention altogether may be the most appropriate choice (e.g., phonics, if there is a decoding deficit).

The effectiveness of RR is supported by the results of multiple studies compiled in several oft-cited reviews. One review by Meyer & Felton (1999, cited over 390 times) summarized the theoretical foundations and history of the development of fluency training, and reported narratively the results of numerous studies. In 2000, the National Reading Panel reviewed strategies for increasing students' reading fluency and concluded that there was support for RR as an effective fluency intervention. Across all experimental studies cited, the review reported an average moderate effect size for RR of $d = 0.41$ for overall reading achievement (NRP, NICHD, 2000). A meta-analysis by Therrien (2004, cited 450 times) reported mean fluency effect sizes for RR of $ES = 0.76$, $SE = .06$ (students without disabilities) and $ES = 0.77$, $SE = .09$ (students with learning disabilities). Effect sizes for RR for both transfer and non-transfer[1] measures were also calculated as well as for conditions under which different intervention components were implemented.

It is notable that in contrast to the aforementioned results, more recent reviews have called into question the historical support for RR, perhaps prompted by the introduction of quality indicators for educational research (Gersten, et al., 2005; Horner et al., 2005; Kratochwill et al., 2010). A number of researchers have since reevaluated some of the RR research according to these more rigorous criteria. First, Chard and colleagues (2009) conducted a review of RR wherein they applied quality indicators proposed for group research (Gersten et al., 2005) and single case research (Horner et al., 2005) to studies that had been conducted with students with disabilities between January 1975 and December 2006. Results indicated that there was insufficient evidence to declare RR an evidence-based practice for this population (Chard, Ketterlin-Geller, Baker, Doabler, &

---

[1] *Transfer* refers to the transfer of thinking and learning to a dissimilar context (Barnett & Ceci, 2002). A transfer passage (sometimes referred to as *far transfer*) is one that is novel to the student; a *non*-transfer passage is one that the student has read previously.

Apichatbutra, 2009). Second, O'Keeffe and colleagues (2012) evaluated the research-base on RR by first using the What Works Clearinghouse system to screen the cohort of studies featured in six major RR reviews (O'Keeffe, Slocum, Burlingame, Snyder, & Bundock, 2014; WWC, 2008, table 8; note that the system was slightly modified for single case research). To the studies that met screening criteria, the authors then applied four review systems: two systems proposed for group research (Gersten et al., 2005; WWC, 2008) and two systems proposed for single case research (Horner et al., 2005; Kratochwill et al., 2010). The review systems yielded *zero* single case studies and just four group studies that met review criteria and could be evaluated. Similar to the conclusions drawn by Chard (2009), results indicated that there was too little support in the research literature to declare RR an evidence-based practice.

In sum, the research support for and against RR as an evidence-based practice depends on the lens through which it is viewed. On one hand, if quality indicators are not stringently applied to the research, there is a body of studies that has long been accepted as evidence for the effectiveness of RR (e.g., Meyer & Felton, 1999; NRP, NICHD, 2000; Therrien, 2004). On the other hand, if quality indicators are stringently applied to the research, the quantity of studies that provide evidence of RR's effectiveness is insufficient to declare it an evidence-based practice. It is encouraging that four group studies cited in the review by O'Keeffe (2012) met WWC criteria (Conte & Humphreys, 1989; Eldredge, Reutzel, Hollingsworth, 1996; Homan, Klesius, & Hite, 1993; Young, et al., 1996) and four additional group studies identified in the Chard et al. (2009) review were considered either high or acceptable quality (Mathes & Fuchs, 1993; O'Shea, Sindelar, & O'Shea, 1987; Rashotte & Torgesen, 1985; Sindelar et al., 1990). Furthermore, all the indicated studies reported positive results for the effects of RR. Thus, although current standards of rigor yield too few studies to unequivocally support RR as evidence-based, I posit that the research on the effectiveness of practice for increasing the fluency of basic skills coupled with the results of studies contained within the reviews conducted by O'Keeffe, Chard, and their colleagues comprise sufficient evidence to substantiate the implementation of RR as a reading fluency strategy that is likely to be effective if matched properly to student need and implemented with procedural integrity. Meanwhile, more high quality research needs to be conducted to build the evidence base for RR.

Intervention Components.

Since RR was first described as a procedure in 1979, researchers have experimented by varying its components and measuring student outcomes to determine which procedures are best for different contexts and populations of students. Some manipulations have included (1) utilizing different interventionists (e.g., adults, peers), (2) altering the number of passage readings, (3) including a performance criterion (4) providing cues to focus on an aspect of fluency (e.g., speed, comprehension), (5) modeling fluent reading, (6) providing corrective feedback, or (7) requiring students to graph their progress, (8) including comprehension questions, (9) providing incentives for progress, and (10) contrasting repeated readings with non-repetitive reading.

This section will review RR research conducted between 1977 and 2015. Included are results reported in a meta-analysis of RR conducted by Therrien and published in

2004. The meta-analysis improved on the National Reading Panel's review of RR literature (NRP, NICHD, 2000) in that it focused exclusively on RR separate from other fluency strategies. There were a total of 18 articles that met the inclusion criteria for Therrien's meta-analysis, including studies that: (1) were published after a chapter on RR written by Dahl (1977) and before June, 2001, (2) were experimental and quantitative, (3) included school-age participants, and (4) provided enough information to calculate standard mean gain effect sizes (Becker, 1988; see Therrien, 2004 for search terms and more specific inclusion criteria).

Results reported in the meta-analysis will be extended by a review of RR literature published after June, 2001 and before January, 2015, the cut-off year specified in the RR meta-analysis (Therrien, 2004). A search was conducted within the database Academic Search Premier (via EBSCOhost) using the search term *repeated reading*. Articles were restricted to those that (1) included *repeated reading* in the abstract and/or title, (2) were published in peer-refereed journals, (3) were experimental and quantitative, and (4) were written in English. Articles must have also reported on studies that (5) were conducted with students in elementary school in the United States, (6) occurred during the school year, (7) implemented RR specifically, or provided enough information to verify that passages were read at least two times, (8) implemented the specified intervention in a school or home setting, and (9) evaluated effects of RR on a measure of ORF. This search produced a total of 16 articles that met the inclusion criteria.

Study findings will be compiled so as to present the evidence for each of the RR components that have been listed. Results from the meta-analysis will be reported in terms of mean fluency effect size for specific components, and results from the present literature review will be reported as they provide relevant examples or evidence related to specific intervention components. A primary focus will be effects on ORF for transfer as opposed to non-transfer measures/passages. The rationale for this focus is that from a pragmatic perspective, the most meaningful effect may be increases in ORF on passages the student has not yet encountered. Although RR is almost always successful for increasing ORF on non-transfer measures, the goal of fluency practice is to help students increase their ability to read *any* text passage they encounter with speed, accuracy, and prosody, not just text passages that have been read repeatedly.

Interventionists. RR interventionists can vary. Students can implement RR with peers, or an adult tutor can implement RR with a student one-to-one, in dyads, or in groups. Studies reviewed in the meta-analysis included interventions that were implemented by either adults or peers. The effect on ORF (mean fluency effect size) for transfer measures was much larger for interventions that were implemented by an adult (*ES* of 1.37, *SE* = .177) compared to those implemented by a peer (*ES* of .36, *SE* = .062; Therrien, 2004).

RR with peers. In the present review, three studies included peer-implemented interventions. Results were similar to those reported in the meta-analysis in that students made gains, although there were no studies that directly compared the effects of peer-implemented to those of adult-implemented interventions. In the studies reviewed, students took turns repeatedly reading paragraphs of text. Oddo and colleagues (2010) conducted a study with sixteen students in fourth grade, for eight weeks (Oddo, Barnett, Hawkins, & Musti-Rao, 2010). Peer-implemented RR occurred in small teacher-supervised groups, three days per week. A multiple baseline design (Baer, Wolf & Risley, 1968) was used to evaluate effects. Results of visual analysis indicated that relative to baseline, level of ORF on non-transfer measures increased for all students (Oddo et al., 2010). Two other studies by Yurick and colleagues (2006) were conducted with pairs of students in third grade (study 1) and fifth grade (study 2). Peer-implemented RR occurred over 11-27 sessions. A multiple baseline design (Baer et al., 1968) was used to evaluate effects. Results of visual analysis indicated that relative to baseline trends (which were relatively flat), the trend of ORF on non-transfer measures increased after RR was implemented. Additionally, the level of ORF on the first reading of a new passage (which can be viewed as a measure of transfer) was often a higher than the baseline mean (Yurick et al, 2006).

Adults with student dyads. In the present review, there were two studies that examined the results of adult-implemented RR interventions with student dyads (Vadasy & Sanders, 2008; 2009). The first study (2008) included a sample of 119 fourth and fifth grade students randomly assigned to a treatment or control group. RR interventions were implemented by paraprofessionals with the treatment group for 15 min daily, for eighteen weeks. Results indicated no effects on ORF on transfer measures for the treatment group, although the treatment group outperformed the control group on measures of vocabulary, word comprehension, and passage comprehension. One possible reason for the lack of effects on ORF may be that students in the treatment group scored one standard deviation below the population mean in reading accuracy, which suggests that RR may not have been an appropriate intervention (Vadasy & Sanders, 2008). The second study (2009) included a sample of 202 students in second and third grade randomly assigned to one of three conditions: paraprofessional-implemented RR, teacher-implemented RR, or a control group. Results indicated that the effect on ORF on transfer measures was slightly higher for teachers (*ES* = 38) than paraprofessionals (*ES* = 37), but suggest that paraprofessionals can achieve gains that are comparable to teachers (Vadasy & Sanders, 2009).

Family members as interventionists. There is a growing literature base to support family member implementation of RR. In the present review, there were two studies where a parent or other family member served as the interventionist (Kupzyk, Daly, & Andersen, 2012; Kupzyk, McCurdy, Hofstader, & Berger, 2011). Multiple baseline designs (Baer et al., 1968) were used to evaluate effects and found that students increased their ORF level on transfer measures relative to baseline after RR was implemented by a family member. Further, the results of one study (Kupzyk et al., 2011) found that students maintained the effects of the intervention at six- and nine-weeks.

Interventionists: Conclusions from the research. Study results contained within the literature on interventionists for RR indicate that a variety of individuals can successfully serve in this role. Although the largest effect sizes are for teachers, other adults are effective as interventionists, including family members. An important finding is that peer-implemented interventions lead to increases in ORF while perhaps making more efficient use of resources than adult-implemented one-to-one, dyad, or small group interventions. Meanwhile, students learning in any of these conditions still have access to high opportunities to respond. However, effect sizes for peer-implemented interventions were smaller than for adult-implemented, which should be taken into account when considering efficiency

Number of Passage Readings. Some RR protocols require a fixed number of passage readings. Studies reviewed in the meta-analysis required students to read passages a fixed number of times [2-4 times (non-transfer) or 2-3 times (transfer)] and effects on ORF fluency were calculated separately for each outcome. There were moderate to large overall effects on ORF (mean fluency effect size) for transfer measures (*ES* of .38, *SE* = .061), and non-transfer measures (*ES* of .81, *SE* = .066). Effect sizes differed depending on the number of times a passage was read. For transfer measures, the mean fluency effect size was higher for three readings (*ES* = .42, *SE* = .091) than two (*ES* = .37, *SE* = .087). For non-transfer measures, the effect size was higher for four readings (*ES* = .95, *SE* = .145) than two (*ES* = .57, *SE* = .141) or three (*ES* = .85, *SE* = .088) readings (Therrien, 2004).

In the present review, thirteen studies specified that students were required to read passages a fixed number of times ranging from two to six, with four passage readings being the most common. Study results contained within the literature on number of passage readings indicate that repetition, regardless of whether students read passages few or many times, generally led to increased performance on non-transfer measures. Effects on transfer measures were also generally positive in that treatment groups outperformed control groups (e.g., O'Connor et al., 2007; Swanson & O'Connor, 2009; Vadasy & Sanders, 2008) and individual students increased ORF performance relative to baseline (e.g., Ardoin, Williams, Klubnick, & McCall, 2009; Kupzyk et al., 2012; Kupzyk, et al., 2011; Lo, Cooke, & Starling, 2011). Just one study directly examined the effects of two different fixed numbers of passage readings (Ardoin et al., 2009). Using a rapid reversal design, four students in second, fourth, and fifth grades alternated between three and six readings within a RR intervention. Both conditions led to substantial increases in ORF on non-transfer and transfer passages relative to baseline. However, ORF increases did not differ significantly by condition, indicating no enhanced effect for doubling the amount of passage reading (Ardoin et al., 2009).

There were also occasional mixed results for studies that utilized a fixed number of passage readings. For instance, Coleman and Heller (2010) conducted a study with four students in third, fourth, and fifth grade with intellectual and physical disabilities, for approximately three to four weeks. Researchers implemented RR one-to-one until students reached a terminal criterion. The intervention included five readings, two of which involved listening to a computer model, combined with error correction. Effects on ORF were positive on non-transfer measures, but mixed on transfer measures. One reason could be that students with higher-level needs may have to develop certain sub-skill fluency before ORF gains on non-transfer measures can generalize to a transfer measure. The frequency and/or duration of the intervention may also have been inadequate to promote generalization of reading fluency (Coleman & Heller, 2010). In another example, Musti-Rao, Hawkins, and Barkley (2009) conducted a study with a twelve students in fourth grade. Students engaged in peer-implemented RR for three 10 min sessions per week, reading each section of text two times. Although effects on ORF were positive on non-transfer measures, they did not generalize to transfer measures. One reason may be the low dosage or duration of the intervention (as little as six weeks), which might have been insufficient for some students (Musti-Rao et al., 2009). Additionally, neither study provided detailed information on the number of errors students made during RR. It is possible that inaccurate reading practice may have inhibited ORF progress or increase off-task behavior, thereby decreasing the amount of reading students did per session (for a discussion on reading at the frustration, instructional, and independent reading levels, see Treptow et al., 2007).

Performance criteria. Some RR protocols include a specified performance criterion. A number of the studies reviewed in the meta-analysis required students to read to a performance criterion before moving on to a new passage. There was a large effect on ORF (mean fluency effect size) on transfer measures (*ES* of 1.70, *SE* = .188)(Therrien, 2004).

In the present review, four studies required that students read to a performance criterion (Chafouleas, Martens, Dobson, Weinstein, & Gardner, 2004; Coleman & Heller, 2010; Yurick et al., 2006, Studies 1 and 2). No studies evaluated the effects of performance criteria directly. Study results contained within the literature that included performance criterion as part of RR indicate that it can be helpful for increasing ORF on non-transfer and transfer measures. In spite of some mixed results on transfer measures in studies conducted by Coleman & Heller, 2010 and Chafouleas et al., 2004, the practice appears to be generally effective.

Cuing. Some RR protocols include cueing to focus students' attention on some aspect of fluent reading. A number of studies reviewed in the meta-analysis included cueing procedures wherein students were told to focus on speed or comprehension prior to reading. There were moderate to large effects on ORF (mean fluency effect size) on non-transfer measures when students were cued for speed (*ES* of .72, *SE* = .185), comprehension (*ES* of .81, *SE* = .096), or both (*ES* of .94, *SE* = .135). Effects on comprehension were observed as well, but tended to be slightly smaller than those found for ORF (*ES* = .66, .75, and .67, respectively; Therrien, 2004).

In the present review, no studies evaluated the effects of cueing directly or contained cuing as part of RR procedures. Study results contained within the literature on cueing indicate that cueing students to focus on both speed and comprehension have a large to moderate effect on ORF on non-transfer measures, but additional research is needed to evaluate effects on transfer measures.

Modeling. Some RR protocols include a model of fluent reading for the student. A number of studies reviewed in the meta-analysis included this practice, and in all cases a peer modeled fluent reading of the passage before having the tutee read it. Effects on ORF (mean fluency effect size) on transfer measures were larger for interventions that included modeling (*ES* of .40, *SE* = .077) than when no modeling occurred (*ES* of .30, *SE* = .104; Therrien, 2004).

In the present review, there were five studies that incorporated some kind of modeling (Coleman & Heller, 2010; Kupzyk et al., 2012; Kupzyk, et al., 2011; Vadasy & Sanders, 2008, 2009). In all cases, teachers or peers modeled fluent reading prior to having the tutee read, with the exception of one study (Coleman & Heller, 2010) that included computer modeling. No studies evaluated the effect of modeling directly. Study results contained within the literature on modeling indicate a larger effect size for modeling versus no modeling.

Corrective feedback. Some RR protocols include the provision of corrective feedback. Corrective feedback was part of the intervention represented by fifteen of the fluency effect sizes calculated for the meta-analysis (3 non-transfer, 12 transfer). Topographies of corrective feedback included correcting mispronunciations, providing the correct sound/word, or prompting a student to re-read/sound out the word. Effects on ORF (mean fluency effect size) on non-transfer measures were lower when students received corrective feedback (*ES* of .68, *SE* = .119) than when they did not (*ES* of .88, *SE* = .075). In contrast, effects on ORF (transfer measures) were higher for students who received corrective feedback (*ES* of .51, *SE* = .06) than when they did not (*ES* of .46, *SE* = .227; Therrien, 2004).

In the present review, no studies evaluated the role of corrective feedback directly, but the majority of the studies included it as part of the RR protocol. Study results contained within the literature on corrective feedback are mixed. Results reported in the meta-analysis indicate that the effects on ORF were higher on non-transfer measures when corrective feedback was not included. One hypothesis is that corrective feedback focused students' attention on reading accurately rather than quickly, which may have slowed down their reading speed. Conversely, the effect on ORF on transfer measures was *larger* when students received corrective feedback. Perhaps providing corrective feedback encourages more accurate fluency practice that inhibits oral reading speed in the short term (as evidenced by lower ORF scores on non-transfer measures) but over time positively impacts ORF on transfer measures. Overall, there is strong support for the use of error correction procedures as part of oral reading intervention (Huebusch & Lloyd, 1998). Additionally, providing error correction makes sense because it can reinforce accurate reading, and reduce the possibility that students will develop consistent inaccurate responses to stimuli.

Graphing progress. Some RR protocols require students to graph their progress (e.g., WRCM or errors). A number of studies reviewed in the meta-analysis included interventions where students graphed their progress. Overall, the effect on ORF (mean fluency effect size) on transfer measures was greater for graphing progress (*ES* of .57, *SE* = .075) than for not graphing progress (*ES* of .40, *SE* = .091). Furthermore, there were large effects for graphing when adults served as interventionists (*ES* of 1.58, *SE* = .208; Therrien, 2004).

In the present review, no studies evaluated the effects of graphing directly. Study results contained within the literature on graphing progress include that the procedure may be effective for increasing students' performance, especially when an adult implements RR.

Comprehension questions. Some RR intervention protocols require that students answer comprehension questions about the passage(s) they read. A number of studies reviewed in the meta-analysis included this component in their protocols. The effect on ORF (mean fluency effect size) on transfer measures was larger for interventions with comprehension questions (*ES* of .39, *SE* = .084) than those without (*ES* = .33, *SE* = .091;

Therrien, 2004).

Study results contained within the literature on the use of comprehension questions as part of a RR protocol indicate that asking comprehension questions has a moderate effect on ORF, although the effect size was not much larger than when no questions were asked. Nevertheless, asking comprehension questions makes sense for reasons that have to do with social validity. Processing the meaning and content of text is the fundamental purpose of reading, lending support for including comprehension questions as part of RR interventions.

Incentives. The use of incentives as part of RR was not evaluated in Therrien's meta-analysis. In the present review, there was one study that examined the use of rewards. Chafouleas and colleagues (2004) evaluated the effects of combining contingent rewards with RR for three students in second grade, one who was diagnosed with a learning disability. Results indicated that contingent rewards did not significantly enhance students' ORF performance on transfer measures (Chafouleas et al., 2004).

Study results contained within the literature on the use of incentives as part of RR are mixed in that some students appeared to benefit from the practice while others did not. Contingent rewards may therefore be useful for some students and contexts.

Repeated versus non-repetitive reading. More recently, some researchers have questioned whether it is necessary to read a passage repeatedly in order to practice reading fluency. Three studies were located that compared the effects of RR with *non-repetitive* reading, which is reading continuously from a text instead of going back and practicing the text via re-reading (O'Connor et al., 2007; Swanson & O'Connor, 2009;

Therrien, Kirk, & Woods-Groves, 2012). Interventions in all three studies were conducted one-to-one by an adult tutor. In all studies, students in each condition read aloud for the same amount of time.

In the first study, a sample group of 37 students in second and fourth grade were randomly assigned to a RR, non-repetitive, or control condition (O'Connor et al., 2007). In the second study, a sample group of 155 students in second and fourth grade were randomly assigned to a RR, non-repetitive, or control condition (Swanson & O'Connor, 2009). Both studies had similar results in that RR and non-repetitive conditions resulted in similar ORF gains on transfer measures. Results from the 2009 study also seemed to indicate that non-repetitive reading was less helpful for comprehension for students who had low working memory and/or word recognition skills, although it could be argued that students with low word-recognition skills should receive intervention to help increase that sub-skill rather than concentrating on fluency (Swanson & O'Connor, 2009). In a third study, a sample of 30 students in third, fourth, and fifth grades were randomly assigned to either a RR or non-repetitive condition (Therrien et al., 2012). Results showed that students in the non-repetitive condition made almost twice the ORF gain on transfer measures as students in the RR condition in terms of words read correctly per minute (*ES* = .64), however the difference in ORF was non-significant. It is uncertain why this result was obtained. It is possible that the study was underpowered due to a small *n* (*n* =15 in each group; Therrien et al., 2012).

Study results contained within the literature on non-repetitive reading are inconclusive but interesting in that they suggest an alternate view that the *quantity* of text reading, not *repeated* text reading contributes to increased ORF, especially on transfer measures (see Allington, 2001; Kuhn, 2004). Differences between the conditions were not statistically significant in any study; still, the results of the study by Therrien and colleagues (2012) were significant from a practical perspective in terms of mean growth for the two groups, which seems to warrant further investigation. Specifically, students in the non-repetitive condition gained an average of 26.89 WRCM after 50 15 min sessions compared to an average of 15.73 WRCM for students in the RR condition. A limitation of the study is its small sample size, therefore future research might attempt to replicate the findings with a larger sample of students who demonstrate a range of fluency abilities at pretest (Therrien et al., 2012).

Implemented During School Breaks.

Regression in reading ability over the school breaks (e.g., summer) can lead to gaps in reading achievement as wide as months or years. This regression is especially evident for reading fluency, a procedural skill that is susceptible to decay during long periods of time without practice (Geary, 1995). One possible way to moderate this effect is to implement interventions such as RR during school breaks in order to provide students with the opportunity to practice their reading fluency skills. However, it is possible that RR might be differentially effective when it is implemented during such programs as opposed to the regular school year. For example, programs conducted during breaks are sometimes taught by teachers who may know very little about individual students, which could make it challenging to target specific needs. Other aspects of such programs could impact student engagement or dosage by serving as antecedents for challenging or off-task behavior. For instance, summer programs can be short in duration and low attendance is often an issue. Fun activities might compete with intervention schedules and make it challenging to maintain consistency. Students may feel a sense of burnout after a long year of school which can make reading fluency practice seem like drudgery, especially when compared to fieldtrips to fun destinations like the swimming pool or zoo. This section will review the research literature on RR implemented during summer months.

A search was conducted within the databases Academic Search Premier (via EBSCOhost), ERIC (via EBSCOhost), and ERIC (via CSA) using the search terms *repeated reading* and *summer*. Articles were restricted to those that (1) included *repeated reading* anywhere in the text and *summer*, *vacation*, or *school break* in the abstract, (2) were published in peer-refereed journals, (3) were experimental and quantitative, and (4) were written in English. The articles also needed to report on studies that (5) were conducted with students in kindergarten through third grade in the United States, (6) implemented RR specifically, or provided enough information to verify that passages were read more than 1x per session, (7) evaluated effects of RR on a measure of ORF, (8) occurred in a school setting, and (9) were implemented during the summer months. This search produced a total of three articles that met the inclusion criteria.

Savaiano & Hatton (2013) conducted a study with three students with visual impairments in third through sixth grade. Two of the participants wore glasses; the third participant did not require any optical devices to read. All three were able to read regular or large-sized print. RR interventions were implemented one-to-one with an adult tutor for 30 minutes daily for four weeks during the summer, at a school for students with visual impairments. During each session, the adult tutor reminded students of the criterion reading rate, and then timed the student while he or she read the passage. Errors were not corrected, but after each reading the student was provided with feedback on the number of words read correctly per minute. Students continued practicing the passage until the criterion reading rate was met or 30 minutes passed, whichever came first. Results indicated that RR was effective for increasing ORF level on transfer passages relative to baseline for one student and a second student increased in both level and trend. A third student did not appear to benefit from the intervention, maintaining a consistent level and trend across baseline and intervention phases. This response was not due to inaccurate practice—in fact, the student had lower errors than the other two participants. The authors posited that the student might have been concentrating on accuracy to the exclusion of increasing his reading rate (Savaiano & Hatton, 2013).

Rafferty (2012) conducted the second study with four second grade students with emotional behavioral disorders during three weeks of a general education summer reading program. Researchers studied the use of a tactile self-prompting device combined with self-monitoring and evaluated effects on on-task behavior during whole and small group instruction during the literacy block. The study also evaluated the effects on ORF of literacy instruction combined with self-monitoring. The literacy block included peer-implemented RR that occurred for an unspecified amount of time daily. Results indicated that self-monitoring improved on-task behavior. Moreover, students'

ORF on transfer measures increased (at least doubled) for all students after the procedure was implemented (Rafferty, 2012).

Manset-Williamson & Nelson (2005) conducted the third study with a group of 21 students in third, fourth, and fifth grade during five weeks of a summer reading clinic conducted in local schools. Students were randomly assigned to conditions that included phonemic awareness/analysis, decoding, and fluency instruction combined with one of two different types of comprehension strategy instruction. The intervention was implemented one-to-one with an adult tutor for 60 min per day, four times a week. Approximately 10 minutes per day were devoted to fluency instruction during which time the student first “shadow-read” the passage with the tutor, and then read the passage a second time. Corrective feedback was provided for errors. Results indicated that ORF on non-transfer and transfer measures increased for both groups, but there were no significant differences between groups (Manset-Williamson & Nelson, 2005).

Taken together, there is small body of research on the effectiveness of RR during school breaks, such as summer programs, and the findings are mixed but overall consistent with studies conducted during the school year.

Conclusions

An important question is whether RR can be considered an evidence-based practice for increasing reading fluency, and research findings are not entire consistent. On one hand, the findings from multiple reviews of research that spans decades have long been accepted as sufficient evidence for RR’s effectiveness for students who are at various ages and stages of reading fluency development (NRP, NICHD, 2000; Therrien, 2004;

ETC). On the other hand, recent reviews have evaluated many of the same and subsequent RR studies by applying rigorous research methodology standards and concluded that there is *in*sufficient evidence to declare RR an evidence-based practice at this time (Chard et al., 2009; O'Keeffe et al., 2012). This question is important to answer considering the long-term consequences of repeated reading regression (Hernandez, 2000; Reder, 2010). Further, an intervention like RR, which provides students with reading fluency practice, could help students maintain, and perhaps even increase, their level of fluency during extended time spent away from school.

Impacting Effectiveness: Dimensions of Implementation

The process of learning to read fluently has remained somewhat enigmatic. There is a large body of research on reading instruction, yet differences in theoretical models and concerns regarding the quality of some of the reading research literature leaves one questioning how to best intervene with struggling readers. Students do not always progress as desired even when instructional approaches and interventions are implemented as intended. Just ask a group of practitioners and you are likely to hear stories of students who developed fluent reading skills almost magically, and other students who struggled. If asked to explain the reasons for the disparity, many would be hard-pressed to provide a definitive answer. Achievement gap statistics are a reflection of these anecdotes, and constitute evidence that more research is needed to better refine reading fluency instruction and intervention.

Addressing the reading achievement gap poses a significant challenge. In the effort to design and implement instructional programs and interventions that promote adequate growth for all students, budget cuts and shrinking resources complicate the task and force practitioners to find a balance between effectiveness and efficiency by targeting their resources carefully and making efficient use of time, materials, and staff (Sugai & Horner, 1999 as cited in McIntosh, Chard, Boland, & Horner, 2006). Educational researchers have responded to these conditions with efforts to determine "what works." Dimensions of implementation such as intervention [treatment] intensity (Warren et al., 2007) and procedural fidelity—as well as variables that impact those dimensions—are potentially critical in this endeavor because they can directly impact the effectiveness of an intervention in general, and with individual students. This section will review both constructs to provide a framework of theoretical support for the examination of variables that are specifically related to the implementation of RR, and potentially it effectiveness.

Intervention Intensity

In 2007, Warren, Fey, & Yoder published a seminal article asserting that intervention [treatment] intensity may be a key factor for optimizing intervention effectiveness. The authors proposed that research on intensity variables had the potential to both enhance and determine the true potential of early intervention in the area of communication and language development (Warren et al., 2007).

The authors also noted that there was no widely accepted definition of intervention intensity—a fact that precluded a systematic study of the construct and its influence on intervention effectiveness. Rather, scholars conceptualized the construct of intervention intensity in multiple ways, often in terms of intervention quality or quantity (Warren et al., 2007). Qualitative versions have included instructional programs that are (1) more highly systematic and explicit, (2) delivered by instructors with high levels of expertise or training (Fuchs & Fuchs, 2006), (3) conducted 1:1 or in small, homogenous groups of students who share similar learning needs (Fuchs & Fuchs, 2006; Graff, Green, & Libby, 1998; O'Connor, 2000; Vaughn, Linan-Thompson, & Kouzekanani, 2003), (4) more individualized (e.g., MTSS or RtI systems, Burns & Gibbons, 2008; see also Barnett, Daly, Jones, & Lentz, 2004; Daly, Lentz & Boyer, 1996), or (5) comprised of layered intervention components (Daly, Martens, Hamler, Dool, & Eckert, 1999; O'Connor, 2000). Quantitative versions have included instructional programs that are specified as (1) the number and duration of weekly sessions (e.g., 12 min sessions for 3x/week; O'Connor, 2000), (2) the hours of treatment per week measured across a specified number of weeks (e.g., .75 hrs/week over 9 weeks; Boettcher, 1983), or (3) the total number/density of sessions over a specified unit of time (e.g., density of intervention sessions over 3 years varied by parent; Ramey & Ramey, 1998).

The current model of dosage. As a solution to the wide range of definitions for intervention intensity, Warren and colleagues proposed terminology for five distinct dimensions of the construct: *dose, dose form, dose frequency, total intervention duration*, and *cumulative intervention intensity* (Warren et al., 2007), referred to collectively as *dosage*. Although the authors were writing specifically about intervention in the area of speech and language intervention, the article proved to be an important catalyst in educational intervention research, and the result was a number of related studies that focused on acquiring a better understanding of differential intervention intensity and its role in optimizing intervention effectiveness. Much of the work has been in the area of communication and language development (e.g., Williams, 2005; Granpeesheh, Dixon, Tarbox, Kaplan, & Wilke, 2009), but there are studies of intervention intensity in other educational intervention areas as well, including early reading (e.g., Ardoin et al., 2009; Begeny, Hawkins, Krouse, & Laugle, 2011; Wanzek & Vaughn, 2008) and math (e.g., Codding, Hilt-Panahon, Panahon, & Benson, 2009; Duhon, Mesmer, Atkins, Greguson, & Olinger, 2009).

The *teaching episode* is at the core of the dosage model proposed by Warren and colleagues, and is defined as “one or more acts, performed by an interventionist, that are believed to produce desired student learning” (Warren et al., 2007, p. 71). Teaching episodes can vary in topography depending on the area of instruction (Yeaton & Sechrest, 1981). In reading, examples of teaching episodes include grapheme stimuli such as individual letters for which a student might be asked to produce a corresponding phoneme, or more complex stimuli such as words to read or comprehension questions to answer.

The concept of the teaching episode is similar to the *learn unit* proposed by Greer and Hogin McDonough (1999). Defined in behavior-analytic terms, the learn unit consists of “the interlocking operants of instruction that incorporate particular student and teacher interactions that predict whether certain student behavior will be controlled by particular stimuli and setting events” (Greer & Hogin McDonough, 1999, p. 6). It is notable that the learn unit specifies that countable teacher-student interactions are necessary to produce learning. Meanwhile, the teaching episode as defined by Warren and colleagues implies but does not explicitly refer o such an interaction between a student and his or her teacher or therapist.

The dimensions of dosage specified by Warren and colleagues (2007) define intervention intensity in terms of the form and occurrence of such teaching episodes, which are intended to take place at some specified frequency, during the course of some specified activity. This specified dose is then scheduled to occur with some measure of frequency over a period of time, and the cumulative intervention intensity is calculated as the product of these factors.

Five terms comprise the dimensions of dosage:

. 1) *Dose* is the number of teaching episodes that occur within an intervention session (e.g., a dose of 25 novel word presentations [i.e., teaching episodes] during a session).

. 2) *Dose form* is "the typical task or activity within which the teaching episodes are delivered" (Warren et al., 2007). Important subcomponents include the average rate of teaching episodes that occur within a specified unit of time (*dosage rate*), intervention session length, and distribution of episodes across the session. Session length should always be specified along with dosage rate, the latter calculated by dividing the *session duration* by the number of teaching episodes (e.g., 15 words read correctly/25 min of intervention = .6/ words read correctly per min of intervention).

