8-2014

# The Effects of Prompts and Comprehension Assessment on Oral Reading: Moderating Effect of Reading Skills

Bethany Evelyn Forbes
*University of Tennessee - Knoxville*, bbowli10@utk.edu

To the Graduate Council:

I am submitting herewith a dissertation written by Bethany Evelyn Forbes entitled "The Effects of Prompts and Comprehension Assessment on Oral Reading: Moderating Effect of Reading Skills." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in School Psychology.

Christopher H. Skinner, Major Professor

We have read this dissertation and recommend its acceptance:

David Cihak, Sherry Bain, Dennis Ciancio

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

# The Effects of Prompts and Comprehension Assessment on Oral Reading: Moderating Effect of Reading Skills

A Dissertation Presented for the
Doctor of Philosophy Degree
The University of Tennessee, Knoxville

Bethany Evelyn Forbes
August 2014

# Dedication

This dissertation is dedicated to my husband, Alan Forbes, for his continuing support and encouragement throughout my graduate school career.

Acknowledgements

I would like to thank my committee members, Dr. Christopher Skinner, Dr. Sherry Bain, Dr. David Cihak, and Dr. Dennis Ciancio for their guidance throughout the process of creating this dissertation. The advice and encouragement my from my committee chair, Dr. Skinner, has left me with skills in writing, research design, and data interpretation that will be invaluable throughout my professional career. I am grateful to Dr. Ciancio for providing his expertise and assistance in analyzing data. I would also like to acknowledge the members of my research group, Emily Taylor, Megan Schall, and Samantha Cazzell who assisted in organizing and collecting data.

Abstract

Words correct per minute (WCPM) scores, derived from oral reading fluency (ORF) assessments, are used, in part, to make decisions regarding special education eligibility. WCPM scores are sensitive to environmental factors such as the presence of a stopwatch, administrator characteristics, and instructions. Using sixth-, seventh-, and eighth-grade middle school students, we replicate and extend previous research on the effects of environmental prompts on ORF scores by instructing students to read fast and investigating the reading skill-by-instructions interaction. We also evaluated how students who had been were instructed to read fast (phase two) responded to subsequent (phase three) standard instructions and standard instructions plus a requirement to answer comprehension questions.

Both Experiment I and II revealed that when students were instructed to read fast, as opposed to read their best, they increased their WCPM and errors. In Experiment I, a two-by-three mixed model ANOVA revealed a significant interaction between reading skill and instructions. When instructed to read fast, those with stronger reading skills had significantly larger increases in WCPM and smaller increases in errors. This interaction was not found in Experiment II. One explanation for these discrepant findings relates to differences in the difficulty level of passages used in the two studies. During Experiment I, harder passages were assigned to the read fast phase. Harder passages may have caused weaker readers more difficulty than stronger readers, which could account for the significant interaction.

During Experiment II, within-subject analyses were used to assess how students who were instructed to read fast during phase two responded to standard instructions and standard instructions plus comprehension questions during phase three. Both groups altered their reading based on the new instructions and their WCPM and error scores decreased, approaching their

phase one (standard instructions) levels. These findings, which showed that students responded to differences in instructions with significant increases and decreases in their WCPM, have applied implications for the administration of ORF assessments within Response to Intervention (RtI) programs. Limitations and directions for future research are discussed.

Table of Contents

List of Tables

List of Figures

# Chapter I

## Literature Review

For approximately 30 years students have been identified as having a specific learning disability (SLD) using discrepancy models (Cahan, Fono, & Nirel, 2012; Francis et al., 2005). With these models, teachers refer students who are struggling in specific academic areas for an evaluation. Once referred, standardized intelligence and achievement tests are administered to determine if a significant difference exists between IQ and academic achievement level. The size of the discrepancy needed between IQ and achievement to qualify for an SLD may be considered arbitrary (Dykeman, 2006). Regardless, if a large enough discrepancy between IQ and achievement in a particular academic area exists, the student can be diagnosed with an SLD and receive special education services.

For a variety of reasons, school psychologists have been dissatisfied with discrepancy models (O'Donnell & Miller, 2011). Students' scores on standardized assessments may not be stable over time, indicating the disorder may not always manifest itself (Dykeman, 2006). Additionally, methods for improving academic skills of students with SLD generally do not differ from methods for improving the skills of low achieving students who do not qualify as having an SLD (Fletcher, Denton, & Francis, 2005; Shapiro, 2011). Teachers' perceptions of their skills in teaching struggling students will also affect the likelihood that a child is referred for an evaluation, which influences which students receive the SLD diagnosis (Meyer, 2000). One of the biggest criticisms of discrepancy models is that students must be failing before any action is taken to conduct a full evaluation and provide services to disabled students. This has been coined as the "wait to fail" phenomenon (O'Donnell & Miller, 2011).

Dissatisfaction with discrepancy models eventually affected educational law. The reauthorization of the Individuals with Disabilities Education Improvement Act (IDEIA; 2004) affirms that states a) must not require the use of a discrepancy between intellectual ability and achievement, b) must permit the use of a process based on a student's response to a scientifically validated intervention, and c) may permit the use of another scientifically validated process for identifying students with SLD [20 U.S.C. 1221e–3; 1401(30); 1414(b)(6); 34 C.F.R. §300.307]. With the reauthorization of IDEIA (2004), many states and school districts have opted to use a process to determine students' responsiveness to scientifically validated interventions, also called Response to Intervention (RtI). One purpose of switching to RtI models is to avoid some of the concerns associated with discrepancy models. Such concerns include requiring students to be failing before intervening, over-identifying students with SLD, and ensuring students are given adequate instruction before being evaluated (McKenzie, 2009).

Implementation of RtI involves screening all students to determine who will receive intervention services and this usually occurs three times per year: fall, winter, and spring (Hughes & Dexter, 2011). Often the student population's performance on the screening measures is used to create local norms. Initial poor performance on screening measures indicates that a student is not responding to the general curriculum, or "tier one." Those who score at or below a predetermined percentage of students on a screening measure (e.g., bottom 15%) receive intervention services, putting them in what is referred to as "tier two." This may involve an intervention specialist working with homogenous, tier two students in small groups, allowing for targeted instructional scaffolding (Justice, 2006). The progress of those receiving intervention services is continuously monitored using brief, fluency-based measures at least once a month, but usually each week (Fuchs & Fuchs, 2006). If a student continues to perform poorly on the

fluency-based measures, he or she receives a more intensive intervention, frequently referred to as “tier three” (Wankez & Vaughn, 2010). RtI may be implemented using two to four tiers, with tier three or four associated with special education placement, as unresponsiveness in earlier tiers gives reason to suspect a disability (Fuchs & Fuchs, 2012; Fuchs, Mock, Morgan, & Young, 2003).

The reading screening measures used during RtI are often referred to as curriculum-based measures (CBM), which typically measure rate of accurate responding. Perhaps the most commonly used measure is often referred to as oral reading fluency (ORF). ORF assessments are used to obtain measures of the number of words read correctly in one minute, called words correct per minute (WCPM). Data obtained from ORF assessments are tracked over time and are used to determine if a student is responding to an intervention (i.e., if the intervention is effective for that student). Evaluating responsiveness can be done in many ways. For example, educators may consider a student’s slope of improvement relative to the average slope of improvement of the rest of the class, or if the student has met or exceeded cut points on screening measures (Fuchs, Fuchs, & Compton, 2004).

ORF assessments are useful because they can be obtained quickly, are relatively cheap, multiple forms are easily created, and are simple to administer (Deno, 2003). ORF passages can come from classroom texts, authentic reading material, or passages developed specifically for ORF assessments. Similar scores have been obtained, regardless of the source of the passage (Griffiths, VanDerHeyden, Skokut, & Lilles, 2009; Hintz, Conte, Shapiro, & Basile, 1997).

When using ORF assessments an examiner sits with a student while the student reads a passage out loud. The examiner records the student’s errors (e.g., omissions and mispronounced words) and tracks the number of seconds it takes to finish the passage or asks the student to stop

reading after 1 minute. The number of correctly read words is then totaled and presented as a WCPM score (Shapiro, 2011).

**Reliability and Validity of ORF Assessments**

Researchers have shown that ORF assessments have adequate predictive and construct validity, discriminant validity between special and general education students and grade levels, and high levels of inter-scorer agreement (Brown-Chidsey, Davis, & Maya, 2003; Fuchs & Fuchs, 2002; Reschly, Busch, Betts, Deno, & Long, 2009). Performance on ORF assessments correlates highly with reading measures on standardized achievement tests, such as the Woodcock-Johnson Reading Mastery Test-Revised (WRMT-R) Word Attack, Word Identification, Passage Comprehension, Basic Skills, and Total Reading scores (Hosp & Fuchs, 2005). Scores on ORF assessments are predictive of future performance on later ORF assessments, as well as the likelihood of meeting or exceeding expectations on the reading portion of end-of-year exams up to two years later (Good, Simmons, & Kame'enui, 2001; Keller-Margulis, Shapiro, & Hintze, 2008). Researchers have shown ORF assessments have good predictive validity across ethnic groups and are valid for students differing in socio-economic background, gender, and race (Hintz, Callahan, Matthews & Tobin, 2002; Knoff & Dean, 1994).

One generally accepted explanation for the relationship between reading rate and other general reading outcome measures relates to processing speed. When one can process text faster it enhances her/his ability to understand the relationship between individual words, thus facilitating comprehension (Peter, Matsushita, & Raskind, 2010). Despite a large amount of evidence suggesting ORF assessments adequately predict comprehension skills, there is some concern that they do not capture all aspects of comprehension (Valencia et al., 2010).

**Passage Equivalency and Variability in ORF Assessments**

One concern associated with ORF assessment is the degree of variance caused by nonequivalent passages. Researchers have found that the standard error of measure caused by passage variability can amount to more than half a grade level (e.g., more than 16 WCPM, see Christ & Ardoin, 2009; Poncy, Skinner, & Axtell, 2005). When students' ORF assessment scores move up by 16-20 WCPM from one week to the next, educators should not conclude that the students improved their reading by over half a grade level. Rather, this amount of change is likely caused by passage variability and other unaccounted for measurement error, which may be the result of non-standardized administration and/or poor scoring procedures (Christ, 2006; Christ & Silberglitt, 2007; Poncy et al., 2005).

**Influences of Various Prompts on ORF Assessment Scores**

ORF scores are highly sensitive to changes in students' reading skills. While this sensitivity is generally desirable, a concern is that these scores are also sensitive to other factors. Whereas passage difficulty and non-standardized administration and scoring procedures may introduce non-systematic error, other factors may systematically influence ORF scores. Systematic influences on ORF assessment scores may hinder our ability to use these scores for making eligibility decisions using across-student comparisons (e.g., initial placement into RtI remedial services) and when making within-student decisions including evaluating interventions or responsiveness (Christ, 2006; Christ & Silberglitt, 2007; Poncy et al., 2005).

**Effect of timing.** ORF assessment procedures involve timing students' oral reading, which is often done in an explicit manner, meaning students are aware the examiner is using a timing device that is visible. The question of whether knowledge that one is being timed increases speed of responding has been studied and results have shown that students will

generally complete more work under explicit timing conditions compared to covert timing conditions (Cates & Rhymer, 2006; Evans-Hampton, Skinner, Henington, Sims, & McDaniel, 2002; Rhymer, Henington, Skinner, & Looby, 1999). Working with three elementary students Cates and Rhymer (2006) compared the number of Dolch word phrases read under explicit and covert timing conditions. Students read Dolch word phrases on flashcards, were required to repeat the phrases they read incorrectly or did not know, and were provided with praise for cooperating and following instructions at the end of 3 minute sessions. Students read more phrases correctly when the teacher explicitly timed them (i.e., showed them a stopwatch and told them she was going to see how fast they could read) compared to the covert timing condition during which the teacher used a wristwatch (Cates & Rhymer, 2006).

Evans-Hampton et al. (2002) administered fluency-based math probes to eighth-grade students and investigated differences in digits correct, digits incorrect, and the percentage of digits correct among African-American and Caucasian students when explicitly and covertly timed. When using covert timing, the stopwatch was held below the table. When using explicit timing, the stopwatch was visible and students were told they were being timed. Similar to Cates and Rhymer's (2006) findings, students wrote more digits correct per minute and decreased their digits incorrect per minute during the explicit timing condition. This finding is somewhat contrary to a similar study by Rhymer et al. (1999) who found timing second-grade students completing simple addition and subtraction fluency-based probes increased rates of completed problems but did not increase accuracy rates. Together, these findings support the hypothesis that students can increase their response rates when timed, but the effect on accuracy rates is less clear.

Although there is evidence suggesting explicitly timing students will increase completed work, the difficulty of the task may moderate this effect. Rhymer et al. (2002) gave sixth-grade students fluency-based math probes that contained problems of varying difficulty (i.e., single-digit addition, three-digit subtraction, and complex multiplication problems). Timing was either covert (i.e., wrist watch) or explicit (i.e., stopwatch was visible and students were told they would be timed). Students completed significantly more problems when being explicitly timed on the addition and subtraction probes, but not on the complex multiplication probe. Additionally, the percentage of correctly answered problems was consistent from the covert to explicit timing conditions for all problem types. This finding suggests that overtly timing students may enhance their rates of accurate responding on tasks that are simple, but have little effect on their performance when tasks are complex.

While task complexity may be important, skill levels may account for the Rhymer et al. (2002) findings. Specifically, when working on tasks that have been mastered, timing may enhance performance. When working on tasks that have not been mastered, timing may not have the same effect. Using third-grade students and fluency-based math assessments targeting addition, subtraction, and multiplication problems, Rhymer, Skinner, Henington, D'Reaux, and Sims (1998) found that explicit timing conditions produced greater completion rates compared to covert timing conditions, but overall lower accuracy rates. This finding is contrary to other findings on explicit timing and accuracy, suggesting completion rates may increase under explicit timing conditions, but accuracy will be unaffected (Evans-Hampton et al, 2002; Rhymer et al., 2002). Further investigation of students' performance in the Rhymer et al. (1998) study showed that those students whose baseline performance was low or average relative to high performers decreased their accuracy during the explicit timing condition, while those who

performed high at baseline maintained their level of accuracy. This supports the notion that the more difficult or complex a task is for an individual, the more difficult it will be to increase or maintain accuracy levels under timed conditions.

**Effect of location, examiner, and timing.** Other assessment characteristics that might influence scores on ORF assessments include the familiarity of the location, the familiarity of the examiner, and the possibility that awareness of being timed interacts with these variables. To test these possible effects, Derr and Shapiro (1989) investigated differences in students' WCPM and percentage of errors (i.e., errors divided by words read) under differing circumstances. Using reading probes taken from classroom texts, third- and fourth-grade students read in either familiar (i.e., a reading group), less familiar (i.e., the teacher's desk), or least familiar (i.e., an office outside the classroom) settings, and also to a familiar (i.e., the teacher) or unfamiliar person (i.e., the school psychologist). Across all these conditions, the effect of timing the students' reading was also tested. Students were either assigned to the timed (i.e., told they had one minute to read aloud and a stopwatch was visible) or untimed condition (i.e., aloud reading was recorded and these audio recordings where used to collect data on time to read). Three passages were read by the students in each condition. WCPM and percentage of errors served as the dependent variables and the median and first probe scores were analyzed for differences across conditions.

When using the teacher as the examiner and comparing differences between reading in the reading group or at the teacher's desk, students read more WCPM when reading in their reading group. However, an interaction analysis suggested that for the first probe, this was only true for untimed students. Timed students also read more WCPM at the teacher's desk than untimed students, but not in the reading group. Students who were timed had a higher

percentage of errors when at the teacher's desk than when in their reading group, and untimed students did not differ on percentage of errors across settings. Timed students also had a higher percentage of errors than untimed students while at the teacher's desk, but not while in the reading group.

When the examiner was the school psychologist, on the first passage students read more words at the teacher's desk than in the office and when timed. When analyzing median scores (middle score of three passages), timed students read more words than untimed students in both settings. No differences were found in the percentage of errors made across settings or timing conditions for the first probe or median score analysis. When comparing the effects of the familiarity of the administrator on median scores, students read more words when reading to their teacher rather than the school psychologist and when timed. When analyzing the first probe scores, only the timed students read more words when reading to the teacher. Differences in the percentage of errors made were only found when analyzing the first probe and timed students had a higher percentage of errors.

These findings emphasize the importance of considering students' familiarity with various aspects of ORF assessment conditions. All analyses suggest that students will read more words when timed, something that is generally present during ORF assessments, and when in a more familiar setting (Derr & Shapiro, 1989). These findings are consistent with research on the effects of timing on fluency-based math probes (Cates & Rhymer, 2006; Rhymer et al., 1999; Rhymer et al., 1998).

It is clear from the Derr and Shapiro (1989) study that ORF assessment conditions can influences students' WCPM scores. Derr-Minneci and Shapiro (1992) attempted to determine if those effects are consistent across students at differing reading skill levels. Third- and fourth-

grade students served as participants and approximately one-third of them read below, one-third read at, and one-third read above grade level. Like the Derr and Shapiro (1989) study, students were assessed in familiar (i.e., reading group) and less familiar (i.e., teacher's desk and school psychologist's office) settings and by familiar (i.e., teacher) and less familiar (i.e., school psychologists) examiners. Comparisons were made between students who were timed and untimed and between students at differing reading skill levels. The median WCPM score served as the dependent variable.

When the teacher was the examiner, students at all reading levels read more WCPM when in their reading groups and when timed. When the school psychologist was the examiner, students read more WCPM at the teacher's desk than in the school psychologist's office. A significant interaction also revealed that this effect was greater for timed weaker and stronger readers. Thus, the setting appeared to make a greater difference for timed stronger and weaker readers compared to untimed stronger and weaker readers when reading to the school psychologist.

