5-2017

An Examination of the Relationship of Oral Reading Fluency, Silent Reading Fluency, Reading Comprehension, and the Colorado State Reading Assessment

Christy L. Bloomquist
Utah State University

AN EXAMINATION OF THE RELATIONSHIP OF ORAL READING FLUENCY, SILENT READING FLUENCY, READING COMPREHENSION, AND THE COLORADO STATE READING ASSESSMENT

by

Christy L. Bloomquist

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY
in
Education

Approved:

Cindy D. Jones, Ph.D. Major Professor	Susan Turner, Ph.D. Committee Member
Nicole Pyle, Ph.D. Committee Member	Margaret M. Lubke, Ph.D. Committee Member
Sarah K. Clark, Ph.D. Committee Member	Mark McLellan, Ph.D. Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2017

ABSTRACT

An Examination of the Relationship of Oral Reading Fluency, Silent Reading Fluency, Reading Comprehension, and the Colorado State Reading Assessment

by

Christy L. Bloomquist, Doctor of Philosophy

Utah State University, 2017

Major Professor: Cindy Jones, Ph.D.
Department: Teacher Education and Leadership

This study evaluated how measures of oral reading fluency (ORF) and silent reading fluency (SRF) compare as predictors of reading comprehension and how these predictors vary as a function of proficiency level for fourth- and fifth-grade students. Additionally, the study sought to examine the relationship between measures of oral reading fluency, silent reading fluency, reading comprehension, and the Transitional Colorado Assessment Program (TCAP) with these students. Participants were 175 fourth- and fifth-grade students from two randomly selected schools in Colorado. A correlational predictive design was used. Results indicated that measures of ORF and SRF were predictors of reading comprehension and that the relationship of measures of ORF and SRF with comprehension changes over time. Regression analysis results indicated that 45.0% of the variance in reading comprehension was accounted for by the ORF measure for the sample population, as compared to 53.0% of the variance accounted for by SRF

measures. Thus, measures of SRF might be a better predictor for maturing readers to determine reading proficiency, monitor student progress, and guide instructional practices.

A structural equation model (SEM) analyzed the relationship of the measure of SRF with reading comprehension as moderated by proficiency level. Analysis for the SRF measure by reading proficiency was conducted at the whole group level. The model accounted for 59.0% of the moderation. Results indicated that reading proficiency level and the SRF measure were both associated with reading comprehension. Reading proficiency level is a significant moderator of the relationship between measures of reading comprehension and SRF.

A SEM mediation model was used to analyze the relationship of measures of ORF, SRF, reading comprehension, and TCAP. The direct effects of the ORF and SRF measures on TCAP were both predictive with 66.0% of the variance accounted for with SRF measure and 66.5% of the variance accounted for with ORF measure.

Results indicated that as grade level increases, the relationship between measures of ORF, SRF, and reading comprehension changes. Additionally, SRF measures can be a viable alternative to ORF measures for upper elementary students as a predictor of reading comprehension and on the TCAP high-stake assessment.

(151 pages)

PUBLIC ABSTRACT

An Examination of the Relationship of Oral Reading Fluency, Silent Reading Fluency, Reading Comprehension, and the Colorado State Reading Assessment

Christy L. Bloomquist, Doctor of Philosophy

The purpose of this study was to evaluate how measures of oral reading fluency and silent reading fluency compare as predictors of reading comprehension and how these vary as a function of proficiency level for fourth- and fifth-grade students. Additionally, the study sought to examine the relationship between measures of ORF, SRF, reading comprehension, and the TCAP with these students. As silent reading fluency is utilized more in the classroom as grade level increases, a silent reading fluency measure might be a better predictor for maturing readers to determine reading proficiency, monitor student progress, and guide instructional practices.

A correlational prediction design with measures for the variables of ORF, SRF, reading comprehension, and TCAP were used in this study that included 175 fourth- and fifth-grade students from two randomly selected schools in Colorado. Linear regression models were used to analyze the relationship of measures of oral reading fluency and silent reading fluency with reading comprehension. The results indicated that measures of ORF and SRF were predictors of reading comprehension, but the relationship changed as students matured from fourth to fifth grade. Thus, as students progress in grade level, measures of SRF might be potentially a better indicator of students' reading comprehension. A structural equation model (SEM) was used to analyze the relationship

of silent reading fluency measures with reading comprehension as moderated by reading proficiency level. Reading proficiency level is a significant moderator of the relationship between reading comprehension and the SRF measure. A SEM mediation model was used to analyze the relationship of measures of ORF, SRF, reading comprehension, and TCAP. The direct effects of the ORF and SRF measures on TCAP were predictive with 66.0% of the variance accounted for with the SRF measure and 65.5% of the variance accounted for with the ORF measure.

Results of this study indicated that as grade level increases, the relationship between measures of ORF, SRF, and reading comprehension changes. As students progress from fourth to fifth grade, the ORF measure has a higher correlation with reading comprehension for fourth-grade students, while the SRF measure has a higher correlation than the ORF measure with reading comprehension for fifth-grade students. Measures of SRF can be a viable alternative to ORF measures for upper elementary students as predictors of reading comprehension and high-stakes assessment.

DEDICATION

This work is dedicated to my wonderful children, Camden and Dawson, who have supported and encouraged me even when they did not understand the reason behind my efforts. They give me my passion for life and make each and every day a joy. My life would not be complete without each of them.