. 3) *Dose frequency* is how many intervention sessions occur per time unit such as day, week, and so forth (e.g., 1x/day, 4x/week).

. 4) *Total intervention duration* is the total span of time a given intervention is implemented (e.g., 4 weeks; 2 months; 3 years).

. 5) *Cumulative intervention intensity* is "the product of dose x dose frequency x total intervention duration" (Warren et al., 2007, p. 72). As an example, a *dose* of 25 word stimuli in a *dose form* such as word reading occurs at a *dose frequency* of four times per week for a *total intervention duration* of 10 weeks. The cumulative intervention intensity is then the product of dose (25) x dose frequency (4) x total intervention duration (10), or 1000 words (Warren et al., 2007).

Figure 1 displays the current dosage model, as defined by Warren and colleagues (2007). It is important to note that the model captures dimensions of intervention intensity as they relate to the intervention's topography, teacher inputs, and time spent in intervention. The role of the student is not explicitly included.

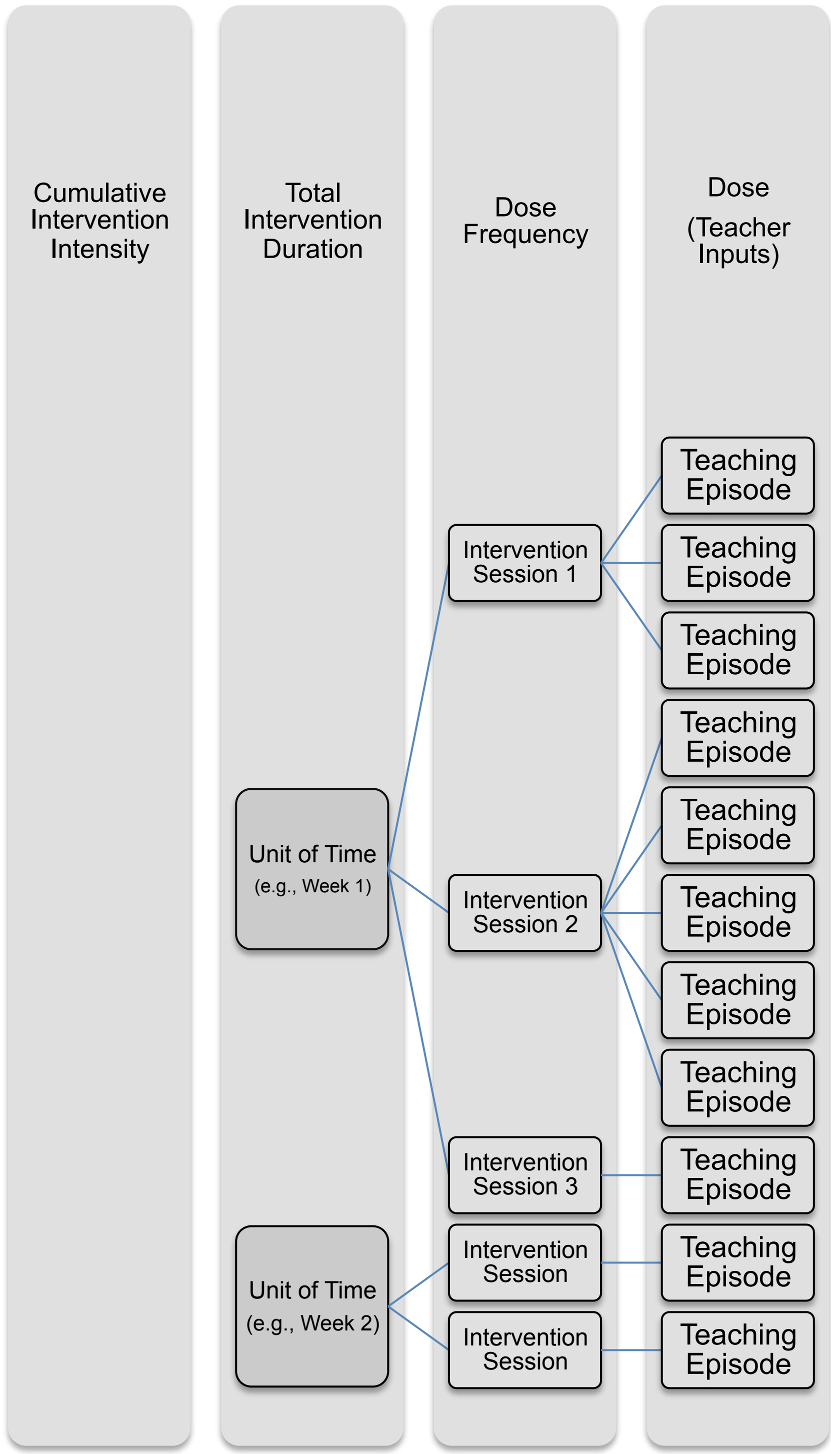


Figure 1. The current dosage model: Dimensions of intervention intensity, as specified by Warren, Fey, & Yoder (2007), illustrating how within given units of time, the number of intervention sessions can vary, as well as the number of teaching episodes.

Procedural Fidelity. Procedural fidelity may be may be another key factor for optimizing intervention effectiveness. Evidence for this assertion can be found in the results of studies indicating that high levels of procedural fidelity correlate with more positive student outcomes (Durlak, & Dupree, 2008 [see reference in Sanetti et al., 2011]; Noell et al., 2005). Procedural fidelity, which has also been referred to as treatment integrity, intervention integrity, intervention implementation, or intervention fidelity, is a multidimensional construct (Dane & Schneider, 1998; DiGenarro-Reed & Codding, 2013; Gresham, 1989, 2014; Sanetti, Kratochwill, & Long, 2013, Warren et al., 2007). Aspects that are frequently noted in the literature include (1) adherence (the accuracy and consistency with which the strategy/intervention is delivered as originally designed; Gresham, 2014), (2) exposure or dosage (amount of instruction/intervention administered; see Warren et. al., 2007, also Bellg et al., 2004), (3) quality of delivery (e.g., competence of the interventionist; Gresham, 2014), (4) treatment differentiation (theoretical distinctions between aspects of treatments and how those distinctions manifest in treatment delivery; Century, Rudnick, & Freeman, 2010; Gresham, 2014), and (5) participant responsiveness (engagement and/or perceived relevance by participant; Dane & Schneider, 1998; Gresham, 2014; Jones et al., 2008), also conceptualized as "client acts" or *student inputs* (e.g., participant producing/practicing the skill; Baker, 2012).

Procedural fidelity is important to maintain because it facilitates learning and enables practitioners to draw accurate conclusions about the effectiveness of instructional programs and interventions (Bellg et al., 2004; Lane, Bocian, MacMillan, & Gresham, 2010; Yeaton & Sechrest, 1981). Unfortunately, schools do not usually evaluate procedural fidelity in any systematic way, in spite of research that verifies it importance (Cochrane & Laux, 2008). Even researchers do not always evaluate procedural fidelity (Barnett et al., 2014; Gresham et al., 1993a, b; Sanetti et al., 2011; O'Donnell, 2008; Swanson, Wanzek, Haring, Ciullo, & McCulley, 2011). Furthermore, evaluations of fidelity that *are* conducted often only assess the amount of time spent (e.g., dose, number of sessions; Sanetti et al., 2011; Warren et al., 2007) or the degree to which the interventionist completed all the steps of an intervention protocol (adherence; Dusenbury, Brannigan, Falco, & Hansen, 2003; Gresham, 2009; Sanetti et al., 2011). However, when interventions do not work as intended, it may be important to look at other factors as well. The current dosage model only accounts for teacher inputs, and not for the contribution of the student (Warren et al., 2007). Yet student responses and behaviors (hereafter referred to as *student inputs*) may contribute significantly to the effectiveness of an intervention for an individual student (Baker, 2012; Jones et al., 2008). Consider that a given student's responses can take different forms such as correct, incorrect, no response—factors that are directly related to the amount and accuracy of responding, as well as the student's engagement. Low levels of any of these variables could help to explain a student's lack of progress, suggesting that differentiating between dose and adherence (as they are typically conceptualized), and *actual* student responding might provide critical information for evaluating and optimizing the effectiveness of an intervention.

The following is an applied example that will illustrate the importance of student inputs, and how the same dose could result in qualitatively and quantitatively different learning experiences for a student—and potentially moderate their progress. Consider the following: a student (Luna) is provided with a specified dose of 25 teaching episodes per 10 min intervention session wherein each word in a phonics word sort is a stimulus-response pair (i.e., teaching episode). Luna's task is to read the words aloud and sort them correctly into columns according to their medial sounds. During three intervention sessions conducted in a week, Luna's number of words read during intervention were as follows: 25/25 in Session 1, 25/25 in Session 2, and 5/25 in Session 3 (Luna read five words, and then became upset and tore up the rest of the word cards). Luna's number of words read correctly were as follows: 12/25 in Session 1, 10/25 in Session 2, and 2/25 in Session 3. Throughout the week, dose (25) and dose frequency (3) were consistent across all three sessions, and Luna was provided with 75 opportunities to respond. However, Luna's performance was *in*consistent and *in*accurate; further, progress-monitoring data indicated little growth—perhaps due to the low levels of accurate responding. This performance leads one to question how an evaluation of effectiveness might better quantify what a student is actually doing during intervention, since the dosage parameters as currently defined account for teacher inputs only, and may not tell the whole story.

The concept of measuring student inputs during interventions is well established in many areas of research. Researchers in the medical field acknowledge that participant behavior—what the individual does during an intervention—is a key factor related to effectiveness, and routinely measure this variable (Bellg et al., 2004). Examples are also found in sports medicine and the arts where the amount and quality of participant practice is known to be essential for developing fluency and expertise. For instance, prior to achieving world status, top performers routinely accrue 10,000 hours of practice (Williams & Hodges, 2005), also known as the 10-year rule (Simon & Chase, 1973, as cited in Williams & Hodges, p. 2).

In contrast, student inputs are rarely measured as an aspect of procedural fidelity in academic intervention research where it is more typical to measure just the inputs of the interventionist or the amount of intervention provided. For example, recent reviews in educational research noted that the most common measures of procedural fidelity were dosage (e.g., number of days and/or sessions of intervention) and adherence (interventionists' completion of steps on a procedural checklist; Sanetti et al., 2011; O'Donnell, 2008; Swanson et al., 2011). There are exceptions, but they are somewhat rare; in one review, Swanson and colleagues noted that out of 50 articles that reported procedural fidelity data, just four included an assessment that was quantified student inputs. In those four studies, dosage was reported in terms of the number of pages read along with the time spent reading, and the quantity of lessons that were completed by the student (Swanson et al., 2011). Similarly, in a review of five effectiveness studies, O'Donnell (2008) noted that two included the degree to which students adhered to components of the intervention (work completion; problems attempted) in addition to adherence by the interventionist. Finally, in a review of 187 studies, 3.2% measured procedural fidelity using a permanent product, which may have included something students completed such as worksheets, or an indirect measure of work completed such as "good behavior" tickets (Barnett et al., 2014).

It is encouraging that student inputs are occasionally assessed as an aspect of procedural fidelity, however the large majority of studies fail to incorporate this measure. Such scant attention to this potentially critical variable might increase if quality indicators for group research such as the What Works Clearinghouse Procedures and Standards Handbook put forth by Gersten and colleagues (2005) recommended the assessment of student inputs as a potential contributor to procedural fidelity and treatment intensity. At the present time however, the manner in which the construct is described and the recommendations articulated focus almost exclusively on the inputs of the interventionist, with only a brief reference to the student (Gersten et al., 2005). In another example, the authors of the What Works Clearinghouse *Practice Guide for Assisting Students Struggling in Reading* assert that reading interventions are most effective if a student receives a dose that includes a session duration of at least 20-40 min per day, 3-5 times per week, but do not recommend any specifics regarding the amount or quality of student inputs (Gersten et al., 2009). The assumption is that most struggling readers will benefit from evidence-based interventions delivered according to these recommendations, but this view is somewhat limited because it does not account for variation in student responding that can impact intervention effectiveness.

Omitting the measurement of student inputs in intervention research, especially in reading, runs contrary to what is known about the relationship between student responding and learning—either by intuition, or as indicated by research. Students need to develop automaticity so decoding and word recognition occur quickly and efficiently during reading tasks, allowing the student to focus attention on the meaning of the text

(LaBerge & Samuels, 1974). Teachers seem to know this on some level; one would be hard-pressed to find a classroom where literacy instruction does not include some kind of reading practice opportunities like reciting letter names, practicing sight words, or reading and re-reading books. Many reading interventions also include some form of repeated practice to enhance learning (e.g., incremental rehearsal, Burns, Zaslofsky, Kanive, & Parker, 2012; word study, Bear, Invernizzi, Templeton, & Johnston, 1996; repeated reading, Therrien, 2004), which has been described as one of the most effective ways to increase fluency in a skill (Chase & Symonds, 1992).

Failure to measure student inputs may potentially lead practitioners to draw erroneous conclusions about the effectiveness of an intervention. Recall Luna, whose inputs varied in amount and accuracy while teacher inputs and dosage remained consistent. As long as the teacher's implementation fidelity was sufficient, and interventions occurred as scheduled, any lack of progress might seem to be the result of some unknown within-student variable, and Luna might then be considered a *non-responder* (McMaster, Fuchs, Fuchs, & Compton, 2005). However, the real problem could be that the Luna did not produce enough quality responses during intervention for learning to occur. In such a case, non-responder is a misnomer because it suggests, in a way, that she *could* not respond better— rather than that she *did* not. Taken together, there is a need for an expanded model of dosage that accounts for the role of student inputs in evaluating and optimizing intervention effectiveness.

A Revised Model of Dosage

C. Everett Coop said, “Drugs don’t work in patients who don’t take them.” (as cited in Ho, Bryson, & Rumsfeld, 2009, p. 3028). Similarly, instructional programs and interventions will not work as expected if students do not participate in ways that are necessary to facilitate desired levels progress. It follows that the conceptualization of intervention dosage should therefore include the responses and behaviors of the *student* as well as the interventionist.

Dose, as it is currently defined in the model put forth by Warren and colleagues (2007), quantifies the actions of the interventionist, but does not account for the level and quality of student inputs, which are essential to outcomes and may be valuable data for assessing intervention effectiveness. Recently there has been some acknowledgement of this omission in the lead article to a scientific forum on the topic of optimizing intervention intensity wherein the author Baker (2012) stated, “exclusive consideration of what a clinician [teacher] does fails to account for the contributions of a client [student] towards an intervention outcome” (Baker, 2012, p. 404; see also Jones et al., 2008). Baker further asserted that determining the optimal intensity of an intervention—that which is most effective for a particular student—necessitates a deeper examination of the active ingredients that contribute to learning. Baker advocated “[looking] inside the teaching unit to isolate and evaluate the quality and quantity of [the actions of the student] (i.e., client acts)” as well as the actions of the teacher (Baker, 2012, pp. 404).

The current dosage model also assumes a high level of procedural fidelity. However, the inputs of the interventionist *and* the student, either alone or in combination, may vary in ways that could degrade procedural fidelity. In turn, an altered level of procedural fidelity has the potential to significantly moderate intervention intensity, and by extension, the intervention's effectiveness. For instance, a dose may be specified at a level that is considered to be appropriately intense, but if an interventionist fails to implement the procedure with fidelity (e.g., too few teaching episodes are presented), the specified dose does not occur. In the same way, an interventionist may implement with a high level procedural fidelity but if student inputs are low or poor quality (e.g., a student responds infrequently or inaccurately), the specified dose also does not occur. The level of student inputs can thus directly impact dosage.

Taken together, I propose a revision of the current dosage model, expanded to include student inputs that may be critical to evaluating the effectiveness of an intervention and specifying dosage. Figure 2 displays this revised model.

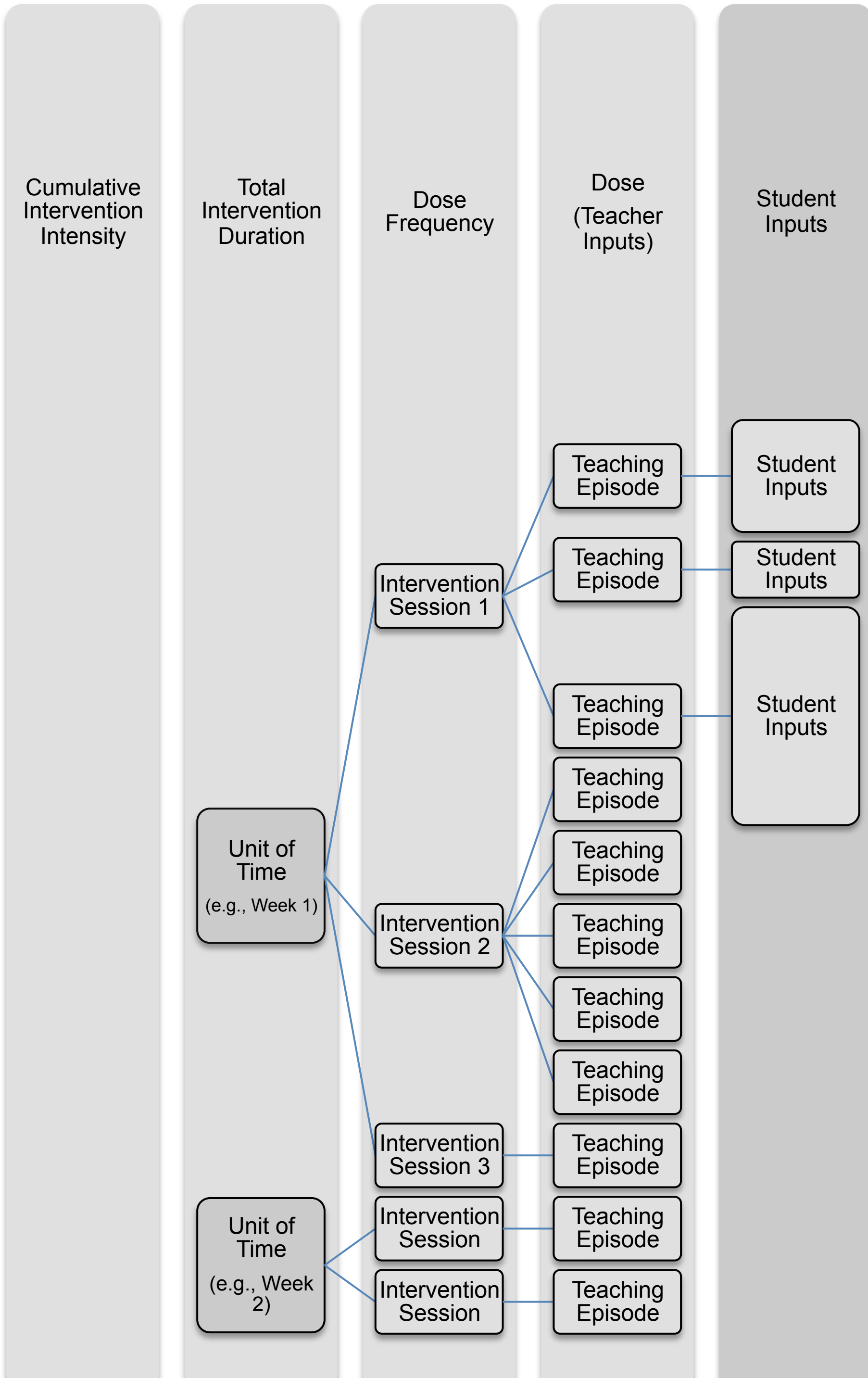


Figure 2. A revised dosage model: Dimensions of intervention intensity expanded to include *student inputs*. Boxes representing student inputs are different sizes to illustrate how quality and quantity can vary and potentially impact intervention effectiveness for individual students, even as teacher inputs in the form of a dose of teaching episodes remains constant.

Summary

In sum, the general question of *how much intervention is enough* is important. Specific to the development of reading fluency, factors such as time spent away from school and student responses and behaviors during instructional programming and intervention have the potential to impact the quantity and quality of students' reading fluency practice, and as a result, their oral reading fluency growth.

Meeting the needs of individual students may require moving beyond simple measures of time spent in intervention or procedural fidelity on the part of the interventionist only. Baker (2012) identified an important omission by noting that the current dosage model proposed by Warren and colleagues (2007) did not account for student inputs. The proposed revision to the model rectifies the omission of student inputs and gives weight to their potential influence on the effects of instructional programs and interventions.

Researchers may find that one way to address the achievement gap is by evaluating effectiveness according to this more comprehensive model. In the area of reading fluency instruction and intervention for instance, student inputs that are critical to effectiveness may include measurement of variables such as time spent reading (e.g., minutes, days, and so forth), level of engagement in the intervention, reading accuracy, and the number of words read. Additionally, it may be essential to provide supplemental ORF practice during school breaks in order to attenuate learning losses that can accumulate over time and increase a student's personal reading achievement gap.

Educational research is progressing toward a more complete understanding of dimensions of implementation and their role in evaluating and optimizing intervention effectiveness, but more work is needed.

Review of Study Purpose and Research Questions

The purpose of the present study was to explore the notion of "enough of what works" by evaluating the effects of a reading fluency intervention implemented with second and third grade struggling readers during a school break, as well as variation in student responses and behavior during intervention (*student inputs*) that could alter the quantity and quality of practice, and potentially contribute to differences in intervention effectiveness for individual students

The following research questions guided the study:

1. What are the effects on oral reading fluency of a repeated reading (RR) intervention implemented during four weeks of summer school with students in 2nd and 3rd grade whose reading is accurate but slow?
2. What dimensions of RR implementation that include student inputs predict and/or correlate with posttest oral reading fluency scores, controlling for students' oral reading fluency at pretest?

CHAPTER THREE

METHOD

Participants, Interventionists, and Setting

The participants in this study were 79 students (42 male) in second grade (n = 44) and third grade (n = 35). The students attended a summer school program in a large urban city in the Midwest, having been invited due to fall and/or winter scores on the Measures of Academic Progress (MAP) assessment for math and/or reading that fell below the district grade level benchmark. Summer school attendance was not mandatory and parents/guardians could choose whether or not to enroll their student.

Two elementary schools that hosted the summer programs within the district were designated as target sites. Summer program attendees were drawn from nine surrounding elementary schools. Eighty-two percent of students qualified for free and reduced-price lunch. Twelve percent of the students spoke a second language at home according to parent report. No students received ESL services during summer school. Information was not available to confirm whether any students received ESL or special education services during the regular school year. Figure 3 displays the race/ethnicity proportions for participants. Table 1 displays additional demographic information.

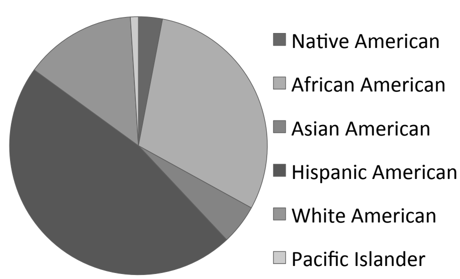


Figure 3. Race/ethnicity proportions for participants.

Table 1

Percentage(number) of Participants by Demographic Category

Category	Group Combined	Intervention	Control
Race/ethnicity			
White American	14(11)	8(6)	6(5)
African American	30(24)	13(10)	18(14)
Hispanic American	47(37)	23(18)	24(19)
Native American	3(2)	1(1)	1(1)
Asian American	5(4)	3(2)	3(2)
Pacific Islander	1(1)	0(0)	1(1)
Male	53(42)	46(23)	38(19)
Grade 2	56(44)	28(22)	28(22)
School 1	18(14)	8(6)	10(8)
School 2	38(30)	20(16)	18(14)
Grade 3	44(35)	23(18)	22(17)
School 1	44(14)	10(8)	8(6)
School 2	27(21)	13(10)	14(11)

Note. [a]Based on combined average of the percentage of students in participating schools who qualify for free/reduced price lunch.

Criteria for Participation

All second and third grade students who attended the summer school program were invited to participate in this study. A Formative Assessment System for Teachers (FAST; Christ, Ardoin, & Eckert, 2011) oral reading fluency (ORF) screening probe was administered to students who gave their assent and whose parent/guardian provided consent. The accuracy and rate data obtained from the screening assessment was used to determine whether students met inclusion criteria, and were also used as the ORF pretest for analysis.

Screening data were used to identify students whose reading was accurate but slow, indicating a need for fluency intervention. A student met initial criteria for the study if his or her median ORF rate in the form of words read correctly per minute (WRCM) fell at or below a pre-determined cut score (i.e., “slow reading rate”). The ORF cut score was based on the 50th percentile grade level spring norms for second and third grade students (114, 136 WRCM, respectively; 2013 FAST norms, T. Christ, personal communication, June 20, 2014). The norms were then adjusted for the standard error of measurement (SEM) that could be expected under relatively controlled conditions when administering three second or third grade CBM-R oral reading fluency passages. The SEM for reliabilities from .95-.97 ranged from 5-9 (median of 7) for a typical sample of second and third grade students (see Christ & Silberglitt, 2007). The median SEM value of 7 was then added to the 50th percentile grade level spring ORF norms, resulting in a cut score of 121 WRCM for students entering third grade in the fall (114 + 7 SEM) and 143 WRCM for students entering fourth grade in the fall (136 + 7 SEM). This calculation served to maximize the inclusion of students whose ORF rate fell at or below the 50th percentile by accounting for possible random measurement error.

A student met secondary inclusion criteria if his or her oral reading accuracy was ≥ 93% on the grade level ORF screening probes (i.e., “accurate reader”). This accuracy benchmark was based on a study conducted by Treptow, Burns, and McComas (2007) wherein increased levels of on-task behavior were observed when students read passages with 93%-97% accuracy (instructional level), and higher reading comprehension for levels of accuracy at or above 93% (Treptow et al., 2007). Students who obtained < 93% accuracy on the grade level ORF screening probes were asked to read three additional probes that were one grade level lower, similar to procedures implemented by Therrien, Kirk, & Woods-Groves (2012). If oral reading accuracy improved to ≥ 93% on the lower grade-level passages, the student was admitted to the study. All students who met these specified inclusion criteria were allowed to participate.

Final Sample

Of the 100 students who met inclusion criteria in second (n = 56) and third (n = 44) grade, 34 students attended School 1 (16 in grade two, 18 in grade three) and 66 students attended School 2 (40 in grade two, 26 in grade three). The final sample of participants consisted of 79 students in second (n = 44) and third (n = 35) grade, randomly assigned to intervention (n = 40) and control (n = 39) groups. Throughout the study, 21 students stopped attending summer school (10 from the intervention group and 11 from the control group). The resulting overall attrition of 21% was within limits specified as a quality indicator for group experimental research (Gersten et al., 2005). Differential attrition was 2%, also within recommended limits according to the What Works Clearinghouse Procedures and Standards Handbook (IES, 2014). Table 2 displays attrition statistics for study participants.

Table 2

Attrition Statistics for Study Participants

Variable	% of total (*N* = 100) Initial participants Treatment (n = 50)	Initial participants Control (n = 50)	Lost to attrition Treatment (n = 10)	Lost to attrition Control (n = 11)	% of total (*N* = 79) Participants remaining Treatment (n = 40)	Participants remaining Control (n = 39)
1. Gender (male)	0.26(26)	0.26(25)	0.03(3)	0.06(6)	0.29(23)	0.24(19)
2. Second grade	0.28(28)	0.28(28)	0.06(6)	0.06(6)	0.28(22)	0.28(22)
3. Third grade	0.22(22)	0.22(22)	0.04(4)	0.05(5)	0.23(18)	0.22(17)
4. All grades combined	0.50(50)	0.50(50)	0.10(10)	0.11(11)	0.51(40)	0.49(39)

Note. Values for variables are percentages (numbers).

Interventionists

VISTAs (Volunteers in Service to America) conducted the repeated reading interventions. VISTAs, many of whom are high school graduates or college students, commit to a service term during which time they volunteer a minimum of 40 hours per week. It is intended that the work VISTAs do remains primarily focused on addressing the literacy and academic achievement needs of students from low-income communities. More information on the VISTA program can be found at http://www.mnliteracy.org/volunteers/become-vista.

Eight VISTAs served as interventionists for this study. Their ages ranged from 19-26 years. Five were female. Two of the VISTAs identified their race as Black and six identified their race as White. Six of the eight VISTAs reported that they had completed one or more years of college education. All but one VISTA reported some experience teaching reading to children through elementary education teaching practicums, tutoring, or mentoring.

Setting

Interventions took place in a quiet area of the school, in a location determined by consulting with the summer school coordinator and the volunteer supervisor at each school. VISTAs at School 1 worked with their students in an empty classroom at small tables spaced approximately 10 feet apart. VISTAs at School 2 worked with their students at small tables in the hallways. Observations of interventions in both schools indicated that students experienced a similar intervention environment, largely free from distractions. Students who received interventions in the classroom at School 1 were subject to a low level of ambient noise from up to three interventions taking place at the same time. Students who received interventions in the hallways at School 2 were subject to occasional short-term noise from passers-by.

Materials

Formative Assessment System for Teachers (FAST) screening probes were used to assess oral reading fluency rate and accuracy (Christ et al., 2011). Pre- and post-tests utilized second grade and third grade probes for students who had completed second and third grade, respectively, in the spring immediately preceding the summer school session (see Appendix A).

FAST progress monitoring probes were used as practice passages during intervention sessions (Christ et al., 2011). The grade level of the practice passages used for typically corresponded to students' previous year of school. For example, a student who had completed second grade in the spring practiced second grade passages during intervention. However, if a student was unable to read grade level practice passages with ≥ 93% accuracy, he or she read practice passages that were one grade level lower (see Appendix B).

Each VISTA was provided with an intervention kit that contained all the materials needed to conduct the intervention. The kit included an intervention log, intervention checklist, an intervention protocol script, two copies of the instructional passage for each student in the group (one copy for the student, one copy for the VISTA to mark the student's data), a sticker chart for each student, one graph for each student (used to record ORF on the first and fourth read of the intervention session), and writing utensils (e.g., pencils, markers). The kit was also outfitted with three timers: one timer was used to time the 1 min readings; the remaining two timers were each labeled with the name of one student in the dyad. During the intervention, the VISTA used the timers to measure the total minutes each student spent in intervention. See Appendices C, D, and E for examples of materials that were included in the intervention kit.

Measures

Several measures were used to answer the research questions and included: (1) the ORF screening measure which determined eligibility for the study, but also functioned as a pretest, (2) ORF posttest, and (3) data were collected on variables related to implementation.

Oral Reading Fluency (ORF) Pre- and Posttest

The primary dependent measure was oral reading fluency (ORF). Multiple studies provide strong theoretical backing (LaBerge & Samuels, 1974) and empirical evidence (Deno, Mirkin, & Chiang, 1982, Fuchs, Fuchs, & Maxwell, 1988; Hintze, Callahan, Matthews, Williams, Tobin, 2002, Jenkins, Fuchs, Van den Broek, Espin, & Deno, 2003) to support that oral reading fluency is appropriate as a screening measure and indicator of overall reading competence, including comprehension. The short length of the summer school program and concern cited by the school district with regard to the duration of time students spent testing precluded the administration of additional assessments (e.g., comprehension, vocabulary).

FAST passages were chosen because their psychometric properties helped to ensure a similar level of difficulty across probes and practice passages, which served to minimize instrumentation threats to internal validity. Alternate form reliability of FAST passages was $r = .90$ (range .87-.92, SEm = 4.97). The median internal consistency (item-total correlation) was $\alpha = .90$ (range .89-.91). The median test-retest reliability for second and third grade was $r = .93$ (range .91-.94; Christ et al., 2014). Additionally, correlations between CBM-R scores and the Minnesota Comprehensive Assessment (MCA) in reading have been found to be statistically significant and strong for third grade students, $r = .68$ (Silberglitt, Burns, Madyun, & Lail, 2006).

ORF was measured in the form of median words read correctly per min (WRCM) and median accuracy (ACC) on three grade level FAST ORF screening passages. Three-passage (as opposed to single-passage) administration was chosen to minimize measurement error (Christ et al, 2014; Jenkins, Zumeta, Dupree, & Johnson, 2005).

Pre- and posttest procedures: The ORF screening passages were administered before interventions began (pretest) and after interventions ceased (posttest). The interval between pre- and posttest administration was five weeks. There is empirical evidence to support that a five-week interval between testing instances does not artificially inflate ORF estimates of reading ability due to passage memory effects (Jenkins et al., 2005).

The lead author and a graduate assistant, along with the VISTAs administered the assessments. The lead author and graduate assistant were doctoral candidates in special education and school psychology, respectively. Both had extensive experience working with children as classroom teachers (19 years, 3 years, respectively), and in the administration of individual reading assessments. Additionally, both had served as interventionists during a three-year study that implemented a multi-tiered system of support targeting early reading, during which time such assessments were part of their graduate assistant duties. VISTAs attended a three-hour training delivered by the lead author and graduate assistant to learn how to conduct the assessment.