All students read more WCPM with the teacher as the examiner rather than the school psychologist and again, timed students read more WCPM than untimed students. Although a higher percentage of stronger and average readers increased their WCPM when reading to their teacher and when timed, differences between the number of timed stronger and average readers who increased their reading speed and timed weaker students who increased their reading speed were not statistically significant.

Generally, findings from the Derr-Minneci and Shapiro (1992) study are consistent with findings from the Derr and Shapiro (1989) study. The additional findings that in some instances students with stronger and weaker reading skills were differentially affected by varying

conditions and timing procedures (overt versus covert) suggests that specific administration procedures may have significant predictable effects on performance based on students' reading skill levels. Furthermore, the implication of familiarity (setting and administrator) suggests these effects may not be consistent within the same student over repeated measures. Thus, scores may change as students gain experience and become familiar with repeated assessments (e.g., every week for students getting remedial RtI services) within the same location (e.g., their classroom) and the same administrator (e.g., their teacher).

**Effect of incentives and feedback.** There are mixed findings related to providing rewards for improved performance on ORF assessments. Three fourth-grade students in a summer program were provided incentives (i.e., tokens) for improving their WCPM scores from one passage to the next when reading increasingly more difficult passages. If incentives were ineffective at improving WCPM scores, modeling and practice were added and if effective in combination, incentives were removed to test the individual effects of modeling and practice. Providing rewards did not make a significant difference for all students included in the study and only contributed to improvement for some students when combined with modeling and practice (Noell et al., 1998). Similarly, providing second- through fifth-grade students with a popsicle party for increasing their digits correct per minutes on fluency-based math probes did not improve performance (Christ & Schanding, 2007).

The type of feedback provided after completing ORF assessments may also affect performance. Eckert, Dunn, and Ardoin (2006) investigated the impact of providing students with feedback on their WCPM or their errors on subsequent ORF assessment performance. Six second-grade students who were reading at a frustrational level were given feedback after completing a baseline phase where they simply read three second-grade level passages.

Unexpectedly, providing feedback on the number of errors made resulted in greater increases in WCPM for most students. Providing feedback on the number of WCPM generally resulted in fewer errors in subsequent assessment sessions. Although students' performance was counter to what was expected in both feedback conditions, evidence of the impact of feedback on performance was found.

**Effect of understanding of instructions.** The manner in which participants interpret the standardized instructions for completing ORF assessments could vary among students. Students are usually instructed to do their "best" reading when completing ORF assessments. Some may interpret this to mean that mistakes should be avoided while others may interpret it to mean that they should read as much of the passage as they can in a fixed period of time (Colón & Kranzler, 2006). To investigate the different effects of telling students to do their "best" reading and telling students to read "as fast as you can without making mistakes," Colón and Kranzler (2006) exposed 50 fifth-grade students to both sets of instructions and observed the effect on students' WCPM and errors. All students were administered two ORF assessments in which they were told to simply read the passages out loud, then in counterbalanced order, half the participants were instructed to read two more passages as fast as possible and then read another two passages while doing their best reading. Students were also administered three reading subtests from the Woodcock-Johnson Tests of Achievement, Third Edition (WJ-III) to determine how well students' WCPM under the differing instructions predicts standardized reading scores.

Colón and Kranzler (2006) found that students read significantly more WCPM and made significantly more errors when they received instructions to read fast. Scores from all three conditions were predictive of scores on the WJ-III reading subtests, and differences in their predictability were not significant. The finding that students made more errors when instructed

to read fast supports the speed-accuracy trade off hypothesis, which suggests that speed and accuracy cannot be maximized at the same time (Colón & Kranzler, 2006). If students interpret the goal of ORF assessments differently, those at similar reading skill levels could obtain significantly different WCPM scores on ORF assessments that do not reflect real reading skill differences. Findings from this study point to the importance of ensuring that students understand what is meant by "best" reading.

**Summary of the effects of environmental prompts.** Researchers demonstrated how merely showing students a stopwatch (Derr & Shapiro, 1989) or telling students to read fast (Colón & Kranzler, 2006) caused them to increase their WCPM during ORF assessments. However, telling them that they would be reinforced for faster reading or providing them feedback on their WCPM did not consistently increase WCPM (Eckert et al., 2006; Noell et al., 1998). When inconsistent effects occur it is likely that we do not have a strong understanding of the processes and variables affecting our outcomes, suggesting the need for additional basic research (Skinner, 2002).

**Stress: A Rational for the Effect of Prompts on ORF Assessment Scores**

The reasons why environmental prompts appear to impact student performance on ORF assessments is unclear. It is possible that some prompts (e.g., stopwatch, unfamiliar examiner, offer of rewards), elicit higher stress in students which may influence students' performance. The Yerkes-Dotson Inverted U hypothesis suggests that a moderate level of arousal will occasion optimal performance (Willingham, 1998). Several studies across a variety of tasks including solving puzzles (Klein & Beith, 1985), creative tasks (Byron, Khazanchi, & Nazarian, 2010), academic performance (Gaeddert & Dolphin, 1991), and a range of sport-related behavior (Suinn, 2005) have supported the Yerkes-Dotson Inverted U hypothesis. A criticism of the

Inverted U hypothesis involves the problem of “arousal” being inadequately defined. Researchers have referred to this construct as “anger,” “arousal,” “sexuality,” “fear,” etc. (Neiss, 1988). One explanation for why various prompts have their effects on ORF assessment performance may relate to the reaction they elicit, which in the case of ORF assessments would probably be best labeled as “arousal” or “stress.”

The second component of the Yerkes-Dodson Inverted U hypothesis includes consideration of the difficulty of the task. If a task is “simple,” the relationship between arousal/stress and performance is linear; as arousal/stress increases, performance will continue to increase, even at very high levels of arousal/stress. A curvilinear relationship between arousal/stress and performance exists for “difficult” tasks where too much stress can cause performance decrements (Diamond, Campbell, Park, Halonen, & Zoladz, 2007). This difference in simple and difficult tasks is thought to exist because with strong emotional reactions, it becomes more difficult to attend to and use cues in the environment. For simple tasks, decreased attention to external cues may enhance performance, but when performing difficult tasks, it is necessary to use a number of cues from the environment, which becomes harder to do with increased emotionality (Easterbrook, 1959).

Researchers have found some indirect support for the differential impact of stress on simple and difficult tasks. Rhymer et al. (1998) and Rhymer et al. (2002) found that accuracy on easy math tasks was not impacted by putting students under time pressure, but accuracy on harder math tasks decreased under time pressure. Time pressure could act to increase arousal/stress, which decreased accuracy on a math task for students categorized as having average or weaker skills on the targeted task (Rhymer et al., 1998). One’s self-perception of their skills in performing a particular task could also account for increased arousal/stress, which

could in turn produce poorer performance.  Consequently, those who perceive themselves as skilled may not become overly aroused/stress from various prompts, allowing them to focus on the necessary cues to perform well.  However, individuals  with lower academic self-concept, greater test anxiety, lower test scores, or high levels of outside pressure (e.g., parental pressure) may respond to prompts with higher arousal/stress levels that prevent them from focusing on the necessary cues to perform well (Gadbois & Sturgeon, 2011; O'Rourke, Smith, Smoll, & Cumming, 2011).

**Implications of Stress as a Moderator Variable**

These studies may have implications for ORF assessments.  Perhaps when conducting ORF assessments, various factors may be anxiety producing, including overtly timing students' performance (Derr & Shapiro, 1989), unfamiliar assessor and/or location (Derr-Minneci & Shapiro, 1992), being instructed to read rapidly (Colón & Kranzler, 2006), and being offered reinforcement for reading faster (Christ & Schanding, 2007; Noell et al., 1998).  The possibility that these stress factors may have a different effect on students dependent upon their confidence in their reading skills and/or their actual reading skills may help explain inconsistent effects across studies and students.  For example, in the Noell et al. (1998) study, the offer of the reward may have caused stronger or more confident readers to improve their ORF, but not the weaker readers.

In implementing RtI, we make decisions based on across-student comparisons (e.g., eligibility to RtI).  If stress has a systematic impact on students' scores that is dependent upon their levels of skill, these decisions may be flawed.  For example, at the beginning of the year, mass benchmark assessments may be conducted by a group of trained assessors who come into the school and administer standardized ORF assessments at a central locale where students are

bought to the assessor. Using trained assessors may allow for more standardized administration and scoring procedures, which should reduce error (Christ, 2006). However, such procedures may also enhance stress, which could improve the performance of stronger readers relative to weaker readers. This could result in inappropriate placement of students and an increase in the number of students identified as eligible for services.

Because local norms are used to make eligibility decisions, any procedure that systematically improves some students' scores relative to others is a concern. However, if stress or arousal associated with ORF assessment procedures are partially responsible for these different systematic effects across students, then decisions based on repeated measures of the same student (e.g., intervention or responsiveness evaluations) may also be impacted. For example, when benchmarks are set it is not uncommon for every student in a school to be assessed three times: fall, winter, and spring (Hughes & Dexter, 2011). During these mass assessments it is more likely that a stranger (e.g., school psychologist or other trained professional) administers assessments in an odd environment (e.g., the school cafeteria). Those who score low may receive remedial services where they are assessed more frequently, in their classroom, by their teacher. Students may find both the location (classroom) and assessor (teacher) less stressful, which may improve the performance of students with weaker skills. Even if these factors do not initially reduce stress, as students are assessed more frequently (e.g., weekly) they will become more familiar with the procedures and it is likely the stress associated with these assessments will decrease. If these decreases in stress cause improved performance, then we may conclude that some remedial procedures are effective when they had little impact on students' skill level.

**ORF Assessment and Reading Comprehension**

Although there is evidence supporting the use of ORF assessments as an indicator of global reading skill and comprehension skills, some are concerned that instructional practices may emphasize reading speed, and educators may begin to view fluency as an end rather than a means (Reschly et al., 2009; Roberts, Good, & Corcoran, 2005). For those teachers focused on enhancing ORF scores, simple prompting procedures (e.g., tell students to read their fastest, show students the stopwatch) may allow educators to increase ORF scores without improving the correlates these scores are designed to measure (i.e., global reading skills, comprehension). Researchers have applied several strategies to address these concerns.

One strategy has been to develop and apply other measures that more directly assess reading comprehension, such as Maze and Cloze assessments procedures. With Cloze, every seventh word is missing and students are to generate correct words as they read (Fuchs & Fuchs, 1992). With Maze assessments, students are given three word options, only one of which makes sense in the sentence. Students are instructed to circle the word that makes sense. Maze and Cloze procedures have demonstrated adequate validity, reliability, and sensitivity (Brown-Chidsey et al., 2003; Espin & Foegen, 1996; Hale et al., 2011; Shin, Deno, & Espin, 2000; Wayman, Wallace, Wiley, Tich´a, & Espin, 2007), and may have more face validity than WCPM because they provide an indirect measure of comprehension (e.g., students must comprehend passages to know what words belong in blank spaces).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS) assessment procedures include informing students they will be asked to tell the examiner about what they read (Bellinger & DiPerna, 2011). The number of words a student says during this portion of the administration is called retell fluency (RTF) and is used as an indicator of comprehension skills.

Roberts et al. (2005) found that RTF has an estimated reliability of .57 and obtained a correlation of .51 with the WJ-III Broad Reading Cluster, explaining 26% of the variance in those scores. However, when added to WCPM scores, RTF added very little to the variance explaining WJ-III Broad Reading Cluster scores. Additionally, Ridel (2007) found stronger correlations between WCPM and performance on the GRADE standardized test of reading ability than any other DIBELS subtest. Other researchers have found weak correlations (i.e, .16-.32) between performance on ORF assessments and RTF (Pressley, Hilden, & Shankland, 2005).

Difficulty in scoring is a major limitation of RTF (Bellinger & DiPerna, 2011), which was addressed by Skinner (1998) who described how researchers collected a measure of reading comprehension rate (RCR) by having students read same length passages (i.e., same number of words) and then answer multiple choice comprehension questions on those passages. The percent of the passage understood per minute of reading could then be calculated by multiplying the percentage of comprehension questions answered correctly by 60 s and dividing by the seconds needed to read the passage. Researchers have found evidence that RCR is a valid measure of global reading skills (Hale et al., 2011; Neddenriep, Hale, Skinner, Hawkins, & Winn, 2007). Neddenriep, Skinner, Wallace, and McCallum (2009) repeatedly measured WCPM, percentage of comprehension questions answered correctly, and RCR when evaluating a peer tutoring intervention. Exploratory analysis showed that RCR correlated more strongly with WCPM than percentage of comprehension questions correct. Others demonstrated that the RCR measure was sensitive enough to detect changes in reading skills occasioned by repeated reading and listening-while-reading (Freeland, Skinner, Jackson, McDaniel, & Smith, 2000; Hale et al., 2005; McDaniel et al., 2001), reinforcement (Freeland, Jackson, & Skinner, 1999), and pre-reading comprehension intervention (McCallum et al., 2011; Ridge & Skinner, 2011; Williams

& Skinner, 2004).

When conducting RCR assessments, students must read aloud to ensure that they have actually done the reading (Freeland et al., 2000; Hale et al., 2005; Neddenriep et al., 2007). Additionally, students must be prompted when they come to an unknown or difficult word, otherwise the time spent reading may be artificially inflated (Skinner, Neddenriep, Bradley-Klug, & Ziemann, 2002). Consequently, RCR assessments are similar to ORF assessments, except that students are informed that they will have to answer comprehension question when they have finished.

Although RCR measures may be a more direct assessment of reading comprehension than ORF, researchers have found that the measure of reading speed embedded within the RCR measure may account for most of the global reading score variance accounted for by the RCR measure (Hale, Skinner, Wilhoit, Ciancio, & Morrow, 2012; Skinner et al., 2009; Williams et al., 2011). In other words, when taking an RCR measure, if assessors merely record the time required to read the passage aloud, they can obtain a strong predictor of global reading skill by disregarding the direct measure of comprehension (i.e., percentage of comprehension questions correct). Despite this finding, researchers have suggested that including a measure of comprehension following ORF assessments may have several advantages. When reading for comprehension, students may read in a different manner than when reading for speed, which may produce a more valid measure of global reading skills (Bellinger & DiPerna, 2011; Skinner, 1998; Skinner et al., 2002; Skinner et al., 2009). Consequently, WCPM measures taken when students have to answer comprehension questions following the ORF assessment may correlate better with global reading skills (Hale et al., 2012).

Another advantage of requiring students to answer comprehension questions following

ORF assessments relates to findings from earlier studies on factors that influence WCPM without improving ORF. Skinner (2011) described an incident that occurred in his classroom where ORF data were collected weekly on all students who were two or more grade levels behind in reading. In his classroom, Skinner had students self-graph their weekly WCPM and made these graphs available to others (i.e, they were posted in a folder) so that well informed visitors to the classroom could view these graphs and praise students for improvement. One day, one student took a deep breath as Skinner was delivering ORF instructions and when he said begin, the student began reading aloud as rapidly as possible, skipping unknown words and not pausing for punctuation. Although his errors went up, his WCPM increased from about 35 to about 55. Clearly, this score should be considered invalid as his reading had not improved that much in one week. However, his WCPM scores had, and it was impossible to make this student return to his typical reading following this incident. Furthermore, following some public praise from classroom visitors who looked at this student's performance graphs, other students began engaging in similar behaviors during ORF assessments.

One instance like this would not be a serious concern, except that previous studies reviewed suggest that merely displaying a stopwatch during ORF assessment or instructing students to read fast, as opposed to reading their best, may occasion similar changes in reading behavior (Colón & Kranzler, 2006; Derr & Shapiro, 1989; Derr-Minneci & Shapiro, 1992). Furthermore, because some educators may see reading speed as an end rather than a means, some may be concerned that educators may intentionally encourage rapid word calling without regard for comprehension or prosody (Reschly et al., 2009; Roberts et al., 2005). Even if teachers are not directly or intentionally informing students that they should read their fastest during ORF assessment, it is hard to imagine that repeated and large-scale ORF assessments (i.e.,

each student in the elementary school assessed three times per year) do not cause many students to assume that these assessments are designed to measure reading speed. Therefore, regardless of directions and instructions, if students interpret that ORF assessments are designed to measure reading speed, then they are likely to read differently (Colón & Kranzler, 2006) than if they were reading for comprehension. Thus, a second reason for including comprehension questions following ORF assessments, even when comprehension performance is not used to measure reading skills, is that including comprehension questions may prevent students from engaging in speed reading-like behaviors (e.g., rapid word barking or word calling, see Samuels, 2007).

## Research Questions

Reading skill levels vary across students; consequently, when all students in a grade are given grade-level passages (e.g., all second-grade students read second-grade level passages), some students will be reading material that they have mastered and some will be reading material at their frustrational level. . If reading skills and/or passage difficulty levels interact with stress or arousal, then procedures that enhance stress or arousal levels are likely to increase the discrepancy of scores across stronger and weaker readers. Such procedures that reduce the scores of weaker readers relative to stronger readers may increase false positives, which could result in some students who do not need remedial services receiving such service. Furthermore, if stress or arousal lessens as weak readers complete more frequent ORF assessment in their classroom, scores may increase, even when interventions are ineffective. Both inappropriate eligibility decisions and inappropriate treatment evaluations can waste valuable resources. Experiment I was designed to determine if changing ORF instructions to emphasize the importance of reading as fast as possible would increase WCPM and errors. Additionally, we

conducted moderation analyses to determine if this change in instructions had different effects on students dependent upon their reading skills.

Specific research questions addressed in Experiment I include:

1. During ORF assessments, will instructing students to read fast and accurately increase students' WCPM and error scores?

2. During ORF assessments, will students' reading skills moderate the effects of instructing them to read fast?

The question of whether or not prompts to read faster would differentially affect stronger, weaker, and average students was also addressed in Experiment II by first identifying students' reading skill level using Maze assessments. Therefore, Experiment II addressed the following question:

3. Will reading skills moderate the effects of instructing students to read fast during ORF assessments?

Once students have been prompted to read aloud at a faster rate it may be difficult to cause them to return to their typical aloud reading speed (Skinner, 2011). Therefore, in addition to replicating the findings from Experiment I, we conducted Experiment II to answer another research question:

4. Following ORF assessments where students are instructed to read fast, does returning to standard instructions and standard instructions plus comprehension questions influence ORF scores?

# Chapter II

## Prompting Faster Reading in Middle School Students during Fluency Assessments: Changes Moderated by Reading Skills

Although originally developed for use with special education students (Deno & Mirkin, 1977), measures of oral reading fluency (ORF) are often used in conjunction with Response to Intervention (RtI) models to identify general education students with reading skill deficits and to evaluate the effects of remedial procedures (Shapiro, 2011). Researchers investigating the psychometric properties of ORF have demonstrated that measures of words read correctly per minute (1) exhibit adequate reliability and validity when compared with other standardized reading measures, (2) can discriminate between differing populations, and (3) provide an adequate estimate of global reading ability (Deno & Fuchs, 1991; Reschley, Busch, Betts, Deno, & Long, 2009). Although these findings support the use of ORF assessments, researchers have identified and sought to control various sources of error associated with ORF measures (Ardion, Roof, Klubnick, & Carfolite, 2008; Christ, Zopluoglu, Long, Monaghen, 2012; Derr-Minneci & Shapiro, 1992; Poncy et al., 2005).

Perhaps the largest source of error in ORF assessments is derived from discrepancies in the reading probes (Poncy et al., 2005). Although researchers have used various procedures in attempt to create equal passages (e.g., readability formulas, Euclidean distance), creating equivalent probes is still a challenge (Ardoin & Christ, 2009; Christ et al., 2012). Thus, the error associated with ORF slope and single point measures is often so large (e.g., standard errors that amount to almost one year of growth) that it hinders or prevents educators from using ORF data to make placement and intervention evaluation decisions (Ardoin & Christ, 2009; Christ et al., 2012; Poncy et al., 2005).

Other factors such as timing procedures, administration instructions, the test administrator, and the setting may systematically influence performance on ORF assessments. Colón and Kranzler (2006) demonstrated how changing ORF instructions by directing fifth-grade students to read fast, as opposed to reading their best, caused significant increases in words correct per minute (WCPM) as well as the number of errors made. Derr and Shapiro (1989) investigated a less direct prompt. During ORF assessments the administrator either conspicuously timed (i.e. displayed the stopwatch) or covertly timed (i.e. no stopwatch present) third- and fourth-grade students' aloud reading. Results showed higher WCPM and percent errors under the conspicuous timing condition, demonstrating that subtle prompts can influence students' WCPM and percent of errors during ORF assessments.

Derr-Minneci and Shapiro (1992) investigated the effect of location, administrator, and explicit timing on third- and fourth-grade students' WCPM. Students read more WCPM in their reading group, to their teacher, and when they were conspicuously timed compared to reading in an unfamiliar setting (e.g., office), to an unfamiliar administrator (e.g., school psychologist), or when they were covertly timed. Discrepancies in setting, administrator, and task demands (timed versus untimed) were substantial with an average difference of (a) 18 WCPM between reading to the teacher in a group versus at the teacher's desk, (b) 12 WCPM when reading to a school psychologist in an office versus at the teacher's desk, and (c) 11 WCPM when reading to the teacher at the teacher's desk versus reading to the school psychologist at the teacher's desk. When one considers that the average yearly growth rate in WCPM is approximately 25 words for third grade and 22 words for fourth grade (Christ, Silberglitt, Yeo, & Cormier, 2010) differences of 11-18 WCPM represents half of a school year or more worth of learning.

While prompts to read faster (e.g., conspicuous timing, instructing students to read as fast as they can) have been shown to enhance WCPM, the effect is inconsistent. Noell et al. (1998) found that three fourth-grade students did not increase their number of WCPM when offered rewards contingent on improving their rate of reading. When Derr-Minneci and Shapiro (1992) examined the impact of timing procedures, they found some evidence that conspicuous timing procedures were more likely to enhance WCPM in skilled readers relative to weaker readers; however, because these findings were not statistically significant, they recommended that future researchers investigate the possibility that reading skill may moderate effects of ORF prompts and conditions.

Researchers investigating performance on computation worksheets found evidence that a skill level-by-task interaction may moderate the impact of prompting students to respond more rapidly. Specifically, researchers found that when working on simple problems, prompting students to work more quickly enhanced their speed of accurate responding; but, when working on more difficult problems these prompts had little effect on rates of accurate responding and enhanced errors (Rhymer et al., 1998; Rhymer et al., 2002). Although researchers offered no causal explanation for their finding, if one assumes that prompting students to work fast enhances arousal, stress, or anxiety then these moderator effects are consistent with previous research. Specifically, human performance researchers have shown that when skills are mastered, increased arousal can enhance performance, but when working on poorly developed skills, increased arousal may hinder performance (Diamond et al., 2007).

**Purpose of Current Study**

Various prompts, whether subtle (e.g. conspicuous timing) or obvious (e.g. instructed to read fast), delivered during ORF assessments can systematically influence WCPM and error

scores (Derr & Shapiro, 1989; Colón & Kranzler, 2006). Furthermore, there is evidence that aptitudes may moderate the effects of these prompts (Derr-Minneci & Shapiro, 1992; Rhymer et al., 1998; Rhymer et al., 2002). With the current study, we replicated and extended previous research on prompts and ORF assessment performance. Specifically, we replicated Colón and Kranzler's (2006) investigation of the effects of altering instructions on WCPM and errors. Consistent with recommendations provided by Derr-Minneci and Shapiro, we extended this research by conducting additional analyses to investigate whether reading skill moderated these effects.

## Method

### Participants and Setting

Participants included 73 seventh- and eighth-grade students (44 females and 29 males) at a rural middle school in the Southeastern United States. Over 75% of students enrolled in this school were considered economically disadvantaged. Our sample was approximately 90% Caucasian, 1% African American, 1% Hispanic, 3% Native American, 4% biracial, and 1% indicated "other." All participants completed the study on one of four days in the middle of the spring semester. Procedures were conducted in a quiet office or hallway.

### Materials and Measures

Modified passages from the seventh-grade level *Timed Reading* series (Spargo, 1989) were used to measure number of WCPM and number of errors. First, the primary experimenter reduced the passages to 120 words. Next, all passages were modified until their Flesh-Kincaid readability scores fell between the seventh- and eighth-grade reading levels. The six passages were randomly sequenced and all participants read each passage in the same sequence. Passages

one, two, and three were read in order during the pretest phase, and passages four, five, and six were read in order during the posttest phase.

Three doctoral students in school psychology administered the procedures. All had prior ORF training and had administered ORF assessments to at least 75 elementary students. Before collecting any data, each experimenter received additional training on the specific procedures applied during the study. Each practiced giving the two sets of instructions. Experimenters practiced starting and stopping their stopwatch in plain view. Additionally, experimenters were instructed to audio record all sessions by starting the recorder and placing it in plain view, on the desk, in front of the participants before delivering instructions.

During ORF assessments, participants read each passage aloud as an experimenter followed along on a copy of the passage, scored errors, supplied words when participants paused for more than 3 s, and re-directed participants when they lost their place. Errors included skipped words, mispronunciations, transposed words, and words provided by an experimenter after a 3-s pause. If a participant self-corrected an error within 3 s, experimenters counted the word as correct. When the participants finished reading the passage the experimenter recorded the number of seconds required to complete the passage. After all assessments were completed, researchers calculated the number of errors made and WCPM for each passage. As each passage contained 120 words, the number of errors did not have to be converted to a percentage. Experimenters calculated WCPM scores by multiplying total words read correctly by 60 s and dividing by the seconds spent reading. The dependent variables analyzed were each student's median errors and WCPM for each phase.

**Design and Procedures**

A pretest-posttest experimental design was used to evaluate the effects of different ORF instructions on participant errors and WCPM. After obtaining institutional support for the study, two middle school teachers, one math teacher and one social studies teacher, sent parental consent forms home with all of their seventh- and eighth-grade students. After receiving consent for 73 students to participate, participants were randomly assigned to groups so that for each student assigned to the control group, two were assigned to the experimental group.

Experimenters scheduled procedures to be conducted during regularly scheduled math and social studies classes. Each participant completed all procedures within the same 7-12 min session. During experimental sessions experimenters escorted participants from their classroom to either an empty office or a quiet hallway with two chairs and a table. Working one-to-one with each participant, an experimenter solicited and obtained child assent to participate from all students whose parents provided consent and attended class the day procedures were run. After providing assent, participants completed a demographic form and six ORF assessments. For all participants, the same six passages were administered in the same sequence.

For both groups, during the pretest phase, the first three passages were administered using standard ORF procedures. During the posttest phase, these standard procedures were repeated with 23 participants randomly assigned to the control group using passages four through six. When reading under standard instructions, experimenters instructed participants by saying

> When I say ‘begin,’ start reading this passage aloud. Read across the page. Try to read each word. If you come to a word you don’t know I’ll tell it to you. Be sure to do your best reading. Are there any questions? Begin. (Shinn & Shinn, 2002a, p.18)

During the posttest phase, the 50 participants assigned to the experimental group received different ORF instructions for passages four through six. Rather than receiving instructions to read their best, reading speed was emphasized as experimenters read the following instructions:

> When I say ‘begin,’ start reading this passage aloud. This time, read as accurately and as fast as you can. Read across the page. Try to read each word. However if you come to a word you do not know, say something before I provide the word and continue reading, both fast and accurately. Before you begin reading take a deep breath. Read in a quiet voice, but loud enough for me to hear you. Read as fast and accurately as you can. Don’t pause for punctuation, or read with expression, instead read as fast as possible. Do you have any questions? Take a deep breath. Begin.

**Analysis**

For each phase (pretest and posttest), each participant's median WCPM and error scores were analyzed. Thus, if a student read 100, 70, and 102 WCPM during the pretest phase, the analyzed pretest score was 100 WCPM. For our initial analysis, for each dependent variable, a two-by-two mixed model ANOVA was used to test for significant differences. The within-subjects factor was phase (pretest or posttest) and the between-subjects factor was group. The control group always received standard instructions and the experimental group received rapid instructions in the posttest phase.

In order to investigate the hypothesis that reading skill moderates the impact of prompting rapid reading, we employed an analytic approach conceptually similar to analysis of attribute-by-treatment interactions (Cronbach & Snow, 1977). Because WCPM is a valid and reliable indicator of global reading skills (see Reschly et al., 2009 for a meta-analysis), we used median pretest WCPM scores as an indicator of aptitude (i.e., reading skills). Using only the

data from the 50 experimental group participants, we correlated median pretest scores (a measure of aptitude) with change scores, a measure of a treatment effect calculated by subtracting median pretest scores from median posttest scores for both dependent variables. Next, we formed three subgroups by ranking our experimental group participants from highest to lowest based on median pretest WCPM scores. The first 17 were assigned to the stronger reading skills subgroup, the next 17 to the average subgroup, and the final 16 to the weaker subgroup. The stronger subgroup's median pretest WCPM scores ranged from 153 to 200 WCPM, the average subgroup's median scores ranged from 125 to 151 WCPM, and the weaker subgroup's WCPM scores ranged from 98 to 124. Finally, we applied mixed model ANOVAs to test for differences across experimental subgroups. The within-subjects factor was phase (pretest or posttest) and the between-subjects factor was subgroup (weaker, average, and stronger reading skills).

**Inter-scorer Agreement**

An experimenter listened to audio recordings of 11 experimental group and 10 control group participants (28%) and independently scored errors and reading time. Next, the experimenter calculated WCPM and errors for each passage. Pearson product-moment correlations between the original dependent variables and those collected by the experimenter listening to the recordings for WCPM were .96 for the pretest phase and .99 for the posttest phase. The correlation between the two raters for errors was .69 for the pretest phase and .80 for the posttest phase. The correlations for WCPM are acceptable. The weaker correlations for errors suggest these data should be interpreted with caution (House, House, & Campbell, 1981).

## Results

### Control versus Experimental Group

Group means and standard deviations, calculated using each member's median WCPM by phase, are presented in Table 1. A two-by-two mixed model ANOVA revealed a non-significant main effect for group (control or experimental) and a significant main effect for phase, $F(1,71) = 5.10$, $p = .027$, with significantly greater WCPM scores during the pretest phase. The group-by-phase interaction depicted in Figure 1 was significant, $F(1, 71) = 18.86$, $p < .000$. Those in the experimental group showed an increase in WCPM when instructions were altered (i.e., posttest), while those in the control group, who continued to receive standard instructions, showed a decrease in WCPM from pretest to posttest. Pairwise comparisons revealed the increase in WCPM across phases was not significant for the experimental group ($p = .068$) and the decrease in WCPM from pretest to posttest for the control group was significant ($p < .000$). This significant decrease suggests the posttest passages were more difficult than the pretest passages. Pairwise comparisons of pretest WCPM revealed a non-significant difference between the experimental and control group ($p = .421$).

The mean number of errors and standard deviations, calculated by using each participant's median number of errors made by phase, is presented in Table 2. A two-by-two mixed model ANOVA revealed a non-significant main effect for group and a significant main effect for phase, $F(1, 71) = 35.33$, $p < .000$, with participants making more errors during posttest. The increases in errors from pretest to posttest were significant for both the experimental ($p < .000$) and control groups ($p = .048$). Again, this decrease in control group performance suggests that the posttest passages were more difficult than the pretest passages. The condition-by-phase interaction depicted in Figure 2 was significant, $F(1, 71) = 6.80$, $p = .011$. The experimental

Table 1

*Words Correct Per Minute (WCPM) Means, Standard Deviations, Change Scores, p-Values, and Effect Sizes Across Groups and Phases*

| | | Phases | | | | |
|---|---|---|---|---|---|---|
| Groups | *n* | Pretest *M*(*SD*) | Posttest *M*(*SD*) | Change Pretest-Posttest | *p*-value | Effect Size[1] |
| Experimental | 50 | 141.16(27.25) | 145.26(40.16) | +4.10 | .068 | .12 |
| Control | 23 | 146.73(27.38) | 133.74(26.31) | -13.02 | .000 | .48 |

[1]All effect sizes were calculated using Cohen's *d* formula: mean one minus mean two, divided by the pooled standard deviation (Cohen, 1988).

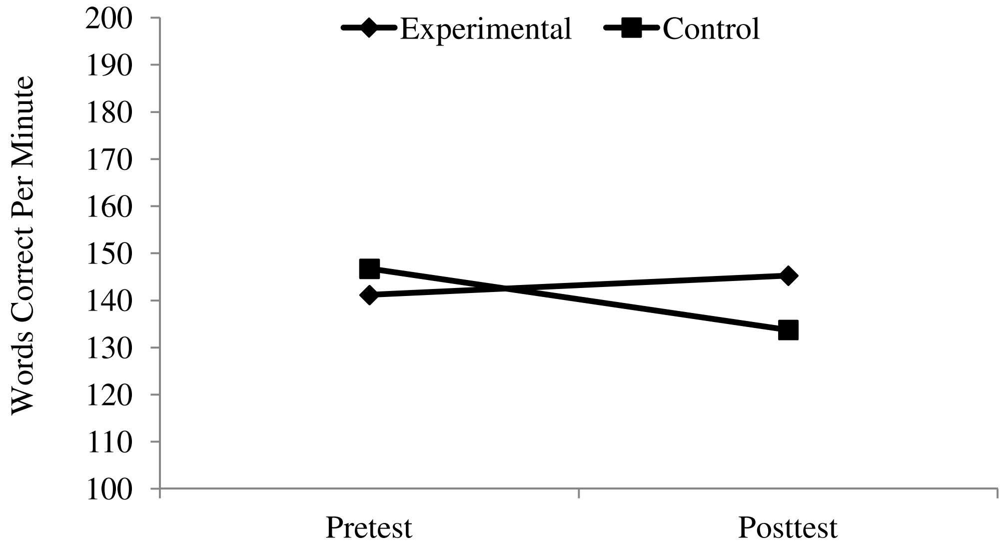


*Figure 1.* Mean words correct per minute (WCPM) by the experimental and control group in pretest and posttest.