The lead author, graduate assistant, and VISTAs administered the oral reading fluency pretests prior to the start of any interventions. The posttests were administered at the end of the summer program, after four weeks of intervention. Both tests utilized identical procedures. See Appendix F for the administration protocol and fidelity checklist. The administration setting was a quiet area of the school with the student and test administrator seated adjacent to each other at a table or desk. During the assessment, the student read three grade-level passages for 1 min each while the test administrator marked correct responses and errors on his/her own copy. A correct response was defined as a word read correctly or self-corrected within 3 s. An error was defined as no response (including omitted words), or a word read incorrectly and not self-corrected, within 3 s. Words read correctly but out of order were also counted as errors (e.g., if the text said, "*cute, fuzzy puppy*" but the student read, "*fuzzy cute puppy*" this amounted to two errors). Words read correctly (WRCM) and errors (ERR) were recorded for each passage. WRCM was calculated as follows:

$$WRCM = Total\ Words\ Read\ in\ 1\ min - ERR$$

A final oral reading fluency (ORF) score was recorded in the form of median WRCM and median ERR. Accuracy was then calculated as follows, using the final ORF score:

$$ACC = WRCM/(WRCM + ERR)$$

To permit aggregating results across grades during analysis, oral reading fluency (ORF) pre- and post- test raw scores were standardized using grade-based norms (T. Christ, personal communication, June 20, 2014) using the following equation:

$$\mathrm{Z} = \frac{\mathrm{X} - \mu}{\sigma}$$

Dimensions of Implementation: Student Input Variables

Data were collected on six variables related to implementation. These variables are referred to as student inputs (see also *client acts*, Baker, 2012), and were believed to capture participant responses that could potentially influence the quality and quantity of student practice, and ORF outcomes by extension.

Student input variables were determined by examining the repeated reading (RR) fluency protocol and identifying aspects of student participation that were essential for reading practice to occur, and that might impact the quantity or quality of that practice.

These essential aspects included attending the scheduled intervention sessions, and engaging in the intervention by completing the 1 min oral readings as directed with a high level of accuracy. The specific variables that were measured to quantify these essential aspects are listed in the paragraphs that follow, along with a brief rationale specifying how each variable is potentially related to oral reading fluency outcomes as part of the RR intervention. Detailed information about how these data were collected can be found in the Procedures section.

(1) Cumulative minutes spent in intervention. This variable captured the amount of time a student spent in intervention. Recommendations on intervention dosage commonly suggest a certain number of minutes per session, and student attendance is directly related to the number of scheduled minutes of intervention received. Notably, this variable related to, but was not a direct measure of, the amount of time the student spent engaged in observable reading during the intervention sessions he or she attended.

(2) Percentage of 1 min readings completed during intervention sessions. This variable captured the percentage of assigned practice trials a student completed across intervention sessions he or she attended. Students were required to complete four 1 min passage readings during each intervention session. It was believed that a student might potentially complete fewer readings per intervention session if he or she arrived late to a session or uncooperative (e.g., refused to participate).

(3) Cumulative number of 1 min readings completed. This variable captured the total number of assigned practice trials a student completed across all intervention sessions. The variable is also a direct measure of the cumulative time a student was engaged in observable reading during all the intervention sessions he or she attended.

(4) Cumulative number of words read correctly. This variable captured the total number of words read correctly across all intervention sessions a student attended, which could be considered the most direct measure of the quantity of practice. The magnitude was larger or smaller depending on a student's reading rate, attendance, and/or refusal/resistance during intervention. For instance, two students who completed the same number of 1 min readings did not necessarily engage in the same amount of practice due to differences in their reading rate. A student who read an average of 50 words per minute and completed 50 1 min readings read only half as many words as a student who read an average of 100 words per minute and completed 50 1 min readings.

(5) Mean oral reading accuracy. This variable captured a student's mean oral reading accuracy, which could also be considered an aspect of practice quality (e.g., level of correct responses), and potentially related to improved reading fluency.

(6) Student engagement. This variable captured the degree to which the student was engaged and cooperative during intervention, as perceived by the VISTAs. Student engagement was assessed using a 0-3 point rating scale.

Data on variables related to implementation were recorded as entries in daily intervention logs and via codes recorded on paper copies of students' individual reading passages. Appendices G and H display examples of how data on these variables were recorded.

Fidelity to Measurement Procedures and Interobserver Agreement.

To ensure accurate measurement of the ORF pre- and post- tests, data on interobserver agreement (IOA) and procedural fidelity to the assessment protocol were collected prior to administering assessments. All individuals who administered the assessments demonstrated the ability to conduct them with high fidelity (*M* = 96.8%, range: 90-100%) as measured by using a checklist comprised of assessment steps that an observer marked as “yes” or “no” to indicate that the step was conducted as specified in the protocol. Percentage of fidelity was calculated as follows:

$$fidelity = \frac{total\ steps\ completed}{total\ steps\ possible}\ X\ 100$$

IOA was assessed and a percentage of IOA was calculated as follows (House, House, & Campbell, 1981):

$$inter - observer\ agreement = \frac{agreements}{agreements + disagreements}\ X\ 100$$

IOA for pre- and post- test assessments determined the degree to which observers agreed on WRCM and errors. As students attempted to read each word in the assessment, observers either agreed or disagreed that the student read the word correctly. IOA was assessed (*M* = 99.9%, range: 99-100%).

IOA data on the measurement of variables related to implementation were gathered weekly for 25% of all intervention sessions conducted and are displayed in Table 3. IOA for all variables, with the exception of student engagement, were calculated as follows (House, House, & Campbell, 1981):

$$inter - observer\ agreement = \frac{agreements}{agreements + disagreements}\ X\ 100$$

IOA was assessed for student engagement according to the percentage of observed sessions where both raters either (1) agreed on a given student’s behavior ratings, or (2) differed by one rating (see Table 3). For example, if the VISTA and the observer both rated a student's behavior as a "3", the raters were in 100% agreement. If the VISTA rated the student's behavior as a "3" and the observer rated the behavior as a "2", the raters were considered one off in their agreement.

Detailed methods for the collection and calculation of all data are described in the Procedures section.

Table 3

Interobserver Agreement (IOA) Percentages for Variables Related to Implementation

Variables	*M*	Range
1. Cumulative minutes spent in intervention	0.98	0.77 – 1.00
2. Mean percentage of 1 min readings completed	1.00	--
3. Cumulative number of 1 min readings	1.00	--
4. Cumulative number of words read correctly	0.99	0.92 – 1.00
5. Mean oral reading accuracy	0.99	0.92 – 1.00
6. Student engagement		
Percentage of ratings with 100% agreement		0.74
Percentage of ratings with one-off agreement		0.26

Note. IOA for variables 1-5 calculated as follows: IOA = (agreements/(agreements + disagreements))100. IOA for variable 6 shows the percentage of ratings that were 100% agreement (VISTA and observer agree) or one-off agreement (VISTA and observer differed by one rating).

Intervention and Control Conditions

Students in both groups (intervention, control) received regular summer school core literacy instruction from their classroom teacher. An environmental assessment was conducted to verify that each classroom contained books at different reading levels, and covered a variety of fiction and non-fiction topics. Classrooms were also observed directly to confirm the nature of "business as usual" core instruction received by students in both conditions (intervention and control). All teachers were observed during their 90 min literacy block at least one time during the summer program. The length of the literacy block was the same at both school sites and consisted of a whole-class mini-lesson, after which students participated in a variety of small group and independent activities related to literacy (e.g., independent reading, response journals, book groups, and so forth).

In addition to core instruction, students in the intervention group also received a supplemental repeated reading (RR) fluency intervention implemented by a trained VISTA. It was hypothesized that RR would be an appropriate intervention for this sample of students whose reading was accurate ($\geq$ 93%) but slow (< 50^{th} percentile). Furthermore, reading fluency is a procedural skill subject to decay over extended periods of time without practice, such as during a school break. RR was thus chosen as a way to maintain and potentially improve students' reading fluency during the summer program.

Students attended intervention sessions once per day, four days per week, in dyads. Session length varied, averaging 21.39 min per day (*SD* = 5.59; range: 9.22 to 37.02 min). The RR intervention protocol required each student to complete four 1 min readings of a passage (aloud), and to follow along silently or whisper-read while the other student in the dyad completed his or her 1 min readings. Students were also provided with a model of fluent reading by the VISTA, error correction, and a graphic representation of their ORF progress. In addition, students answered four general comprehension questions. See Appendix I for intervention protocol. Detailed intervention procedures are described in the *procedures* section.

Procedural Fidelity.

Procedural fidelity data were gathered weekly for 25% of all intervention sessions conducted by the VISTAs. The first author and graduate assistant assessed fidelity using a checklist comprised of intervention steps, which an observer marked as “yes” or “no” to indicate that the step occurred. All components of the intervention protocol were considered important, however it was hypothesized that some components would be more critical to developing students’ ORF than others. For instance, repeatedly reading the passage (Samuels, 1979, Therrien, 2004) and receiving error correction (Therrien, 2004) were considered more critical to developing ORF than answering a question that required a prediction about what might happen at the end of the story. Thus, procedural fidelity was assessed using a weighted checklist to monitor the degree to which critical components occurred. In the weighted checklist, points were allocated such that the VISTA could not receive a score of more than 80% if he or she did not complete the critical components of the intervention (see Appendix J for specifics on how components were weighted). The first author and graduate assistant indicated the completion or non-completion of each step/component and calculated a percentage as:

$$fidelity = \frac{total\ steps\ completed}{total\ steps\ possible}\ X\ 100$$

Fidelity percentages were reported to VISTAs on the same day the session was observed.

These discussions were always supplemented with specific positive feedback on components of the intervention or interactions with the student that the VISTA did well, along with any areas that needed to be addressed in order to improve fidelity   Mean weighted fidelity percentages were above 95% across all four weeks: 96% (range: 80%-100%), 98% (range: 98%-100%), 98% (range: 81%-100%), 99.49 (range: 97%-100%), respectively.

Experimental Design

A between-groups experimental design with matched (pair) random sampling was utilized to evaluate the effects of the RR intervention. All participants who met inclusion criteria were first stratified by grade, and then rank-ordered by their pretest median WRCM. If two or more students had an identical score, those students were further rank ordered according to their percentage of oral reading accuracy (ACC). Students were then placed in matched pairs and randomly assigned to either the intervention or control condition (see Figure 4). Interventions were implemented for four weeks of the summer program. At the end of the program, the posttest was administered to all students in both groups.

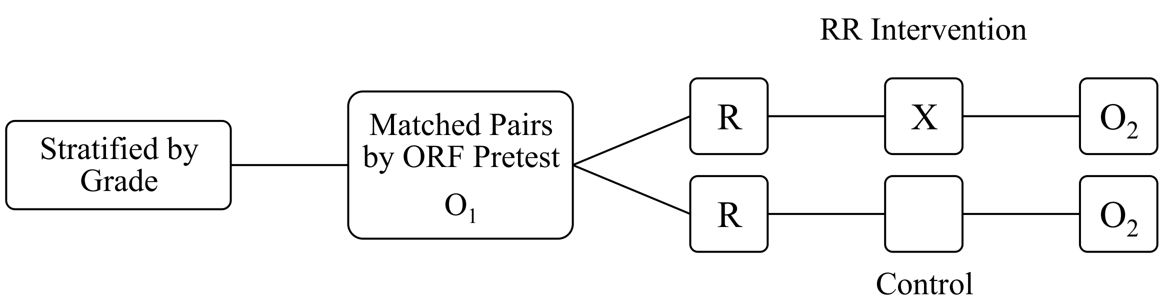


Figure 4. Experimental design where ORF = oral reading fluency, O_1 = ORF pretest, R = random assignment, X = repeated reading (RR) intervention, and O_2 = ORF posttest.

Examining Group Differences

Independent samples t-tests were conducted to compare the standardized pretest scores for condition (intervention, control), and grade (second, third). Mean differences in ORF pretest scores for intervention (*M* = -1.07, *SD* = 0.97) and control (*M* = -1.01, *SD* = 0.81) groups were non-significant, *t* (77) = -.317, *p* = .752, α = .05, *Hedge's g* = -0.07 (CI = -0.51 to .37). Differences in ORF pretest scores for second grade intervention (*M* = -.83, *SD* = .83) and control (*M* = -.89, *SD* = 0.73) were non-significant, *t* (42) = .249, *p* = .805, α = .05, *Hedge's g* = 0.08 (CI = -0.51 to .67). Differences in ORF pretest scores for third grade intervention (*M* = -1.36, *SD* = 1.07) and control (*M* = -1.16, *SD* = 0.91) groups were also non-significant, *t* (33) = -.607, *p* = .548, α = .05, *Hedge's g* = -0.20 (CI = -0.86 to .47).

Differences in attendance by condition (intervention, control), and by grade (second, third) that might have mediated treatment outcomes were also evaluated. Descriptive statistics indicated that the distributions of attendance for intervention and control groups (grades 2 and 3 separate and combined) were negatively skewed, with somewhat large kurtosis values (e.g., 2.96 and 5.12; see Table I4- 1). Means for attendance were thus compared using the non-parametric Mann Whitney-U test. Results indicated that attendance was not greater for intervention (*Mdn* = .88) than for control (*Mdn* = .94) groups, *U* = .82, *p* > .05. Attendance was also not greater for second grade students in intervention (*Mdn* = .94), than control (*Mdn* = .94) groups, *U* = .90, *p* > .05 or for third grade students in intervention (*Mdn* = .88) than control (*Mdn* = .94) groups, *U* = .66, *p* > .05. Thus, groups were comparable with respect to attendance.

Data Analysis

Separate analyses were conducted to answer both a priori research questions. In addition, post hoc questions emerged in the course of data analysis. This section describes the analyses that were conducted for each research question, and relevant descriptive statistics, including the results of assumptions testing.

Research Question 1

A one-way analysis of covariance (ANCOVA) was conducted to determine whether there were statistical differences in the posttest WRCM of students in control and intervention groups after controlling for pretest levels. Data for this analysis consisted of independent observations, independent variables were categorical, and the dependent variable and covariate were measured on a continuous scale. Data were also screened for violations of additional assumptions.

Linearity. A scatterplot of the dependent variable (ORF posttest) and the covariate (ORF pretest) indicated that the assumption of linearity was reasonable. That is, there were concomitant increases in posttest values as pretest values increased (see Figure K1). The relation was also found to be significant, $F(1, 77) = 788.21$, $p = <.001$, $\alpha = .05$.

Normality. Examination of boxplots indicated no extreme outliers and a relatively normal distribution for the dependent variable (ORF posttest) and also the covariate (ORF pretest) at each level of the factor (condition; see Figures K2, K3). Skewness and kurtosis values were within accepted limits of ±2.00 (Ghasemi & Zahediasl, 2012; see Tables K1, K2).

Homogeneity of variance. Non-significant results of Levene's test, $F(1, 77) = .41$,

$p = .523$, $\alpha = .05$, indicated that the assumption of homogeneity of variance was met.

Homogeneity of regression slopes. Examination of scatterplots and models of covariate x factor interactions indicated that the relation between the dependent variable and covariate at each level of the factor (intervention, control) was mostly parallel and in the same direction, with slight heterogeneity $t(3,75) = 2.24$, $p = .028$. However, the ANCOVA is generally robust if sample sizes are equal (or mostly so), thus the analysis was conducted (Hamilton, 1977).

Post-Hoc Questions and Analyses

In the course of examining unstandardized (raw) pre- to post- test results for practical consideration, unexpected differences in the growth of second and third grade students were detected. First, relative to the control group, the mean change from pre- to post-test was more pronounced for second than for third graders. Second, the mean change in WRCM was similar for third grade students in both intervention and control groups. These grade level differences led to the following post-hoc research question:

1. Is grade level a factor with regard to the effects on oral reading fluency of the repeated reading intervention, implemented during four weeks of summer school with students in 2^{nd} and 3^{rd} grade whose reading was accurate but slow?

Analysis: Post hoc question 1: To answer the question, a two-way ANCOVA was conducted to determine whether there were statistical differences in the posttest WRCM of second and third grade students in control and intervention groups after controlling for pretest levels, including a condition x grade interaction. Data for this analysis consisted of independent observations, independent variables were categorical, and the dependent variable and covariate were measured on a continuous scale. Data were also screened for violations of additional assumptions.

Linearity. A scatterplot of the dependent variable (ORF posttest) and the covariate (ORF pretest) indicated that the assumption of linearity was reasonable. That is, there were concomitant increases in posttest values as pretest values increased (see Figure K1). The relation was also found to be significant, $F(1, 77) = 788.21$, $p = <.001$, $\alpha = .05$.

Normality. Examination of boxplots indicated no extreme outliers and a relatively normal distribution for the dependent variable (ORF posttest) and also the covariate (ORF pretest) at each level of the factors (condition, grade; see Figures K4, K5). Skewness and kurtosis values were within accepted limits of ±2.00 (Ghasemi & Zahediasl, 2012; also see Tables K4, K5), suggesting that normality was a reasonable assumption.

Homogeneity of variance. Non-significant results of Levene's test, $F(3, 75) = 1.43$, $p = .241$, $\alpha = .05$, indicated that the assumption of homogeneity of variance was met.

Homogeneity of regression slopes. Examination of scatterplots and models of covariate x factor interactions indicated that the relation between the dependent variable and covariate at each level of the factors (condition, grade) were mostly parallel and in the same direction. Once again, an exception was slight heterogeneity of the regression slopes for condition, $t(3,75) = 2.24$, $p = .028$. However, the ANCOVA is generally robust if cell sizes are equal (or mostly so), thus the analysis was conducted (Hamilton, 1977).

Results of the two-way ANCOVA indicated that there were no significant effects for grade, or the interaction of grade x condition, which led to a second post-hoc research question:

2. Is there a subset of students for whom the effects on oral reading fluency of a repeated reading intervention are greater?

Analysis: Post hoc question 2. To answer this question, a two-way ANCOVA was conducted to determine whether there were statistical differences in the posttest WRCM of relatively high and low level readers in control and intervention groups after controlling for pretest levels, including a condition x grade interaction. Specifically, students were designated as either relatively high-level readers (26-50th grade level percentile) or low-level readers (0-25th grade level percentile). A three way ANCOVA was also conducted to examine effects for condition, grade, and level, including all possible interactions. Data for this analysis consisted of independent observations, independent variables were categorical, and the dependent variable and covariate were measured on a continuous scale. Data were also screened for violations of additional assumptions.

Linearity. A scatterplot of the dependent variable (ORF posttest) and the covariate (ORF pretest) indicated that the assumption of linearity was reasonable. That is, there were concomitant increases in posttest values as pretest values increased (see Figure K1). The relation was also found to be significant, $F(1, 77) = 788.21$, $p = <.001$, $\alpha = .05$.

Normality. Examination of boxplots indicated that there were generally no extreme outliers and a relatively normal distribution for the dependent variable (posttest) and also the covariate (pretest) at each level of the independent variables (condition, grade, level; K6-K9). Exceptions included the posttest scores for the second grade low-level control group, which contained two extreme outliers. Examination of the data confirmed that the scores were accurate. Further, there was no evidence to indicate that the scores represented a unique subset of students. Taken together, the decision was made to retain the scores as part of the dataset. Skewness and kurtosis values were within accepted limits of ±2.00 (Ghasemi & Zahediasl, 2012), with the exception of posttest scores for the third grade relatively high-level intervention group and low-level control group, where kurtosis values were > 2.00 (-2.50, 3.01, respectively; see Tables K6-K8). ANCOVA is robust to slight non-normality however, thus no transformations of the data were conducted (Hopkins & Weeks, 1990).

Homogeneity of variance. Non-significant results of Levene's tests, $F(3, 75) = .41$, $p = .748$ [two-way], $F(7, 71) = .95$, $p = .474$ [three-way], $\alpha = .05$, indicated that the assumption of homogeneity of variance was met.

Homogeneity of regression slopes. Examination of scatterplots and ANCOVA models including covariate x factor interactions indicated that the relation between the dependent variable and covariate at each level of the factors (condition, grade, level) was mostly parallel and in the same direction. Once again, an exception was slight heterogeneity of the regression slopes for condition, $t(3,75) = 2.24$, $p = .028$. However, the ANCOVA is generally robust if cell sizes are equal (or mostly so), thus the analysis was conducted (Hamilton, 1977).

Research Question 2

Correlations and multiple linear regression analysis were conducted to determine which variables related to implementation accounted for the most variance in, and were most predictive of, posttest WRCM, controlling for pretest levels. Data for this analysis consisted of independent observations, independent variables were continuous, and the dependent variable was measured on a continuous scale. Data were also screened for violations of additional assumptions.

Linearity. Scatterplots of the relation between ORF posttest and variables related to implementation as well as ORF pretest indicated that the assumption of linearity was reasonable (Figures K10-K14).

Normality. Examination of boxplots indicated that except for the distributions of student engagement scores of "1" and "0", there were no extreme outliers and a relatively normal distribution for ORF pre- and posttest scores, as well as variables related to implementation (see Figures I15-I24. Further, skewness and kurtosis values were within accepted limits of ±2.00 (Ghasemi & Zahediasl, 2012) with the exception of the student engagement score of "0", which had a skewness value of 2.61 and a kurtosis value of 5.59 (see Table K9, K10).

Homoscedasticity. Studentized residuals plotted against standardized predicted values show a mostly even scatter pattern (Figures K25-K27).

Absence of autocorrelation. Absence of autocorrelation was verified by Durbin Watson values ranging from d = 1.99-2.15.

Little or no multicollinearity. Acceptable levels of multicollinearity were verified by variance inflation factors < 10 (O'Brien, 2007), and bivariate correlations < .80. One exception was the significant and large correlation between ORF pretest and posttest. Partial correlations were thus calculated to control for pretest level and found to be < .80.

Student Input Variables

Descriptive statistics for student input variables were examined prior to conducting the multiple regression analysis. Results led to the exclusion of two variables from the models. First, *mean percentage of 1 min readings completed during intervention* was excluded because there was no variability in the scores; values were 100% for all students (except for a single student on one day). Second, it was questioned whether *student engagement* would add meaningfully to the analysis. For example, the distribution of the lowest behavior rating (0) was somewhat non-normal due to relatively large skewness and kurtosis values (2.61 and 5.59, respectively; see Table K10), likely attributable to the fact that there were few instances where student behavior was rated as "0" (see Appendix L for operational definitions of behavior ratings). VISTAs maintained that they typically rated students as "often" or "always" on task, and that challenging behavior did not reduce task completion for students. Further, the mode of student engagement scores was uniformly high (see Figure 5). Student engagement was therefore examined by calculating a Spearman's rank-order correlation to test for a possible relation between students' behavior ratings and ORF posttest scores. The result was non-significant, $r_s(38)$. = -.133, p = .414.

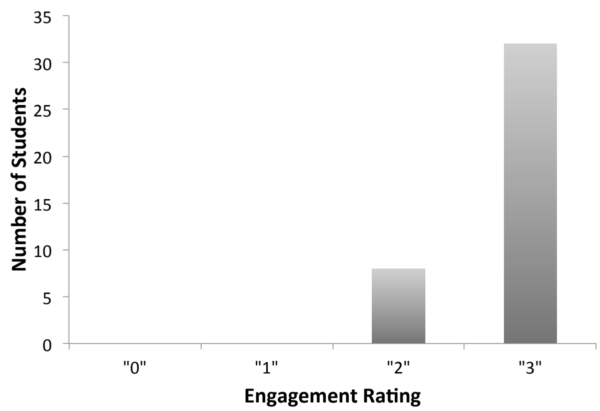


Figure 5. Number of students with a mode engagement rating of 0, 1, 2, or 3.

The results of the Spearman's rank-order test combined with the lack of variability in student engagement scores and the large proportion of high scores (suggesting a ceiling effect) informed the decision to exclude *student engagement* from the multiple regression models. Analyses were conducted with the remaining variables: (1) cumulative minutes spent in intervention, (2) cumulative number of 1 min readings, (3) cumulative number of words read correctly, and (4) oral reading accuracy.

Procedures

VISTA Training

VISTAs were trained before the start of summer school by the first author and graduate assistant to conduct the repeated reading intervention. The training included modeling and practice opportunities so that each interventionist had the opportunity to conduct the steps of intervention, record data on variables related to implementation, and ask any questions they had about the protocol. The VISTAs were also encouraged to practice conducting the intervention with each other and independently.

Pretest

Before intervention sessions began, the first author and graduate assistant administered the pre-test to each participant who returned his or her consent form and provided assent. WRCM and ERR were recorded for each of three passages read. A final ORF score was recorded as median WRCM and median ERR. Accuracy was also calculated and recorded.

Participant Identification and Assignment to Conditions

Pre-test scores were used to identify students who met inclusion criteria (ORF < 50^{th} percentile; accuracy ≥ 93% on the grade level pre-test probe). Any student who did not achieve the accuracy criterion was directed to read three additional probes one grade level lower. If accuracy improved to the criterion level, he or she was admitted to the study. Students were then randomly assigned to intervention and control conditions according to methodology described in the *experimental design and data analysis* section.

Intervention Sessions

Students in the intervention group were assigned to a dyad with another student in the same grade who read at a similar speed and the same level of difficulty. Teacher schedules and opinions about groupings were also considered on a case-by-case basis. Each intervention session followed the same procedure and took place during the independent work portion of the literacy block, in a quiet area of the school.

Just before the scheduled intervention time, the VISTA gathered his or her intervention kit and dyad of students, and proceeded to the space allocated for interventions. The VISTA greeted the students and positively affirmed their attendance as they got settled into their seats. Timers were then started (one designated for each student in the dyad) in order to keep track of the time spent in intervention. If at any point during the intervention a student left (e.g., to use the restroom, eloped) the VISTA paused that student's timer until he or she returned.

The VISTA then conducted the steps of the repeated reading intervention according to the specified protocol (see Appendix I). First, the objective was stated, which was to increase reading fluency in terms of rate, accuracy, and expression. Second, baseline data were collected. To do so, the VISTA took turns conducting a 1 min timed reading (READ 1) with each student while recording WRCM and ERR. Corrective feedback was provided after each reading, which consisted of pointing in turn to each word the student missed and saying, "This word is ___. What word is this?" After the student correctly read the word, the VISTA said, "Yes, that word is ___. Please re-reread from here." (gesturing). The student reread the sentence containing the missed word. The VISTA also provided feedback on the student's expression, graphed WRCM and errors on the student's fluency graph, and showed the results to the student. Meanwhile, the student who was not currently doing the baseline reading sat approximately 1-2 meters to the side, and was given a high-interest book to peruse. Third, the VISTA modeled fluency by reading a short section of the passage aloud after which students briefly stated what the passage was mostly about. Fourth, students completed three more 1 min readings (READS 2, 3, 4). Specifically, the students took turns reading the passage aloud in 1 min intervals, and were encouraged to try and read more than they did the time before. While one student read aloud, the VISTA marked WRCM and ERR while the other student read along silently or in a whisper. After each reading, corrective feedback, as well as feedback on expression and WRCM and ERR were provided to the student in the same manner as for the baseline reading, except for graphing WRC and ERR, which was only done again after READ 4. The graph served to provide students with a graphic representation of their progress in the intervention at that point, as students were able to see how much they improved from READ 1 to READ 4. Prior to the READ 4, students were also prompted to think about the most important thing they learned from the passage; after READ 4 this was discussed briefly. Fifth, students were asked to predict what the rest of the story would be about after which the VISTA withdrew the passages, read the remainder of the passage aloud, and students discussed whether their predictions were accurate and what the passage made them think about.

After the intervention concluded, the VISTA immediately stopped the timers for both students, recorded the number of minutes each student spent in intervention, and rated their behavior as a 0, 1, 2, or 3. Students then chose a sticker, placed it on their attendance chart, and were returned to class.

Data Collection for Student Input Variables

The following section describes how data on student input variables were collected and recorded. See Appendices G and H.

1. Cumulative minutes spent in intervention. The VISTA used timers, each labeled with the name of one student in the dyad, to record the total minutes of intervention during each session. At the start of a session, the VISTA started the stopwatches designated for each student present. If a student came late, the VISTA started that student's stopwatch when they arrived. Similarly, if a student left the group, the VISTA stopped that student's stopwatch until he or she returned. When the session concluded, the VISTA stopped both stopwatches and recorded the total minutes of intervention for each student in the intervention log. The total minutes of intervention for each session were summed for each student to yield the total minutes of RR intervention that he or she received across all intervention sessions the student attended.

2. Percentage of 1 min readings completed. Each student was expected to complete four 1 min readings of the assigned passage during each intervention session. This included one baseline read and three additional practice readings. Students also "whisper-read" during their partner's 1 min readings, but these were not recorded or monitored for errors. The number of 1 min readings for which data were recorded in the intervention logs established the total number of readings. The percentage of 1 min readings completed across all intervention sessions the student attended was calculated as follows:

$$\%\ 1\ min\ readings\ completed = \frac{total\ readings\ completed}{(4\ x\ total\ sessions\ attended)}\ x\ 100$$

3. Cumulative number of 1 min readings completed. The number of readings for which data were recorded in the intervention logs was summed to yield the cumulative number of 1 min readings the student completed across all intervention sessions he or she attended

4. Cumulative number of words read correctly. Each student read from his or her own copy of the reading passage during intervention sessions. As he or she did so, the

VISTA recorded intervention performance for that student. This was accomplished by marking codes on a separate paper copy of the passage that was labeled with the students' name. The total number of words read (TWR, inclusive of errors) in 1 min was coded by making a small bracket after the last word read and writing and then encircling a 1, 2, 3, or 4 next to the bracket (to denote Read 1, Read 2, etc.). For example, the VISTA wrote and circled a "1" next to the bracket for the first reading, wrote and circled a "2" next to the bracket for the second reading, and so forth. The VISTA also denoted errors (ERR) students made during each of their passage readings, utilizing the same operational definition of errors as for the pre- and post- ORF assessments. ERRs were coded by writing a 1, 2, 3, or 4 above any word the student missed (to denote Read 1, Read 2, etc.). For example, the VISTA wrote a "1" over ERRs made during the first reading of the passage, a "2" over ERRs made during the second reading, and so forth. See Appendix A for a sample of how passages were coded. When the session concluded, the total number of words read correctly (WRC) for each passage reading was calculated as follows and written in the intervention log:

$$WRC = TWR - ERR$$

The total number of WRC for each passage reading during an intervention session was summed, and totaled across all intervention sessions the student attended in order to yield the cumulative words read correctly.

5. Mean oral reading accuracy. The researchers used WRC and ERRs recorded in intervention logs to calculate a percentage of daily oral reading accuracy as follows:

$$\begin{matrix} daily\ mean\ \% \\ oral\ reading\ accuracy \end{matrix} = \frac{total\ WRC\ of\ all\ reads}{total\ WRC\ of\ all\ reads\ + total\ ERR\ of\ all\ reads}\ x\ 100$$

The percentages of daily oral reading accuracy for each intervention session were used to determine a mean oral reading accuracy across all intervention sessions the student attended, calculated as follows:

$$mean\ \%\ oral\ reading\ accuracy = \frac{sum\ of\ all\ mean\ \%\ daily\ oral\ reading\ accuracy}{total\ sessions\ attended}$$

6. Student engagement. Student engagement utilized a 0-3 rating scale. Operational definitions for each level of the rating scale were provided to aid the VISTA in making this judgment: 0 = not engaged (present in intervention or intervention area, but does not read at all, not looking at materials during the intervention); 1 = minimally engaged; present in the intervention area, reads 1x through the passage, is sometimes looking at the materials (e.g., while partner reads); 2 = mostly engaged; present in the intervention area, reads through the passage 2-3x, is mostly looking at the materials (e.g., while partner reads); 3 = fully engaged; present in the intervention area, reads through the passage 4x, is always looking at the materials (e.g., while partner reads). Each day, VISTAs recorded in the intervention logs an individual rating of student engagement. Student engagement was recorded as the mode of the VISTA's ratings across all intervention sessions the student attended.

Posttest

After the last intervention session conducted during the summer program, the first author and graduate assistant administered the posttest to each participant in the intervention and control groups, using procedures that were identical to the pre-test. WRCM and ERR were recorded for each of three passages read. A final ORF score was recorded as median WRCM and median ERR. Accuracy was also recorded.

CHAPTER FOUR

RESULTS

Purpose and Research Questions

The purpose of this chapter is to report the results of the current investigation. The chapter commences with a review of the research questions that were posed, and proceeds to review the results of analyses and relevant data pertaining to each question. The study addressed the following a priori research questions:

1. What are the effects on oral reading fluency of a repeated reading intervention implemented during four weeks of summer school with students in 2^{nd} and 3^{rd} grade whose reading is accurate but slow?
2. What dimensions of RR implementation that include student inputs predict and/or correlate with posttest oral reading fluency scores, controlling for students' oral reading fluency at pretest?

   The following variables were analyzed:

   a. Cumulative minutes spent in intervention
   b. Cumulative number of 1 min readings completed
   c. Cumulative number of words read correctly
   d. Mean oral reading accuracy

Information gathered in the course of answering the a priori research questions also led to two post-hoc questions:

1. Was grade level a factor with regard to the effects on oral reading fluency of the repeated reading intervention, implemented during four weeks of summer

school with students in 2nd and 3rd grade whose reading is accurate but slow?

2. Is there a subset of students for whom the effects on oral reading fluency of a repeated reading intervention are greater?

Research Question 1

The first research question examined the effects on oral reading fluency (ORF) of a RR intervention implemented with students in 2nd and 3rd grade whose reading was accurate but slow, for four weeks of a summer program. A one-way analysis of covariance (ANCOVA) was conducted to answer the research question using the following equation:

$$Y_{ij} = \mu + \beta X_{ij} + \alpha_i + \epsilon_{ij}$$

The results of the ANCOVA [between subjects factor: condition (intervention, control); covariate: pretest score] are displayed in Tables 4 and 5. There was a significant main effect for condition on levels of ORF posttest scores, controlling for differences in pretest scores, $F(1, 76) = 4.17$, $p = .045$, $\alpha = .05$, *n*p2 = 0.05, Hedge's g = .08. That is, accounting for initial pretest level, ORF scores of students who received the RR intervention increased more than those of students in the control group. Effect size was also calculated to facilitate judgment of the overall impact of repeated reading interventions. As a reference, partial eta-squared effect sizes of 0.01 are considered small and effect sizes of 0.06 are considered moderate. However, partial eta-squared can over-estimate effects. Therefore, Hedge's g effect size was also computed, which adjusts for sample size. Hedge's g effect sizes of 0.2, 0.5, and 0.8 are considered to be small, medium, and large, respectively (Cohen, 1988). The partial eta-squared effect size for condition was medium while the Hedge's g effect size was very small.

The means of the ORF posttest were ordered as expected across conditions, with higher mean ORF posttest scores for the intervention group, after adjusting for initial differences in pretest ORF. The intervention group had the highest mean (*M* = -.77) while the control group had the lowest mean (*M* = -.91), which was a significant difference according to Fisher's Least Significant Difference (LSD) procedures (Fisher 1977).

Table 4

Standardized ORF Pre- and Posttest Means as a Function of Condition

		ORF Pretest (WRCM)		ORF Posttest (WRCM)	
Condition	*n*	*M*	*SD*	*M*	*SD*
Intervention	40	-1.07	.97	-.80	.97
Control	39	-1.01	.81	-.87	.83

Note. ORF = oral reading fluency. WRCM = words read correctly per minute. *M* = mean. *SD* = standard deviation.

Table 5

Analysis of Covariance of ORF Posttest Scores as a Function of Condition (Intervention, Control), With ORF Pretest Scores as Covariate

Source	*df*	*SS*	*MS*	*F*	*p*	η_p^2	*Hedge's g (CI)*
Covariate (ORF)	1	69.73	69.73	823.77	<.000	.92	
Condition	1	.35	.35	4.17	.045	.05	.08(-.37-.08)
Error	76	6.43	.05				
Total	79	131.47					

Note. ORF = oral reading fluency.

Examining Unstandardized Student Data

In addition to examining the statistical differences in the standardized ORF rates (WRCM) for students in intervention and control groups overall, unstandardized (raw) results for both groups were also examined for practical consideration. Figure 6 and Table 6 display the unstandardized ORF scores for intervention and control groups at pre- and post-test. Students in second and third grades are generally expected to make a weekly gain of about 1.5 WRCM during the academic year (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). On average, students who received repeated reading interventions increased their oral reading rate by approximately twice as many WRCM as students in the control group. For second and third grade students receiving interventions, there was a mean change of 8.15 WRCM over four weeks, translating to an average of 2.04 words per week, which is above the expected 1.5 WRCM. For second and third grade students in the control group, there was a mean change of 4.13 WRCM of over four weeks, translating to an average of 1.03 words per week, which is below the expected 1.5 WRCM (see Table 6 and Figure 7).

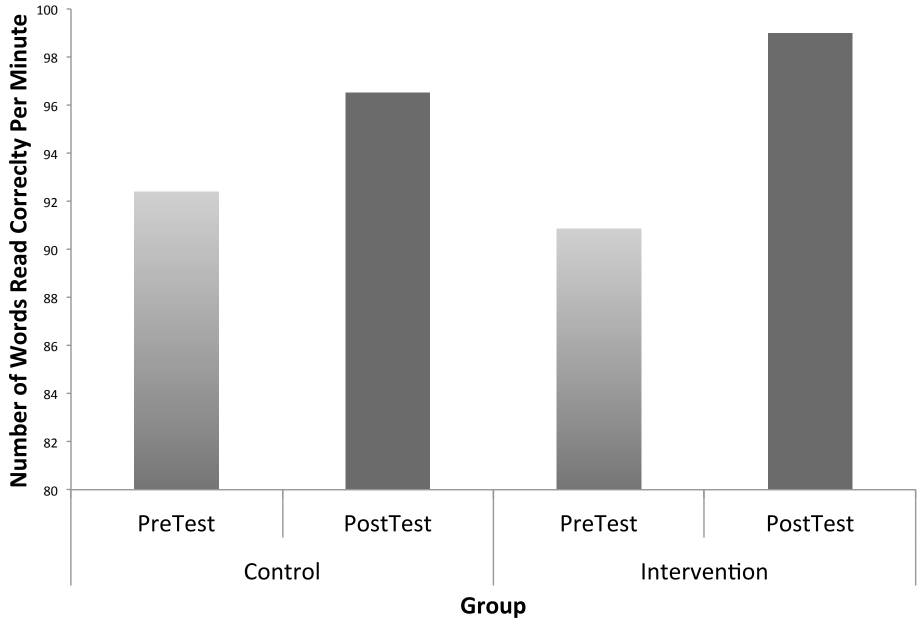


Figure 6. Comparison of mean unstandardized pre- and post- test scores in the form of number of words read correctly per minute, for intervention and control group.

Table 6

Unstandardized ORF Pre- and Posttest Means as a Function of Condition

		ORF Pretest (WRCM)		ORF Posttest (WRCM)		Total change in WRCM	Mean change in WRCM/week[a]
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*M*
Intervention	40	90.85	28.64	99.00	33.17	8.15	2.04
Control	39	92.38	25.60	96.51	27.21	4.13	1.03

Note. ORF = oral reading fluency. WRCM = words read correctly per minute. *M* = mean. *SD* = standard deviation.
[a]Interventions were conducted 4x weekly for four weeks.

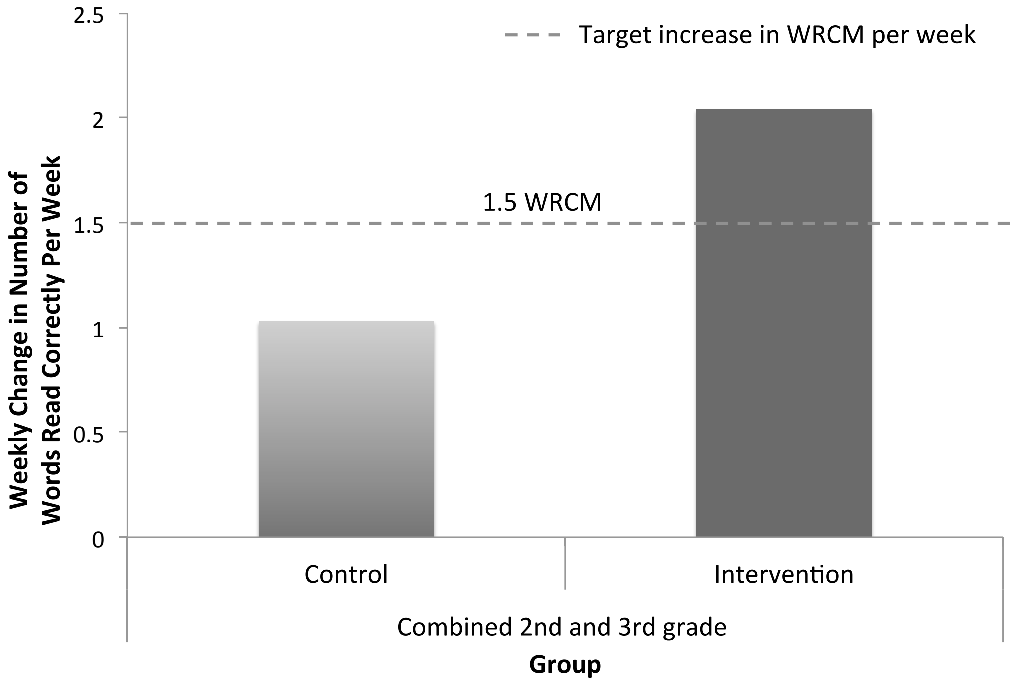


Figure 7. Comparison of unstandardized mean weekly change from pre- to post- test in the form of number of words read correctly per minute, for intervention and control groups. Realistic/ambitious standards of weekly growth for grade 2, 3 respectively: 1.5/2.0; 1.0/1.5 (Fuchs et al., 1993).

The percentage of students in intervention and control groups who maintained or increased their ORF rate was also calculated. Results indicate that students in the intervention group were more likely to increase their oral reading fluency rate during summer school. Specifically, 36 out of 40 students who received interventions (90%) maintained or increased their oral reading rate compared to 28 out of 39 students in the control group (72%). Put another way, nearly one-third of the students in the control group experienced a decrease in their oral reading fluency over the course of the 4-week summer program, compared to one-tenth of the students who received intervention.

Post-hoc Analysis

In the course of disaggregating the data to examine unstandardized gains in WRCM per week by grade, the results led to a series of post-hoc questions that required additional analyses. Specifically, ANCOVA results constituted evidence that RR was effective overall, yet there were unexpected differences in the unstandardized growth of second grade students relative to the control group when compared to the unstandardized growth of third grade students relative to the control group. First, the mean change from pre- to post-test was more pronounced for second than for third graders. Second, the mean change in WRCM was fairly similar for third grade students in intervention and control groups (see Table 7 and Figure 8). Taken together, these results seemed to suggest differences in the effects of RR by grade.

Table 7

Unstandardized ORF Pre- and Posttest Means as a Function of Condition and Grade

		ORF Pretest (WRCM)		ORF Posttest (WRCM)		Total change in WRCM	Mean change in WRCM/week[a]
Group	*n*	*M*	*SD*	*M*	*SD*	*M*	*M*
Second grade							
Intervention	22	87.45	25.83	96.36	27.33	8.91	2.23
Control	22	85.64	22.57	87.91	23.95	2.27	0.57
Third grade							
Intervention	18	95.00	31.99	102.22	39.75	7.22	1.81
Control	17	101.12	27.28	107.65	27.79	6.53	1.63

Note. ORF = oral reading fluency. WRCM = words read correctly per minute. *M* = mean. *SD* = standard deviation.
[a]Interventions were conducted 4x weekly for four weeks.

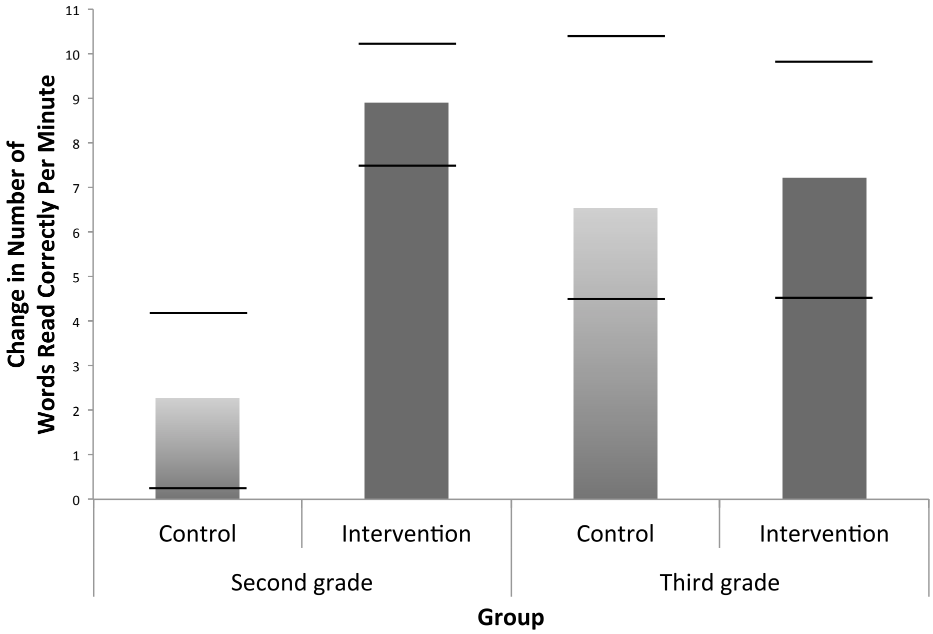


Figure 8. Comparison of unstandardized mean change from pre- to post- test in the form of number of words read correctly per minute, for intervention and control groups. Horizontal lines represent standard error.

Post-Hoc Research Question 1

The first post-hoc research question examined whether effects on oral reading fluency of the RR intervention differed for students by grade. A two-way ANCOVA was conducted to answer this question.

The results of the two-way ANCOVA [between subjects factors: condition (intervention, control), grade (2, 3); covariate: pretest score] are displayed in Tables 8 and 9. For this ANCOVA model, there was a non-significant main effect for condition on levels of ORF posttest scores, controlling for differences in pretest scores, $F(1, 74) = 3.54$, $p = .064$, $\alpha = .05$, np2 = 0.05. This change in significance for condition (compared to the results of the one-way ANCOVA) may be due to a loss of degrees of freedom, which can impact the precision of the model. There was also no significant effect for grade, $F(1, 74) = 1.34$, $p = .252$, $\alpha = .05$, $n\text{p}2 = 0.02$, and no significant interaction of condition x grade, $F(1, 74) = 1.73$, $p = .193$, $\alpha = .05$, $n\text{p}2 = 0.02$.

The means of the ORF posttest were ordered as expected across conditions, with higher ORF posttest scores for students in intervention group, and in third grade, after adjusting for initial differences in pretest ORF. The intervention group had the highest mean ($M = -.77$) while the control group had the lowest mean ($M = -.89$) and the third grade students had the highest mean ($M = -.79$) while second grade students had the lowest mean ($M = -.87$). Additionally, second and third grade students in the intervention group had higher means ($M = -.77, -.78$, respectively) while second and third grade students in the control group had lower mean ($M = -.98, -.81$, respectively). However, none of these differences were significant. In sum, results did not support that effects on ORF were dependent on students' grade in school.

Table 8

Standardized ORF Pre- and Posttest Means as a Function of Condition and Grade

		ORF Pretest		ORF Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*
Intervention					
Second Grade	22	-.83	.83	-.83	.88
Third Grade	18	-1.36	1.07	-1.12	.88
Control					
Second Grade	22	-.89	.73	-.82	.77
Third Grade	17	-1.16	.91	-.94	.77

Note. ORF = oral reading fluency. *M* = mean. *SD* = standard deviation.

Table 9

Analysis of Covariance of ORF Posttest Scores as a Function of Condition (Intervention, Control) and Grade (2, 3), with ORF Pretest Scores as Covariate

Source	*df*	*SS*	*MS*	*F*	*p*	η_p^2	*Hedge's g (CI)*
Covariate (ORF)	1	66.55	66.55	797.18	<.001	.92	
Condition	1	.30	.30	3.54	.064	.05	.07(-.37-.51)
Grade	1	.11	.11	1.34	.252	.02	.35(-.09-.80)
Cond. x grade	1	.14	.14	1.73	.193	.02	
Error	74	6.17	.08				
Total	79	131.47					

Note. ORF = oral reading fluency.

Post-Hoc Research Question 2

The results of the first post-hoc question indicated that there was no significant effect for grade on oral reading fluency outcomes, suggesting that differential results may only have *appeared* to be grade-related. It was plausible that observed differences were actually due to another factor such as reading level, or to any interaction between the factors of level, grade, or condition. Thus, the second post-hoc research question examined whether the effects on oral reading fluency of a repeated reading intervention were more robust for students according to their reading level. Students were designated as relatively high-level readers (WRCM = 26^{th}-50^{th} percentile) or low-level readers (WRCM = 0-25^{th} percentile) using grade-based norms (T. Christ, personal communication, June 20, 2014). Two-way and three-way ANCOVAs were conducted to explore this question.

The results of the two-way ANCOVA [between subjects factors: condition (intervention, control), reading level (relatively high, low); covariate: pretest score] are displayed in Tables 10 and 11. The predicted main effect of condition was significant, $F(1, 74) = 7.53$, $p = .008$, $\alpha = .05$, $n\text{p}2 = 0.09$. There was not a significant effect for level, $F(1, 74) = .53$, $p = .469$, $\alpha = .05$, $n\text{p}2 = 0.01$. However, the interaction of condition x level was significant, $F(1, 74) = 5.82$, $p = .018$, $\alpha = .05$, $n\text{p}2 = 0.07$ (see Figure 9), indicating that relative to the control group and controlling for pretest differences, there were larger posttest scores for relatively high-level readers who received the RR intervention than for low-level readers who received the same intervention. Thus, the effect of condition was not uniform across all participants, but dependent on reading level.

The means of the ORF posttest were ordered as expected for condition, after adjusting for initial differences in pretest ORF. The intervention group had the highest mean ($M = -.76$) while the control group had the lowest mean ($M = -.94$), and the difference was statistically significant according to LSD procedures. The means of the ORF posttest were not ordered as expected for reading level however. *Low*-level students had the highest mean ($M = -.81$) and *relatively high*-level students had the lowest mean ($M = -.89$), although the difference was small, and not significant. Additionally, there was evidence of an interaction between reading level and condition such that relatively high-level readers appeared to benefit more from the RR intervention. Specifically, relatively high-level readers in the intervention group had the highest mean ($M = -.72$) while relatively high-level readers in the control group had the lowest mean ($M = -1.06$), and the difference was statistically significant. The pattern was somewhat similar for low-level readers, wherein low-level readers in the intervention group had the highest mean (*M* = -.80), low-level readers in the control group had the lowest mean (*M* = -.83), however the difference was *not* statistically significant.

Table 10

Standardized ORF Pre- and Posttest Means as a Function of Condition and Reading Level

Group	*n*	ORF Pretest *M*	ORF Pretest *SD*	ORF Posttest *M*	ORF Posttest *SD*
Relatively high level					
Intervention	16	-.18	.33	.21	.37
Control	13	-.15	.33	-.10	.46
Low level					
Intervention	24	-1.66	.78	-1.48	.93
Control	26	-1.44	.62	-1.26	.70

Note. ORF = oral reading fluency. *M* = mean. *SD* = standard deviation. Reading level: relatively high = ORF rate at the 26-50th percentile, low = ORF rate at the 0-25th percentile.

Table 11

Analysis of Covariance of Median Number of ORF Posttest Scores as a Function of Condition (Intervention, Control) and Reading Level (Relatively high, Low), with ORF Pretest Scores as Covariate

Source	*df*	*SS*	*MS*	*F*	*p*	η_p^2
Covariate (ORF)	1	30.92	30.92	396.70	<.001	.84
Condition	1	.60	.60	7.53	.008	.09
Level	1	.04	.04	.53	.469	.01
Cond. x level	1	.47	.47	5.82	.018	.07
Error	74	5.92	.08			
Total	79	131.47				

Note. ORF = oral reading fluency. Reading level: relatively high = ORF rate at the 26-50th percentile, low = ORF rate at the 0-25th percentile.

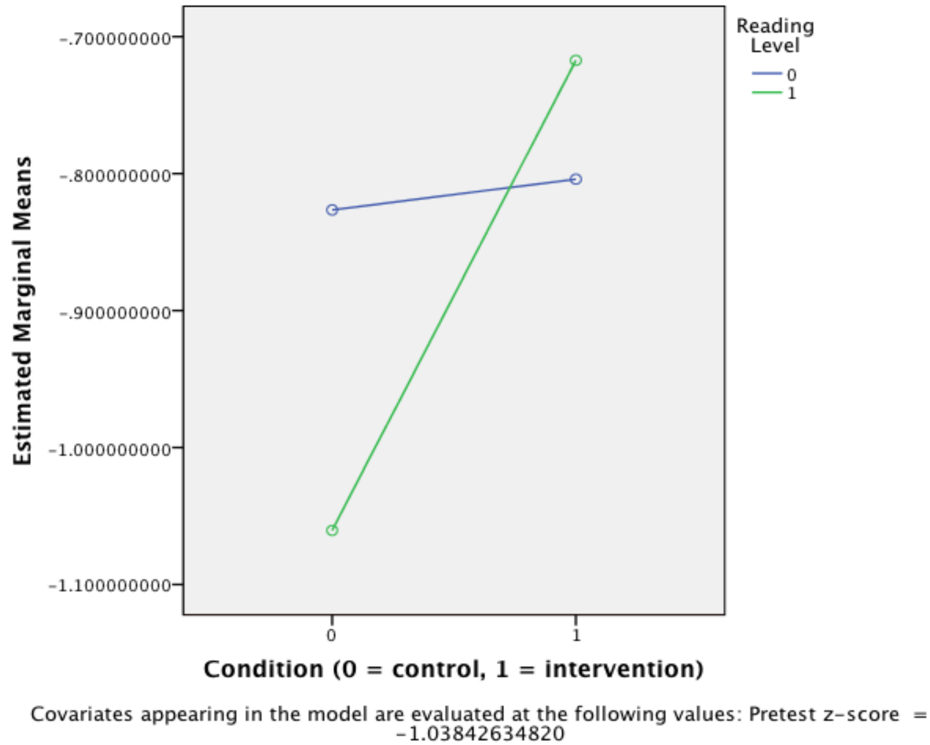


Figure 9. Estimated marginal means of oral reading fluency posttest z-scores for condition, by reading level. Reading level: 1 = relatively high; ORF rate at the 26-50th percentile, 0 = low; ORF rate at the 0-25th percentile.

The results of the three-way ANCOVA [between subjects factors: condition (intervention, control), grade (2, 3), reading level (relatively high, low); covariate: pretest score] are displayed in Tables 12 and 13. Interactions were also examined, including condition x grade, condition x level, grade x level, and condition x grade x level. Similar to the results of the two-way ANCOVA, the predicted main effect of condition was again significant, $F(1, 70) = 7.91$, $p = .006$, $\alpha = .05$, np2 = 0.10, as well as the interaction of condition x level, $F(1, 70) = 6.28$, $p = .015$, $\alpha = .05$, np2 = 0.08. There were no significant effects for level, $F(1, 70) = .26$, $p = .613$, $\alpha = .05$, $n\text{p}2 = 0.01$, grade, $F(1, 70) = 2.60$, $p = .112$, $\alpha = .05$, $n\text{p}2 = 0.04$, or for the interactions of condition x grade, $F(1, 70) = .53$, $p = .469$, $\alpha = .05$, $n\text{p}2 = 0.01$, grade x level, $F(1, 70) = 1.49$, $p = .227$, $\alpha = .05$, $n\text{p}2 = 0.02$, or condition x grade x level, $F(1, 70) = .93$, $p = .337$, $\alpha = .05$, $n\text{p}2 = 0.01$.

Like the previous analyses, the means of the ORF posttest were once again ordered as expected for condition, after adjusting for initial differences in pretest ORF. The intervention group had the highest mean ($M = -.74$) while the control group had the lowest mean ($M = -.94$), and the difference was statistically significant according to LSD procedures. The interaction between reading level and condition also remained, wherein relatively high-level readers in the intervention group had the highest mean ($M = -.68$) while relatively high-level readers in the control group had the lowest mean ($M = -1.04$), and the difference was statistically significant; low-level readers in the intervention group had the highest mean ($M = -.80$) and readers in the control group had the lowest mean ($M = -.82$), and the difference was *not* statistically significant. These results provide additional support that effects on ORF for condition were not uniform, but depended on students' reading level.

Table 12

Standardized ORF Pre- and Posttest Means as a Function of Condition, Reading Level, and Grade

		ORF Pretest		ORF Posttest	
Group	*n*	*M*	*SD*	*M*	*SD*
Relatively high level, grade 2					
Intervention	11	-.18	.38	.15	.38
Control	8	-.15	.35	-.16	.55
Relatively high level, grade 3					
Intervention	5	-.18	.23	.35	.37
Control	5	-.15	.32	.01	.33
Low level, grade 2					
Intervention	11	-1.48	.60	-1.24	.64
Control	14	-1.31	.50	-1.19	.62
Low level, grade 3					
Intervention	13	-1.82	.90	-1.69	1.10
Control	12	-1.58	.72	-1.33	.80

Note. WRCM = words read correctly per minute. *M* = mean. *SD* = standard deviation. Reading level: relatively high = ORF rate at the 26-50th percentile, low = ORF rate at the 0-25th percentile.

Table 13

Analysis of Covariance of ORF Posttest Scores as a Function of Condition (Intervention, Control), Reading Level (Relatively high, Low), and Grade (2, 3), with ORF Pretest Scores as Covariate

Source	*df*	*SS*	*MS*	*F*	*p*	η_p^2
Covariate (ORF)	1	29.78	29.78	379.43	<.001	.84
Condition	1	.62	.62	7.91	.006	.10
Level	1	.02	.02	.26	.613	.00
Grade	1	.20	.20	2.60	.112	.04
Cond. x level	1	.49	.49	6.28	.015	.08
Cond x grade	1	.04	.04	.53	.469	.01
Level x grade	1	.12	.12	1.49	.227	.02
Cond. x level x grade	1	.07	.07	.93	.337	.01
Error	70	5.49	.08			
Total	79	131.47				

Note. ORF = oral reading fluency. Reading level: relatively high = ORF rate at the 26-50th percentile, low = ORF rate at the 0-25th percentile.

Research Question 2

The second research question examined the degree to which student input variables related to implementation predicted and/or correlated with posttest oral reading fluency scores while controlling for students oral reading fluency level at pretest. The specific variables that were examined included: (1) cumulative minutes spent in intervention, (2) cumulative number of 1 min readings completed, (3) cumulative number of words read correctly, and (4) mean oral reading accuracy. Note that two variables, *student engagement* and *percentage of 1 min readings completed,* were eliminated from the analysis (see Student Input Variable section of Method for specifics). Definitions and rationales for examining each variable as a potential predictor of oral reading fluency are specified in Appendix C.

Multiple linear regression analysis was conducted to analyze the research question, and models were fitted where combinations of students' pretest ORF and variables related to implementation predicted posttest ORF.

Pearson's product-moment correlations were first calculated and are displayed in Table 14. There were significant moderate to strong correlations between all variables and ORF pretest and posttest scores, and a very strong correlation between pre- and posttest scores (r = 0.97, p = <.001).

Table 14

Descriptive Statistics and Bivariate Correlations for all Variables

Variables	*M*	*SD*	1 Pretest	2 Posttest	3 Cum. min	4 Cum. 1 min readings	5 Cum. WRC	6 Mean acc.
1. Pretest	-1.07	1.13	--					
2. Posttest	-0.80	1.13	0.97 <.001 .	--				
3. Cumulative minutes intervention	297.18	75.81	0.34 .031 .	0.34 .031	--			
4. Cumulative 1 min readings	55.30	8.92	0.32 .042	0.32 .048	0.63 <.001	--		
5. Cumulative words read correctly	5738.05	2000.49	0.76 <.001	0.78 <.001	0.22 .177	0.59 <.001	--	
6. Mean accuracy	0.99	0.01	0.57 <.001	0.60 <.001	-.0.15 .349	0.02 .912	0.58 <.001	--

Note. α = .05. Cum. = cumulative. Min = minutes. WRC = words read correctly. ACC. = accuracy.

Partial correlations were therefore conducted where ORF pretest scores were controlled on the association between ORF posttest scores and variables related to implementation. Results indicate that there were no significant correlations with variables related to implementation and ORF posttest scores (Table 15). However, computing the square of each Pearson's correlation coefficient provides some information on the patterns of these relations. Specifically, cumulative number of words read correctly and oral reading accuracy accounted for more variance than cumulative minutes spent in intervention and cumulative number of 1 min readings (6.7%, 3.8%, respectively versus 0.16%, 0.0036%, respectively), although the relation was non-significant. There were also significant correlations between the variables related to implementation. For example, there were statistically significant strong positive correlations between cumulative 1 min readings and cumulative minutes in intervention ($r = .59$, $p = <.001$), and also with cumulative words read correctly ($r = .55$, $p = <.001$). Notably, cumulative 1 min readings can be conceptualized as a measure of attendance because in this investigation, the data indicated students always completed four readings per session. These results therefore indicate that attendance was closely related to the total number of minutes of intervention and the total number of words read. There was also a statistically significant moderate negative correlation between cumulative minutes in intervention and oral reading accuracy ($r = -.45$, $p = .004$), indicating that as accuracy decreased, cumulative minutes of intervention increased.

Table 15

Partial Correlations Between Oral Reading Fluency Posttest Scores and Variables Related to Implementation, Controlling for Oral Reading Fluency Pretest Scores

Variables	1 Posttest	2 Cum. min.	3 Cum 1 min readings	4 Cum. WRC	5 Mean acc.
1. Posttest	--				
2. Cumulative minutes intervention	.04 .809	--			
3. Cumulative 1 min readings	.01 .970	.59 <.001	--		
4. Cumulative words read correctly	.26 .111	-.07 .687	.55 <.001	--	
5. Oral reading accuracy	.20 .231	-.45 .004	-.22 .188	.27 .100	--

Note. α = .05. Cum. = cumulative. Min = minutes. WRC = words read correctly. ACC. = accuracy.

This analysis comprised three multiple linear regression models that were fitted to predict posttest ORF based on variables related to implementation as well as pretest ORF. Model 1 was fitted where students' pretest ORF predicted students' posttest ORF. The resulting regression equation indicated that the model explained a significant 94% of the variance, $R^2 = 0.941$, $F(1, 38) = 625.91$, $p < .001$ (see top panel, Table 16). Pretest ORF had a significant positive regression weight, $\beta = .97$, $p = <.001$, and was a significant predictor of ORF posttest scores.

Model 2 was fitted where students' pretest ORF and mean oral reading accuracy predicted students' posttest ORF. The resulting regression equation indicated that the model explained a significant 95% of the variance, $R^2 = 0.945$, $F(2, 37) = 317.68$, $p < .001$ (see middle panel, Table 16). Similar to Model 1, pretest ORF had a significant positive regression weight, $\beta = .94$, $p = <.001$, and significantly predicted posttest ORF. Accuracy had a non-significant positive regression weight, $\beta = .06$, $p = .231$, and was not a significant predictor of posttest ORF.

Model 3 was fitted where students' pretest ORF, mean oral reading accuracy, and the remaining variables related to implementation predicted posttest ORF. All variables were entered into the model simultaneously. The resulting regression equation indicated that the model explained a significant 95% of the variance, $R^2 = 0.953$, $F(5, 34) = 138.62$, $p < .001$ (see bottom panel, Table 16). Similar to Models 1 and 2, pretest ORF had a significant positive regression weight, $\beta = .80$, $p = <.001$, and was a significant predictor of posttest ORF while accuracy had a non-significant positive regression weight, $\beta = .04$, $p = .453$, and was not a significant predictor of posttest ORF. Of the remaining variables, only cumulative words read correctly had a significant positive regression weight, $\beta = .20$, $p = .031$, thus the number of words read correctly by students across all intervention sessions was a significant predictor of posttest ORF. Cumulative minutes in intervention had a non-significant positive regression weight, $\beta = .11$, $p = .069$ and cumulative 1 min readings had a non-significant negative regression weight, $\beta = -.13$, $p = .074$, thus neither was a significant predictor of posttest ORF.

Table 16

Regression Analysis Summary for Variables Related to Implementation Predicting Oral Reading Fluency Posttest Scores, Models 1, 2, and 3

Variable	B (SE)	β	t	p	95% CI
	Model 1				
1. Pretest	1.13(.05)	.97	25.02	<.001	[1.04-1.22]
	Model 2				
1. Pretest	1.09(.06)	.94	19.94	<.001	[.98-1.20]
2. Accuracy	6.00(4.93)	.06	1.22	.23	[-3.98-15.99]
	Model 3				
1. Pretest	.93 (.09)	.80	10.99	<.001	[.76-1.10]
2. Accuracy	4.27 (5.62)	.04	.76	.45	[-7.15-15.68]
3. Cumulative WRC/1500[a]	.17 (.08)	.20	2.26	.03	[.02-.32]
4. Cumulative min in intervention/80[b]	.14 (.07)	.11	1.88	.07	[-.01-.28]
5. Cumulative 1 min readings/16[c]	-.27(.15)	-.13	-1.84	.07	[-.60-.03]

Note. Adjusted R^2 = .941 for Model 1: ΔR^2 = .002 for Model 2 ΔR^2 = .008 for Model 3 (*p*s < .05). CI = confidence interval for *B*. WRC = words read correctly. Variables were scaled to improve comprehension of the unstandardized beta coefficients.
[a]Cumulative WRC was divided by 1500 (average number of WRC/week = ~1500)
[b]Cumulative min in intervention was divided by 80 (average intervention minutes/week = ~80).
[c]Cumulative 1 min readings was divided by 16 (number of assigned 1 min readings/week = 16).

Examining Student Input Variables

Raw data on student input variables related to implementation were also examined descriptively to supplement information derived from the multiple regression analysis.

This section of the analysis provides information regarding differences in levels of each variable for individual students.

Cumulative words read correctly. Cumulative words read correctly was the only significant predictor of posttest ORF, and levels for individual students differed substantially. Totals ranged from 2012-10,230 words, which translates into a 134 % difference between most and fewest words read. Figure 10 is a dual-axis graph that displays students' raw values for the cumulative number of words read across the four-week summer program, in rank order from least to greatest (gray bars, scale on left). Also included is each student's change in WRCM from pretest to posttest (red bars, scale on right side of graph). The slope line indicates that on average, greater positive changes in oral reading fluency were associated with higher levels of cumulative words read.