Table 2

*Error Means, Standard Deviations, Change Scores, p-Values, and Effect Sizes Across Groups and Phases*

| Groups | *n* | Phases<br>Pretest<br>*M*(*SD*) | <br>Posttest<br>*M*(*SD*) | <br>Change Pretest-<br>Posttest | *p*-value | Effect Size |
|---|---|---|---|---|---|---|
| Experimental | 50 | 1.86(1.58) | 4.20(3.20) | +2.34 | .000 | .98 |
| Control | 23 | 1.61(1.08) | 2.52(2.11) | +.91 | .048 | .57 |

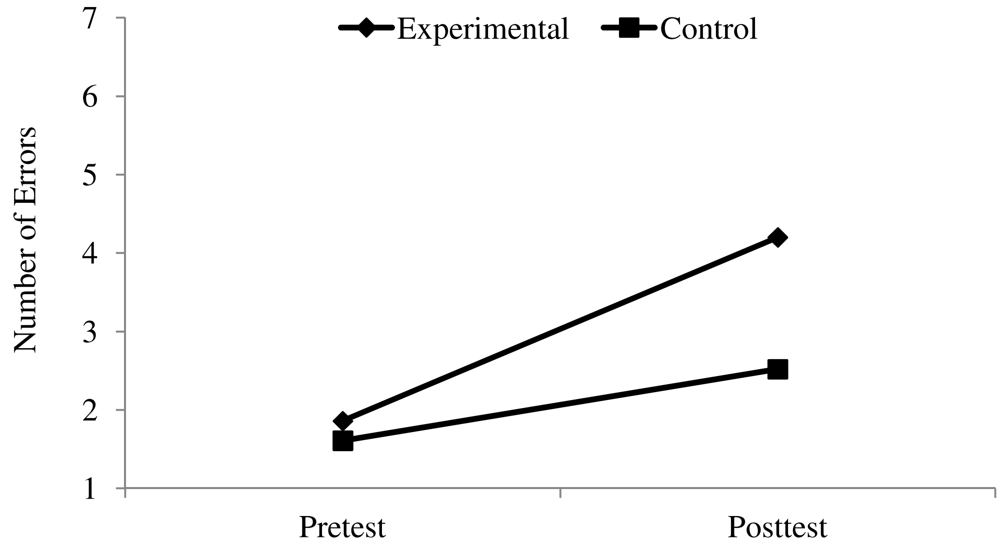


*Figure 2.* Mean errors by experimental and control groups in pretest and posttest.

group increased their mean number of errors from pretest to posttest by approximately 2.34 errors, while the control group increased their mean number of errors by approximately .91 errors. The significantly greater increase in errors by the experimental group may have been caused by the change in instructions prompting participants to read as fast as they can. Pairwise comparisons showed the mean number of errors made by the experimental and control groups during pretest did not differ significantly ($p$ = .491).

**Differences Among Stronger, Average, and Weaker Readers**

To test the hypothesis that students at varying reading levels will be differentially affected by the read fast instructions, several analyses with experimental group participants were run. For the experimental group, our correlation between number of WCPM in pretest and the difference in WCPM from pretest to posttest (i.e., median posttest WCPM - median pretest WCPM) was significant, $r(50) = .61$, $p < .000$. Thus, those who read more WCPM during standard instructions (pretest) exhibited a greater increase in WCPM when prompted to read as fast as they could during posttest.

Table 3 provides summary WCPM statistics for the three subgroups (stronger, average, and weaker readers) that we formed based on median pretest WCPM across phases. A two-way mixed model ANOVA revealed a significant main effect for subgroup (stronger, average, weaker), $F(2, 47) = 90.11$, $p < .000$, but not phase. Pairwise comparisons revealed that all three subgroups differed from one another at the $p < .000$ level on both pre- and posttest assessments, which suggests that our attempt to separate participants into subgroups based on WCPM scores was successful.

Table 3

*WCPM Means, Standard Deviations, Change Scores, p-Values, and Effect Sizes Across Stronger, Average, and Weaker Subgroups and Study Phases*

| Phases | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Pretest | | Posttest | | | | | |
| Subgroups | *n* | *M*(*SD*) | Range | *M*(*SD*) | Range | Change | *p*-value | Effect Size |
| Stronger | 17 | 172.97(12.75) | 153-200 | 187.38(30.35) | 146-264 | +14.41 | .000 | .67 |
| Average | 17 | 137.50(8.43) | 125-151 | 140.45(19.23) | 108-183 | +2.95 | .439 | .21 |
| Weaker | 16 | 111.25(7.81) | 98-124 | 105.64(13.48) | 87-131 | -5.61 | .155 | .53 |

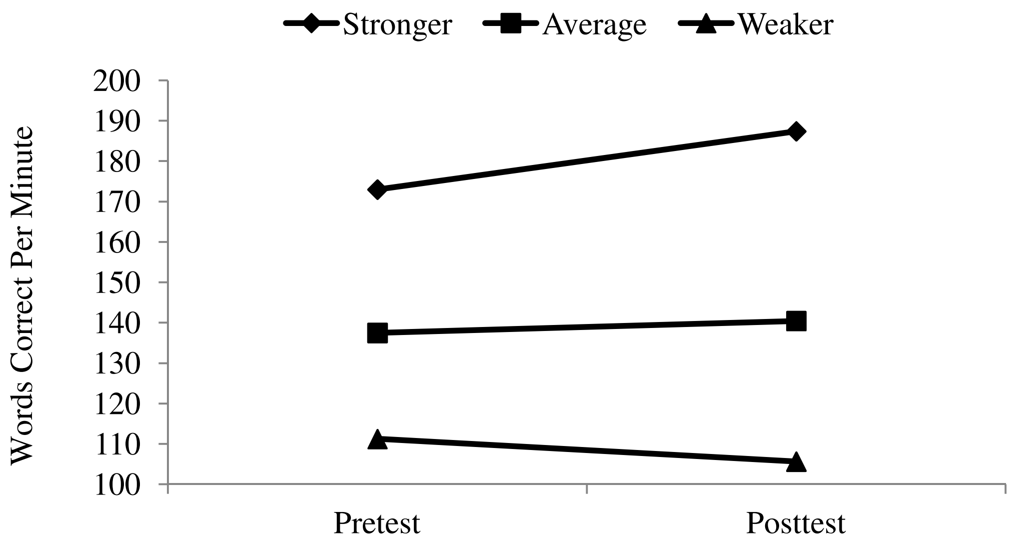


*Figure 3*. Mean words correct per minute (WCPM) by the stronger, average, and weaker groups in pretest and posttest.

The two-way mixed model ANOVA depicted in Figure 3 revealed a phase-by-subgroup interaction effect, $F(2, 47) = 6.921$, $p = .002$. Within subgroups, across phase pairwise comparisons revealed the stronger subgroup significantly increased their WCPM ($p < .000$), while changes in the average and weaker subgroups' WCPM across phases were not statistically significant. Pretest to posttest pairwise comparisons also revealed the stronger subgroup's 14.41 average increase in WCPM differed significantly from the weaker subgroup's 5.62 average decrease in WCPM, $p = .002$. The average subgroup's 2.94 WCPM increase from pretest to posttest did not differ significantly from the change in WCPM of the stronger or the weaker subgroups.

The correlation between median pretest WCPM and the difference in median pretest and posttest errors was significant, $r(50) = -.43$, $p = .002$. Those participants who achieved higher WCPM in the pretest phase tended to make fewer additional errors in the posttest phase when instructed to read fast. We used the same stronger, average, and weaker subgroups to test for significant differences in errors (see descriptive statistics in Table 4). A two-way mixed model ANOVA, depicted in Figure 4, revealed a significant main effect for subgroup, $F(2, 47) = 5.80$, $p = .006$, with the stronger subgroup differing significantly from the weaker subgroup ($p = .004$). The average subgroup did not significantly differ from the stronger or the weaker subgroups. There was also a significant main effect for phase, $F(1, 47) = 67.82$, $p < .000$, with more errors among participants during posttest. Pairwise comparisons revealed that all three subgroups significantly increased their mean number of errors from pretest to posttest, with the stronger subgroup increasing from 1.18 to 2.65 ($p = .005$), the average subgroup increasing from 2.06 to 3.77 ($p =.001$), and the weaker subgroup increasing from 2.38 to 6.31 ($p < .000$). Analysis revealed a significant phase-by-subgroup interaction, $F(2, 47) = 7.30$, $p = .002$. Pairwise

Table 4

*Error Means, Standard Deviations, Change Scores, p-Values, and Effect Sizes Across Stronger, Average, and Weaker Groups and Study Phases*

| | | Phases | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Pretest | | Posttest | | | | |
| Subgroups | *n* | *M*(*SD*) | Range | *M*(*SD*) | Range | Change | *p*-values | Effect Size |
| Stronger | 17 | 1.18(1.19) | 0-4 | 2.65(2.15) | 0-8 | +1.47 | .005 | .88 |
| Average | 17 | 2.06(1.30) | 0-5 | 3.76(1.80) | 1-7 | +1.70 | .001 | 1.10 |
| Weaker | 16 | 2.38(2.00) | 0-7 | 6.31(4.18) | 1-18 | +3.93 | .000 | 1.27 |

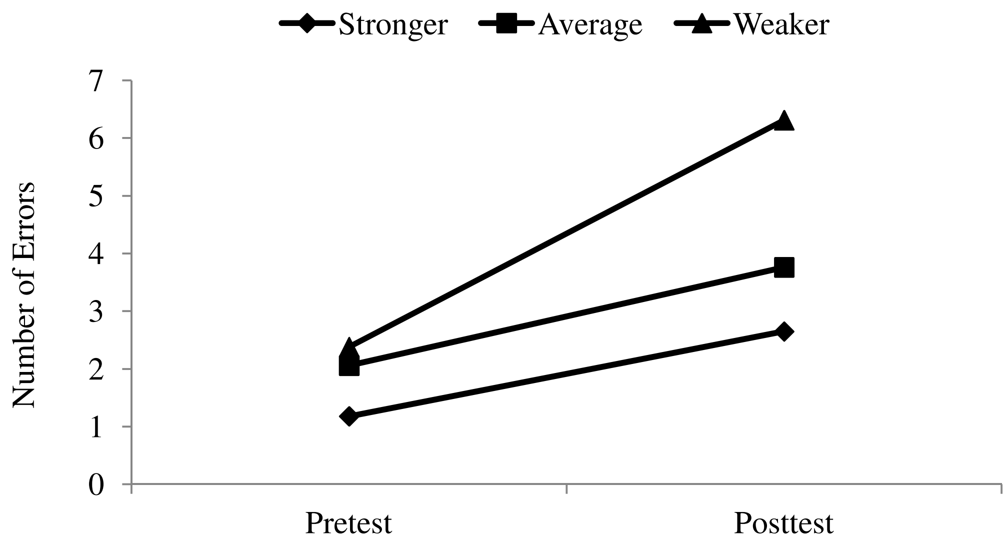


*Figure 4*. Mean errors by the stronger, average, and weaker groups in pretest and posttest.

comparisons showed that the weaker subgroups' increase in errors from pretest to posttest was significantly greater than the average ($p$ = .009) and stronger ($p$ = .003) subgroups' increases in errors.

## Discussion

For the control group, standard oral reading fluency (ORF) instructions were provided during both the pretest and posttest phases; therefore, the significant increases in errors and the significant decreases in WCPM from pretest to posttest suggest that posttest passages were more difficult than pretest passages. These differences in passage difficulty across phases likely reduced the observed increase in the experimental groups' WCPM (an insignificant increase of approximately 4 WCPM) after they were instructed to read as fast as they could. This limitation did not influence our phase-by-group interaction, which suggests that altering instructions to prompt students to read fast results in significantly higher WCPM than standard instructions. With respect to errors, our significant phase-by-group interaction suggests that instructing students to read fast also increases the number of errors made.

Our current results support previous findings which suggest that when administering ORF assessments encouraging students to read fast, either directly (e.g., with instructions) or indirectly (e.g., by showing them the stopwatch), may increase both WCPM and errors (Colón & Kranzler, 2006; Derr & Shapiro, 1989; Derr-Minneci & Shapiro, 1992). We expanded this research with our analysis that suggested reading skill moderated the effects of prompting students to read fast. Specifically, we found evidence that prompting students to read fast caused more improvement in WCPM and less of an increase in errors for stronger readers relative to the weaker readers. These findings have both applied and theoretical implications that must be considered in light of limitations.

## Limitations

There are a number of limitations associated with the current study. From pretest to posttest, the significant decrease in WCPM and increase in errors by the control group suggest that passages in the two phases were not equivalent. This finding supports those of others who suggested that readability formulas will not produce adequately equivalent passages (Brown-Chidsey et al., 2003; Poncy et al., 2005). Had the passages been equivalent, the increase in WCPM by the experimental group in posttest may have been more pronounced, thus providing clearer evidence of the effects of instructing students to read fast. Although our interaction analysis provided some control for these non-equivalent passages, researchers conducting similar studies should pretest passages to ensure a greater degree of equivalency.

There are also limitations associated with our measures. Our inter-scorer agreement correlation for errors ($r = .69$ and $.80$) were low and suggest scoring inconsistencies which may have limited our ability to find significant differences. Another limitation associated with the current study is related to our analyses of moderator effects. Although WCPM may serve as an adequate measure of global reading skills, we used pretest WCPM to form our stronger, average, and weaker subgroups, and we used this measure to calculate WCPM change scores. Consequently, we derived our change score using our measure of our moderator. Future researchers could address this limitation by establishing pre-reading scores using an independent measure (e.g., standardized reading achievement test).

Several external validity limitations associated with the current study should be addressed by future researchers. A relatively homogeneous sample of 73 seventh- and eighth-grade, mostly Caucasian students was used for this study. Similar studies should be conducted with larger and more diverse samples. Because ORF measures are often used as part of RtI programs in

elementary schools, similar studies should be conducted with students in grades one through five to determine if they respond to such prompts in the same manner. Researchers should conduct similar studies across different settings, using a variety of prompts to encourage rapid reading. Also, researchers should determine if reading skill by passage reading level interacts with prompts to read faster (Rhymer et al., 2002). Perhaps even strong readers will exhibit considerable increases in errors when asked to read passages as fast as they can if the passages are above their grade level.

**Implications and Future Research**

Despite these limitations, there are several theoretical implications associated with our findings that may provide direction for future research. Researchers should attempt to explain why students with stronger reading skills tended to show greater increases in WCPM and smaller increases in errors than those with weaker reading skills. Perhaps because they *can* read faster, merely prompting more rapid reading causes students with stronger skills to read faster.

Researchers investigating human performance have found evidence that skill level and arousal or stress may interact (Lars & Molander, 1986; O'Rourke et al., 2011). In our study, having a stranger (i.e., experimenter) instructing participants to read as fast as possible may have enhanced stress or arousal which enhanced performance (WCPM) in stronger readers, but not in weaker readers. An alternative cognitive-behavioral explanation may be that students who feel that they have weak skills become anxious when pressured to respond as fast as they can, and this anxiety hinders their oral reading performance (Abrahamsen, Roberts, Pensgaard, & Ronglan, 2008). Future researchers could test these hypotheses by running similar studies where students who are strong readers (two years above grade level) and weak readers (two years below grade level) are exposed to similar conditions while reading passages at or slightly below

their grade level and other passages several years above their grade level. If the effect is caused by a skill level by task interaction, then both groups should respond similarly to grade level and above grade level tasks. However, if self-perception (e.g., reading esteem) is causally related, then the weaker readers should be more adversely affected than the stronger readers across all passages.

There are applied implications associated with the current study that should be investigated. Some use ORF assessments within an RtI model to collect benchmark data by assessing all students within a school over a brief period of time. During these mass administrations, ORF assessments are often administered outside the classroom (e.g., cafeteria) by someone other than the student's teacher, which may influence WCPM scores (Derr & Shapiro, 1989; Derr-Minneci & Shapiro, 1992). Regardless, those who are performing poorly relative to their peers are granted access to remedial services (Hughes & Dexter, 2011).

Our results suggest that demand characteristics associated with ORF assessments can systematically suppress weaker readers' (those likely to receive RtI services) beginning-of-the-year benchmark WCPM scores, relative to the scores achieved by their peers. Future studies designed to explain why this may occur could have applied implications. For example, it is not difficult to imagine students experiencing increased arousal or stress as they wait in line outside the cafeteria for their turn to sit with a stranger and "do their best reading." If anxiety, arousal, or stress interacts with ORF performance dependent upon students' reading skills, by using strangers to conduct benchmark assessments in large rooms, we may be causing students with weaker reading skills to score even lower, relative to their peers.

Because all measures contain error, the lowest scoring subgroup from a sample frequently will exhibit improvement upon re-assessment (Hsu, 1995). This phenomenon, known

as regression to the mean, suggests that when subsequent assessments are used to determine whether those lowest scoring students are improving or responding to interventions, results will be affirmative, even if no skill development has occurred. Thus, both regression to the mean and demand characteristics associated with RtI benchmark testing may produce artificially lower scores in the lowest performing students during initial benchmark RtI assessments, which may result in inappropriate placement into remedial service. Additionally, subsequent assessments conducted by teachers in the classroom may result in higher scores because the students are less anxious or aroused as they have become more accustomed to assessment procedures as they are repeated. Thus, it is possible that educators are concluding that their remedial procedures have been effective, when in fact no change in skills has occurred. Instead, regression to the mean and lower levels of anxiety and arousal may result in higher scores on post-benchmark assessments. As both inappropriate placements and inappropriate evaluations of interventions can waste valuable resources and hinder student learning (Skinner, 2008), future researchers should continue to investigate the interactions of demand characteristics associated with ORF assessment procedures and within-student factors including reading skills, anxiety, and reading esteem.

**Conclusions: Standardized Administration may not be Sufficient**

Given that even subtle variations in assessment procedures can influence WCPM scores, (Derr & Shapiro, 1989) it is likely that other unidentified factors systematically influence WCPM scores. Our current study and past findings support applying standardized assessment procedures (Colón & Kranzler, 2006; Derr-Minneci & Shapiro, 1992). However, standardization may not be enough as we found evidence of an inconsistent and systematic impact of assessment procedures across students which could hinder our ability to make placement decisions based on

relative WCPM.  Furthermore, the impact of specific ORF assessment procedures may not be consistent within students.  If anxiety moderates the impact of some ORF assessment procedures even when standardized procedures are applied, weaker readers' WCPM scores may increase as students become accustomed to and less aroused or stressed by ORF assessment procedures during repeated assessments designed to evaluate their responsiveness or intervention effectiveness.  Given these applied implications researchers should continue this line of investigation.