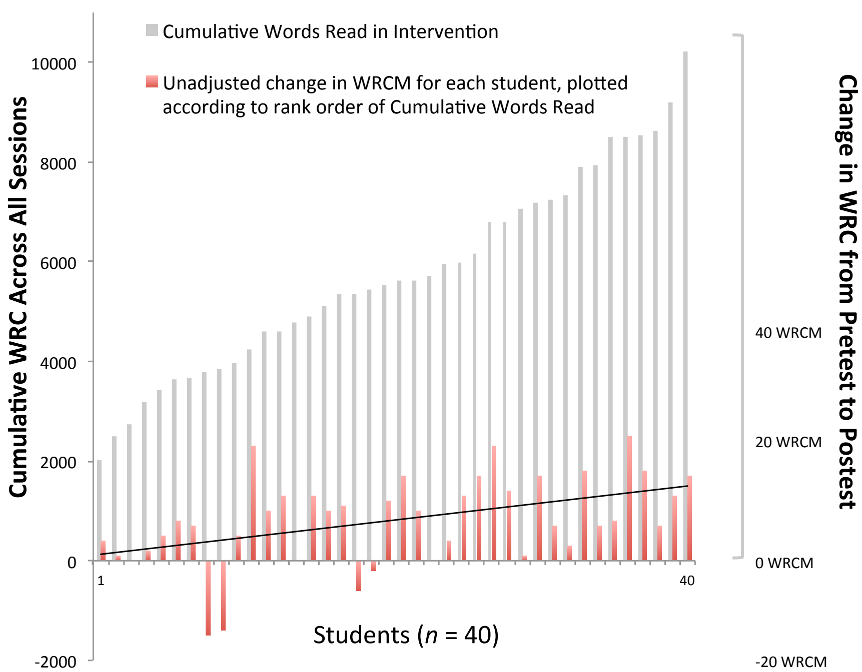


Figure 10. Individual students' cumulative number of words read correctly during intervention in rank order from least to greatest (gray bars); students' change in WRCM from pretest to posttest (red bars).

On average, students read 411 words correctly during intervention sessions. The average for individual students ranged from 212 to 682 words per session. This translated into a 105-percentage difference between the highest and lowest average words read. See Figure 11.

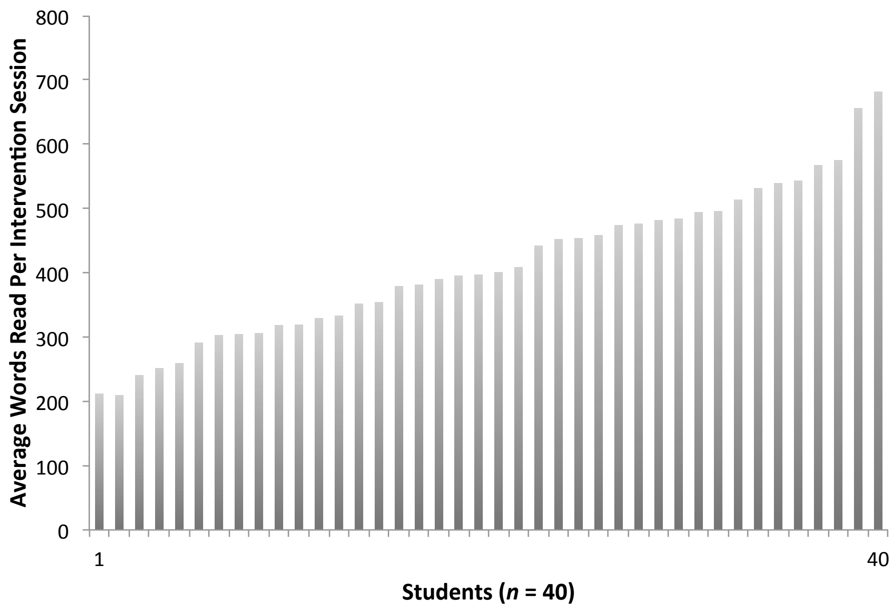


Figure 11. Individual students' average number of words read correctly per intervention session in rank order, from least to greatest.

Cumulative minutes in intervention. Cumulative minutes in intervention differed for individual students. Students received a total of 164-460 minutes of intervention, which translated into a 95% difference between most and fewest minutes of intervention. Figure 12 displays students' cumulative minutes of intervention across the four-week summer program, in rank order from least to greatest.

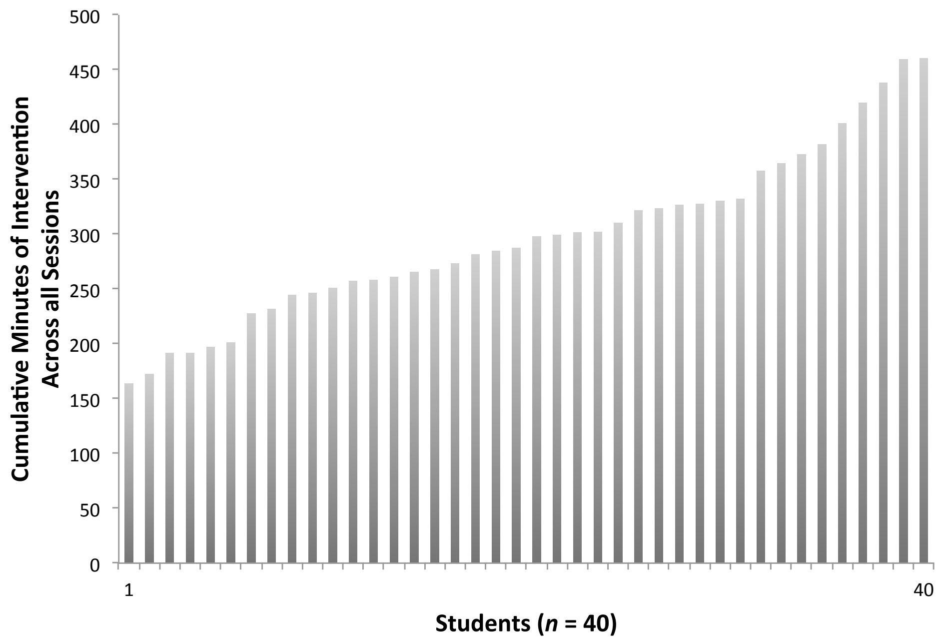


Figure 12. Individual students' cumulative minutes of intervention in rank order, from least to greatest.

Cumulative number of 1 min readings. Cumulative number of 1 min readings differed for individual students. The variable was essentially a measure of attendance in that students were assigned four 1 min readings during each intervention session and the data revealed that virtually all students read the passage four times if they attended the intervention session. The number of readings completed across all intervention sessions attended ranged from 28 readings (44% attendance; 7 sessions) to 64 readings (100% attendance; 16 sessions), which translated into a 78% difference between most and fewest 1 min readings. Figure 13 displays students' cumulative number of 1 min readings across the four-week summer program, in rank order from least to greatest.

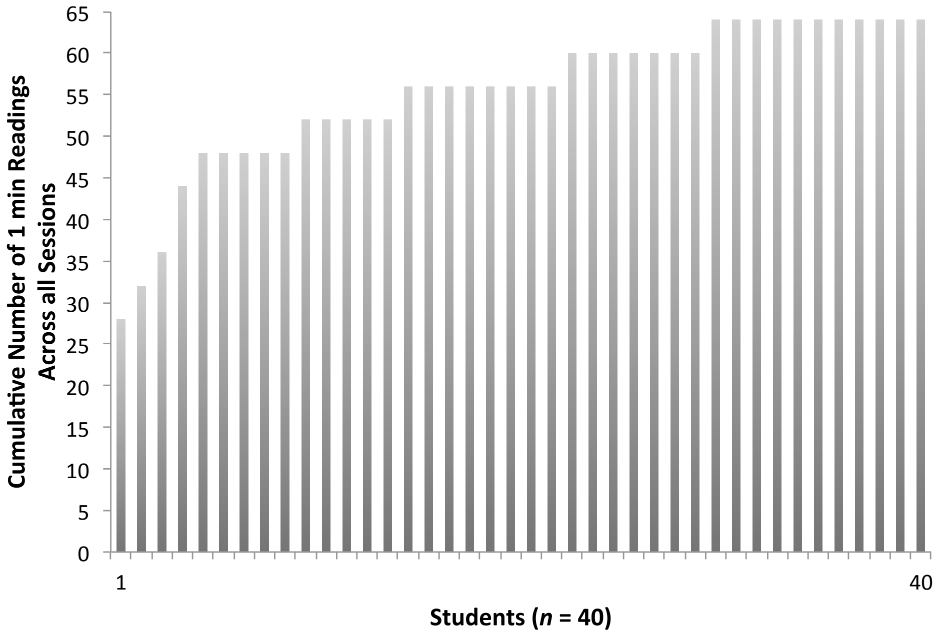


Figure 13. Individual students' cumulative number of 1 min readings in rank order, from least to greatest.

Mean oral reading accuracy. Mean oral reading accuracy differed very little for individual students. Mean accuracy for individual students ranged from 95.30% to 99.95% accuracy, which translated into a 5% difference between lowest and highest mean accuracy. Figure 14 displays students' mean oral reading accuracy across the four-week summer program, in rank order from least to greatest.

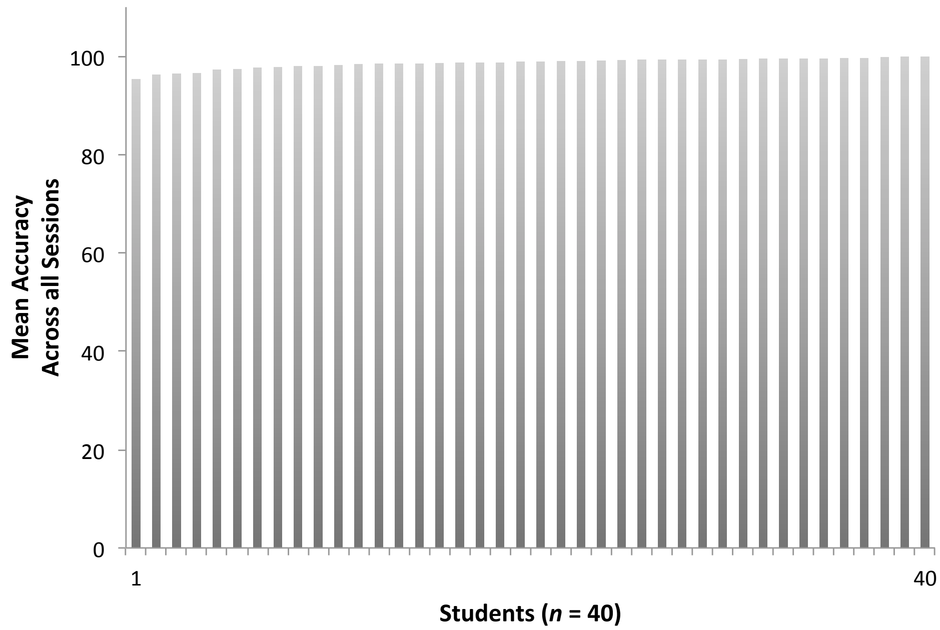


Figure 14. Students' mean oral reading accuracy in rank order, from least to greatest.

CHAPTER FIVE

DISCUSSION

Organization of the Chapter

The purpose of this chapter is to synthesize the results of the present study. First, the purpose of the study is reviewed, along with a priori and post hoc research questions. Second, results for each research question are discussed and contextualized within the existing research literature. Third, limitations of the study and directions for future research are proposed. The chapter closes with implications for practice and a final conclusion.

Review of Study Purpose and Research Questions

A troubling number of students that comprise a variety of demographic groups, both with and without disabilities, currently lag behind their more skilled peers in reading. These discrepancies in progress indicate a need to improve the effectiveness of instructional approaches and interventions so they are potent enough to *accelerate* progress and close reading achievement gaps for individual students, as well as provide a buffer for learning loss that often occurs during school breaks.

It has been proposed that one approach to addressing the reading achievement gap may be to ensure that students get "enough of what works". The underlying concept is that instructional programs and interventions that are matched to student needs will be most effective when they are implemented with a high degree of procedural fidelity and at a sufficient level of treatment intensity (i.e., dosage)—a conclusion that can be drawn logically, but for which there is also support in the research literature. Multiple studies have reported findings that substantiate the relation between the development of a practiced skill (such as reading proficiency) and the quantity and quality of practice (Bryan & Harter, 1897; Ericsson et al., 1993; Chase & Simon, 1973; Williams & Hodges, 2005).

Related to the notion of "enough of what works", two additional factors warrant consideration because they potentially impact the quantity and quality of reading practice for individual students. One factor is time spent away from school (e.g., during the summer months) when students may engage in few literacy activities. A second factor is student inputs (i.e., responses and behaviors of the student) during reading instruction and intervention. Measurement of student inputs may thus be critical to accurately evaluate the effectiveness of instructional programs and intervention. The importance of this dimension of implementation has been noted by a number of scholars (Baker, 2012; Dane & Schnieder, 1998; Jones et al., 2008; Lichstein et al., 1994), however few studies have included, mentioned, or assessed the construct (see review by Barnett et al., 2014).

The purpose of the present study was to contribute to the reading fluency literature. One focus was a well-known reading fluency practice intervention, *repeated reading* (RR) and its effects on a particular form of reading fluency, *oral reading fluency*, that can serve as an indicator of overall reading competence (Fuchs et al., 2001, p. 239). The RR intervention was implemented during a summer program.

A second focus was to examine student inputs as a dimension of implementation related to treatment intensity and procedural fidelity. Specifically, it was posited that variance in responses and behaviors during intervention could potentially alter the quantity and quality of practice, and contribute to differences in intervention effectiveness for individual students.

This investigation evaluated the effects of RR on oral reading fluency (i.e., oral reading *rate* in the form of words read correctly per minute, referred to as oral reading fluency or ORF) during four weeks of a summer program. The investigation also evaluated dimensions of RR implementation that included student inputs potentially related to or predictive of those effects. Further, there was hope of a collateral gain in terms of knowledge that could inform further study of RR and its implementation in order to increase levels of progress for *all* students for whom the intervention is deemed appropriate.

The purpose of the study was thus two-fold, with the following a priori research questions addressed:

1. What are the effects on oral reading fluency of a repeated reading intervention implemented during four weeks of summer school with students in 2nd and 3rd grade whose reading is accurate but slow?
2. What dimensions of RR implementation that include student inputs predict and/or correlate with posttest oral reading fluency scores, controlling for students' oral reading fluency at pretest?

In addition to the a priori research questions, two post hoc questions were generated in the course of analyzing the study data:

1. Is grade level a factor with regard to the effects on oral reading fluency of the repeated reading intervention, implemented during four weeks of summer

school with students in 2^{nd} and 3^{rd} grade whose reading is accurate but slow?

2. Is there a subset of students for whom the effects on oral reading fluency of a repeated reading intervention are greater?

Discussion of Results

The RR intervention was implemented with a sample of participants who were enrolled in schools where the majority of students were from families with low socioeconomic status. Care was taken to ensure that the repeated reading (RR) intervention was an appropriate match for the abilities of the participants, limiting the sample to students whose oral reading fluency rate was at or below the 50^{th} percentile for grade level, but who were able to read the practice passages with at least 93% accuracy. The RR intervention was implemented under conditions of high procedural fidelity with mean weighted percentages that remained above 95% across all four weeks of implementation.

Research Question 1: What are the effects on oral reading fluency of a repeated reading intervention implemented during four weeks of summer school with students in 2^{nd} and 3^{rd} grade whose reading is accurate but slow?

The first research question examined the effectiveness of RR during a four-week summer program. Results from a one-way ANCOVA indicated that overall, RR was an effective intervention for increasing the ORF of second and third grade students when implemented by non-licensed instructors and conducted with pairs of students over four weeks of a summer program. There were statistically significant differences in posttest ORF for students who received RR intervention in addition to their core literacy instruction, compared to students in the control group. For this sample of students whose reading was accurate but slow, RR intervention produced greater ORF growth than core literacy instruction alone. In order to make one year's progress over the course of the school year, it is recommended that students' ORF increases by 1.5 WRCM each week (Fuchs et al., 1993). On average, students who received RR interventions met this criterion during the summer program, increasing their ORF by 2.04 WRCM per week compared to students in the control group whose ORF increased by just 1.03 WRCM per week.

These results build on previous RR research (Kupzyk et al., 2011; Lo et al., 2011; Vadasy & Sanders, 2009) indicating that intervention comprised of four 1 min readings can be effective for maintaining or increasing second and third grade students' ORF. Furthermore, RR potentially enhances core instruction as indicated by the 90% of students in the intervention group who maintained or increased their ORF compared to 70% of the control group.

Results also build on previous research (Kupzyk et al., 2011) indicating that RR intervention is effective when implemented by non-licensed teachers. Moreover, sessions were fairly compact, averaging just over 20 minutes per day, but impacted two students. A typical 1:1 RR intervention dosage is 20 minutes. The outcomes obtained suggest that it may be more efficient to implement RR with pairs of students than 1:1.

Finally, this study provides evidence that RR can be effective even in the short term; results were obtained in just four weeks.

Post Hoc Question 1: Is grade level a factor with regard to the effects on oral reading fluency of the repeated reading intervention, implemented during four weeks of summer school with students in 2nd and 3rd grade whose reading is accurate but slow?

In the course of reviewing the raw changes in ORF from pre- to posttest, second grade students in this sample seemed to benefit more from the supplemental RR intervention than third grade students. Weekly ORF growth was robust for second grade students in the intervention group, averaging 2.25 WRCM, which exceeded the recommended weekly growth of 1.5 WRCM (Fuchs et al., 1993). In contrast, weekly ORF growth for second grade students in the control group averaged just 0.57 WRCM. Meanwhile, weekly ORF growth for third grade students was higher for students in the intervention group, averaging 1.81 WRCM—but not by much when compared to the control group, which averaged 1.63 WRCM. Both third grade groups exceeded the recommended level of 1.50 WRCM.

The apparent differential results for second versus third graders led to the first post hoc research question that analyzed whether grade level was a significant factor in ORF growth. A two-way ANCOVA was conducted to analyze this question and results confirmed that there was a non-significant effect for grade. The interaction between intervention condition and grade was also tested, and the result was a non-significant effect.

One alternate explanation for the appearance of an effect for grade may have been that core instruction was stronger in third grade classrooms. Unfortunately, there was insufficient evidence to either confirm or refute this assertion. Anecdotal observations were conducted to ascertain the content and topography of core instruction, but no quantitative measures were collected that would allow for examining how core instruction was related to ORF growth.

Post Hoc Question 2: Is there a subset of students for whom the effects on oral reading fluency of a repeated reading intervention are greater?

The analysis conducted to answer post hoc question1 confirmed that there only *appeared* to be an effect for grade. Instead, maybe RR was more effective for a different subset of students, such as relatively high-level readers (ORF rate at the 26th-50th percentile) or low-level readers (ORF rate at the 0-25th percentile). Results of a two-way ANCOVA confirmed that there was *not* a significant effect for reading level alone. However, when the interaction between intervention condition and reading level was tested, the result was significant. An examination of marginal means further illustrated that the relatively high-level readers who received the RR intervention made significantly more gains than relatively high-level readers in the control group, low-level readers in both intervention and control groups made similar gains. These same results were also confirmed by the results of the three-way ANCOVA that tested condition, grade, and reading level, as well as all possible interactions.

Taken together, a unique contribution of this study is the notion that relatively high-level readers may derive more benefit from RR than low-level readers. Just one study was located (Sindelar, Monda, & O'Shea, 1990) that directly contrasted the effects of RR for students with different reading rates, similar to the relatively high-level and low-level readers in the present study. Sindelar and colleagues found significant differences in ORF gains for students according to their pretest fluency level. That is, students with higher fluency rates (mastery level) improved their ORF rates on the practice passages more than students with lower fluency rates (instructional level). However the study only tested effects of RR on the non-transfer (practice) passages (Sindelar et al, 1990), whereas the present study measured the effects of RR on transfer passages.

Why did relatively high-level readers benefit more from RR in this study? One rationale is that the fixed duration readings, wherein every student read four times for 1 min, privileged students with higher reading rates. Even though session duration was held constant across all students, the amount of practice differed for individual students according to their reading speed. For example, a student who completed four 1 min readings at a speed of 100 WRCM engaged in twice as much practice during an intervention session as a student with a reading speed of 50 WRCM. These findings are particularly intriguing when considered within the context of the reading achievement gap. It is plausible that for some students, a simple prescription for developing fluent reading may be a larger dose of intervention.

Research Question 2: What dimensions of RR implementation that include student inputs predict and/or correlate with posttest oral reading fluency scores, controlling for students' oral reading fluency at pretest?

The variables initially selected for examination were chosen because they represented student inputs that potentially moderated the effectiveness of the RR intervention (see Appendix M). These variables included (1) cumulative minutes spent in intervention, (2) percentage of 1 min readings completed, (3) cumulative number of 1 min readings completed, (4) cumulative number of words read correctly, (5) oral reading accuracy, and (6) student engagement were evaluated. Notably, two variables were excluded from the final analysis. Percentage of 1 min readings completed was excluded due to near-perfect correlation with cumulative 1 min readings. Student engagement was excluded due to a lack of variability; students were most commonly rated as being *always* or *often* on task (see Appendix K).

Multiple linear regression analysis was used to determine which of the remaining variables (oral reading accuracy, cumulative words read correctly, cumulative minutes in intervention, and cumulative number of 1 min readings) predicted posttest ORF scores. As shown in Model 3 (Table 16), the cumulative number of words read correctly was the only variable that was a significant predictor of posttest ORF.

Why does the number of words read correctly better predict ORF posttest outcomes than other variables related to implementation? One possible explanation is that cumulative number of words read correctly is the most direct measure of the quantity of students' practice across all the intervention sessions attended. Meanwhile, oral reading accuracy, the cumulative number of minutes spent in intervention, and the cumulative number of 1 min readings completed were *related* to the number of words read correctly, but more distal as a measure of practice quantity that may not have as much influence on outcomes.

These results extend previous work that examined the amount of reading done by students. One study was located that counted the number of words students read, however the study did not examine the relation between quantity of reading and reading proficiency. Allington (1984) measured the amount of reading students did during one week of reading lessons. First, third, and fifth grade classroom teachers submitted lesson logs, indicating the books and number of pages that students read, and also designated the students as "good" or "poor" readers. Researchers then quantified the number of words by counting them directly or estimating them using a ten-line estimation method (see Allington, 1981 for specific methods). Results indicated that the amount of reading third grade students did was extremely varied, from 416 to 7,544 words within a week's time. However, missing was any relation between the amount of reading that students did, and how that variability was related to reading outcomes.

These results are also interesting to consider in the context of a study by Vaughn & Wanzek (2008) who compared the effects on reading outcomes for different amounts of time spent in intervention. Students were assigned to a control group, one 30 min dose, or two 30 min doses of a reading intervention package. Results indicated that students who received interventions made more gains than students in the control group, but two doses of intervention were not significantly more effective than a single dose. Data from fidelity observations suggest that the results may have been influenced by student inputs. The authors noted that the fluency aspect of the intervention (which occurred near the end of the session) was often not completed due to off-task behavior, and there were anecdotal reports that students were fatigued during the second daily intervention session.

Limitations and Directions for Further Research

The findings that have been presented, along with several limitations to this study, merit consideration and underscore the need for additional research in the area of reading fluency intervention. Limitations are grouped into two categories:(a) sample limitations and (b) methodological limitations. This section will present those limitations, suggest improvements to sampling or method as relevant, and propose additional research.

Sample Limitations

Participants comprised a convenience sample of struggling readers who attended summer school, returned consent forms, and provided their assent. As a result, characteristics of the participants may differ in meaningful ways from the larger national population of second and third grade students and limit the generalizability of the findings. For example, students came primarily from low SES families for whom learning loss is typically more prevalent. It is possible that if participants were middle-class students whose out-of-school time tends to include more literacy-related activities, significant results may not have been obtained. Students whose parents provided their consent and ensured that their child attended summer school may also provide a different level of literacy support in the home than students whose parents did not provide consent or send their child to summer school. Furthermore, students who gave their assent may have been more accepting of supplemental reading help in general. There were very few behavior issues during this study; most students were very cooperative during the intervention. Levels of compliant behavior do not necessarily represent the population of second and grade students in general and prevent generalization of conclusions about the relation between behavior and ORF outcomes.

Another limitation of the sample was its size. Although sufficiently powered for aggregating the results across both grades, the cell sizes for second and third grades and for relatively high- and low-level readers were relatively small. Reduced power reduces the sensitivity of statistical analyses, and limits the ability to detect effects when they exist. Moreover, the regression analysis would have been stronger if the sample was at minimum $N \geq 50 + m$, where m is the number of variables included in the model (as many as five in this study). It has been suggested that regression analyses conducted with sample sizes that are too small may over-fit the data and limit generalization of the results (Harris, 1975, as cited in Lance & Vandenberg, 2009). Additional research should utilize a larger, nationally representative sample.

Methodological Limitations

The first methodological limitation of this study was that it used a single outcome measure (ORF) to evaluate the effectiveness of repeated reading. ORF is considered a general outcome measure, and its results can predict performance on other more comprehensive measures of reading (e.g., comprehensive states tests such as the Minnesota Comprehensive Assessment [MCA], Silberglitt et al., 2006), however the use of multiple outcome measures has been established as a quality indicator for experimental and quasi-experimental group research (Gersten et al., 2005). The decision to utilize ORF as a single outcome measure was based on concerns raised by school district administration about the amount of instructional time children spent taking pre- and posttests. The summer program was relatively short (six weeks, four days per week) and many students did not attend all days. The amount of time used for assessment was minimized to help maximize students' instructional time and minimize disruption of the summer program. Nevertheless, the purpose of developing fluency is to read for meaning (comprehension). Additional research on RR implemented in the context of a summer break should incorporate additional measures of generalized performance in reading, such as a measure of reading comprehension.

A second methodological limitation is that ORF data collection was limited to a pre- and posttest administered at the beginning and end of the summer program, which prevents drawing conclusions about the impact of RR on summer reading loss. The results of this study showed that RR was effective for preventing summer reading loss for 90% of the participants *during the summer program*. Additional research on RR conducted during summer break should include (at minimum) an ORF measure at the end of the spring term and beginning of the fall term in order to increase understanding about the effects of the RR intervention over time.

A third methodological consideration was the high level of procedural fidelity that was maintained, which can be considered as both a strength and a limitation of the study. On one hand, adherence to procedural fidelity reduced threats to internal validity because interventions were implemented as intended, which in turn strengthens the ability to draw a causal inference about the effects of RR. On the other hand, procedural fidelity would be unlikely to be maintained at such high levels in a real-world school environment, especially since it tends to be assessed much less frequently than it was in this study (Cochrane & Laux, 2008). It is therefore not possible to ascertain how effective this intervention would be under typical conditions, where implementation oversight (e.g., fidelity assessment) might be infrequent, or even non-existent. Additionally, it is possible that the weekly observations of intervention sessions that were conducted contributed to high levels of procedural fidelity in this study. Future studies of RR should assess fidelity in a more covert manner when the goal is to determine the effects of implementing the intervention under typical conditions.

A fourth methodological limitation is that this study design does not permit comparing the effects of an intervention comprised of a fixed amount of time spent reading versus a fixed number of words read. Future research should examine how prescribing dosage using each method impacts students' slope of growth. This contrast would be especially illuminating if conducted in such a way as to compare effects for relatively high- and low-level readers. A compelling question to answer is whether it is possible to identify a "dose" of reading practice that would be likely to close an individual student's reading achievement gap.

A final methodological limitation is that there was no variation in the percentage of 1 min readings completed, and little variation in oral reading accuracy and levels of student engagement. With the exception of one student on one day, all students completed all assigned readings during the intervention sessions they attended. Accuracy and student engagement were also uniformly high. Additional research might directly examine how different levels of these variables influence posttest ORF.

Implications for Practice

In the conceptualization of this study, there was hope of a collateral gain in terms of knowledge that could inform further study of RR implementation to increase levels of progress for *all* students for whom the intervention is deemed appropriate. The results reported build on previous research supporting the effectiveness of RR in general, and when implemented with students whose reading is accurate but slow. Furthermore, the procedure was effective as a supplemental intervention implemented during a short summer program when some students—particularly those in families with low socioeconomic status—might otherwise have limited access to literacy activities. Under conditions where interventionists were trained and procedural fidelity was maintained at high levels, the majority of students (90%) who received the RR intervention maintained or increased their oral reading fluency rates. However, 10% of the students still did not make expected levels of progress, an issue that is important to investigate when the goal is to address the achievement gap.

One possible explanation—and perhaps the most parsimonious—is that some students simply needed more intervention than they received. There is evidence to support this assertion in that the cumulative number of words read correctly was the only significant predictor of posttest ORF, after controlling for pretest ORF. In addition, relatively high-level readers appeared to benefit more from the RR intervention than low-level readers, possibly due to a confluence of higher oral reading rate and a protocol wherein students read for a fixed number of minutes—a condition that would have privileged relatively high-level readers in terms of their quantity of practice.

Taken together, the analyses conducted in the course of this examination support what has been posited, that "more of what works" may be a useful approach to addressing the reading achievement gap. It may be especially useful to operate from this perspective when considering interventions to improve oral reading fluency, a skill comprised of procedural sub-skills including rapid letter-sound correspondence and decoding (see Geary, 1995; LaBerge & Samuels, 1974). For such skills, student responses in the form of presence/absence at intervention sessions, on-/off-task behavior, high/low reading rate, or accurate/inaccurate responding may have a significant effect on dosage in terms of the quantity and quality of practice—and by extension, reading outcomes. When interventions address such skills, it may be more critical to ensure that a student engages in sufficient quantities of high quality practice than to focus on the number of intervention minutes assigned or received.

At present, intervention programming is typically planned as a designated number of intervention minutes. However, such a metric may not be optimal if the goal is to ensure that students receive an intervention dosage that is appropriately intense with regard to their needs. In this study for example, the average number of minutes spent in intervention varied across students and by session, but the fixed number of 1 min practice readings meant that additional time in intervention did not equal more practice. One suggested change is to ensure that fluency intervention protocols are designed so students continue to engage in the "active ingredients" for as long as possible during their session minutes.

It may also be useful to consider a student's oral reading rate when prescribing intervention dosage. Students could be assigned to read a certain number of *words* per intervention session rather than a certain number of intervention *minutes*. It should be noted that such a practice would necessarily introduce variable session length, which could be problematic in settings where practitioners need to maintain a tight schedule.

One solution is to plan intervention dosage based on reading level, given that students with low-level reading rates need more time to read the same number of words as a student with a relatively high-level reading rate. Schedules could then be arranged to include extended time for those who need it most.

Finally, the results of this study help build a case for measuring student inputs as an aspect of procedural fidelity in addition to measuring interventionists' adherence to procedural fidelity. Such a change would improve the precision with which procedural fidelity and intervention intensity are evaluated, which could greatly enhance what we understand about the effectiveness of interventions for individual students.

Conclusion

In conclusion, the dire economical and social consequences for failing to ensure that all students learn to read proficiently and on time underlies the importance of continued efforts to develop and refine evidence-based practices in the area of reading instruction and intervention. Given the magnitude and persistence of the reading achievement gap, it is critical to increase our understanding of how different variables contribute to intervention effectiveness for individual students. This examination built on the evidence-base for RR as a practice that can be effective for improving and maintaining students' reading fluency in general, and during extended school breaks—especially for students in families with low socioeconomic status who might otherwise engage in fewer literacy activities during time way from school than their peers in families with higher incomes. This examination also illustrated that it may be important to focus attention on understanding the contributions of student inputs to procedural fidelity and treatment intensity. Greater knowledge in this area would support practitioners in their efforts to instruct and intervene in ways that closely match students' needs, are optimally effective, and make efficient use of available resources. The results that have been reported, along with the suggestions for additional research, provide some direction for such future endeavors—with a primary end goal of improving the effectiveness of instructional approaches and interventions so they are potent enough to accelerate progress and close reading achievement gaps for individual students

References

Alexander, K. L., Entwisle, D. R., & Olson, L. S. (2001). Schools, achievement, and inequality: A seasonal perspective. *Educational Evaluation and Policy Analysis*, *23*(2), 171-191.

Allington, R. L. (1984). Content coverage and contextual reading in reading groups. *Journal of Literacy Research*, *16*(2), 85-96.

Allington, R. L., McGill-Franzen, A., Camilli, G., Williams, L., Graff, J., Zeig, J., ... & Nowak, R. (2010). Addressing summer reading setback among economically disadvantaged elementary students. *Reading Psychology*, *31*(5), 411-427.

Ardoin, S. P., Williams, J. C., Klubnik, C., & McCall, M. (2009). Three versus six rereadings of practice passages. *Journal Of Applied Behavior Analysis, 42*(2), 375-380.

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of applied behavior analysis*, *1*(1), 91-97.

Baker, E. (2012). Optimal intervention intensity. *International Journal of Speech-Language Pathology*, *14*(5), 401-409.

Banks, J. A., Cookson, P., Gay, G., Hawley, W. D., Irvine, J. J., Nieto, S., ... & Stephan, W. G. (2001). Diversity within unity: Essential principles for teaching and learning in a multicultural society. *Phi Delta Kappan*, 196-203.

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological bulletin*, *128*(4), 612-637.

Barnett, D. W., Daly, E. J., Jones, K. M., & Lentz, F. E. (2004). Response to intervention: Empirically based special service decisions from single-case designs of increasing and decreasing intensity. *The Journal of Special Education*, *38*(2), 66-79.

Barnett, D., Hawkins, R., McCoy, D., Wahl, E., Shier, A., Denune, H., & Kimener, L. (2014). Methods used to document procedural fidelity in school-based intervention research. *Journal of Behavioral Education*, *23*(1), 89-107.

Invernizzi, M., Templeton, S., & Johnston, F. (2000). *Words their way: Word study for phonics, vocabulary, and spelling instruction* (2 nd ed.). Upper Saddle River, NJ: Merrill.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathmatical and Statistical Psychology, 41,* 257–278.

Begeny, J. C., Hawkins, A. L., Krouse, H. E., & Laugle, K. M. (2011). Altering instructional delivery options to improve intervention outcomes: Does increased instructional intensity also increase instructional effectiveness? *Psychology in the Schools, 48*(8), 769-785. Retrieved from http://login.ezproxy.lib.umn.edu/login?url=http://search.proquest.com/docview/964173390?accountid=14586

Begeny, J. C., Krouse, H. E., Ross, S. G., & Mitchell, R. C. (2009). Increasing elementary-aged students' reading fluency with small-group interventions: A comparison of repeated reading, listening passage preview, and listening only strategies. *Journal Of Behavioral Education, 18*(3), 211-228.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., ... & Czajkowski, S. (2004). Enhancing treatment fidelity in health behavior change studies: Best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, *23*(5), 443.

Boettcher, J.V. (1983). Computer-based education: Classroom application and benefits for the learning disabled student. *Annals of Dyslexia, 33*, 201-219.

Boykin, A. W., & Noguera, P. (2011). *Creating the opportunity to learn: Moving from research to practice to close the achievement gap*. Ascd.

Brody, G. H., Lahey, B. B., & Combs, M. L. (1978). Effects of intermittent modeling on observational learning. *Journal of Applied Behavior Analysis*, *11*(1), 87-90.

Bryan, W. L., & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. *Psychological Review*, *4*(1), 27.

Bureau of Labor Statistics (2014). Retrieved from: http://www.bls.gov/news.release/hsgec.t01.htm

Burns, M. K., & Gibbons, K. (2008). *Implementing response-to-intervention in elementary and secondary schools: Procedures to assure scientific-based practices*. CRC Press.

Burke, M. D., Hagan-Burke, S., & Sugai, G. (2003). The efficacy of function-based interventions for students with learning disabilities who exhibit escape-maintained problem behaviors: Preliminary results from a single-case experiment. *Learning Disability Quarterly*, *26*(1), 15-25.

Burns, M. K., Zaslofsky, A. F., Kanive, R., & Parker, D. C. (2012). Meta-analysis of incremental rehearsal using phi coefficients to compare single-case and group designs. *Journal of Behavioral Education*, *21*(3), 185-202.

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, *31*(2), 199-218.

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*(2), 67-81.

Chard, D. J., Ketterlin-Geller, L. R., Baker, S. K., Doabler, C., & Apichatabutra, C. (2009). Repeated reading interventions for students with learning disabilities: Status of the evidence. *Exceptional Children*, *75*(3), 263-281.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive psychology*, *4*(1), 55-81.

Christ, T. J., Ardoin, S. A., & Eckert, T. (2011). *Formative Assessment System for Teachers: Curriculum Based Measurement for Reading* (CBM-R). Minneapolis, MN: Formative Assessment System for Teachers

Christ, T. J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review*, *36*(1), 130.

Cochrane, W. S., & Laux, J. M. (2008). A survey investigating school psychologists' measurement of treatment integrity in school-based interventions and their beliefs about its importance. *Psychology in the Schools*, *45*(6), 499-507.

Codding, R. S., Hilt-Panahon, A., Panahon, C. J., & Benson, J. L. (2009). Addressing mathematics computation problems: A review of simple and moderate intensity interventions. *Education and Treatment of Children*, *32*(2), 279-312.

Coleman, M. B., & Heller, K. W. (2010). The use of repeated reading with computer modeling to promote reading fluency with students who have physical disabilities. *Journal Of Special Education Technology, 25*(1), 29-41.

Conte, R., Humphreys, R. (1989). Repeated readings using audio-taped materials enhances oral reading in children with reading difficulties. *Journal of Communications Disorders, 22*, 65-79.

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*(3), 227-268.

Dahl, P. R. (1977). An experimental program for teaching high-speed word recognition and comprehension skills. In J. Burton, T. Lovitt, & T. Row- lands (Eds.), *Communication research in learning disabilities and mental retardation* (pp. 33–65). Baltimore: University Park Press.

Daly III, E. J., Lentz Jr, F. E., & Boyer, J. (1996). The instructional hierarchy: A conceptual model for understanding the effective components of reading interventions. *School Psychology Quarterly*, *11*(4), 369.

Daly III, E. J., 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*(2), 203-218.

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.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control?. *Clinical psychology review*, *18*(1), 23-45.

Deno, S. L., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional children*.

Digennaro-Reed, F. D., Codding, R., Catania, C. N., & Maguire, H. (2010). Effects of video modeling on treatment integrity of behavioral interventions. *Journal of applied behavior analysis*, *43*(2), 291-295.

Downey, D. B., von Hippel, P. T., & Broh, B. (2004). Are schools the great equalizer? School and non-school sources of inequality in cognitive skills. *American Sociological Review*, *69*(5), 613-635.

Duhon, G. J., Mesmer, E. M., Atkins, M. E., Greguson, L. A., & Olinger, E. S. (2009). Quantifying intervention intensity: A systematic approach to evaluating student response to increasing intervention frequency. *Journal of Behavioral Education*, *18*(2), 101-118.

Durkin, D. (1978). What classroom observations reveal about reading comprehension instruction. *Reading research quarterly*, 481-533.

Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health education research*, *18*(2), 237-256.

Eldredge, J. (1990). Increasing the performance of poor readers in the third grade with a group-assisted strategy. *The Journal of Educational Research*, *84*(2), 69-77.

Eldredge, J. L., Reutzel, D. R., & Hollingsworth, P. M. (1996). Comparing the effectiveness of two oral reading practices: Round-robin reading and the shared book experience. *Journal of Literacy Research, 28*, 201-225.

Entwisle, D. R., & Alexander, K. L. (1992). Summer setback: Race, poverty, school composition and math achievement in the first two years of school. *American Sociological Review, 57,* 72-84.

Entwisle, D. R., & Alexander, K. L. (1994). The gender gap in math: Its possible origins in neighborhood effects. *American Sociological Review, 59*, 822-838.

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological review*, *100*(3), 363.

Fuchs, D., & Fuchs, L. S. (2006). Introduction to response to intervention: What, why, and how valid is it?. *Reading Research Quarterly*, *41*(1), 93-99.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect?. *School Psychology Review*, *22*, 27-27.

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*(3), 239-256.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education*, *9*(2), 20-28.

Gall, J. P., Gall, M. D., & Borg, W. R. (1999). Applying educational research: A practical guide. Longman Publishing.

Gast, D. L., & Ledford, J. (Eds.). (2010). *Single-subject research methodology in behavioral sciences*. Routledge.

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: Implications for mathematical development and instruction. *American Psychologist*, *50*(1), 24.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., & Santoro, L. (2009). Linan-Thompson, S. 2009. *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide*. IES

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional children*, *71*(2), 149-164.

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, *10*(2), 486–489. http://doi.org/10.5812/ijem.3505

Graff, R. B., Libby, M. E., & Green, G. (1998). The effects of reinforcer choice on rates of challenging behavior and free operant responding in individuals with severe disabilities. *Behavioral Interventions*, *13*(4), 249-268.

Granpeesheh, D., Dixon, D. R., Tarbox, J., Kaplan, A. M., & Wilke, A. E. (2009). The effects of age and treatment intensity on behavioral intervention outcomes for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 3*(4), 1014-1022.

Greer, R. D., & McDonough, S. H. (1999). Is the learn unit a fundamental measure of pedagogy?. *The Behavior Analyst*, *22*(1), 5.

Gresham, F. M. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review*.

Gresham, F. M. (2009). Evolution of the treatment integrity concept: Current status and future directions. *School Psychology Review*, *38*(4), 533.

Gresham, F. M. (2014). Measuring and analyzing treatment integrity data in research. In L. M. H. Sanetti & T. R. Kratochwill (Eds), *Treatment integrity: A foundation for evidence-based practice in applied psychology*. The American Psychological Association.

Hamilton, B.L. (1977). An empirical investigation of the effects of heterogeneous regression slopes in analysis of covariance. *Educational and Psychological Measurement, 37,* 701–702.

Haring, N. G., Lovitt, T C., Eaton, M.D., & Hansen, C. L. (1978). The fourth R: Research in the classroom. Columbus, OH: Merrill.

Hayes, D. P., & Grether, J. (1969, April). *The school year and vacations: When do students learn?* Paper presented at the Eastern Sociological Association

Convention, New York.

Hayes, D. P., & Grether, J. (1983). The school year and vacations: When do students learn? *Cornell Journal of Social Relations, 17*(1), 56-71.

Hernandez, D. J. (2011). *Double Jeopardy: How Third-Grade Reading Skills and Poverty Influence High School Graduation.* Annie E. Casey Foundation.

Heubusch, J. D., & Lloyd, J. W. (1998). Corrective feedback in oral reading. *Journal of Behavioral Education, 8*(1), 63-79.

Heyns, B. (1987). Schooling and cognitive development: Is there a season for learning? *Child Development, 58*, 1151-60.

Hintze, J. M., Callahan, J. E., Matthews, W. J., Williams, S. A., & Tobin, K. G. (2002). Oral reading fluency and prediction of reading comprehension in African American and Caucasian elementary school children. *School Psychology Review*, *31*(4), 540-553.

Ho, P. M., Bryson, C. L., & Rumsfeld, J. S. (2009). Medication adherence: Its importance in cardiovascular outcomes. *Circulation*, *119*(23), 3028-3035.

Homan, S. P., Klesius, J. P., & Hite, C. (1993). Effects of repeated readings and nonrepetitive strategies on students' fluency and comprehension. *Journal of Educational Research, 87*, 94-99.

Hopkins, K. D., & Weeks, D. L. (1990). Tests for normality and measures of skewness and kurtosis: Their place in research reporting. *Educational and Psychological Measurement, 50*, 717-729.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Exceptional Children*, *71*(2), 165-179.

House, A. E., House, B. J., & Campbell, M. B. (1981). Measures of interobserver agreement: Calculation formulas and distribution effects. *Journal of Behavioral Assessment*, *3*(1), 37-57.

Huang, L. V., Nelson, R. B., & Nelson, D. (2008). Increasing Reading Fluency through Student-Directed Repeated Reading and Feedback. *California School Psychologist,* 1333-40.

Institute of Educational Science (2014). *What Works Clearinghouse Procedures and Standards Handbook 3.0.* Retrieved from http://ies.ed.gov/ncee/wwc/documentsum.aspx?sid=19

Jenkins, J. R., Fuchs, L. S., Van Den Broek, P., Espin, C., & Deno, S. L. (2003). Accuracy and fluency in list and context reading of skilled and RD groups: Absolute and relative performance levels. *Learning Disabilities Research & Practice*, *18*(4), 237-245.

Jenkins, J. R., Zumeta, R., & Dupree, O. (2005). Measuring gains in reading ability with passage reading fluency. *Learning Disabilities Research & Practice*, *20*(4), 245-253.

Jeynes, W. H. (2008). A meta-analysis of the relationship between phonics instruction and minority elementary school student academic achievement. *Education And Urban Society*, *40*(2), 151-166.

Jones, H. A., Clarke, A. T., & Power, T. J. (2008). Expanding the concept of intervention integrity: A multidimensional model of participant engagement. *Balance*, *23*, 4-5.

Kame'enui, E. J., & Simmons, D. C. (2001). Introduction to this special issue: The DNA of reading fluency. *Scientific Studies of Reading*, *5*(3), 203-210.

Kamil, M. L., Pearson, P. D., Moje, E. B., & Afflerbach, P. (Eds.). (2011). *Handbook of reading research* (Vol. 4). Routledge.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical documentation. *What Works Clearinghouse*.

Kuhn, M. (2004). Helping students become accurate, expressive readers: Fluency instruction for small groups. *The Reading Teacher*, *58*(4), 338-344.

Kuhn, M. R., Schwanenflugel, P. J., Morris, R. D., Morrow, L. M., Woo, D. G., Meisinger, E. B., & ... Stahl, S. A. (2006). Teaching children to become fluent and automatic readers. *Journal Of Literacy Research, 38*(4), 357-387.

Kuhn, M. R., & Stahl, S. A. (2003). Fluency: A review of developmental and remedial practices. *Journal of educational psychology*, *95*(1), 3.

Kupzyk, S., Daly, E. I., & Andersen, M. N. (2012). Preparing teachers to train parents to use evidence-based strategies for oral reading fluency with their children. *Contemporary School Psychology, 16*, 129-140.

Kupzyk, S., McCurdy, M., Hofstadter, K. L., & Berger, L. (2011). Recorded readings: A taped parent-tutoring intervention. *Journal Of Behavioral Education, 20*(2), 87-102.

Ladson-Billings, G. (2006). From the achievement gap to the education debt:

Understanding achievement in US schools. *Educational researcher*, *35*(7), 3-12.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive psychology*, *6*(2), 293-323.

Lahey, B. B. (1971). Modification of the frequency of descriptive adjectives in the speech of Head Start children through modeling without reinforcement. *Journal of applied behavior analysis*, *4*(1), 19-22.

Lance, C. E., & Vandenberg, R. J. (2009). *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. Taylor & Francis.

Lane, K. L., Bocian, K. M., MacMillan, D. L., & Gresham, F. M. (2004). Treatment integrity: An essential—but often forgotten—component of school-based interventions. *Preventing School Failure: Alternative Education for Children and Youth*, *48*(3), 36-43.

Lichstein, K. L., Riedel, B. W., & Grieve, R. (1994). Fair tests of clinical trials: A treatment implementation model. *Advances in Behaviour Research and Therapy*, *16*(1), 1-29.

Lo, Y., Cooke, N. L., & Starling, A. P. (2011). Using a repeated reading program to improve generalization of oral reading fluency. *Education And Treatment Of Children, 34*(1), 115-140.

Logan, G. D. (1997). Automaticity and reading: Perspective from the instance theory of automatization. *Reading and Writing Quarterly, 13,* 123–146.

Macaruso, P., & Rodman, A. (2011). Efficacy of computer-assisted instruction for the development of early literacy skills in young children. *Reading Psychology*, *32*(2), 172-196.

Manset-Williamson, G., & Nelson, J. M. (2005). Balanced, strategic reading instruction for upper-elementary and middle school students with reading disabilities: A comparative study of two approaches. *Learning Disability Quarterly*, *28*(1), 59-74.

Marty, G. (2013). *Working in the Achievement Gap*. Retrieved from http://www.cehd.umn.edu/Connect/2013Spring/Gap.html

Mathes, P. G., & Fuchs, L. S. (1993). Peer-mediated reading instruction in special education resource rooms. *Learning Disabilities Research & Practice*.

McComas, J., Hoch, H., Paone, D., & El-Roy, D. (2000). Escape behavior during academic tasks: A preliminary analysis of idiosyncratic establishing operations. *Journal of Applied Behavior Analysis*, *33*(4), 479-493.

McIntosh, K., Chard, D. J., Boland, J. B., & Horner, R. H. (2006). Demonstration of combined efforts in school-wide academic and behavioral systems and incidence of reading and behavior challenges in early elementary grades. *Journal of Positive Behavior Interventions*, *8*(3), 146-154.

McMaster, K. L., Fuchs, D., Fuchs, L. S., & Compton, D. L. (2005). Responding to nonresponders: An experimental field trial of identification and intervention methods. *Exceptional children*, *71*(4), 445-463.

Mellard, D., McKnight, M., & Jordan, J. (2010). RTI tier structures and instructional intensity. *Learning Disabilities Research & Practice*, *25*(4), 217-225.

Meyer, M. S., & Felton, R. H. (1999). Repeated reading to enhance fluency: Old approaches and new directions. *Annals of dyslexia*, *49*(1), 283-306.

Musti-Rao, S., Hawkins, R. O., & Barkley, E. A. (2009). Effects of repeated readings on the oral reading fluency of urban fourth-grade students: Implications for practice. *Preventing School Failure, 54*(1), 12-23.

National Assessment for Educational Progress (2011). *Score Gaps*. Retrieved from: http://www.nationsreportcard.gov/reading_math_2015/#reading/gaps?grade=4

National Reading Panel (US), National Institute of Child Health, & Human Development (US). (2000). *Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups.* National Institute of Child Health and Human Development, National Institute of Health.

National Center for Educational Statistics (2015). *Data Tools*. Retrieved from https://nces.ed.gov/datatools/index.asp?DataToolSectionID=4

Noell, G. H., & Gansle, K. A. (2014). The use of performance feedback to improve intervention implementation in schools. In L. M. H. Sanetti & T. R. Kratochwill (Eds), *Treatment integrity: A foundation for evidence-based practice in applied psychology*. The American Psychological Association.

O'Brien, R. M., (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*, 673-690.

O'Connor, R. (2000). Increasing the Intensity of Intervention in Kindergarten and First Grade. *Learning Disabilities Research & Practice (Lawrence Erlbaum)*, *15*(1), 43-54.

O'Connor, R. E., White, A., & Swanson, H. L. (2007). Repeated reading versus continuous reading: Influences on reading fluency and comprehension. *Exceptional Children, 74*(1), 31-46.

Oddo, M., Barnett, D. W., Hawkins, R. O., & Musti-Rao, S. (2010). Reciprocal peer tutoring and repeated reading: Increasing practicality using student groups. *Psychology In The Schools, 47*(8), 842-858.

O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of Educational Research*, *78*(1), 33-84.

O'Keeffe, B. V., Slocum, T. A., Burlingame, C., Snyder, K., & Bundock, K. (2012). Comparing results of systematic reviews: Parallel reviews of research on repeated reading. *Education and Treatment of Children*, *35*(2), 333-366.

O'Shea, L. J., Sindelar, P. T., & O'Shea, D. J. (1985). The effects of repeated readings and attentional cues on reading fluency and comprehension. *Journal of Literacy Research*, *17*(2), 129-142.

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

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). *Listening to children read aloud.* Washington, DC: Office of Educational Research and Improvement, U. S. Department of Education.

Rafferty, L. A. (2012). Self-monitoring during whole group reading instruction: effects among students with emotional and behavioral disabilities during summer school intervention sessions. *Emotional & Behavioural Difficulties*, *17*(2), 157-173.

doi:10.1080/13632752.2012.672866

Ramey, C. T., & Ramey, S. L. (1998). Early intervention and early experience. *American psychologist*, *53*(2), 109.

Rashotte, C. A., & Torgesen, J. K. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly*, 180-188.

Reder, S. (2010). *Adult Literacy Development and Economic Growth.* National Institute for Literacy.

Ross, S. G., & Begeny, J. C. (2011). Improving Latino, English language learners' reading fluency: The effects of small-group and one-on-one intervention. *Psychology In The Schools*, *48*(6), 604-618.

Rumberger, R. W., & Lamb, S. P. (2003). The early employment and further education experiences of high school dropouts: A comparative study of the United States and Australia. *Economics of Education Review*, *22*(4), 353-366.

Samuels, S. J. (1979). The method of repeated readings. *The reading teacher*, 403-408.

Samuels, S., & Farstrup, A. E. (2006). *What research has to say about fluency instruction*. International Reading Association.

Sanetti, L. M., Gritter, K. L., & Dobey, L. M. (2011). Treatment integrity of interventions with children in the school psychology literature from 1995 to 2008. *School Psychology Review*, *40*(1), 72.

Sanetti, L. M., Kratochwill, T. R., & Long, A. C. (2013). Applying adult behavior change theory to support mediator-based intervention implementation. *School Psychology Quarterly*, *28*(1), 47.

Savaiano, M. E., & Hatton, D. D. (2013). Using Repeated Reading to Improve Reading Speed and Comprehension in Students with Visual Impairments. *Journal Of Visual Impairment & Blindness*, *107*(2), 93-106.

Saunders, W. B. (1994). *Dorland's illustrated medical dictionary.* The Curtis Center, Pennsylvania.

Schreiber, P. A. (1980). On the acquisition of reading fluency. *Journal of Literacy Research*, *12*(3), 177-186.

Schreder, S. J., Hupp, S. A., Everett, G. E., & Krohn, E. (2012). Targeting reading fluency through brief experimental analysis and parental intervention over the summer. *Journal of Applied School Psychology, 28*(2), 200-220.

Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of educational psychology*, *96*(1), 119.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Wadsworth Cengage learning.

Shany, M. T., & Biemiller, A. (1995). Assisted reading practice: Effects on performance for poor readers in grades 3 and 4. *Reading research quarterly*, 382-395.

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in Schools, 43*(5), 527-535.

Sindelar, P. T., Monda, L. E., & O'Shea, L. J. (1990). Effects of repeated readings on instructional-and mastery-level readers. *The Journal of Educational Research*, *83*(4), 220-226.

Stahl, S. A., & Kuhn, M. R. (2002). Making it sound like language: Developing fluency. *The Reading Teacher*, *55*(6), 582.

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

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading research quarterly*, 360-407.

Stevens, R. J., Van Meter, P. N., Garner, J., Warcholak, N., Bochna, C., & Hall, T. (2008). Reading and integrated literacy strategies (RAILS): An integrated approach to early reading. *Journal Of Education For Students Placed At Risk*, *13*(4), 357-380.

Sugai, G., & Horner, R. H. (1999). Discipline and behavioral support: Preferred processes and practices. *Effective School Practices*, *17*(4), 10-22.

Sum, A., Khatiwada, I., McLaughlin, J., & Palma, S. (2009). *The consequences of dropping out of high school.* Center for Labor Market Studies Publications.

Swanson, H. L., & O'Connor, R. (2009). The role of working memory and fluency practice on the reading comprehension of students who are dysfluent readers. *Journal of Learning Disabilities, 42*(6), 548-575.

Swanson, E., Wanzek, J., Haring, C., Ciullo, S., & McCulley, L. (2013). Intervention fidelity in special and general education research journals. *Journal of Special*

Education, *47*(1), 3-13.

Symonds, P. M., & Chase, D. H. (1992). Practice vs. motivation. *Journal of Educational Psychology*, *84*(3), 282.

Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading a meta-analysis. *Remedial and special education*, *25*(4), 252-261.

Therrien, W. J., Kirk, J. F., & Woods-Groves, S. (2012). Comparison of a reading fluency intervention with and without passage repetition on reading achievement. *Remedial And Special Education, 33*(5), 309-319.

Thurlow, R., & van den Broek, P. (1997). Automaticity and inference generation during reading comprehension. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *13*(2), 165-181.

Topping, K. J., Samuels, J., & Paul, T. (2007). Does practice make perfect? Independent reading quantity, quality and student achievement. *Learning and Instruction*, *17*(3), 253-264.

Torgesen, J. K. (1998). Catch them before they fall. *American Educator*, *22*, 32-41.

Treptow, M. A., Burns, M. K., & McComas, J. J. (2007). Reading at the frustration, instructional, and independent levels: The effects on students' reading comprehension and time on task. *School Psychology Review*, *36*(1), 159-166.

Uwah, C., McMahon, H., & Furlow, C. (2008). School belonging, educational aspirations, and academic self-efficacy among African American male high school students: Implications for school counselors. *Professional School Counseling*, *11*(5), 296-305.

Vadasy, P. F., & Sanders, E. A. (2008). Code-oriented instruction for kindergarten students at risk for reading difficulties: A replication and comparison of instructional groupings. *Reading And Writing: An Interdisciplinary Journal*, *21*(9), 929-963.

Vadasy, P. F., & Sanders, E. A. (2009). Supplemental fluency intervention and determinants of reading outcomes. *Scientific Studies Of Reading, 13*(5), 383-425.

Vadasy, P. F., & Sanders, E. A. (2010). Efficacy of supplemental phonics-based instruction for low-skilled kindergarteners in the context of language minority status and classroom phonics instruction. *Journal Of Educational Psychology*, *102*(4), 786-803.

Vaughn, S., Linan-Thompson, S., & Kouzekanani, K., (2003). Reading instruction grouping for students with reading difficulties. *Remedial and Special Education, 24*, 301 – 315.

Wanzek, J., & Vaughn, S. (2008). Response to varying amounts of time in reading intervention for students with low response to intervention. *Journal of Learning Disabilities*, *41*(2), 126-142.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental retardation and developmental disabilities research reviews*, *13*(1), 70.

West S.G., Finch J.F., & Curran P.J. (1996) *Structural equation models with nonnormal variables: problems and remedies*. In: Hoyle R. H. (ed) Structural equation modeling: Concepts, issues and applications. Newbery Park, CA: Sage; 1995. pp.

56-75.

Williams, A. M., & Hodges, N. J. (2005). Practice, instruction and skill acquisition in soccer: Challenging tradition. *Journal of sports sciences*, *23*(6), 637-650.

Yeaton, W. H., & Sechrest, L. (1981). Critical dimensions in the choice and maintenance of successful treatments: strength, integrity, and effectiveness. *Journal of consulting and clinical psychology*, *49*(2), 156.

Young, A. R., Bowers, P. G., & McKinnon, G. E. (1996). Effects of prosodic modeling and repeated reading on poor readers' fluency and comprehension. *Applied Psycholinguistics, 17*, 59-84.

Yurick, A. L., Robinson, P. D., Cartledge, G., Lo, Y., & Evans, T. L. (2006). Using Peer-Mediated Repeated Readings as a Fluency-Building Activity for Urban Learners. *Education And Treatment Of Children*, *29*(3), 469-493.

Appendix A

Screening Probe

Administrator Copy **Grade 2 Screening Form 1**

Denny

One morning, Denny was getting dressed for school. She could 10
not find one of her shoes. She needed to find it, or she would be 25
late. Her dog ran over to the window and started barking. Denny 37
looked and saw the school bus. She grabbed the first shoe she 49
could find. Then she ran out of the house. 58

Denny looked down at the shoes on her feet. She was wearing 70
one red shoe and one blue shoe. She was nervous about wearing 82
different colored shoes. Maybe the children on the bus would tease 93
her. Her best friend saw the different colored shoes. She told Denny 105
how much she liked them. The other children on the bus liked them 118
too. 119

After Denny got off the bus, she walked to her classroom. The 131
children in the hall liked her different colored shoes too. Some of the 144
children switched shoes with each other to be like Denny. Denny's 155
teacher saw how much the children liked exchanging shoes. She 165
told the class they would have a funniest shoes contest. The class 177
was happy and started to cheer. 183

When Denny got home, she told her mother what had 193
happened. Her mother laughed at the funny story. While they were 204
talking, her dog ran into the room. He had something red in his 217
mouth. He had found the missing shoe! Denny wondered if he had 229
taken it that morning. She hugged him and went outside to play. 241

Total Words Read: ___________ **- # of Errors:** _____________ **= WRC/min:** _________________

Student: ____________________ **Teacher:** ________________ **Grade:** ________ **Date:** _________

Appendix B

Sample Reading Passage

Administrator Copy **Grade 2 Form 3 (1 per Occasion)**

Jim

Jim's grandfather was a mountain climber. He climbed many 9
different mountains in all parts of the world. Jim wanted to share his 22
adventures and climb with him. He asked his grandfather if he could 34
go with him sometime. His grandfather said he was climbing the 45
next day. He agreed that Jim could come if he followed the rules. 58
Jim agreed and said he would be ready to go. 68

Jim woke up early and got dressed for the hike. His 79
grandfather picked him up and they drove to the mountain. They 90
started on the trail by the side of the mountain. They started to walk 104
very slowly and Jim wondered why. His grandfather told him it was 116
because the mountain was tall. They would need to save energy to 128
finish. But Jim did not feel tired and wanted to climb fast. 140

Before long, Jim grew restless with the slow pace. He wanted 151
to walk faster and he did. He passed his grandfather and 162
disappeared out of site. His grandfather called for him to stay on the 175
path. Jim called back to say that he would. Both of them 187
continued to hike. 190

Jim's legs felt heavy and he began to feel tired. He sat down 203
on a big rock to rest. Time went by and his grandfather caught up 217
to him. His grandfather still had lots of energy. He sat down and 230
waited for Jim to feel better. When Jim felt better, they hiked 242
together, very slowly. Later that day, they reached the top of the 254
mountain together. 256

Total Words Read: ___________ **- # of Errors:** _____________ **= WRC/min:** _________________

Student: ____________________ **Teacher:** ________________ **Grade:** ________ **Date:** _________

Appendix C

Intervention Log

<table>
<tr><th colspan="5">Key</th></tr>
<tr><td>Attendance</td><td colspan="2">P=student present for any portion of intervention</td><td colspan="2">A=student absent for entirety of intervention</td></tr>
<tr><td>Intervention Duration</td><td colspan="4">Record the number of minutes in intervention for each student</td></tr>
<tr><td rowspan="2">Engagement</td><td colspan="4">Task Related Behavior
• Responsive to teacher and peers
• Used eye contact (student’s eyes were on the reading passage or VISTA, as directed)
• Read audibly (when directed to read aloud; actively following along during whisper-reading)
• Asked text related questions (limited chatty behavior)
• Conversation was focused on the text/reading

Physical Behavior
• Remained in intervention area
• Assumed a posture that allowed for easy reading (sat up straight, faced toward the reading passage)
• Hands engaged with reading materials or in lap (e.g., following along with finger)</td></tr>
<tr><td>0-Never/Rarely</td><td>1-Sometimes</td><td>2-Often</td><td>3-Always</td></tr>
<tr><td>First Read WRC</td><td colspan="2">For research purposes only.</td><td colspan="2">Do not fill in.</td></tr>
<tr><td>Fourth Read WRC</td><td colspan="2">For research purposes only.</td><td colspan="2">Do not fill in.</td></tr>
</table>

Date	**Are there any behavior/engagement difficulties with which you need help?**	**Behavior/Engagement Plan**

Monday	Date:		VISTA:					
Student Name	**Attendance P/A**	**Intervention Duration**	**Engagement**	**WRC/ERR**				
				1	**2**	**3**	**4**	**total**
			0 1 2 3					
Was timer started/stopped immediately at the beginning/end of intervention?		□ YES □ NO	If no, please explain.					
			0 1 2 3					
Was timer started/stopped immediately at the beginning/end of intervention?		□ YES □ NO	If no, please explain.					
Notes								

Tuesday	Date:		VISTA:					
Student Name	**Attendance P/A**	**Intervention Duration**	**Engagement**	**WRC/ERR**				
				1	**2**	**3**	**4**	**total**
			0 1 2 3					
Was timer started/stoppped immediately at the beginning/end of intervention?		□ YES □ NO	If no, please explain.					
			0 1 2 3					
Was timer started/stopped immediately at the beginning/end of intervention?		□ YES □ NO	If no, please explain.					
Notes								

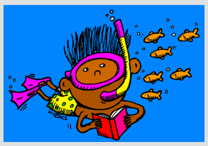
Appendix D

Fluency Graph

WRC									
165									
160									
155									
150									
145									
140									
135									
130									
125									
120									
115									
110									
105									
100									
95									
90									
85									
80									
75									
70									
65									
60									
55									ERR
50									10
45									9
40									8
35									7
30									6
25									5
20									4
15									3
10									2
5									1
	Read 1	Read 4	Read 1	Read 4	Read 1	Read 4	Read 1	Read 4	
	Monday		Tuesday		Wednesday		Thursday		

Appendix E

Attendance Chart



SUMMER READING ROCKS!			

Appendix F

Screening Protocol and Fidelity Checklist

Directions

☐ 1. Place the paper copy of the passage in front of the student.

☐ 2. Read **BOLDED** directions aloud to the student.

☐ 3. Point to the title and say:

☐ 4. "**This is a story about____** (title). **When I say, "BEGIN," start reading aloud on the top of the page. READ ACROSS THE PAGE** (point across the first line and then go to the next line). **Try to read EACH WORD. If you come to a word that you DON'T KNOW, I'll tell it to you. Be sure to do your BEST READING. Okay?**" (pause).

   a. For every other passage in the series that day, point to the title and say: "**This is a story about____ (title). Be sure to do your BEST reading.**"

   (pause).

☐ 5. "**Ready?** (pause) **Begin.**"

☐ 6. Start the timer when the student says the first word.

   a. If the student fails to say the first word after 3 seconds, say the word, mark it as incorrect, then start the timer.

☐ 7. If the student starts speed reading, pause the timer immediately and say,

   a. **"Stop. Do not do your fastest reading. Be sure to do your BEST reading. Let's start again. Ready, begin."**

   b. Resume the timer when the student reaches the point where they were interrupted.

☐ 8. At 1 min - the buzzer sounds - say: "**STOP**".

☐ 9. Mark the LAST WORD the child said within 1 min.

☐ **Total number steps completed.**

Instructions for assessing fidelity:

Mark a step as completed by marking a 1 (yes) or 0 (no) in the box next to the step. Write the total number in the box provided.

Appendix G

Sample Intervention Log, Filled Out

Key				
Attendance	P=student present for any portion of intervention		A=student absent for entirety of intervention	
Intervention Duration	Record the number of minutes in intervention for each student			
Engagement	Task Related Behavior • Responsive to teacher and peers • Used eye contact • Read audibly • Asked text related questions • Conversation focused on the text/reading Physical Behavior • In intervention area • Reading posture that allows for easy reading • Hands engaged with reading materials			
	1-Never/Rarely	2-Sometimes	3-Often	4-Always
First Read WRC	For research purposes only.		Do not fill in.	
Fourth Read WRC	For research purposes only.		Do not fill in.	