## Chapter III

## The Effect of Prompts and Comprehension Questions on Oral Reading Fluency Scores: Is Reading Skill a Moderator?

As response to intervention (RtI) models are implemented in schools, it has become common practice to use oral reading fluency (ORF) assessments to gauge students' general reading abilities (Fuchs & Vaughn, 2012). ORF assessments provide a measure of the number of words students can read in one minute, called words correct per minute (WCPM). ORF scores account for a large percentage of variance in performance on other reading skill measures, can predict future achievement, and have discriminate validity among grade levels and special and general education students (Brown-Chidsey et al., 2003; Fuchs & Fuchs, 2002; Kim, Petscher, Schatschneider, Foorman, 2010; Reschly et al., 2009). Correlations between ORF scores and criterion outcome measures are generally in the .6 to .7 range (Reschly et al., 2009). WCPM scores and slope of improvement in WCPM has been a significant predictor of scores on global reading assessments, such as the Woodcock-Johnson Tests of Achievement, Third Edition (WJ-III) Broad Reading Cluster, the TerraNova Achievement Test, Second Edition, the Stanford Achievement Test, Tenth Edition, and end-of-year achievement exams (Baker et al., 2008; Keller-Margulis et al., 2008; Williams et al., 2011).

Although researchers have demonstrated that ORF assessments are good predictors of general reading skills, performance during ORF assessment is sensitive to environmental prompts that may be present during testing. Such prompts may include students being instructed to read fast, students being told they will be timed, or students being shown the timing device (e.g., stopwatch). Researchers have shown that applying these prompts can cause significant increases in WCPM and error scores (Colón & Kranzler, 2006; Derr & Shapiro, 1989).

Increases in WCPM caused by prompts may vary by reading skill level. Relative to weaker readers, average and stronger readers may show larger increases in their reading speed given certain environmental prompts, such as seeing a stopwatch (Derr-Minneci & Shapiro, 1992; Rhymer et al., 1999). This inconsistent effect across students suggests prompting faster reading may cause weaker readers to have even lower WCPM scores relative to stronger readers, which could lead to over-identification. If these differential effects are caused by arousal or anxiety interacting with skill levels, then within-student effects may be unstable. Consequently, less skilled readers who receive RtI remedial services and are assessed weekly may become accustomed to timed ORF assessment procedures, which may cause an increase in WCPM that is not indicative of reading skill improvement or remediation (see Experiment I).

Researchers attempting to develop a more direct measure of comprehension fluency have conducted assessments of ORF and reading comprehension by having students read passages aloud and then asking students to answer comprehension questions. Reading comprehension rate is calculated by dividing the percentage of correctly answered problems by the number of seconds spent reading, and then multiplying by 60 s (Skinner, 1998). Skinner et al. (2009) and Neddenriep et al. (2007) found that reading comprehension rate was a significant predictor of WJ-III Broad Reading Cluster scores for fourth- and fifth-grade students. Using sixth-, seventh-, and eighth-grade students, Hale et al. (2011) found that reading comprehension rate (RCR), Maze accurate response rate, and WCPM all significantly correlated with WJ-III Broad Reading Cluster scores, with RCR having the strongest relationship.

Although RCR has strong face validity, Skinner et al. (2009) found that the variance in WJ-III Broad Reading Cluster scores accounted for by RCR could be accounted for by reading speed alone. Thus, these researchers concluded that the variable that enhanced the face validity

of the RCR measure (the measure of comprehension in the numerator) did little to enhance the quantitative concurrent validity. Despite this finding, Skinner et al. (2009) suggested that asking comprehension questions following ORF assessments may discourage students from speed reading (i.e., rapid aloud word calling without comprehension). If environmental prompts cause increases in students' WCPM and error scores, administering a measure of reading comprehension may reduce those increases. This may be particularly true for students who have a history of responding to ORF assessments by reading as fast as they can, which may have been caused by their being timed or by explicit instructions to read fast (Colón & Kranzler, 2006; Derr & Shapiro, 1989; Derr- Minneci & Shapiro, 1992).

With the current study we replicated and extended research on prompting rapid reading. Specifically, we attempted to replicate aptitude-treatment interaction studies, which have provided some evidence that weaker and stronger readers would respond differently to prompts to read rapidly. Also, we attempted to extend this research by investigating the effect of instructing participants who read fast in the second phase of the study to do their best reading or do their best reading because they would answer comprehension questions in the third phase of the study.

## Method

### Participants and Setting

Participants included 37 sixth- and 36 eighth-grade students at two rural middle schools in the Southeastern United States. The sample included 37 males and 36 females; 57 were Caucasian, 11 were Hispanic, 1 was African-American, and 4 were multi-racial students. The percentage of students considered economically disadvantaged is equal to 60% at one school and 76% at the other school. Participants completed all procedures within two days in the late fall

and early spring semesters. Experimenters conducted Maze procedures in a group setting in the participants' classroom and ORF assessment procedures in a quiet office or hallway.

**Materials and Measures**

To measure participants' reading skills, three seventh-grade level Maze assessments were administered class wide, 1-6 weeks prior to running the experiment. During Maze assessments, students were provided with passages with three word options provided for every seventh word, only one of which makes sense in the sentence. Participants were instructed to read the passages silently and circle the word they believe makes sense. Several researchers have found that the number of correctly circled missing words per minute spent reading on Maze assessments is a strong predictor of reading comprehension and global reading skills (Hale et al., 2011; Shin et al., 2000; Wayman et al., 2007). After 3 min, participants were instructed to stop reading and hold their pencils in the air. For each participant, the median number of correctly circled words per minute was used to assess reading skill level.

Passages from the seventh-grade level *Timed Reading* series (Spargo, 1989) were used to obtain measures of WCPM and number of errors. Passages were modified by the primary experimenter to be exactly 120 words. Performance on these passages by a control group containing 23 seventh- and eighth-grade participants during Experiment I was used as an indicator of passage difficulty level. During Experiment I, the control group participants scored an average of 146.37 WCPM (range = 144.98-147.57) with an average standard deviation of 29.35 (range = 27.50-30.66) on the three easier passages. On the three harder passages they scored an average of 132.56 WCPM (range = 131.95-133.34) with an average standard deviation of 27.06 (range = 24.99-30.29). For Experiment II, we paired one easier passage with one harder passage, creating three sets of hard-easy ORF assessment sets. The average WCPM and standard

deviations obtained by the 23-participant control group during Experiment I was roughly equivalent across the three ORF assessment sets: 139.16 (30.48), 139.76 (27.90), and 139.27 (26.33) WCPM, respectively.

Provided in the *Timed Reading* series are inferential and factual multiple-choice questions for each passage. During Experiment II, some participants received two inferential and two factual questions immediately following completion of three ORF assessments. Answers to these questions were not dependent variables; rather, these questions served as an independent variable as we were interested in determining if including these questions would discourage participants from reading as fast as they can without regard for comprehension.

School psychology Ph.D. students administered the ORF assessment procedures. All had prior experience administering ORF assessments. Each experimenter received additional training on implementing the study procedures and practiced delivering each set of instructions. Experimenters reviewed the standards for what qualifies as an error and practiced administering the assessments by having another experimenter read and intentionally make errors. Experimenters also practiced scoring by listening to audio recordings of a participant reading the passages. This was done so experimenters had experience scoring using audio recordings prior to obtaining estimates of inter-scorer agreement. Experimenters were instructed to audio record all participants reading the passages by starting a voice recorder and placing it in plain view on the table in front of the participants before delivering instructions. Experimenters were also instructed to visibly start and stop a stopwatch during all assessments, without making any attempts to hide the stopwatch or draw the participants' attention to it.

During ORF assessments, experimenters followed along on a copy of the passages, scoring errors, supplying words when participants paused for 3 s, and redirecting participants if

they lost their place as they read aloud.  Skipped words, mispronunciations, transposed words, and words provided by an experimenter after a 3-s pause counted as errors.  Errors corrected within 3 s were scored as correct.  For each passage, after the student finished reading, the experimenter recorded the number of seconds spent reading.  After reading all passages, experimenters counted the number of errors made on each passage and calculated WCPM by subtracting errors from the number of words in the passages, multiplying the number of words read correct by 60 s, and dividing by the number of seconds spent reading.  The average number of errors made and WCPM within each phase of the experiment served as the dependent variables.

**Procedures**

After obtaining institutional support for the study, one sixth- and one eighth-grade teacher from each school sent parental consent forms home with all of their students. After receiving consent from at least 63 students to participate, the primary experimenter came to the participants' classrooms to obtain child assent, ask students to complete a demographics form, and administer the Maze assessments.  When completing the Maze assessments, the experimenter provided the following instructions:

> When I say 'Begin' turn to the first story and start reading silently. When you come to a group of three words, circle the one word that makes the most sense. Some of the words are replaced with a group of three words. Your job is to circle the one word that makes the most sense in the story. Work as quickly as you can without making mistakes. If you finish the page before I say stop, raise your hand and wait for further instructions. Do not turn to the next story until I tell you to. Do you have any questions? Begin. (Shinn & Shinn, 2002b, p. 14).

The experimenter kept time on a stopwatch and told participants to stop after 3 min. These procedures were repeated three times for each Maze assessment passage. After the third passage, all assessments were collected by the experimenter. A participant completed an assessment before 3 min elapsed on three occasions, two of which were by the same participant. In these instances the experimenter recorded time spent reading and calculated accurate response rate per minute by multiplying correct responses by 60 s and dividing by the number of seconds spent reading. Once the Maze assessments were scored, participants were placed into stronger (highest 33%), average (middle 33%), and weaker (lowest 33%) reading skill groups based on their median Maze assessment scores. During Maze assessments, the experimenter used a procedural integrity form and self-recorded steps as they were completed (see Appendix A).

Experimenters returned to the participants' classroom on another day to individually administer the ORF assessments. Stratified random assignment, based on Maze assessment scores, was used to form groups. Instructions provided for completing the ORF assessments in each phase varied by group. Table 5 displays the order in which each group received instructions across the three phases. One group received standard instructions, followed by instructions to read fast, followed by standard instructions plus an instruction informing them that they would answer comprehension questions. This group is referred to as the standard, fast, questions (SFQ) experimental group. Another group received standard instructions, followed by instructions to read fast, followed by standard instructions again. This group is referred to as the standard, fast, standard (SFS) experimental group. The control group received standard instructions during all ORF assessments and they are referred to as the standard, standard, standard (SSS) control group. All participants received the same passages in the same order for each phase of the study.

Table 5

*Instructions Provided to Each Group During Each Phase*

| | Phase One (Passages One and Two) | Phase Two (Passages Three and Four) | Phase Three (Passages Five and Six) |
|---|---|---|---|
| SFQ Experimental Group | Standard | Read Fast | Comprehension Questions |
| SFS Experimental Group | Standard | Read Fast | Standard |
| SSS Control Group | Standard | Standard | Standard |

Each participant completed all ORF assessments in a quiet hallway or office outside their classrooms in one session, which lasted approximately 10 min. When completing phase one of the study, all groups received standard instructions. Performance during this phase provided a baseline measure of participants' average WCPM and error scores. Participants included in the SSS control group continued to read under these same standard instructions for phases two and three. When providing standard instructions experimenters read the following:

> When I say 'begin,' start reading this passage aloud. Read across the page. Try to read each word. If you come to a word you don't know I'll tell it to you. Be sure to do your best reading. Are there any questions? Begin. (Shinn & Shinn, 2002a, p.18)

Participants included in the SFQ and SFS experimental groups received read fast instructions during the second phase. Read fast instructions prompted students to read as fast as possible.

> When I say 'begin,' start reading this passage aloud. Read across the page. Try to read each word. If you come to a word you don't know I'll tell it to you. Read as fast and accurately as you can. Don't pause for punctuation, instead read as fast as possible. Are there any questions? Begin.

SFQ experimental group participants were administered four comprehension questions after they finished the second passage in phase two. Note, these participants were not informed that these questions would be delivered. We included these questions following our collection of WCPM data in phase two to ensure that these participants had experience answering comprehension questions, a requirement in phase three. Participants in the SFQ experimental group were then given comprehension question instructions for completing phase three. Experimenters provided the following when giving comprehension question instructions:

When I say ‘begin,’ start reading this passage aloud. Read across the page. Try to read each word. If you come to a word you don’t know I’ll tell it to you. Be sure to do your best reading because when you are finished you will answer questions on what you just read. Are there any questions? Begin.

During phase three, SFS experimental group participants were not provided with comprehension questions; rather, they received standard instructions. Thus, our goal was to determine if SFS experimental group participants would continue reading more rapidly even when instructed to do their best. Researchers anticipated that after receiving the comprehension questions instructions or receiving standard instructions again, some participants in the SFQ and SFS experimental groups may ask for clarification (e.g., ask if they should attempt to read fast). In these instances the researchers were trained to repeat the key phrase of the instructions (i.e., do your best reading, do your best reading because you will have to answer questions). All participants received comprehension questions after completing the second and final passage in phase three. This was done for potential exploratory analyses.

**Analysis**

Performance under the varying instructions was analyzed using the average WCPM and error scores across the two passages included in each phase. For each dependent variable, a three-by-three mixed model ANOVA was used to test for significant differences. The within-subjects factor was phase (first, second, and third) and the between-subjects factor was group (SFQ experimental, SFS experimental, and SSS control).

Also, using only WCPM and error scores by the SFQ and SFS experimental groups in phases one and two, we extended our analysis of moderator variables. First, we correlated Maze assessment scores with WCPM and error change scores across phases one and two. We then ran

a two-by-three mixed model ANOVA to test for differences in stronger, average, and weaker readers' reading performance when prompted to read fast. In Experiment I, participants were assigned to reading skill groups based on baseline WCPM scores, which were also used to calculate change scores from phase one to phase two. In the current study, participants were assigned to reading skill groups using pre-experimental Maze assessment scores.

**Inter-scorer Agreement and Procedural Integrity**

To obtain an estimate of inter-scorer agreement for participants' Maze assessment scores, a second experimenter independently scored 27% of these assessments. Going item-by-item, the number of agreements and disagreements were calculated and the number of agreements was divided by the sum of agreements and disagreements and multiplied by 100. Inter-scorer agreement for Maze assessment scores was 100% and the procedural integrity form suggested that experimenters accurately completed 100% of the Maze assessment procedures. To obtain an estimate of inter-scorer agreement on ORF assessment scores, two experimenters collectively listened to approximately 27% of the audio recordings of participants' readings and independently scored their errors. Pearson product-moment correlations between the original experimenter and second experimenter were obtained for WCPM and error scores. The correlation between the two raters was .998 for passages in phase one, .998 for passages in phase two, and .986 for passages in phase three. For errors, the correlation between the two raters for passages in phase one was .993, .959 for passages in phase two, and .927 for passages in phase three. While collecting ORF assessment inter-scorer agreement data, the experimenters collected procedural integrity data by recording completed steps using a procedural integrity form (see Appendix B). Experimenters accurately completed 100% of the ORF assessment procedures.

# Results

## Words Correct Per Minute (WCPM)

Group WCPM means and standard deviations for each phase are presented in Table 6. A three-by-three mixed model ANOVA revealed a non-significant main effect for group and a significant main effect for phase, $F(2, 69) = 25.06$, $p < .000$. Participants read significantly more WCPM in phase two compared to phase one ($p < .000$), and phase three ($p < .000$). The number of WCPM did not significantly differ in phases one and three.

The group-by-phase interaction, depicted in Figure 5, was significant, $F(4, 140) = 7.35$, $p < .000$. Effect sizes and *p*-values for within-group differences in WCPM across phases are displayed in Table 7. Both the SFQ and SFS experimental groups significantly increased their WCPM from phase one to phase two ($p < .000$ for both groups) and significantly decreased their WCPM from phase two to phase three ($p < .000$ for SFQ experimental group and $p = .002$ for SFS experimental group). When comparing phases one and three, differences on WCPM were non-significant for the SFQ and SFS experimental groups. No significant across- phase WCPM differences were found for the SSS control group.

The differences in phase change WCPM scores between groups and associated *p*-values are displayed in Table 8. The increases in WCPM from phase one to two by the experimental groups was significantly larger than the increase by the SSS control group ($p < .000$ for SFQ experimental group comparison and $p = .001$ for SFS experimental group comparison). The decrease in WCPM by experimental groups from phases two to three differed significantly from the small increase in WCPM by the SSS control group ($p < .000$ for SFQ experimental group comparison and $p = .009$ for SFS experimental group comparison). Across-phase comparisons of WCPM change scores revealed no other between group significant differences.

Table 6

*Mean WCPM and Standard Deviations for Each Group Across Phases*

| Group | Phase One<br>Mean WCPM (SD) | Phase Two<br>Mean WCPM (SD) | Phase Three<br>Mean WCPM (SD) |
|---|---|---|---|
| SFQ Experimental Group (n=27) | 137.70 (43.06) | 159.48 (49.42) | 142.03 (38.08) |
| SFS Experimental Group (n=24) | 141.33 (44.05) | 155.76 (46.47) | 143.75 (43.80) |
| SSS control group (n=22) | 136.76 (44.70) | 136.80 (42.55) | 138.00 (40.40) |

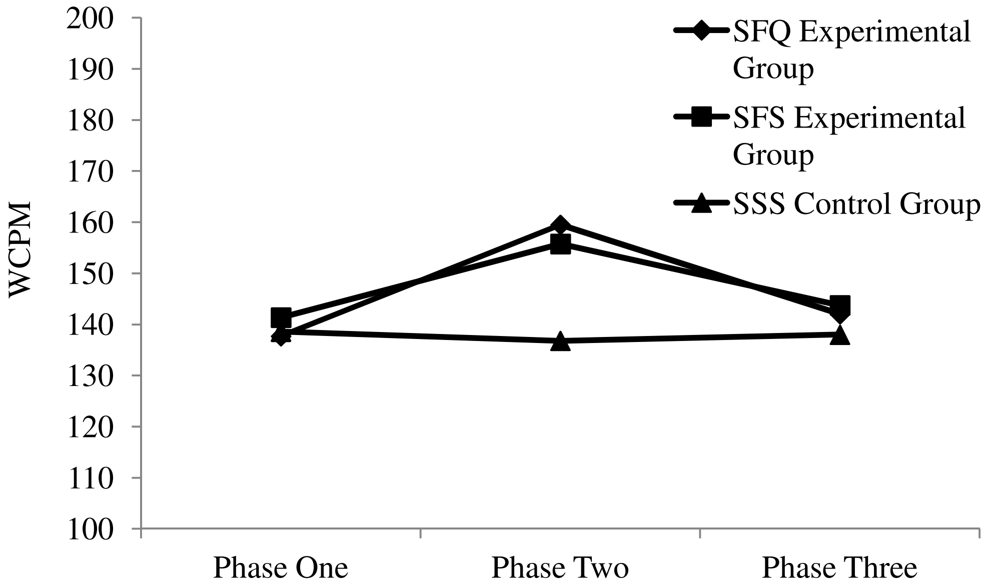


*Figure 5.* Mean WCPM by all groups across phases.