Date	**Are there any behavior/engagement difficulties with which you need help?**	**Behavior/Engagement Plan**
	Lily had a great day! No issues.	
	N/A	
	N/A	
	N/A	

Monday	Date:		VISTA:					
Student Name	Attendance P/A	Intervention Duration	Engagement	WRC/ERR				
				1	2	3	4	total
Lily	P	23:47	0 1 2 (3)	57/2	63/1	75/0	81/0	276/3
Was timer started/stopped immediately at the beginning/end of intervention?		☒ YES ☐ NO	If no, please explain.					
			0 1 2 3					
Was timer started/stopped immediately at the beginning/end of intervention?		☐ YES ☐ NO	If no, please explain.					
Notes								

Tuesday	Date:		VISTA:					
Student Name	Attendance P/A	Intervention Duration	Engagement	WRC/ERR				
				1	2	3	4	total
			0 1 2 3					
Was timer started/stoppped immediately at the beginning/end of intervention?		☐ YES ☐ NO	If no, please explain.					
			0 1 2 3					
Was timer started/stopped immediately at the beginning/end of intervention?		☐ YES ☐ NO	If no, please explain.					
Notes								

Appendix H

Sample Coded Reading Passage

Administrator Copy **Grade 2 Form 3 (1 per Occasion)**

Jim

Jim's grandfather was a mountain climber. He climbed many 9
different mountains in all parts of the world. Jim wanted to share his 22
adventures and climb with him. He asked his grandfather if he could 34
go with him sometime. His grandfather said he was climbing the 45
next day. He agreed that Jim could come if he followed the rules. 58
Jim agreed and said he would be ready to go. 68

Jim woke up early and got dressed for the hike. His 79
grandfather picked him up and they drove to the mountain. They 90
started on the trail by the side of the mountain. They started to walk 104
very slowly and Jim wondered why. His grandfather told him it was 116
because the mountain was tall. They would need to save energy to 128
finish. But Jim did not feel tired and wanted to climb fast. 140

Before long, Jim grew restless with the slow pace. He wanted 151
to walk faster and he did. He passed his grandfather and 162
disappeared out of site. His grandfather called for him to stay on the 175
path. Jim called back to say that he would. Both of them 187
continued to hike. 190

Jim's legs felt heavy and he began to feel tired. He sat down 203
on a big rock to rest. Time went by and his grandfather caught up 217
to him. His grandfather still had lots of energy. He sat down and 230
waited for Jim to feel better. When Jim felt better, they hiked 242
together, very slowly. Later that day, they reached the top of the 254
mountain together. 256

Total Words Read: ___________ **- # of Errors:** _____________ **= WRC/min:** ________________

Student: ____________________ **Teacher:** ________________ **Grade:** ________ **Date:** _________

Appendix I

Repeated Reading Protocol

Step	Say:
1. START TIMER FOR EACH STUDENT	
2. ARTICULATE OBJECTIVE Explain that students will be reading a passage several times,	**"Today, we'll read (*title of passage*) multiple times to work on increasing our fluency. Fluency is the rate, accuracy, and expression that we read—not just how fast we can read. At the end of the passage, we'll answer some questions together."**
3. EXPLAIN GAME OR ACTIVITY Explain to each student that they will be reading together for one minute,	**"You will read this passage for one minute. While you read, your partner will keep track of the errors you make. After you read, your partner will help you correct the errors. Then you will switch roles. Each time that you read, you'll try and read more that the last time, I will be here to monitor the reading and help error correct."**
4. CHECK FOR STUDENT UNDERSTANDING	**"Does anyone have any questions?"**
5. STUDENT COLD READ (READ #1)	**"You will read this story for one minute while (student 1/2) reads this other story quietly"** Hand other story to the second student. Start with the first student: ***"Begin."*** Start stopwatch. At one minute, say: **"Stop"** Mark the last word said by the students.
6. Provide corrective feedback on expression; words student missed; graph WRC/ERR;	**"This word is ___. What word is this? Yes, that word is ___. Please re-read from here** *(indicate a starting point so the student reads a short phrase that includes the word missed)* *(graph WRC/ERR)* *(give student feedback on expression)*
7. Repeat steps 5 and 6 for the second student.	
8. ***MODEL THE ACTIVITY OR GAME*** Model reading 2-3 paragraphs from the passage to the students, explaining to then that they will be working in partnerships to practice rate, accuracy, and expression,	**"Follow along as I model reading the passage fluently. When I am finished you will take turns reading the passage with your partner, helping each other error correct."**
9. Ask the students what the passage is mostly about	**"What is the passage mostly about?"**
10. Allow students to respond. If answers are reasonable, continue. If the answers are not reasonable, provide a brief error correction,	**"I think this passage is mostly about ____."**

11. ***Provide guided practice* (Read # 2)** After model-reading the passage to the students, explain that it is their turn to continue practice reading. Start with the first student(s),	**Read #2**	**Read #3**	**Read #4**
	"Now, you'll read the same passage again and try to read more than you read last time. ***Begin."***		**"Now you'll read the passage a fourth time. When you're done, we'll talk about the most important thing you learned (the who or what)."**
	Start stopwatch. At one minute, say: **"Stop"** Mark the last word said by the students.		
12. Provide corrective feedback on expression; words student missed; graph WRC/ERR;	**"This word is ___. What word is this? Yes, that word is ___. Please re-read from here** *(indicate a starting point so the student reads a short phrase that includes the word missed)* *(graph WRC/ERR)* *(give student feedback on expression)*		
13. Ask students about the most important thing they learned. Remember to include think time so every student has a chance to think about the answer,	**"Now, I want you each to think to yourselves (in your head so do not say it out loud) what the most important thing was you learned (**provide wait time**). Now, who can tell me?"**		
14. Have each student give an answer. Model a correct answer if needed.			
15. Collect the passages and explain to students that we are going to make a prediction,	**"Now, let's think about what we think will happen next or predict what the rest of the passage will be about."** **"I want you each to think to yourselves (in your head so do not say it out loud), based on what we have read so far, what do you predict the rest of the passage will be about?"** Ask the students to share their predictions.		
16. Read the rest of the passage to the students,	**"Now, listen as I read the rest of the passage so we can find out if your predictions are accurate."**		
17. Discuss whether students' predictions were accurate. Provide feedback.			
18. If time allows, have students make a connection to the passage,	**"What does this passage make you think about?"**		
19. **STOP TIMER FOR EACH STUDENT**			

Appendix J

Repeated Reading Fidelity Checklist

Date: ___________ Observer: ____
Interventionist: _________________________________
School: _____________________________________
Student 1 Fidelity: ________________: _____# of YES/ 29 = _____% of YES
Student 2 Fidelity: ________________: _____# of YES/ 29 = _____% of YES
Grade (CIRCLE): 2 3

Tier 2
Intervention Protocol
Fluency: Repeated Reading (with comprehension)

Big Five Reading Target: Fluency (Transitional)

Student 1	**Student 2**	**Intervention Procedure**	**Item Weight**
		1. START TIMER FOR BOTH STUDENTS (YES/NO)	
		2. ARTICULATE OBJECTIVE:	2
		3. EXPLAIN GAME OR ACTIVITY:	1
		4. CHECK FOR STUDENT UNDERSTANDING: (Week 1 only)	1
		5. Student Cold Read (READ #1): • One minute long – cold read	3
		5. ***MODEL THE ACTIVITY OR GAME:*** • Reads about 2 paragraphs • Ask, **"What is the passage mostly about?"**	2
		6. ***PROVIDE GUIDED PRACTICE (READ #2):*** • One minute long	3
		7. GAVE SPECIFIC FEEDBACK:	3
		8. PROVIDED PRACTICE (READ #3):	3
		9. GAVE SPECIFIC FEEDBACK:	3
		10. CUE THINKING ABOUT "THE MOST IMPORTANT THING:	.5
		11. PROVIDED PRACTICE (READ #4):	3
		12. GAVE SPECIFIC FEEDBACK:	3
		13. ASK ABOUT "THE MOST IMPORTANT THING:	.5
		14. MAKE A PREDICTION:	.5
		15. COLLECT PASSAGES FROM STUDENTS (YES/NO):	
		16. READ THE PASSAGE AND DISCUSS THE PREDICTION:	.5
		17. TURN OFF THE TIMER FOR BOTH STUDENTS (YES/NO)	
			Total Points: 29

Appendix K

Descriptive Statistics

Table K1
Descriptive Statistics for Percentage of Attendance

	n	*M*	*SD*	σ^2	*SE*	Min	Max	Skew	Kurt
Grade 2 - all	44	.89	.12	.02	.02	.44	1.00	-1.48	2.96
Intervention	22	.89	.13	.02	.03	.44	1.00	-1.96	5.12
Control	22	.89	.11	.01	.02	.69	1.00	-.73	-.76
Grade 3 - all	35	.84	.15	.02	.02	.50	1.00	-.98	.78
Intervention	18	.84	.15	.02	.03	.50	1.00	-.95	.51
Control	17	.85	.15	.02	.04	.50	1.00	-1.13	.14
Total	79	.87	.13	.02	.02	.44	1.00	-1.23	1.15
Intervention	40	.87	.14	.02	.02	.44	1.00	-1.37	1.86
Control	39	.87	.13	.02	.02	.50	1.00	-1.01	.43

Note. M = Median; *SD* = Standard Deviation; σ^2 = Variance; *SE* = Standard Error of mean; Min = Minimum; Max = Maximum; Kurt = Kurtosis

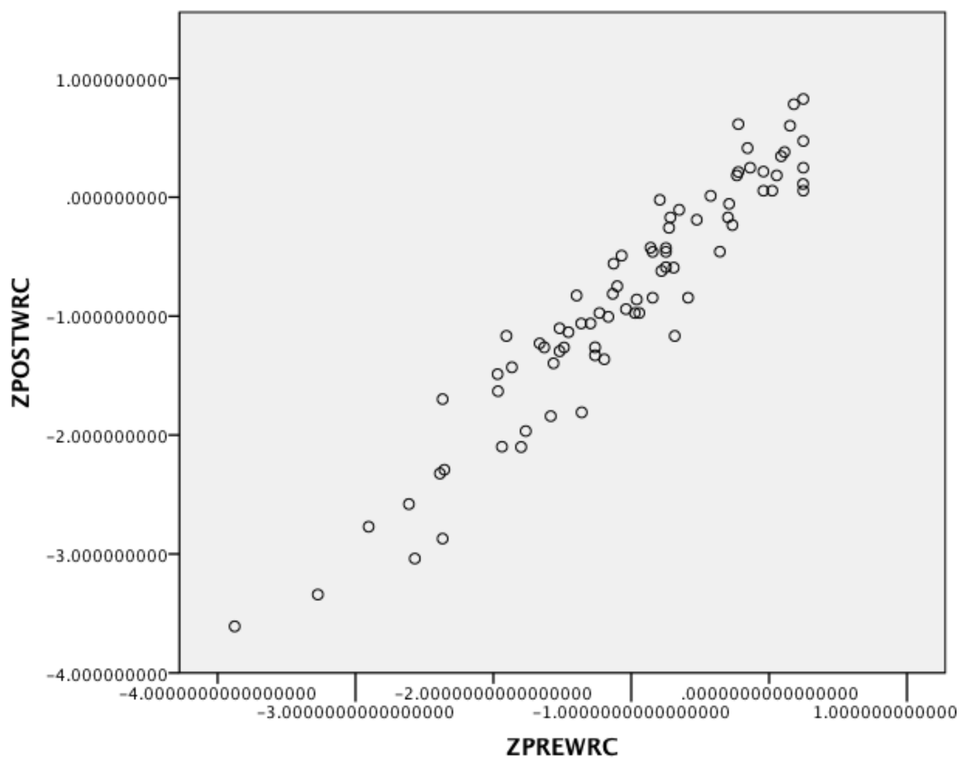


Figure K1. Plot of the linear relation between the covariate (ZPREWRC) and dependent (ZPOSTWRC) variable, for intervention and control groups.

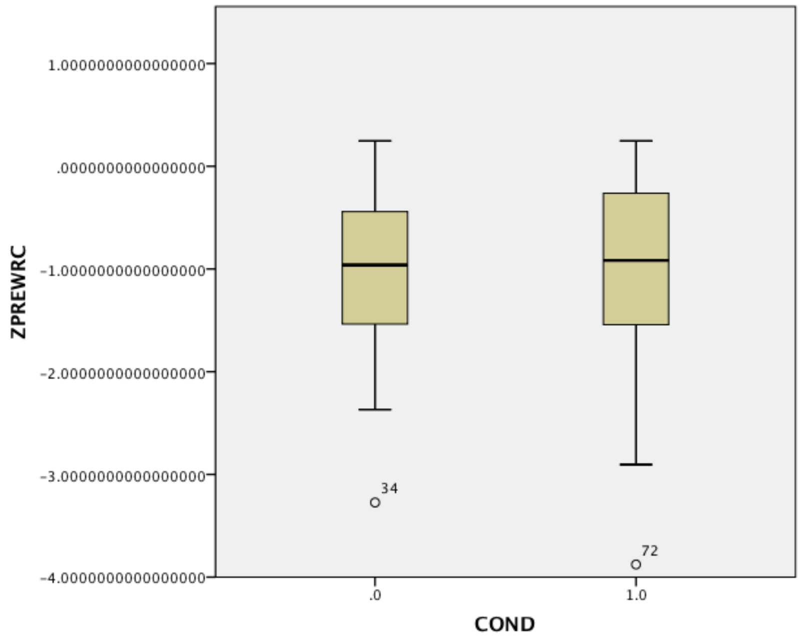


Figure K2. Boxplots of pretest (ZPREWRC) for condition (0 = control, 1 = intervention).

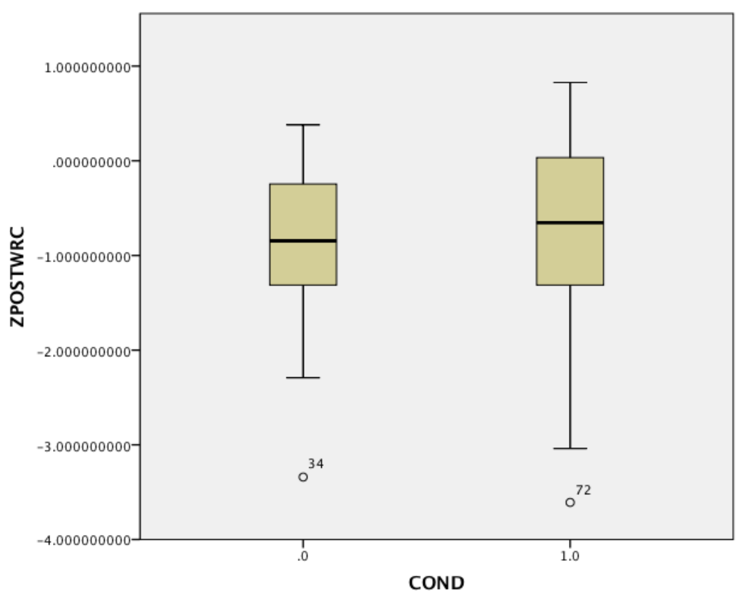


Figure K3. Boxplot of posttest scores (ZPOSTWRC) for condition (0 = control, 1 = intervention).

Table K2

Descriptive Statistics for Unstandardized Pre- and Post-test Scores for All Students by Condition

	n	*M*	*SD*	σ^2	*SE*	Min	Max	Skew	Kurt
Pretest									
Intervention	40	90.85	28.64	820.03	.15	20	141	-.53	-.16
Control	39	92.38	25.60	655.14	4.01	38	143	-.14	-.34
Posttest									
Intervention	40	99.00	33.17	1099.90	5.24	28	159	-.34	-.41
Control	39	96.51	27.21	740.57	4.36	36	147	-.32	-.28

Note. M = median; *SD* = standard deviation; σ^2 = variance; *SE* = standard error of mean; Min = minimum; Max = maximum; Kurt = kurtosis.

Table K3

Descriptive Statistics for Standardized Pre- and Post-test Scores (z-scores) for all Students by Condition

	n	*M*	*SD*	σ^2	*SE*	Min	Max	Skew	Kurt
Pretest									
Intervention	40	-1.07	.97	.95	.15	-3.88	.25	-.77	.42
Control	39	-1.01	.81	.66	.13	-3.27	.25	-.46	.20
Posttest									
Intervention	40	-.80	.97	.95	.15	-3.88	-.25	-.71	-.12
Control	39	-.87	.83	.70	.13	-3.34	.38	-.64	.58

Note. M = median; *SD* = standard deviation; σ^2 = variance; *SE* = standard error of mean; Min = minimum; Max = maximum; Kurt = kurtosis.

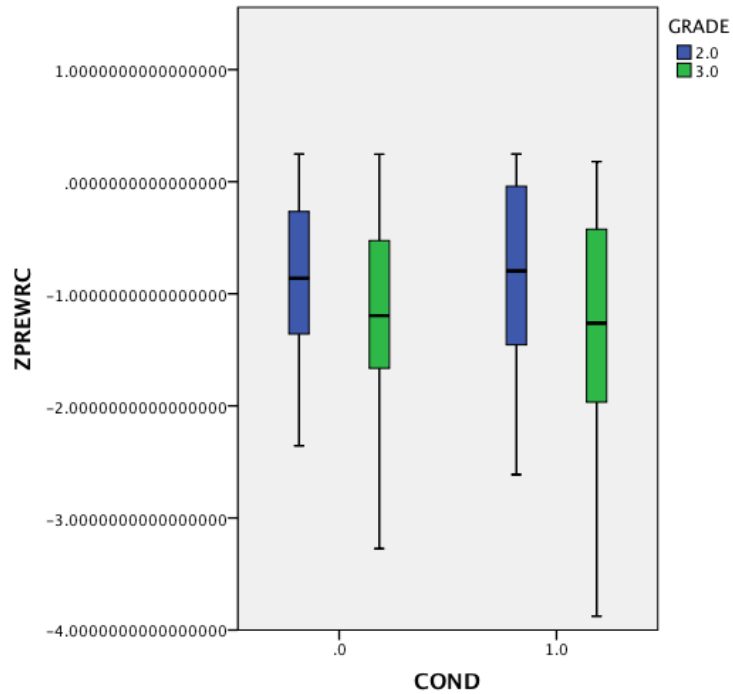


Figure K4. Boxplot of pretest scores (ZPREWRC) for condition (0 = control, 1 = intervention) and grade.

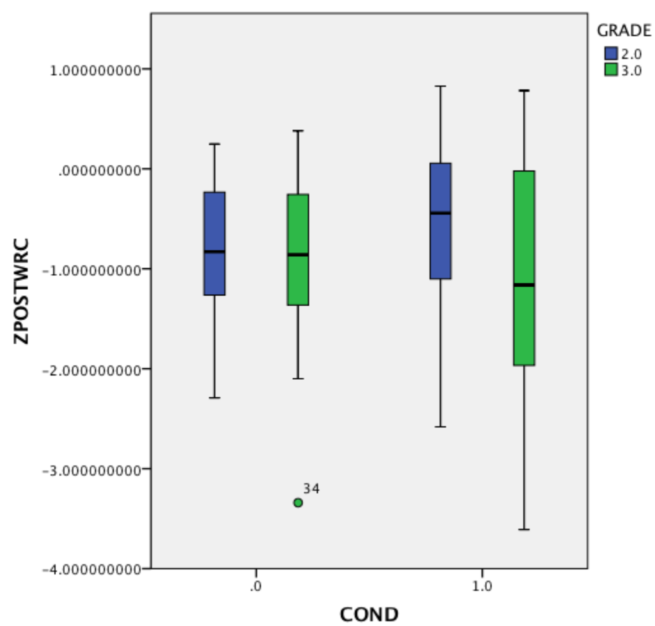


Figure K5. Boxplot of posttest scores (ZPOSTWRC) for condition (0 = control, 1 = intervention) and grade.

Table K4

Descriptive Statistics for Unstandardized Pre- and Post-test Scores for Students by Condition and Grade

	n	*M*	*SD*	σ^2	*SE*	Min	Max	Skew	Kurt
Pretest									
Grade 2									
Intervention	22	87.45	25.83	667.40	5.51	32	121	-.54	-.36
Control	22	85.64	22.57	509.20	4.81	40	121	-.17	-.56
Grade 3									
Intervention	18	95.00	31.99	1023.65	7.54	20	141	-.72	.23
Control	17	101.12	27.28	743.99	6.62	38	143	-.51	.33
Posttest									
Grade 2									
Intervention	22	96.36	27.33	747.10	5.83	33	139	-.66	.30
Control	22	87.91	23.95	573.71	5.10	42	121	-.27	-.74
Grade 3									
Intervention	18	102.22	39.75	1580.42	9.37	28	159	-.36	-.91
Control	17	107.65	27.79	772.37	.04	36	147	-.90	1.44

Note. M = median; *SD* = standard deviation; σ^2 = variance; *SE* = standard error of mean; Min = minimum; Max = maximum; Kurt = kurtosis.

Table K5

Descriptive Statistics for Standardized Pre- and Post-test Scores (z-scores) for Students by Condition and Grade

	n	*M*	*SD*	σ^2	*SE*	Min	Max	Skew	Kurt
Pretest									
Grade 2									
Intervention	22	-.83	.83	.69	.18	-2.16	.25	-.54	-.36
Control	22	-.89	.73	.53	.15	-2.36	.25	-.17	-.56
Grade 3									
Intervention	18	-1.36	1.07	1.15	.25	-3.88	.18	-.72	.23
Control	17	-1.16	.91	.84	.22	-3.27	.25	-.51	.33
Posttest									
Grade 2									
Intervention	22	-.83	.88	.77	.19	-2.58	.83	-.66	.30
Control	22	-.82	.77	.50	.16	-2.29	.25	-.27	-.74
Grade 3									
Intervention	18	-1.12	1.33	1.78	.31	-3.61	.78	-.36	-.91
Control	17	-.94	.93	.87	.23	-3.34	.38	-.90	1.44

Note. M = Median; *SD* = Standard Deviation; σ^2 = Variance; *SE* = Standard Error of mean; Min = Minimum; Max = Maximum; Kurt = Kurtosis

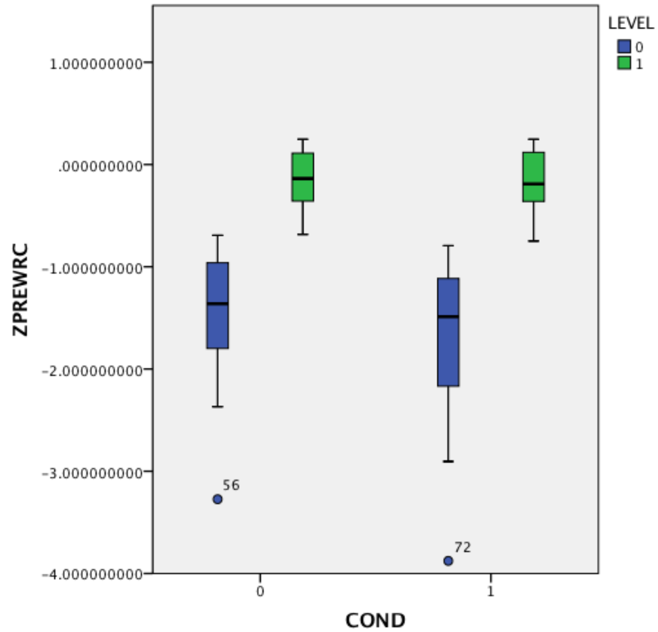


Figure K6. Boxplot of pretest scores (ZPREWRC) for condition (0 = control, 1 = intervention) and level in terms of words read correctly per minute (WRCM); relatively high = WRCM at the 26-50th percentile; low = WRCM at the 0-25th percentile.

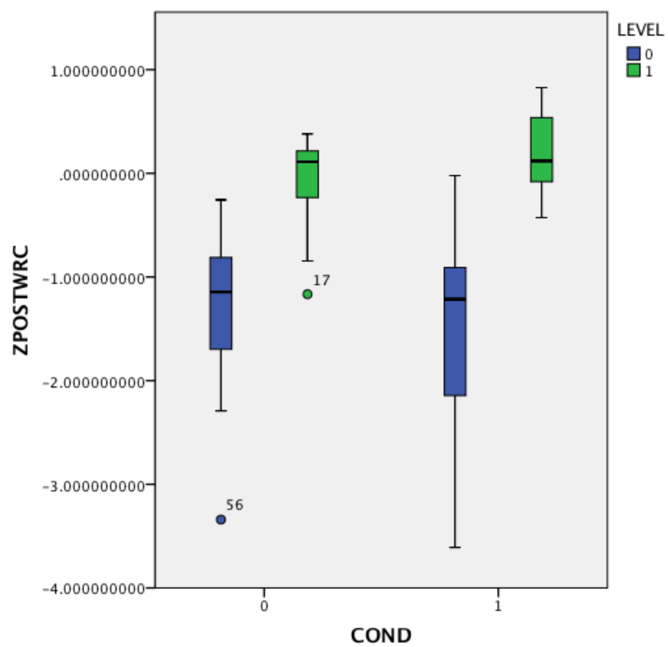
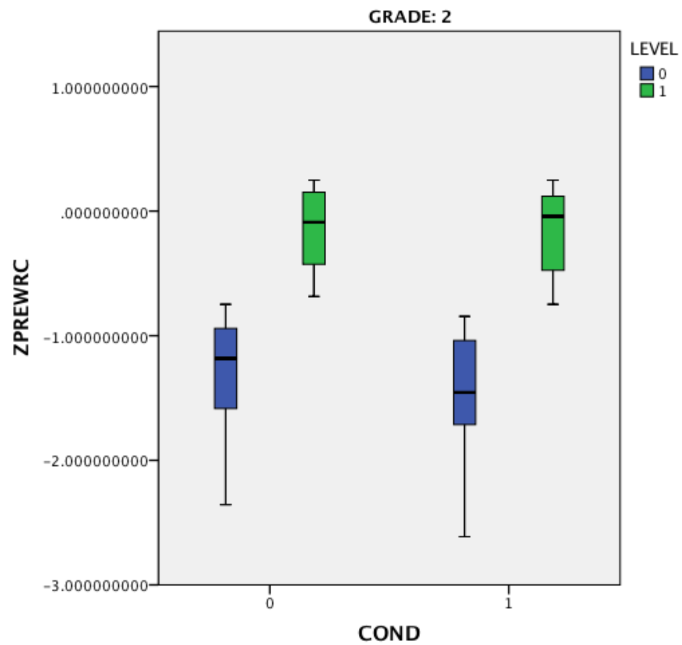


Figure K7. Boxplot of posttest scores (ZPOSTWRC) for condition (0 = control, 1 = intervention) and level in terms of words read correctly per minute (WRCM); relatively high = WRCM at the 26-50th percentile; low = WRCM at the 0-25th percentile.



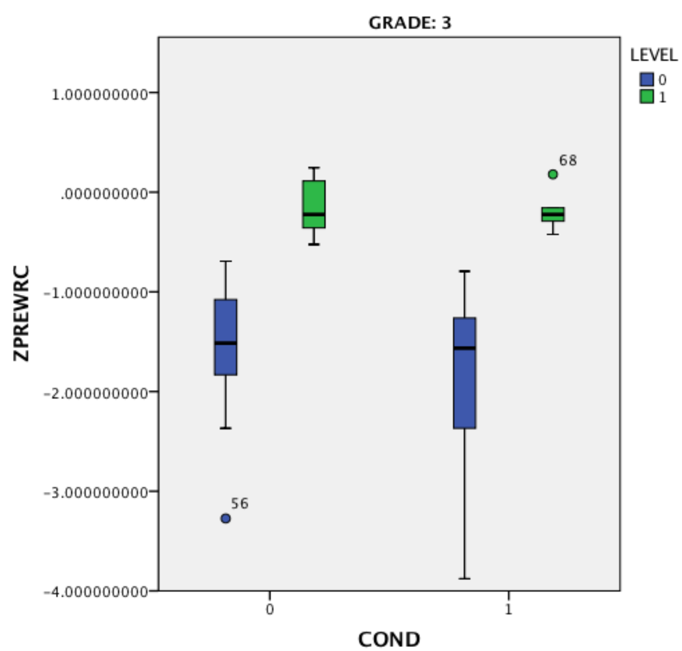
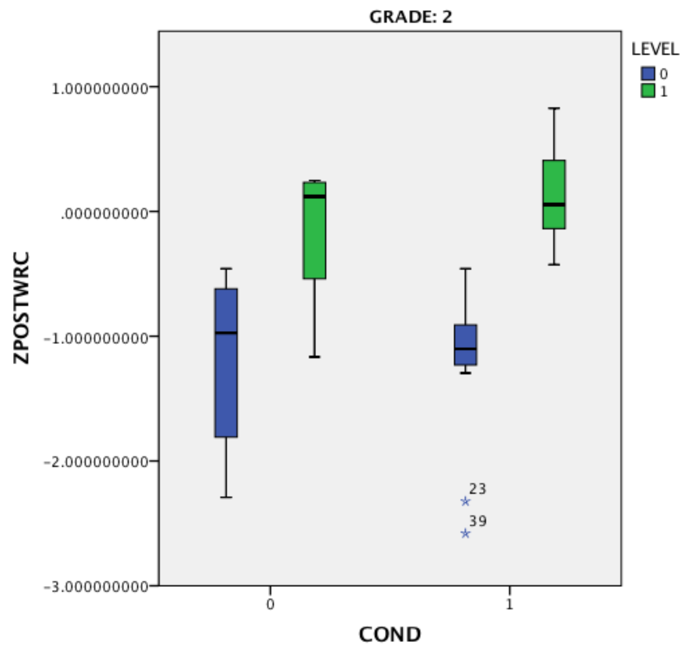


Figure K8. Boxplots of pretest scores (ZPREWRC) for condition (0 = control, 1 = intervention), grade, and level in terms of words read correctly per minute (WRCM); relatively high = WRCM at the 26-50th percentile; low = WRCM at the 0-25th percentile.



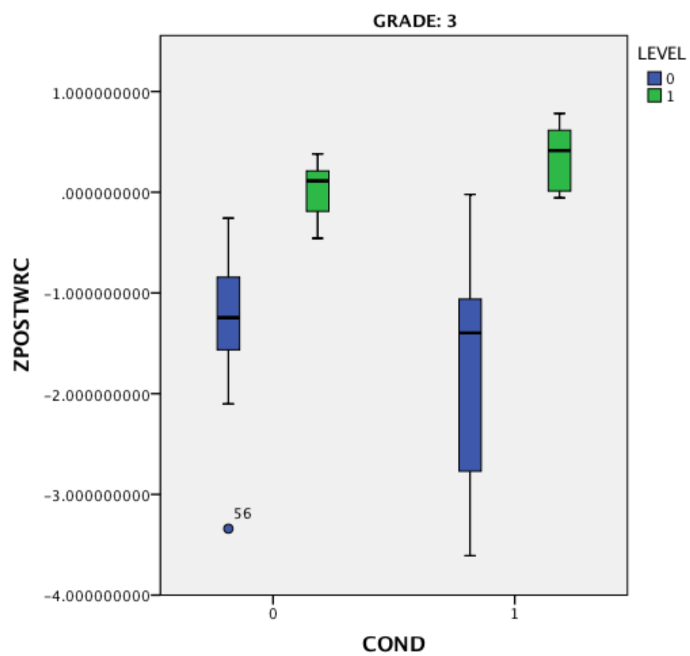


Figure K9. Boxplots of posttest scores (ZPOSTWRC) for condition (0 = control, 1 = intervention), grade, and level in terms of words read correctly per minute (WRCM); relatively high = WRCM at the 26-50th percentile; low = WRCM at the 0-25th percentile.

Table K6

Descriptive Statistics for Standardized Pre- and Post-test Scores (z-scores) by Condition and Level

	n	*M*	*SD*	σ^2	*SE*	Min	Max	Skew	Kurt
Pretest									
Relatively high level									
Intervention	16	-.18	.33	.11	.08	-.75	.25	-.42	-.91
Control	13	-.15	.33	.11	.09	-.68	.25	-.27	-1.21
Low level									
Intervention	24	-1.66	.78	.61	.16	-3.88	-.79	-1.19	1.28
Control	26	-1.44	.62	.34	.12	-3.27	-.69	-1.14	1.74
Posttest									
Relatively high level									
Intervention	16	.21	.37	.14	.09	-.43	.83	.17	-1.03
Control	13	-.10	.46	.22	.13	-1.17	.38	-1.34	.98
Low level									
Intervention	24	-1.48	.93	.87	.19	-3.61	-.02	-.76	-.20
Control	26	-1.26	.70	.49	.14	-3.34	-.26	-1.12	1.73

Note. M = Median; *SD* = Standard Deviation; σ^2 = Variance; *SE* = Standard Error of mean; Min = Minimum; Max = Maximum; Kurt = Kurtosis. Level in terms of words read correctly per minute (WRCM); relatively high = WRCM at the 26-50th percentile; low = WRCM at the 0-25th percentile.

Table K7

Descriptive Statistics for Standardized Pre- and Post-test Scores (z-scores) by Level and Grade

	n	*M*	*SD*	σ^2	*SE*	Min	Max	Skew	Kurt
Pretest									
Relatively high level									
Second grade	19	-.16	.36	.13	.08	-.75	.25	-.49	-1.21
Third grade	10	-.17	.27	.07	.08	-.52	.24	.46	-1.05
Low Level									
Second grade	25	-1.39	.54	.29	.11	-2.61	-.75	-.83	-.20
Third grade	25	-1.70	.81	.66	.16	-3.88	-.69	-1.01	.93
Posttest									
Relatively high level									
Second grade	19	.02	.47	.22	.11	-1.17	.83	-.91	1.41
Third grade	10	-.18	.38	.14	.12	-.46	.78	-.04	-.45
Low level									
Second grade	25	-1.21	.62	.38	.12	-2.58	-.46	-.86	-.22
Third grade	25	-1.52	.97	.93	.19	-3.61	-.02	-.71	-.19

Note. M = Median; *SD* = Standard Deviation; σ^2 = Variance; *SE* = Standard Error of mean; Min = Minimum; Max = Maximum; Kurt = Kurtosis. Level in terms of words read correctly per minute (WRCM); relatively high = WRCM at the 26-50th percentile; low = WRCM at the 0-25th percentile.