Table 7

*WCPM p-Values and Effect Sizes for Within Group Differences Across Phases*

| | Comparison | Change | *p* Value | Effect Size |
|---|---|---|---|---|
| SFQ Experimental Group | Phase One to Two | +21.78 | .000* | 0.47 |
| | Phase Two to Three | -17.45 | .000* | 0.40 |
| | Phase One to Three | +4.33 | .377 | 0.11 |
| SFS Experimental Group | Phase One to Two | +14.42 | .000* | 0.32 |
| | Phase Two to Three | -12.01 | .002* | 0.27 |
| | Phase One to Three | +2.42 | 1.00 | 0.06 |
| SSS control group | Phase One to Two | +.034 | 1.00 | 0.00 |
| | Phase Two to Three | +1.18 | 1.00 | 0.03 |
| | Phase One to Three | +1.22 | 1.00 | 0.03 |

* Indicates significant at the $p < 0.05$ level

Table 8

*Comparisons of Differences in WCPM Change Scores Between Groups Across Phases*

| | Phase One to Two | | Phase Two to Three | | Phase One to Three | |
|---|---|---|---|---|---|---|
| | Difference | *p*-value | Difference | *p*-value | Difference | *p*-value |
| SFQ v SFS Experimental Groups | 7.35 | .076 | 5.44 | .248 | 1.91 | .640 |
| SFQ Experimental Group v SSS Control Group | 21.75 | .000* | 18.63 | .000* | 3.11 | .458 |
| SFS Experimental Group v SSS Control Group | 14.39 | .001* | 13.19 | .009* | 1.20 | .780 |
| Collapsed SFQ and SFS Experimental Groups v SSS Control Group | 18.064 | .000* | -- | -- | -- | -- |

*Indicates significant at $p < .05$ level

**Errors**

The mean number of errors and standard deviations made by the SFQ experimental group, SFS experimental group, and the SSS control group across the three phases are displayed in Table 9. A three-by-three mixed model ANOVA was used to compare mean differences in errors. The analysis revealed a non-significant main effect for group and significant main effect for phase, $F(2, 69) = 4.29$, $p = .016$. When comparing errors across phases one and two and phases one and three, there were no significant differences. Participants made significantly more errors in phase two compared to phase three ($p = .013$).

The group-by-phase interaction, depicted in Figure 6, was significant, $F(4, 140) = 3.49$, $p = .009$. Effect sizes and *p*-values for within group error differences across phases are displayed in Table 10. No significant across-phase differences for the SFQ experimental group and the SSS control group were found. The SFS group's mean errors in phase two was significantly greater than their mean errors in phase one ($p = .012$), and their mean errors in phase three ($p = .017$). Their mean errors in phases one and three did not significantly differ.

The differences in phase-change error scores between groups and associated *p*-values are displayed in Table 11. From phase one to two, both the SFQ and SFS experimental groups made significantly greater increases in errors relative to the control group (p = .016 for SFQ experimental group comparison and p = .004 for SFS experimental group comparison). Changes in error scores from phase two to three did not differ significantly between any groups. The difference in error change scores from phase one to three by the SFQ and SFS experimental groups differed significantly from the change by the SSS control group ($p = .012$ for SFQ experimental group comparison and $p = .038$ for SFS experimental group comparison). The SSS control group had a steady decline in errors across the three phases of the experiment, which

Table 9

*Mean Errors and Standard Deviations for Each Group Across Phases*

| Group | Phase One Mean Errors (SD) | Phase Two Mean Errors (SD) | Phase Three Mean Errors (SD) |
|---|---|---|---|
| SFQ Experimental Group (n=27) | 2.87 (2.86) | 4.13 (4.11) | 3.56 (4.49) |
| SFS Experimental Group (n=24) | 3.00 (3.46) | 4.75 (5.80) | 3.37 (4.04) |
| SSS Control Group (n=22) | 4.56 (6.50) | 3.80 (4.87) | 3.30 (3.78) |

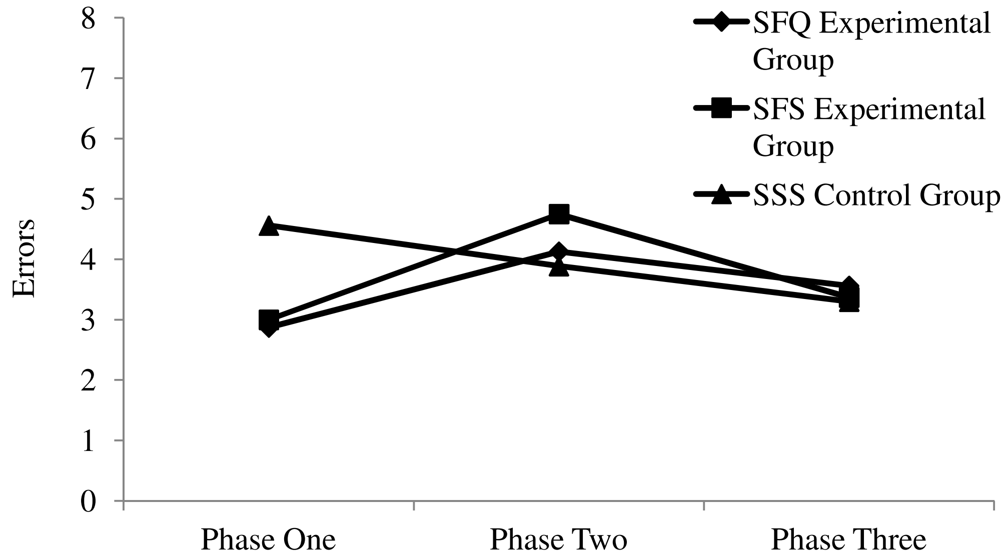


*Figure 6.* Mean errors by all groups across phases.

Table 10

*Error p-Values and Effect Sizes for Within Group Differences Across Phases*

| | Comparison | Change | *p* Value | Effect Size |
|---|---|---|---|---|
| SFQ Experimental Group | Phase One to Two | +1.26 | .078 | 0.36 |
| | Phase Two to Three | -.57 | .626 | 0.13 |
| | Phase One to Three | .69 | .543 | 0.19 |
| SFS Experimental Group | Phase One to Two | +1.75 | .012* | 0.38 |
| | Phase Two to Three | -1.38 | .017* | 0.28 |
| | Phase One to Three | +.38 | 1.00 | 0.10 |
| SSS control group | Phase One to Two | -.77 | .634 | 0.13 |
| | Phase Two to Three | -.50 | .966 | 0.12 |
| | Phase One to Three | -1.27 | .080 | 0.25 |

* Indicates significant at the $p < 0.05$ level

Table 11

*Comparisons of Differences in Error Change Scores Between Groups Across Phases*

| | Phase One to Two | | Phase Two to Three | | Phase One to Three | |
|---|---|---|---|---|---|---|
| | Difference | *p*-value | Difference | *p*-value | Difference | *p*-value |
| SFQ v SFS Experimental Groups | .49 | .545 | .81 | .229 | .31 | .676 |
| SFQ Experimental Group v SSS Control Group | 2.03 | .016* | .07 | .913 | 1.96 | .012* |
| SFS Experimental Group v SSS Control Group | 2.52 | .004* | .88 | .212 | 1.65 | .038* |
| Collapsed SFQ and SFS Experimental Groups v SSS Control Group | 2.27 | .003* | -- | -- | -- | -- |

*Indicates significant at $p < .05$ level

likely accounts for this difference.

**Moderating Effects of Reading Skill**

Table 12 displays the correlation between Maze assessment scores and WCPM change scores from phase one to phase two for both groups who received instructions to read fast. Correlations were non-significant for both changes in WCPM and changes in errors, $r(51) = -.16$, $p = .205$. A mixed model ANOVA was also run to investigate increases in WCPM by participants with stronger and weaker reading skills.

The SFQ and SFS experimental groups were collapsed in order to compare performance by readers at different reading skill levels in phases one and two. Participants were classified as stronger, average or weaker readers based on relative median Maze accurate response rates. Table 13 displays the Maze assessment data for each of the three reading skill level groups. The highest scoring third were considered stronger readers, the middle third average readers, and the lowest third weaker readers.

The mean number of WCPM and standard deviations by each reading skill group in phases one and two are displayed in Table 14. A two-by-three mixed model ANOVA revealed a significant main effect for group, F(2, 48) = 21.02, p < .000. The stronger readers' WCPM scores were significantly greater than the average readers' (p < .000) and the weaker readers' (p < .000), and the average readers' WCPM scores were significantly greater than the weaker readers' WCPM scores (p = .012). There was also a significant main effect for phase with participants reading more WCPM in phase two, $F(1, 48) = 63.72$, $p < .000$. Table 14 also displays the mean WCPM increase by each reading skill group, the associated *p*-value, and effect size. All three groups significantly increased in WCPM from phase one to phase two ($p < .000$ for all groups). The reading skill-by-phase interaction, depicted in Figure 7, was non-significant,

Table 12

*Correlations for Criterion Variables and Change Scores*

| | WCPM Phase One | Maze Accurate Response Rate | Phase Three Passage Two RCR |
|---|---|---|---|
| WCPM Phase One-Two, SFQ and SFS | .10 | .14 | .23 |
| WCPM Phase Two-Three, SFQ | -.30 | -.05 | -.09 |
| WCPM Phase Two-Three, SFS | -.47* | -.25 | -.12 |

*Indicates significant at the $p$ = .05 level.

Table 13

*Mean Maze Accurate Response Rate and Standard Deviations for Each Reading Skill Level Group*

| Reading Skill Group | Mean Maze Accurate Response Rate (SD) |
|---|---|
| Stronger (n = 16) | 12.61 (1.71) |
| Average (n = 18) | 8.82 (1.02) |
| Weaker (n = 17) | 5.45 (.90) |

Table 14

*Mean WCPM and Standard Deviations for WCPM for Stronger, Average, and Weaker Readers in the SFQ and SFS Experimental Groups in Phases One and Two*

| Group | Phase One Mean WCPM (SD) | Phase Two Mean WCPM (SD) | Change | *p* Value | Effect Size |
|---|---|---|---|---|---|
| Stronger (n=16) | 178.09 (26.86) | 198.13 (37.15) | +20.04 | .000* | 0.63 |
| Average (n=18) | 137.49(37.24) | 156.69 (41.14) | +19.21 | .000* | 0.49 |
| Weaker (n=17) | 105.05 (30.14) | 120.80 (30.17) | +15.75 | .000* | 0.52 |

* Indicates significant at the $p < 0.05$ level

$F(2, 48) = .32$, $p = .725$, indicating the three reading skill groups did not differ in their increases in WCPM scores from phase one (standard instructions) to phase two (read fast instructions).

The mean number of errors and standard deviations made by each group in phases one and two are displayed in Table 15. Error change scores from phase one to phase two and their corresponding *p*-values and effect sizes are also displayed in Table 15. A two-by-three mixed model ANOVA revealed a significant main effect for group, $F(2, 48) = 6.69$, $p = .003$. The stronger readers made significantly fewer errors than the weaker readers ($p = .002$). The stronger readers did not significantly differ from the average readers, and the average readers did not significantly differ from the weaker readers. There was a significant main effect for phase with participants making more errors in phase two, $F(1, 48) = 12.44$, $p = .001$. The weaker readers significantly increased in number of errors made from phase one to phase two ($p = .002$), while the average and stronger groups showed non-significant increases; however, the reading skill-by-phase interaction, depicted in Figure 8, was non-significant, $F(2, 48) = 1.12$, $p = .336$.

**MAZE Variance Accounted for by ORF Assessments**

Correlations between Maze assessment scores and WCPM can be found in Table 16. All correlations between Maze assessment scores and WCPM under varying instructions were significant. Furthermore, different instructions yield similar correlations between Maze assessments and WCPM (range = .79 to .82). We converted the correlations to Z-scores and found no significant differences between them (Cohen & Cohen, 1983). Similar correlations suggest performance on ORF assessments under standard, read fast, and read fast plus comprehension questions is predictive of performance on a general reading skill measure.

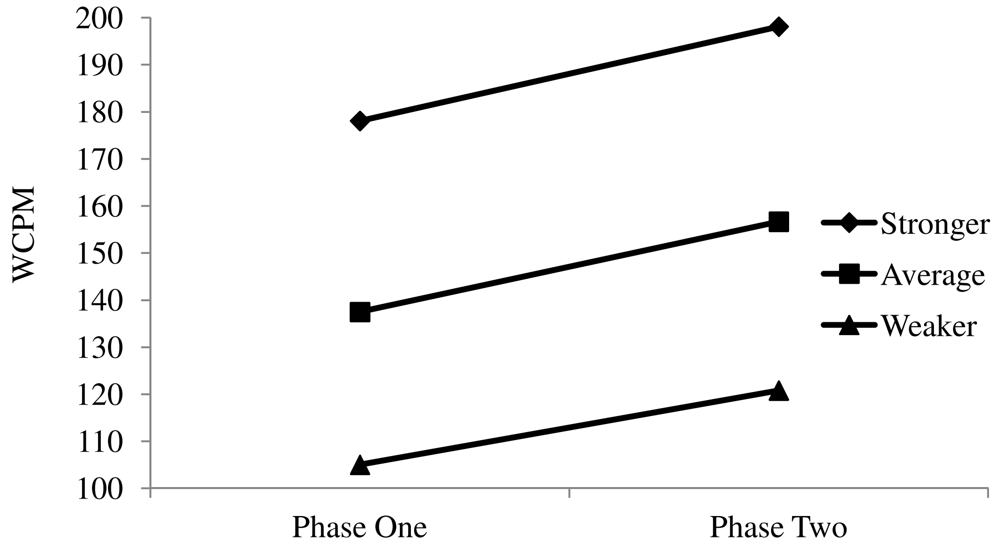


*Figure 7.* Mean WCPM by the stronger, average, and weaker reading skill level groups across phases.

Table 15

*Mean errors and Standard Deviations for Stronger, Average, and Weaker Readers in the SFQ and SFS Experimental Groups in Phases One and Two*

| Group | Phase One Mean Errors (SD) | Phase Two Mean Errors (SD) | Change | *p* Value | Effect Size |
|---|---|---|---|---|---|
| Stronger (n=16) | 1.00 (1.25) | 2.06 (1.75) | +1.06 | .165 | 0.71 |
| Average (n=18) | 2.94 (2.26) | 3.97 (2.39) | +1.03 | .155 | 0.44 |
| Weaker (n=17) | 4.74 (4.02) | 7.12 (7.32) | +2.38 | .002* | 0.42 |

* Indicates significant at the $p < .05$ level

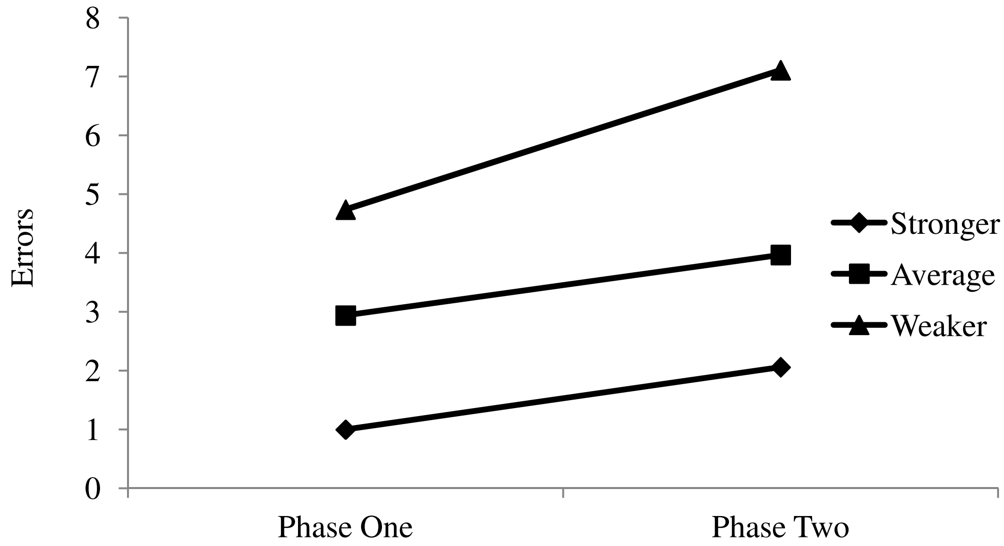


*Figure 8.* Mean errors by the stronger, average, and weaker reading skill level groups across phases.

Table 16

*Correlations Between Maze Assessment Scores and WCPM Across Phases*

| | Standard Instructions (Phase One Average For All Groups) | Read Fast Instructions (Phase Two Average For SFQ and SFS Groups) | Return to Standard Instructions (Phase Three Average For SFS Group) | Return to Standard Instructions Plus Comprehension Questions (Phase Three Average for SFQ Group) |
|---|---|---|---|---|
| Maze Accurate Response Rate | .79** | .76** | .82** | .80** |

**Indicates significant at the $p < .01$ level.

## Discussion

Consistent with results from Experiment I, instructing participants to read fast increased their WCPM and errors scores (Colón & Kranzler, 2006; Derr & Shapiro, 1989; Derr-Minneci & Shapiro, 1992). The SFQ and SFS experimental groups increased their WCPM by approximately 22 words correct and 14 words correct, respectively; a large increase given that the average yearly WCPM growth for sixth-grade students is approximately 29 words (Christ et al., 2010). The finding that participants in the SFS experimental group significantly increased in errors from phase one to phase two suggests that readers may lose accuracy when attempting to increase speed of reading.

In phase three, when participants were instructed to do their best reading (SFS experimental group) or to do their best reading because they would answer questions (SFQ experimental group), mean WCPM and error scores returned to levels similar to the initial standard instructions condition (phase 1) and those obtained by the SSS control group in phase three. Additionally, the decrease in WCPM and errors by participants in the SFQ and SFS experimental groups did not significantly differ. These findings suggest that prompting students to do their best reading may be enough to encourage students to read in the manner they would when attempting to comprehend.

Correlations between Maze accurate response rate and change scores from phase one to phase two were non-significant for the SFQ and SFS experimental groups. Additionally, the interactions between reading skill and instructions were non-significant, suggesting that reading skill did not moderate the effects of being instructed to read fast. This finding is contrary to findings from Experiment I. Other researchers have found that when students are prompted to work quickly, those with stronger math skills increased their speed of accurate responding more

than those with weaker math skills (Rhymer et al., 1998; Rhymer et al., 2002). Weaker readers significantly increased their errors from phase one to two and stronger and average readers did not, suggesting weaker readers may have struggled more to increase their reading speed while maintaining accuracy. However, the interaction was non-significant. Limitations associated with this study may provide some insight as to why we failed to find evidence of reading skill-by-instructions interactions.

**Limitations and Future Research**

Previous researchers who prompted increased speed of responding have found evidence of a skill level-by-treatment interaction when students were working on more difficult math tasks (Rhymer et al., 2002). In Experiment II, the passages administered may not have been difficult enough to allow for an interaction between reading skill and instructions. Future researchers should provide students with passages well above the appropriate grade level. If readers struggle to increase their reading speed when given passages well above their grade level, the interaction effect between skill level and prompts that increase responding could be due to passage difficulty.

Readers may experience stress when instructed to read fast. This stress could have a greater impact on weaker readers as it would likely be more difficult for them to increase their reading speed. We may have found a skill level-by-instructions interaction had the stress weaker readers experienced been paired with additional stress brought on by the administration of harder passages. Perhaps interactions between skill level and stress when prompted to respond faster are found under circumstances in which stress brought on by a prompt and by the difficulty level of the task are both present (Diamond et al., 2007). Thus, future researchers should administer passage on and above participants' grade level and administer measures of anxiety and stress

when attempting to determine if skill level and prompts to increase or improve responding interact.

The sample included mostly Caucasian, sixth- and eighth-grade participants. Similar studies should be conducted with larger, more diverse samples at younger grade levels. Most RtI programs are implemented at the elementary school level. The effect of prompting students to read fast may have different effects on students at younger grade levels.

**Implications**

We found that students at all reading levels significantly increased their WCPM scores when prompted to read fast. This finding has implications for using students' WCPM scores to make decisions regarding intervention services and determining if students are responding to interventions. If students are capable of increasing their WCPM scores, some students may attempt to do this when completing ORF assessments and others may not. This creates challenges in determining which students perform low enough on ORF assessments to qualify for intervention services. A student who is in need of intervention services, but attempts to read fast during ORF assessments could obtain a higher WCPM score than students with average reading skills. Additionally, students who are undergoing interventions may attempt to read fast on some ORF assessment occasions and not others. This could create large fluctuations in WCPM scores over time, making it difficult to evaluate intervention effectiveness (Poncy et al., 2005).

After prompting participants to read fast, prompting them to do their best reading, or to do their best reading because they would answer questions had similar effects on WCPM and error scores; thus, requiring students to complete measures of comprehension may be unnecessary. This is an encouraging finding in that it suggests telling students to do their best

reading may be sufficient for prompting reading for understanding. However, it is still important that all students interpret ORF assessment instructions the same way and attempt to complete the assessments in the same manner. Other prompts (e.g, stopwatch or those not yet identified by researchers) that increase WCPM and error scores may still influence students to read fast (Derr & Shapiro, 1989; Derr-Menneci & Shaprio, 1992). Because comprehension measures may help to lessen the effects of such known and unknown prompts, facilitating our ability to make comparisons between students and within students over time, future researchers should continue this line of research. For example, researchers could determine if students who are instructed to read fast moderate their reading speed when they are also told they will have to answer comprehension questions.

Analysis of the relationship between Maze assessment scores and WCPM across phases revealed strong correlations. These data suggest that standard, fast, and comprehension question instructions all produce valid measures of reading skills. Correlations were all approximately, suggesting performance under varying instructions may be equally predictive of general reading skills.

**Conclusion**

When provided with passages that are controlled for difficulty level, students at all reading skill levels can significantly increase their WCPM scores when prompted to read fast. Because WCPM scores on ORF assessments are now being used in part when making placement decisions, the finding that these scores can be easily inflated suggests that administrators should ensure all students are prompted to complete ORF assessments in the same manner. Findings from this study also suggest instructing students to do their best reading and giving students comprehension measures are equally effective in slowing reading speed; therefore, it may be

unnecessary to provide comprehension questions. However, because comprehension measures could work to reduce the effect of other types of prompts (e.g., viewing the stopwatch, information provided about the importance of reading fast before the assessments) on WCPM and error scores, future researchers should conduct studies investigating these interactive effects.

# Chapter IV

## Conclusion

Student performance on ORF assessments is used in decision-making processes for identifying students in need of reading intervention services, determining if students have responded to interventions, and evaluating remedial procedures (Shapiro, 2011). Researchers have found that during ORF assessments, students will increase their WCPM and errors when exposed to certain prompts (Colón & Kranzlers, 2006; Derr & Shapiro, 1989). Also, there is some evidence that prompt-induced changes in ORF performance may be moderated by students' reading skill level; those with stronger skills may increase their rates of responding more than those with weaker skills (Derr-Minneci & Shapiro, 1992; Diamond, et al., 2007; Rhymer et al., 1998). Asking comprehension questions after ORF assessments could reduce the effect of these prompts, allowing for a more accurate representation of students' ORF skills (Hale et al., 2012).

We attempted to replicate Colón and Kranzler's (2006) research in Experiment I by instructing students to read fast and accurately. We extended this research by analyzing our data for a reading skill-by-instructions interaction. In Experiment II we repeated procedures carried out in Experiment I and added a phase to our study during which participants who read fast in phase two were instructed to read their best or do their best reading because they would answer questions in phase three. In both Experiments, we found that students increased their WCPM scores and also made more errors when instructed to read fast.

In Experiment I we found evidence of an aptitude-by-treatment interaction. Specifically, the correlation between baseline WCPM (a measure of global reading skills) and WCPM change scores from the standard instructions phase (pretest) to the read fast phase (posttest) was significant. Also, when we divided the participants into three groups (stronger, average, and

weaker readers), our mixed model ANOVA revealed a significant reading skill group-by-instructions interaction. Students in the stronger group increased their WCPM scores and made fewer errors when instructed to read fast compared to students with weaker reading skills. In Experiment II, participants at all skill levels significantly increased their WCPM scores when instructed to read fast and correlations between Maze scores and WCPM change scores were not significant.

During phase three of Experiment II, participants who were instructed to read fast in phase two were given standard instructions or standard instructions plus comprehension questions. These participants showed decreases in their WCPM scores back to levels similar to those obtained in phase one. This finding suggests that after students' reading has been altered by instructing them to read fast, it is possible, through instructions, to decrease students' WCPM and error scores to levels typically found when provided with standard instructions and when reading for understanding.

**Implications, Limitations, and Future Research**

**Increases in WCPM and errors after read fast instructions.** In Experiments I and II, we found that participants significantly increased their WCPM and error scores when given instructions to read fast. WCPM scores are used in the RtI process to make decisions regarding students' placements and special education eligibility. If WCPM scores can be quickly and easily inflated, making comparisons on WCPM across students and within students over time may not always be appropriate. To ensure all students are completing ORF assessments in the same manner, ORF administrators should ensure that all students are receiving the same standard instructions every time they are assessed. How students interpret the meaning of "best reading" could also affect the manner in which they complete ORF assessments (Colón & Kranzlers,

2006). If some students interpret "best reading" to mean that they should read carefully and for understanding, and other students interpret it to mean they should read as many words as possible, making accurate comparisons across students will be difficult.

Consistent with previous research, exploratory analysis in Experiment II revealed that Maze assessment scores were significantly correlated with WCPM scores when participants were instructed to do their best reading in phase one and when participants were instructed to read fast in phase two (Colón & Kranzlers, 2006). If WCPM scores are significantly related to scores on a reading comprehension measure regardless of instructional prompts provided to participants, it may be more important to ensure that all participants receive the same prompts and attempt to complete ORF assessments in the same manner than ensuring that all participants do their "best reading."

Participants included in Experiments I and II were in middle school, mostly Caucasian, and from the Southeastern U.S. RtI programs have been mostly implemented in elementary school settings (Hughes & Dexter, 2011). The effect of instructing students to read fast may differ for younger students and students from differing cultural and SES backgrounds. Future researchers should replicate and extend this research using younger, more diverse samples of students.

The potential effects of other prompts should also be taken into consideration. We assessed the effect of one prompt on ORF assessment scores, directly instructing students to read fast. Other prompts such as administrator characteristics, location of administration, and unknown factors can also impact WCPM and error scores (Derr & Shapiro, 1989; Derr-Minneci & Shapiro, 1992). In the current study, a stopwatch was consistently present during all ORF assessments for all participants. This somewhat subtle prompt may have also inflated

participants WCPM and error scores. The effect of telling participants to read fast may have been less pronounced had participants been unaware of the stopwatch. Future researchers should investigate the effects of having the combination of various prompts present or absent during ORF assessments.

**Reading skill-by-instructions interactions.** The finding from Experiment I, which suggests those with stronger reading skills increased their reading speed more than those with weaker reading skills, has implications for RtI implementation. Encouraging faster reading may be detrimental to weaker readers' WCPM scores and may increase stronger readers' WCPM scores. Based solely on this finding, one might conclude that ORF administrators should discourage students from doing their fastest reading and attempt to minimize prompts that might encourage students to read fast.

Reading skill-by-instructions interactions in Experiment II were non-significant, suggesting that when instructed to read fast, stronger readers did not increase their WCPM or errors more than weaker readers. This finding suggests that readers at all skill levels can increase their WCPM when instructed to read fast. If this is the case, prompting students to read fast may allow for comparisons across scores, as long as all students are consistently completing ORF assessments in the same manner. The significant reading skill group-by-instructions interaction in Experiment I suggests against prompting students to read fast. The discrepant findings regarding the reading skill-by-instructions interaction in Experiments I and II imply that no applied recommendation can be made regarding instructions (i.e., instructing all to read fast or all to read their best). Limitations associated with the studies may provide some insight into these discrepant findings.

In Experiment II, we attempted to improve upon Experiment I by using passages with greater equivalency. Performance by the SSS control group in Experiment II across phases suggests we were successful. Table 17 displays the WCPM means obtained by the control group for passages in pre and posttest during Experiment I. These means suggest that all three passages used during the read-fast phase were more difficult than all three passages used during the pretest phase.

The difficulty level of passages in Experiment I may have accounted for the significant reading skill-by-instructions interaction. Administering more difficult passages during the read-fast phase may have allowed for an interaction between reading skills and instructions. Although administering passages at the appropriate difficulty level is not intuitively a limitation of Experiment II, in attempting to find a relationship between increases in reading speed and reading skills, future researchers may want to administer passages that are well above participants' grade level. If passage difficulty accounts for the interaction found in Experiment I, it is important to ensure passages administered within an RtI framework are at the appropriate difficulty level as students with stronger and weaker reading skills may respond differently if provided with prompts to increase their reading speed.

Students may experience stress when provided with reading material that is well above their reading level. Additionally, the instruction to read fast may enhance stress. During Experiment I, weaker readers may have experienced more anxiety because of the combination of stress associated with more difficult passages and the prompt to increase reading speed. Failure to find a reading skill-by-instructions interaction in Experiment II could be because weaker readers experienced less stress and anxiety when reading one easier and one harder passage in phase two. Future researchers should consider assessing stronger and weaker

Table 17.

*Experiment I WCPM Scores and Standard Deviations by Control Group Participants who Received Standard Instructions for all Phases*

| Phase | Passage | Mean | SD | Instructions Provided to Control Group | Instructions Provided to Experimental Group |
|---|---|---|---|---|---|
| Pretest | 1 | 144.98 | 30.662 | Standard | Standard |
| | 2 | 147.57 | 29.893 | Standard | Standard |
| | 3 | 146.57 | 27.650 | Standard | Standard |
| Posttest | 4 | 140.75 | 28.702 | Standard | Read Fast |
| | 5 | 132.38 | 24.986 | Standard | Read Fast |
| | 6 | 122.97 | 25.642 | Standard | Read Fast |

readers' level of anxiety and stress after completing ORF assessments with easy and hard passage and with instructions to read their best and read their fastest.

**Comprehension measures.** In Experiment II, we attempted to slow the reading speed of participants who were prompted to read fast in phase two by providing instructions to either do their best reading, or do their best reading because they would answer questions in phase three. These instructions worked to significantly decrease the WCPM scores obtained by participants in both groups, suggesting that telling students to do their best reading may be sufficient for prompting reading for comprehension. Although the instruction to do your best reading was effective in slowing participants' reading speed, providing students with comprehension questions after completing ORF assessments may help to reduce the effect of other known and unknown prompts that encourage fast reading.

The effects of other potential prompts that can increase WCPM and error scores were either not present or consistently present in the current study (i.e. stopwatch). Therefore, we could not conclude if providing students with comprehension questions reduces the effect of such prompts to a greater degree than instructing students to do their best reading. Future researchers should investigate the differential effects of telling students to do their best reading and telling students to do their best reading because they will answer comprehension questions under varying ORF assessment conditions. For example, researchers could investigate the effect of explicitly telling students they will be timed when they are required to answer comprehension questions and when they are instructed to do their best reading.

Comprehension measures may also increase the utility of ORF assessments when used for decision-making purposes. Unlike students who complete ORF assessments within an RtI framework, participants in the current study were unaffected by the outcome of their

performance. If students are aware that they will have to continue with intervention services, or could receive intervention services if they read slowly, the instruction to "do your best reading" may be insufficient. Researchers should assess students' understanding regarding how they should complete ORF assessments. If students report that they think they should read fast during ORF assessments, researchers should inquire about why students believe they should read fast (e.g. been instructed to, see a stopwatch). If students attempt to read fast because they believe obtaining higher WCPM scores will prevent them from being placed in intervention services, or will remove them from participating in intervention services, providing comprehension measures could encourage reading for comprehension, perhaps resulting in a more valid measure of reading speed for some students.

**Conclusion**

Findings across Experiments I and II suggest that students increase their WCPM and error scores when instructed to read fast, which has implications for how students are instructed to complete ORF assessments within an RtI framework. Specifically, ORF administrators should ensure that all students are provided standardized instructions and attempt to complete ORF assessments in the same manner. Researchers should investigate how students interpret the instruction to "do your best reading" and other prompts that may result in fast reading.

It is unclear whether there is an interaction with reading skills when students are instructed to read fast. Results from Experiment I suggest that when instructed to read fast, stronger readers can increase their WCPM scores more than weaker readers; however, results from Experiment II suggest students at all reading skill levels can increase their WCPM when instructed to read fast. Passages included in the read fast phase of Experiment I were more difficult than those included in Experiment II, which could account for the different findings.

Also, the stress of being instructed to read fast coupled with the stress caused by harder passages may account for the significant skill level-by-instructions interaction found in Experiment I. Finally, our participants were middle school students who typically have strongly developed reading skills. Future researchers should investigate the effect of stress caused by difficult passages and prompts to increase speed of responding alone and in combination on students with stronger and weaker reading skills.

Findings from Experiment II suggest that after students have been instructed to increase their reading speed, WCPM scores can be decreased by instructing participants to do their best reading or to do their best reading because they will have to answer questions. Because participants in both groups decreased in WCPM, providing students with comprehension measures may be unnecessary to encourage reading for understanding. However, other prompts that are often present during ORF assessments may encourage fast reading, and implementation of comprehension measures may counteract these effects.

To allow for appropriate comparisons of scores, it is important to ensure all students are receiving the same standardized instructions during ORF assessments. Additionally, it is unclear if prompts that result in faster reading differentially affect students with stronger and weaker reading skills. Providing students with the prompt to do their best reading may be sufficient for encouraging reading for comprehension. However, more research on the implementation of comprehension measures during ORF assessments should be conducted as these measures may work to reduce the effects of other prompts.

**List of References**

Abrahamsen, F. E., Roberts, G. C., Pensgaard, A. M., & Ronglan, L. T. (2008). Perceived ability and social support as mediators of achievement motivation and performance anxiety. *Scandinavian Journal of Medicine and Science in Sports, 18*, 810-821.