Table K8

Descriptive Statistics for Standardized Pre- and Post-test Scores (z-scores) by Level, Grade, and Condition

	n	*M*	*SD*	σ^2	*SE*	Min	Max	Skew	Kurt
Pretest									
Relatively high level, Grade 2									
Intervention	11	-.18	.38	.15	.12	-.75	.25	-.53	.1.35
Control	8	-.15	.35	.12	.12	-.68	.25	-.50	-1.00
Low level, Grade 2									
Intervention	11	-1.48	.60	.36	.18	-2.61	-.84	-.85	-.27
Control	14	-1.31	.50	.25	.13	-2.36	-.75	-.78	-.24
Relatively high level, Grade 3									
Intervention	5	-.18	.23	.05	.10	-.42	.18	1.17	2.03
Control	5	-.15	.32	.10	.14	-.52	.25	-.58	-.96
Low level, Grade 3									
Intervention	13	-1.82	.90	.81	.25	-3.88	-.79	-1.01	.76
Control	12	-1.58	.72	.52	.21	-3.27	-.70	-1.10	1.70
Posttest									
Relatively high level, Grade 2									
Intervention	11	.15	.38	.14	.11	-.43	.83	.36	-.57
Control	8	-.16	.55	.30	.19	-1.17	.25	-1.24	.04
Low level, Grade 2									
Intervention	11	-1.24	.64	.41	.19	-2.58	-46	-1.37	1.26
Control	14	-1.19	.62	.38	.16	-2.29	-.46	-.54	-1.01
Relatively high level, Grade 3									
Intervention	5	.35	.37	.14	.16	-.06	.78	-.10	-2.50
Control	5	.01	.33	.11	.15	-.46	.38	-.58	-.96
Low level, Grade 3									
Intervention	13	-1.69	1.10	1.21	.31	-3.61	-.02	-.30	-.91
Control	12	-1.33	.80	.64	.23	-3.34	-.26	-1.42	3.01

Note. M = Median; *SD* = Standard Deviation; σ^2 = Variance; *SE* = Standard Error of mean; Min = Minimum; Max = Maximum; Kurt = Kurtosis. Level in terms of words read correctly per minute (WRCM); relatively high = WRCM at the 26-50th percentile; low = WRCM at the 0-25th percentile.

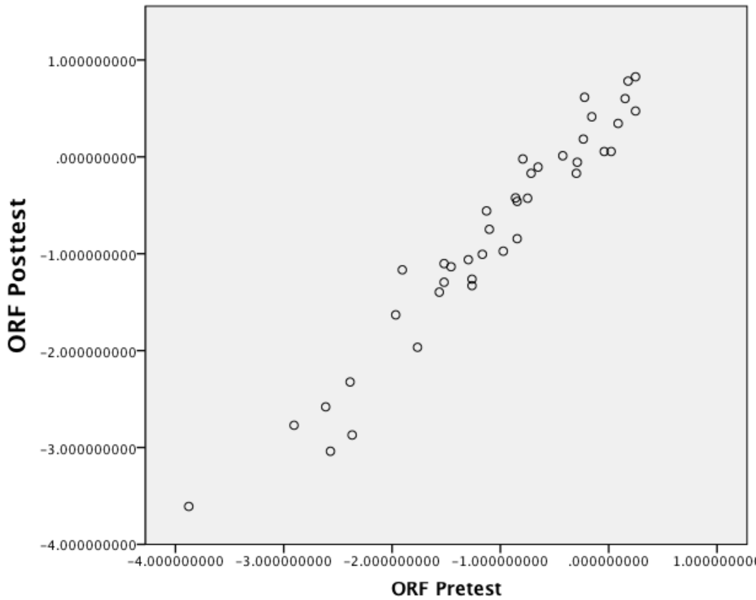


Figure K10. Plot of the linear relation between oral reading fluency (ORF) posttest and pretest scores for the intervention group.

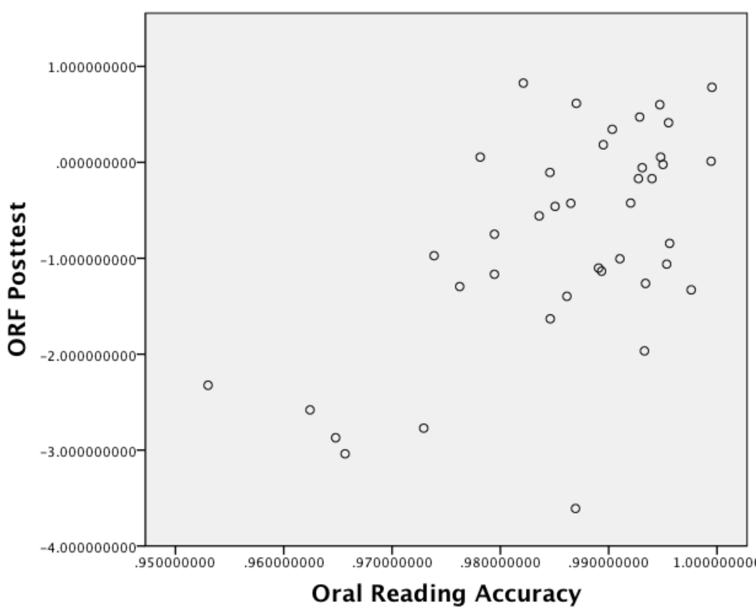


Figure K11. Plot of the linear relation between oral reading fluency (ORF) posttest and oral reading accuracy for the intervention group.

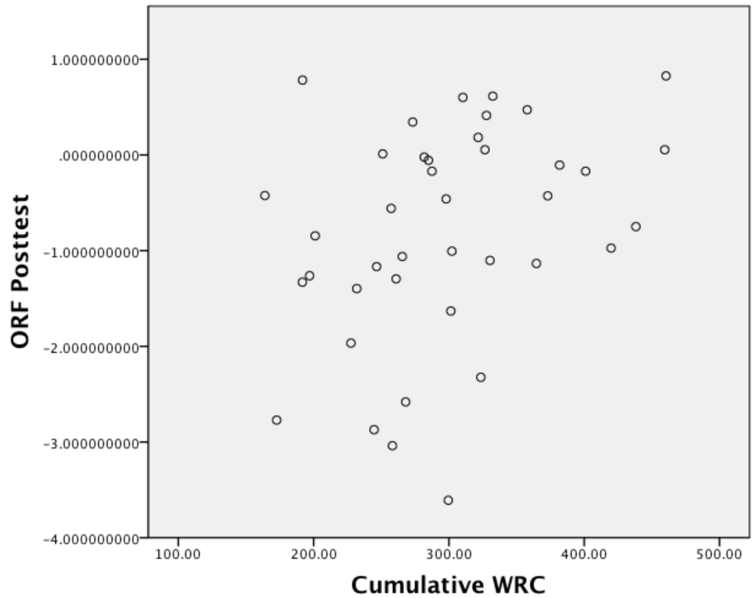


Figure K12. Plot of the linear relation between oral reading fluency (ORF) posttest and cumulative words read correctly (WRC) for the intervention group.

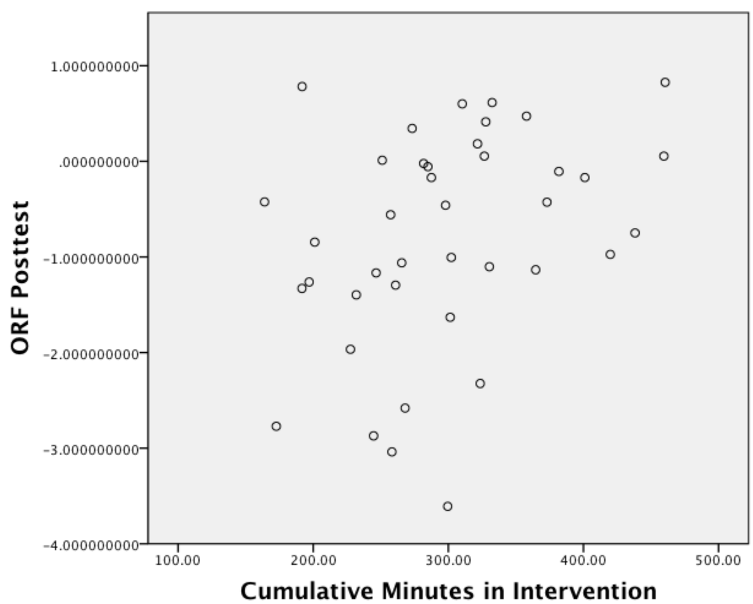


Figure K13. Plot of the linear relation between oral reading fluency (ORF) posttest and cumulative minutes in intervention for the intervention group.

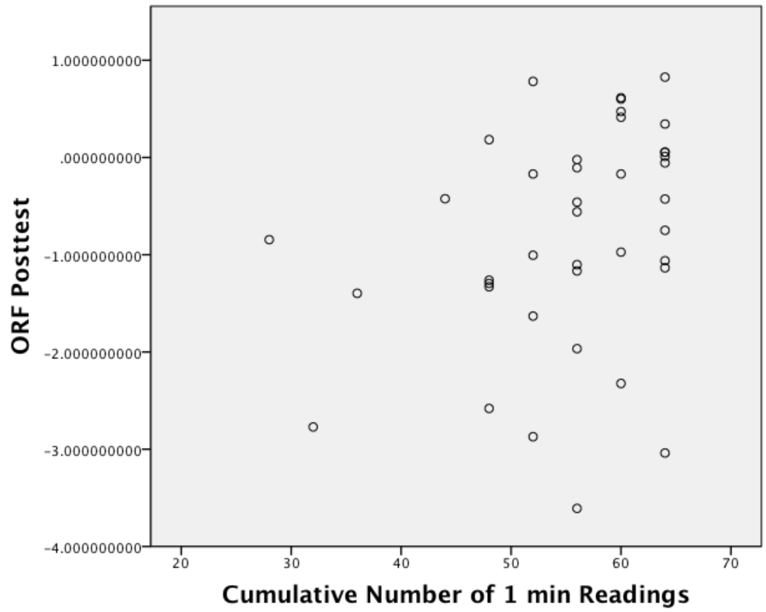


Figure K14. Plot of the linear relation between oral reading fluency (ORF) posttest and cumulative number of 1 min readings for the intervention group.

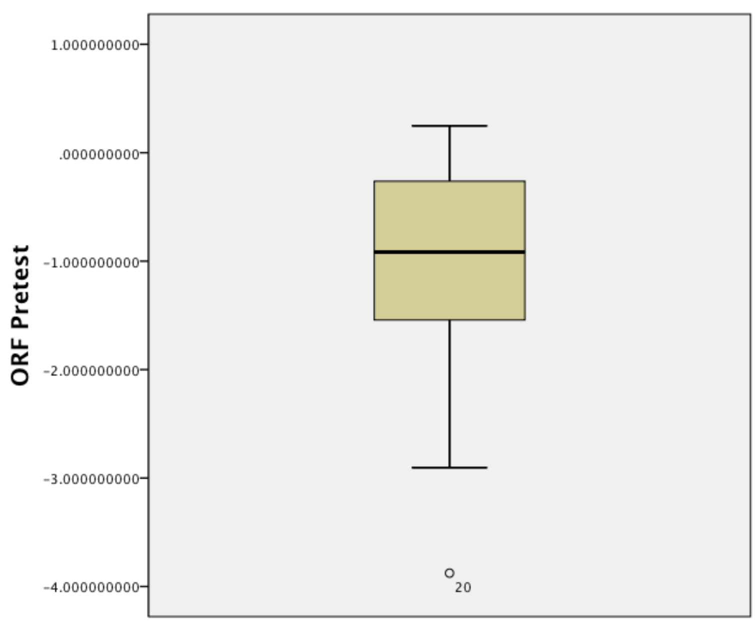


Figure K15. Boxplot of oral reading fluency (ORF) pretest scores for the intervention group.

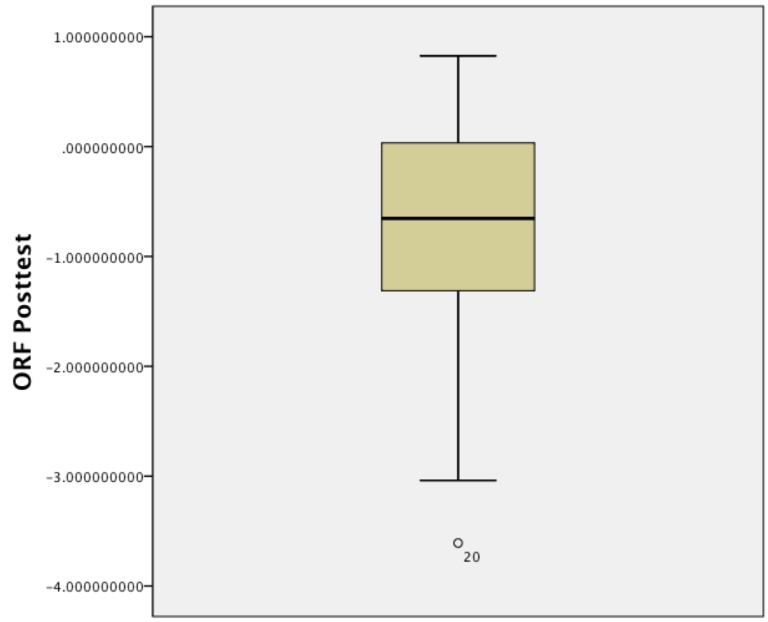


Figure K16. Boxplot of oral reading fluency (ORF) posttest scores for the intervention group.

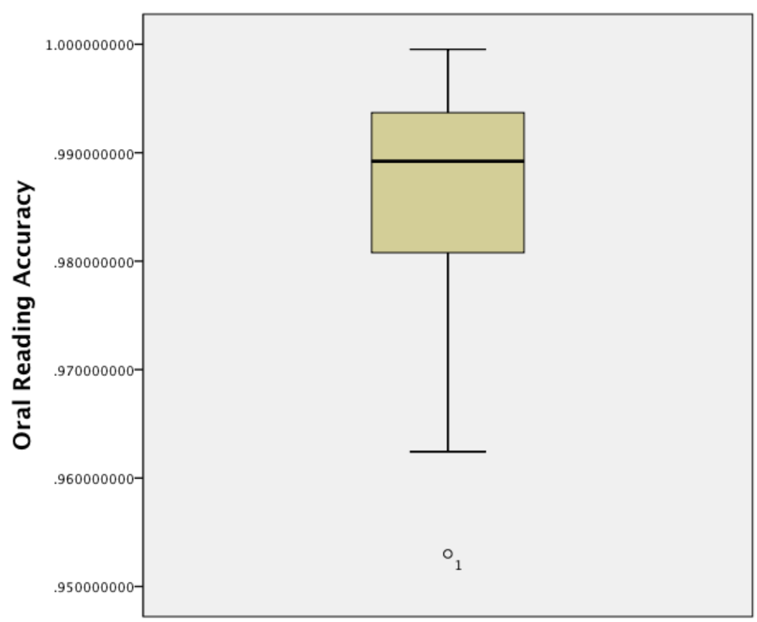


Figure K17. Boxplot of mean oral reading accuracy.

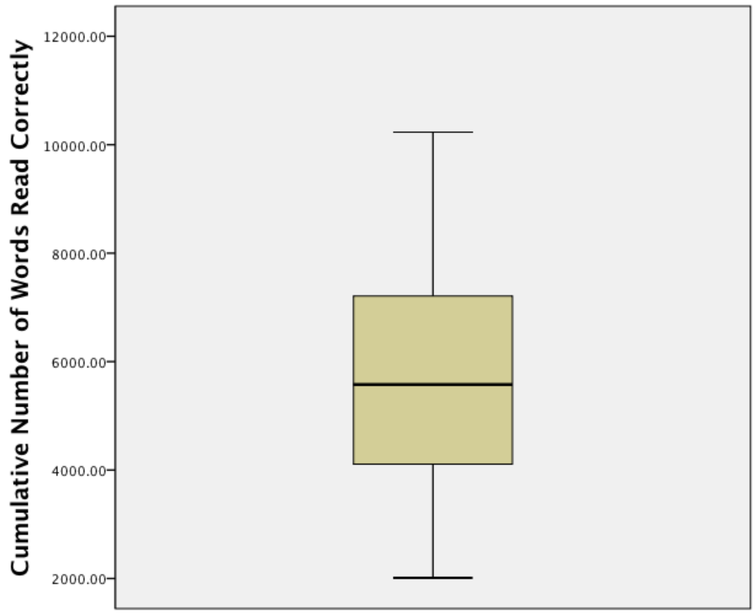


Figure K18. Boxplot of mean cumulative words read correctly.

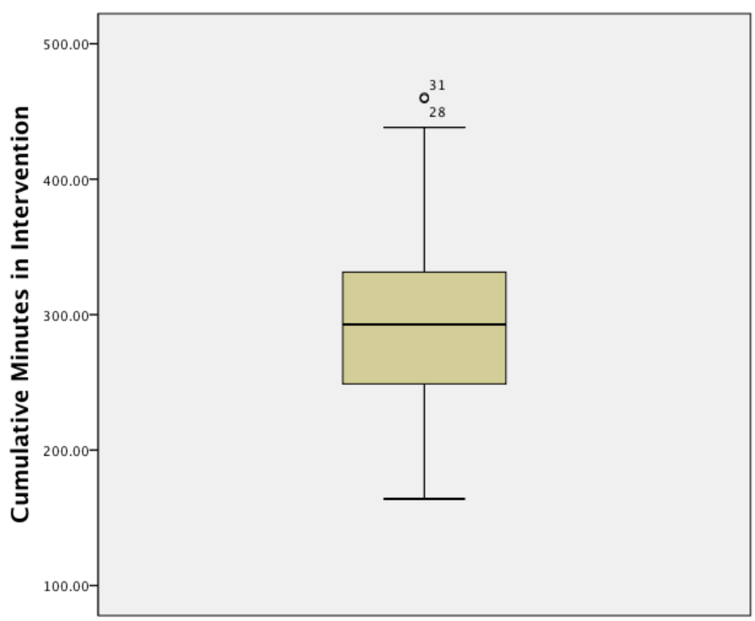


Figure K19. Boxplot of cumulative minutes in intervention.

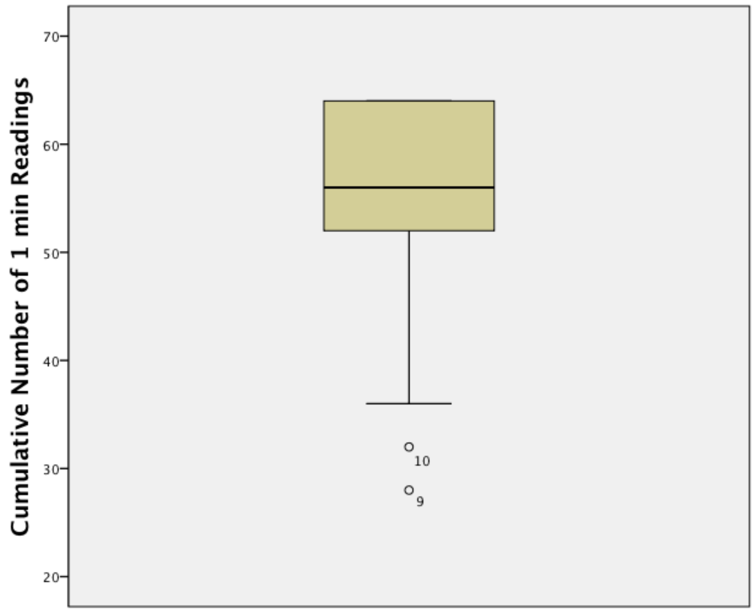


Figure K20. Boxplot of cumulative number of 1 min readings.

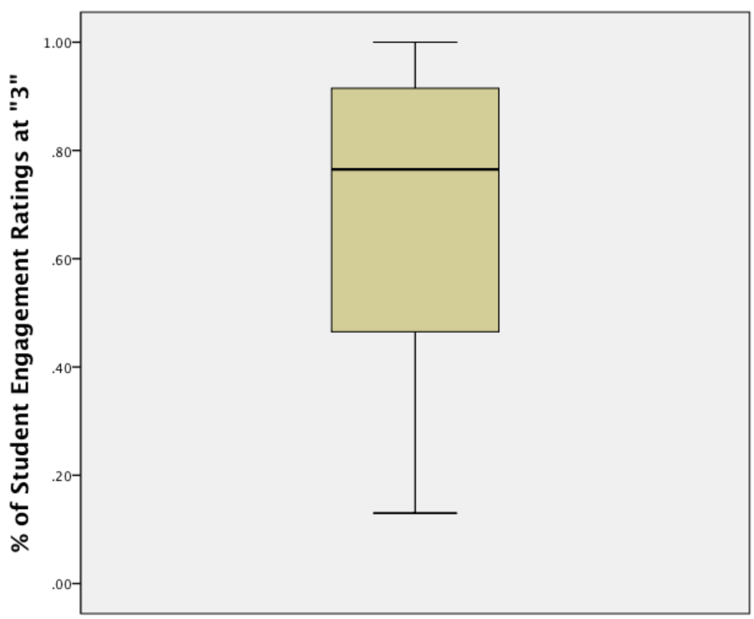


Figure K21. Boxplot of the percentage of student engagement ratings at “3”.

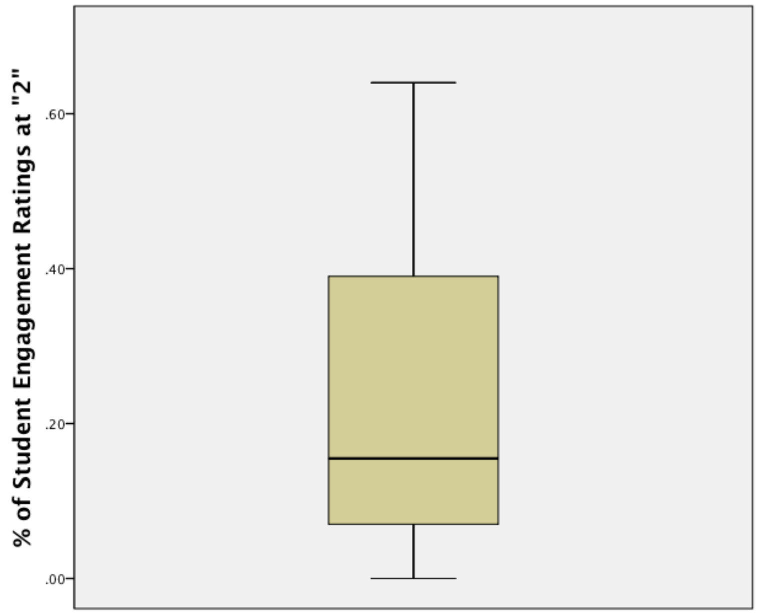


Figure K21. Boxplot of the percentage of student engagement ratings at “2”.

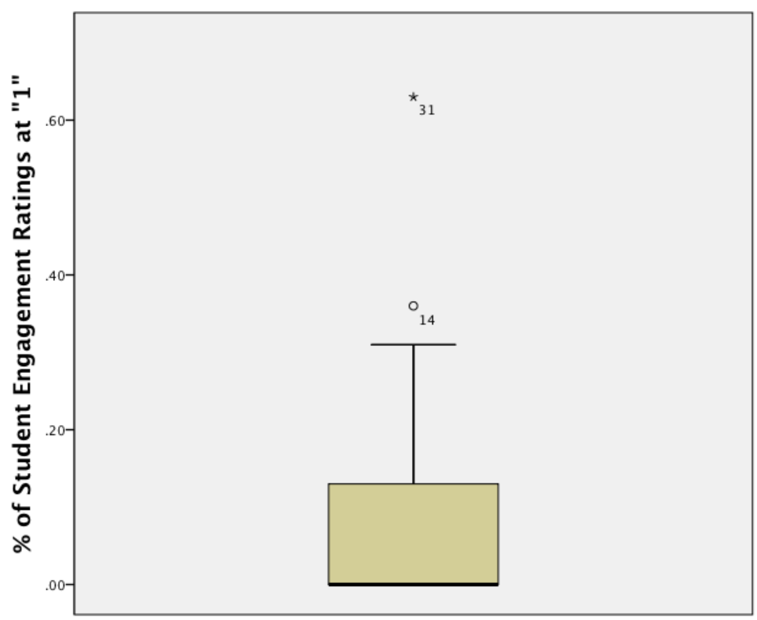


Figure K23. Boxplot of the percentage of student engagement ratings at “1”.

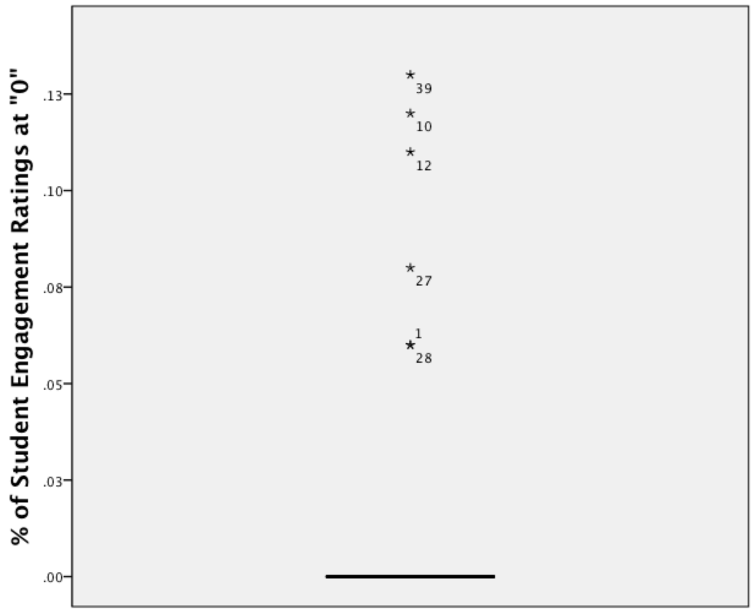


Figure K24. Boxplot of the percentage of student engagement ratings at “0”.

Table K9

Descriptive Statistics for ORF Pre- and Post-test z-scores for All Students in Intervention Groups

	n	*M*	*SD*	σ^2	*SE*	Min	Max	Skew	Kurt
Pretest									
Intervention	40	-1.07	.97	.95	.15	-3.88	.25	-.77	.42
Posttest									
Intervention	40	-.80	.97	.95	.15	-3.88	-.25	-.71	-.12

Note. ORF = oral reading fluency. *M* = Median; *SD* = Standard Deviation; σ^2 = Variance; *SE* = Standard Error of mean; Min = Minimum; Max = Maximum; Kurt = Kurtosis

Table K10

Descriptive Statistics for Variables Related to Implementation

Variable	*n*	*M*	*SD*	σ^2	*SE*	Min-Max	Skew	Kurt
1. Cumulative min spent in intervention	40	297.18	75.81	5746.93	11.99	164.00-460.44	.40	-.24
2. Mean % of 1 min readings completed per session	40	1.00	.00	.00	.00	1.00-1.00	.00	.00
3. Cumulative number of 1 min readings	40	55.30	8.92	79.50	1.41	28-64	-1.36	1.90
4. Cumulative # of words read correctly	40	5738.05	2000.49	4001954.61	316.31	2012.00-10,230.00	.23	-.62
5. Mean oral reading accuracy	40	.99	.01	.00	.00	.95-1.00	-1.56	.45
6. Student engagement score[a]								
0	40	.01	.03	.00	.01	.00-.13	2.61	5.59
1	40	.06	.10	.01	.02	.00-.36	1.50	1.34
2	40	.24	.20	.04	.03	.00-.64	.40	-1.26
3	40	.68	.27	.07	.04	.13-1.00	-.60	-.92

Note. M = median; *SD* = standard deviation; σ^2 = variance; *SE* = standard error of the mean; Min = minimum; Max = maximum; Kurt = kurtosis.

[a]Calculated on students' percentages of ratings at 0, 1, 2, and 3.

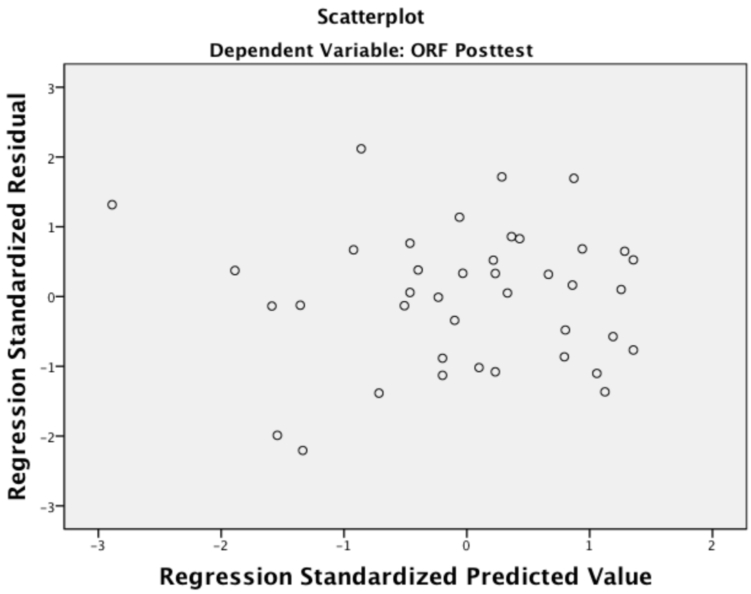


Figure K25. Plot of studentized residual and standardized predicted values for Model 1

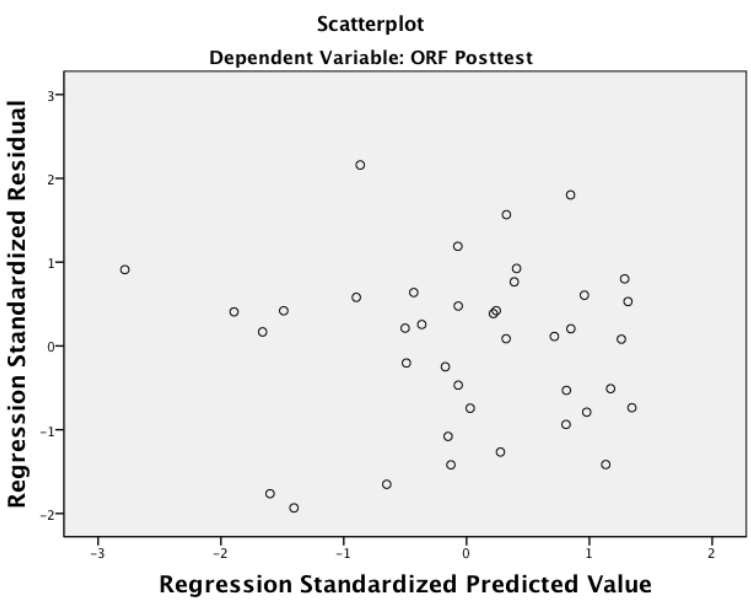


Figure K26. Plot of studentized residual and standardized predicted values for Model 2

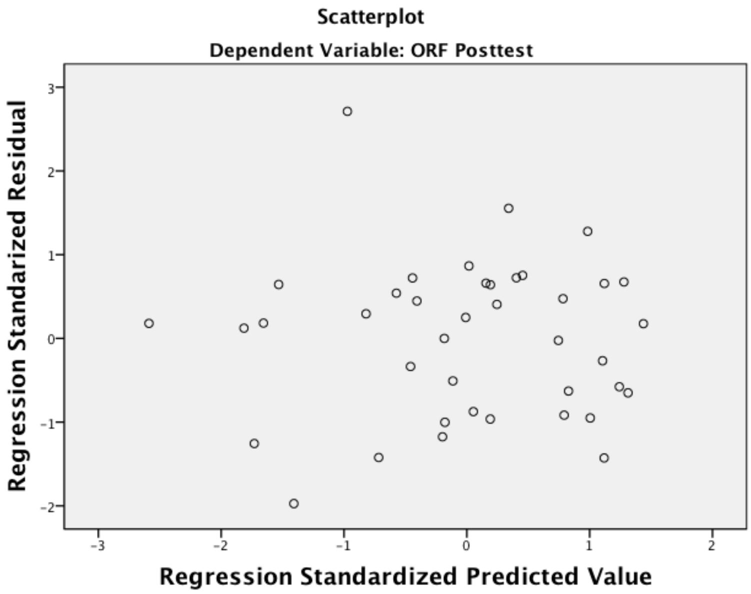


Figure K27. Plot of studentized residual and standardized predicted values for Model 3

Appendix L

Operational Definitions and Rating Levels for Measuring Student Engagement

VISTAs recorded a single daily rating of each student's level engagement in the areas of task-related and physical behavior. VISTAs utilized the following scale to indicate that task-related behaviors as indicated below occured: never/rarely (0), sometimes (1), often (2), or always (3).

Task Related Behavior

1. Responsive to teacher and peers
2. Used eye contact (student's eyes were on the reading passage or VISTA, as directed)
3. Read audibly (when directed to read aloud; actively following along during whisper-reading)
4. Asked text related questions (limited chatty behavior)
5. Conversation was focused on the text/reading

Physical Behavior

1. Remained in intervention area
2. Assumed a posture that allowed for easy reading (sat up straight, faced toward the reading passage)
3. Hands engaged with reading materials or in lap (e.g., following along with finger)