Ardoin S. P. & Christ, T. J. (2009). Curriculum-based measurement of oral reading: Standard errors associated with progress monitoring outcomes from DIBELS, AIMSweb, and an experimental set. *School Psychology Review, 38*, 266-283.

Baker, S. K., Smolkowski, K, Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Beck, C. T. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37*(1), 18-37.

Bellinger, J. M. & DiPerna, J. C. (2011). Is fluency-based story retell a good indicator of reading comprehension? *Psychology in the Schools, 48*, 416-426.

Brown-Chidsey, R., Davis, L, & Maya, C. (2003). Sources of variance in curriculum-based measures of silent reading. *Psychology in the Schools, 40*, 363-377.

Byron, K., Khazanchi, S., & Nazarian, D. (2010). The relationship between stressors and creativity: A meta-analysis examining theoretical models. *Journal of Applied Psychology, 95*, 201-212.

Cates, G. L. & Rhymer, K. M. (2006). Effects of explicit timing on elementary students' oral reading rates of word phrases. *Reading Improvement, 43*, 148-156.

Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35*, 128-133.

Christ, T. J. & Ardoin, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe set development. *Journal of School Psychology, 27*, 55-75.

Christ, T. J. & Schanding, G. T. (2007). Curriculum-based measurement of computational skills: A comparison of group performance in novel, reward, and neutral conditions. *School Psychology Review, 36*, 147-158.

Christ, T. J. & Silberglitt, B. (2007). Estimates of the standard error of measurements for curriculum-based measures of oral reading fluency. *School Psychology Review, 36*, 130-146.

Christ, T. J., Silberglitt, B., Yeo, S., & Cormier (2010). Curriculum-based measurement of oral reading: An evaluation of growth rates and seasonal effects among students served in general and special education. *School Psychology Review, 39*, 447-462.

Christ, T. J., Zopluoglu, C., Long, J. D., & Monaghen, B. D. (2012). Curriculum-based measurement of oral reading: Quality of progress monitoring outcomes. *Exceptional Children, 78*, 356-373.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.).* Hillsdale, NJ: Erlbaum.

Cohen, J. & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hinsdale, NJ: Erlbaum.

Colón, E. P & Kranzler, J. H. (2006). Effects of instruction on curriculum-based measurement of reading. *Journal of Psychoeducational Assessment, 24*, 318-328.

Cronbach, L. & Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.

Deno, S. L. (2003). Curriculum-based measures: Development and perspectives. *Assessment for Effective Interventions, 28*, 3-12.

Deno, S., L. & Fuchs, L. S. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*, 488-500.

Deno, S. L. & Mirkin, P. K. (1977). *Data-based program modification: A manual.* Reston, VA: Council for Exceptional Children.

Derr, T. F. & Shapiro, E. S. (1989). A behavioral evaluation of curriculum-based assessment of reading. *Journal of Psychoeducational Assessment, 7*, 148-160.

Derr-Minneci, T. F. & Shapiro, E. S. (1992). Validating curriculum-based measurement in reading from a behavioral perspective. *School Psychology Quarterly, 7*, 2-16.

Diamond, D. M., Campbell, A. M., Park, C. R., Halonen, J., & Zoladz, P. R. (2007). The temporal dynamics model of emotional memory processing: A synthesis on the neurobiological basis of stress-induced amnesia, flashbulb and traumatic memories, and the Yerkes-Dodson Law. *Neural Plasticity.* Advanced online publication. doi:10.1155/2007/60803

Dykeman, B. F. (2006). Alternative strategies in assessing special education needs. *Education, 127*, 265-273.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*, 183-201.

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

Espin, C. A. & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children, 62*, 497-514.

Evans-Hampton, T. N., Skinner, C. H., Henington, C., Sims, S., & McDaniel, C. E. (2002). An investigation of situational bias: Conspicuous and covert timing during curriculum-based measurement of mathematics across African American and Caucasian students. *School Psychology Review, 31*, 529-539.

Fletcher, J. M., Denton, C., & Francis, D. J. (2005). Validity of alternative approaches for the identification of learning disabilities: Operationalizing unexpected underachievement. *Journal of Learning Disabilities, 38*, 554-525.

Francis, D. J., Fletcher, J. M., Stuebing, K. K., Lyon, G. R., Shaywitz, B. A., & Shaywitz, S. E. (2005). Psychometric Approaches to the identification of LD: IQ and achievement scores are not sufficient. *Journal of Learning Disabilities, 38*, 98-108.

Freeland, J.T., Jackson, B., & Skinner, C.H. (1999, November). *The effects of reinforcement on reading rate of comprehension.* Paper presented at the twenty-sixth annual meeting of the Mid-South Educational Research Association, Point Clear, AL.

Freeland, J.T., Skinner, C.H., Jackson, B., McDaniel, C.E., & Smith, S. (2000). Measuring and increasing silent reading comprehension rates: Empirically validating a repeated readings intervention. *Psychology in the Schools, 37,* 415-429.

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, D. & Fuchs, L. S. (2012). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*, 93-99.

Fuchs, L. S. & Fuchs, D. (1992). Identifying a measure for measuring student progress. *School Psychology Review, 21*, 45-58.

Fuchs, L. S. & Fuchs, D. (2002). Curriculum-based measurement: Describing competence, enhancing outcomes, evaluating treatment effects and identifying treatment non-responders. *Peabody Journal of Education, 77*, 64-84.

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness to instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27*(4), 216-227.

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

Fuchs, L. S. & Vaughn, S. (2012). Responsiveness-to-intervention: A decade later. *Journal of Learning Disabilities, 45*(3), 195-203.

Gadbois, S. A. & Sturgeon, R. D. (2011). Academic self-handicapping: Relationships with learning specific and general self-perceptions and academic performance over time. *British Journal of Educational Psychology, 81*, 207-222.

Gaeddert, W. P. & Dophin, W. D. (1991). Effects of facilitating and debilitating anxiety on performance and study effort in mastery-based and traditional courses. *Psychological Reports, 48*, 827-833.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading, 5*, 257-288.

Griffiths, A. J., VanDerHeyden, A. M., Skokus, M., & Lilles (2009). Progress monitoring in oral reading fluency within the context of RTI. *School Psychology Quarterly, 24*, 13-23.

Hale, A. D., Henning, J. B., Hawkins, R. O., Sheeley, W., Shoemaker, L., Reynolds, J. R., & Moch, C. (2011). Reading assessment methods for middle-school students: An investigation of reading comprehension rate and maze accurate response rate. *Psychology in the Schools, 48*, 28-36.

Hale, A. D., Skinner, C. H., Wilhoit, B., Ciancio, D., & Morrow, J. A. (2012). Variance in broad reading accounted for by measures of reading speed embedded within Maze and comprehension rate measures. *Journal of Psycho-educational Assessment, 30*, 539-554.

Hale, A. D., Skinner, C. H., Winn, B. D., Oliver, R., Allin, J. D., & Molloy, C. C. M. (2005). An investigation of listening and listening-while-reading accommodations on reading comprehension levels and rates in students with emotional disorders. *Psychology in the Schools, 42*, 39-52.

Hosp, M. K. & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with the grade? *School Psychology Review, 34*, 9-26.

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

Hsu, L. M. (1995). Regression toward the mean associated with measurement error and the identification of improvement and deterioration in psychotherapy. *Journal of Consulting and Clinical Psychology, 63*, 141-144.

Hughes, C. A. & Dexter, D. D. (2011). Response to intervention: A researched-based summary. *Theory into Practice, 50*, 4-11.

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No. 108-446, 118 Stat. 2706 (2004).

Justice, L. M. (2006). Evidence-based practice, response to intervention, and the prevention of reading difficulties. *Language, Speech, and Hearing Services in Schools, 37*, 284-297.

Keller-Margulis, M. A., Shapiro, E. S., & Hintze, J. M. (2008). Long-term diagnositic accuracy of curriculum-based measures in reading and mathematics. *School Psychology Review, 37*, 374-390.

Kim, Y.-S., Petscher, Y., Schatschneider, C., & Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology, 102*, 652-667.

Klein, K. & Beith, B. (1985). Re-examination of residual arousal as an explanation of aftereffects: Frustration tolerance versus response speed. *Journal of Applied Psychology, 70*, 642-650.

Knoff, H. M. & Dean, K. R. (1994). Curriculum-based measures of at-risk students' reading skills: A preliminary investigation of bias. *Psychological Reports, 75*, 1355-1360.

Lars, B. & Molander, B. (1986). Effects of adult age and level of skill on the ability to cope with high-stress conditions in a precision sport. *Psychology and Aging, 1*, 334-336.

McCallum, R. S. Krohn, K. R., Skinner, C. H., Mounger, A., Hopkins, M., Waller, S., & Polite, F. (2011). Improving reading comprehension of at-risk High School students: The ART of reading program. *Psychology in the Schools*, *48,* 78-86.

McDaniel, C. E., Watson, T. S., Freeland, J. T., Smith, S. L., Jackson, B., & Skinner, C. H. (2001, May). *Comparing silent repeated reading and teacher previewing using silent reading comprehension rate.* Paper presented at the Annual Convention of the Association for Applied Behavior Analysis, New Orleans.

McKenzie, R. G. (2009). Obscuring vital distinctions: The oversimplification of learning disabilities within RTI. *Learning Disability Quarterly, 32*, 203-215.

Meyer, M. S. (2000). The ability–achievement discrepancy: Does it contribute to an understanding of learning disabilities? *Educational Psychology Review, 12*, 315-337.

Neddenriep, C. E., Hale, A. D., Skinner, C. H., Hawkins, R. O., & Winn, B. D. (2007). A preliminary investigation of the concurrent validity of reading comprehension rate: A direct, dynamic measure of reading comprehension. *Psychology in the Schools, 44*, 373-388.

Neddenriep, C. E., Skinner, C. H., Wallace, M. A., & McCallum, E. (2009). Classwide peer tutoring: Two experiments investigating the generalized relationship between increased oral reading fluency and reading comprehension. *Journal of Applied School Psychology, 25*, 244-269.

Neiss, R. (1988). Reconceptualizing arousal: Psychobiological states in motor performance. *Psychological Bulletin, 103*, 345-366.

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

O'Donnell, P. S. & Miller, D. N. (2011). Identifying students with specific learning disabilities: School psychologists' acceptability of the discrepancy model versus response to intervention. *Journal of Disability Policy Studies 22*, 83–94.

O'Rourke, D. J., Smith, R. E, Smoll, F. L., & Cumming, S. P. (2011). Trait anxiety in young athletes as a result of parental pressure and motivational climate: Is parental pressure always harmful? *Journal of Applied Sports Psychology, 23*, 398-412.

Peter, B., Matsushita, M., & Raskind, W. H. (2010). Global processing speed in children with low reading ability and in children and adults with typical reading ability: Exploratory factor analytic models. *Journal of Speech Language and Hearing Research, 54*, 885-899.

Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment, 23*, 326-338.

Pressley, M., Hilden, K., & Shankland, R. (2005). *An evaluation of end-grade-3 Dynamic Indicators of Basic Early Literacy Skills (DIBELS): Speed reading without comprehension, predicting little.* Unpublished manuscript, Literacy Achievement: Research Center, Michigan State University, East Lansing, MI.

Reschly, A. L., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology, 47*, 427-469.

Ridge, A. D., & Skinner, C. H. (2011). Using the TELLS pre-reading procedure to enhance comprehension levels and rates in secondary students. *Psychology in the Schools*, *48,* 46-58.

Roberts, G., Good, R., & Corcoran, S. (2005). Story retell: A fluency-based indicator of reading comprehension. *School Psychology Quarterly, 20*, 546-567.

Rhymer, K. N., Henington, C., Skinner, C. H., & Looby, E. J. (1999). The effects of explicit timing of mathematics performance in second grade Caucasian and African American students. *School Psychology Quarterly, 14*, 397-407.

Rhymer, K. N., Skinner, C. H., Henington, C., D'Reaux, R. A., & Sims (1998). Effects of explicit timing on mathematics problem completion rates in African-American third-grade elementary students. *Journal of Applied Behavior Analysis, 31*, 673-677.

Rhymer, K. N., Skinner, C. H., Jackson, S., McNeill, S., Smith, T., & Jackson, B. (2002). The 1 minute explicit timing intervention: The influence of mathematics problem difficulty. *Journal of Instructional Psychology, 29,* 305-311.

Samuels, S. J. (2007). The DIBELS tests: Is speed of barking at print what we mean by reading fluency? *Reading Research Quarterly, 42*, 563-566.

Shapiro, E. S. (2011). *Academic Skills Problems: Direct Assessment and Intervention* (4th ed.). New York: Gilford Press.

Shin, J., Deno, S. L. & Espin, C. (2000). Technical adequacy of the Maze task for curriculum based measurement of reading growth. *The Journal of Special Education, 34*, 164-172.

Shinn, M. R. & Shinn, M. M. (2002a). *AIMSweb training workbook: Administration and scoring of reading curriculum-based measurement (R-CBM) for use in general outcome measurement*. Retrieved from http://www.aimsweb.com.

Shinn, M. R. & Shinn, M. M. (2002b). *AIMSweb training workbook: Administration and scoring of reading Maze for use in general outcome measurement*. Retrieved from http://www.aimsweb.com.

Skinner, C.H. (1998). Preventing academic skills deficits. In T. S. Watson & F. M. Gresham (Eds.), *Handbook of child behavior therapy* (pp. 61-82). New York: Plenum Press.

Skinner, C. H. (2002). An empirical analysis of interspersal research: Evidence, implications and applications of the discrete task completion hypothesis. *Journal of School Psychology, 40,* 347-368.

Skinner, C. H. (2008). Theoretical and applied implications of precisely measuring learning rates. *School Psychology Review, 37*, 309-315.

Skinner, C. H. (2011, May). *Increasing the probability of student engaging in assigned work: The additive interspersal procedure and the discrete task complete hypothesis.* Paper presented at the annual convention of the Association for Applied Behavior Analysis-International, Denver, CO.

Skinner, C. H., Neddenriep, C. E., Bradley-Klug, K. L., & Ziemann, J. M. (2002). Advances in curriculum-based measurement: Alternative rate measures for assessing reading skills in pre- and advanced readers. *The Behavior Analyst Today, 3*, 270-281.

Skinner, C. H., Williams, J. L., Morrow, J. A., Hale, A. D., Neddenriep, C. E., & Hawkins, R. O. (2009). The validity of reading comprehension rate: Reading speed comprehension and comprehension rate. *Psychology in the Schools, 48*, 1036-1047.

Spargo, E. (1989). *Timed Readings* (3rd ed., Book 4). Providence, RI: Jamestown Publishers.

Suinn, R. M. (2005). Behavioral interventions for stress management in sports. *International Journal of Stress Management, 12*, 343-362.

Valencia, S. W., Smith, A. T., Reece, A. M., Li. M., Wixson, K. K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly, 45*, 270-291.

Wankez, J. & Vaughn, S. (2010). Tier 3 interventions for students with significant learning problems. *Theory into Practice, 49*, 305-314.

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

Williams, A., & Skinner, C.H. (2004, November). *Using measures of reading comprehension rate to evaluate the effects of a previewing strategy on reading comprehension*. Paper presented at the twenty-eighth annual meeting of the Mid-South Educational Research Association, Gatlinburg, TN.

Williams, J. L., Skinner, C. H., Floyd, R. G., Hale, A. D., Neddenriep, C., & Kirk, E. P. (2011). Words correct per minute: The variance in standardized reading score accounted for by reading speed. *Psychology in the Schools, 48*, 87-101.

Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review, 105*, 558-584.

## Appendices

## Appendix A

Maze Assessments Integrity Form

| Period | | Read names from Consents | Hand out packets | Read assent form, have students sign and date | Complete demographics as a group. Tell students DO NOT turn page | Inform students to not write their name on the packets again | Read directions BEFORE telling students to turn the page | Tell students to turn the page and begin reading | Start timer | Note completion time for students who finish early | Stop students after 3 min | Record completion time on form of students who finish early |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1st | 1 | | | | | | | | | | | |
| | 2 | | | | | | | | | | | |
| | 3 | | | | | | | | | | | |
| 2nd | 1 | | | | | | | | | | | |
| | 2 | | | | | | | | | | | |
| | 3 | | | | | | | | | | | |
| 3rd | 1 | | | | | | | | | | | |
| | 2 | | | | | | | | | | | |
| | 3 | | | | | | | | | | | |
| 4th | 1 | | | | | | | | | | | |
| | 2 | | | | | | | | | | | |
| | 3 | | | | | | | | | | | |
| 5th | 1 | | | | | | | | | | | |
| | 2 | | | | | | | | | | | |
| | 3 | | | | | | | | | | | |
| 6th | 1 | | | | | | | | | | | |
| | 2 | | | | | | | | | | | |
| | 3 | | | | | | | | | | | |
| Completion time for participants who finish early (time and description): | | | | | | | | | | | | |

## Appendix B

ORF Assessment Integrity Form

| Research Number | Group | Phase 1- standard instructions? | Phase 2- Correct instructions? | Gave passage 2-2 questions to correct participants? | Phase 3- Correct instructions? | Gave 3-1 questions to correct participants? | Gave 3-2 questions? |
|---|---|---|---|---|---|---|---|
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |
| | | | | | | | |

## Vita

Bethany Forbes was born in Lexington, Kentucky and grew up in Richmond, Kentucky. She graduated with a B.S. in Psychology from Eastern Kentucky University in 2009. In 2009, Bethany attended the University of Tennessee's School Psychology Ph.D. Program. She graduated with a M.S. in Applied Educational Psychology from the University of Tennessee in December of 2012. Bethany will receive her Ph.D. in School Psychology in August of 2014 following completion of a year-long internship with Tennessee Internship Consortium in Knoxville, TN.