To one who is closest to my heart, who continually gave subtle reminders and questions to keep me going while on the path to achieve the goal I set for myself. You have taught me to enjoy life to the fullest and never take anything for granted. One life, one chance, one ticket…absorb every moment.

Finally, my parents, without them, none of this would have been possible. The continued unconditional love and support through the years can never be forgotten or replaced.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my chair for allowing me to grow and flourish under her guidance. The support, encouragement, and knowledge provided have been my continued light as I worked toward this goal and accomplishment of my doctoral degree. You have been invaluable through the process and always encouraged the "high road" and always held the highest expectations. I cannot thank you enough.

My committee members have provided the professional encouragement and support throughout the research. Each one has been there to answer any question and provide additional support when needed. You each have a gift that I feel fortunate to have learned from.

I would also like to thank the students for their participation and the school district for allowing me to complete the study. The districts' willingness to support research and improve student achievement is greatly appreciated.

Finally, I would like to thank my children who have watched the process and asked numerous questions along the way. Their patience, love, and encouragement have helped this dream become a reality.

Christy L. Bloomquist

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Literacy is the foundation of learning and considered a discipline rather than a subject (Dole, 2003; Paris & Hamilton, 2009). "The abilities to listen, speak, read, and write are basic to academic success in any language" (Colorado Department of Education [CDE], 2013a, p. 9). Armbruster, Lehr, and Osborn (2001) noted

> …in today's schools, too many children struggle with learning to read. As many teachers and parents will attest, reading failure has exacted a tremendous long-term consequence for children's developing self-confidence and motivation to learn, as well as for their later school performance. (p. ii)

The National Research Council (2000) stated that to achieve reading excellence "requires an understanding of why these disparities exist as well as serious, informed efforts to address them" (p. 5). Given the crucial importance of reading proficiency, laws and mandates have been initiated to increase effectiveness of literacy instruction and to reduce the number of struggling readers nationwide (Ardoin, Witt, & Suido, 2004; California State Board of Education, 2006; Good, Simmons, & Kame'enue, 2001; Hosp & Fuchs, 2005; Kansas State Department of Education, 2010; Valencia et al., 2010). Thus, educators are relying on accurate measures to aid identification of students at risk, monitor student progress, and guide instructional practices (Buly & Valencia, 2002; Goffreda & DiPerna, 2010; Kamil, Afflerbach, Pearson, & Moje, 2011; Pyle & Vaughn, 2012). Data have become the driving force in evaluating program effectiveness and student progress as educators are reviewing and analyzing data on a regular basis (Denton, 2012; Fuchs & Fuchs, 2006; Gersten & Dimino, 2006; Hale et al., 2011). In

fact, educational accountability and high-stakes assessment are at the forefront of educational agenda (Good et al., 2001). The desire to perform well on high-stakes assessment led educators to seek a progress monitoring measure that would sufficiently predict student progress and identify students at-risk of not meeting these set expectations. Educators flocked to the Dynamic Indicators of Basic Early Literacy Skills Oral Reading Fluency (DIBELS ORF), and this measure has become the standard reading progress monitoring tool in schools across the nation (Denton et al., 2011; Riedel, 2007).

In fact, oral reading fluency has long been used as an indicator of reading skills. In the late 1970s, Deno and colleagues at the University of Minnesota worked to create curriculum-based measures which were reliable and valid, simple and efficient, easily understood, and inexpensive (Deno, 1985).

> The primary goal of the research program was to develop measurement and evaluation procedures that teachers could use routinely to make decisions about whether and when to modify a student's instructional program. (p. 221)

Based on the work from the University of Minnesota, in the late 1980s, researchers at the University of Oregon began initial studies on the implementation of Dynamic Indicators of Basic Early Literacy Skills (DIBELS). DIBELS became immediately popular with schools across the nation as a predictive measure and after the National Reading Panel report emphasized subtests that were included with DIBELS (Riedel, 2007). Additionally, DIBELS was free to schools, quick and easy to implement, and a component of the national Reading First Initiative (Goodman, 2006). This was an important component as many other reading assessments were expensive, time consuming to administer, lacked the ability to show growth, and limited instructional

value. After the National Reading Panel (2000) emphasized the importance of fluency and as causal determinant of reading comprehension, fluency became a focus and a key element in many reading programs (Rasinski, Rikli, & Johnston, 2009). Now in its seventh edition, DIBELS ORF measure is the most widely assessment used to monitor student progress of literacy growth (Goodman, 2006; Riedel, 2007; Schilling, Carlisle, Scott, & Zeng, 2007) to monitor student progress of literacy growth. Nationwide, educators placed a heavy reliance on the DIBELS ORF measures to predict student achievement on high-stakes tests. But, do ORF measures adequately reflect the developmental growth and authentic reading of older children? Assumptions have been made that ORF measures will operate with older students as they do with younger children, but these assumptions have not been substantiated (Denton et al., 2011).

Statement of the Problem

Although the DIBELS Oral Reading Fluency measure is widely used in grades one through six, researchers are questioning if this assessment adequately reflects the developmental growth of older children (Denton, 2012). The intended purpose of this assessment is to predict a student's reading proficiency, but it is criticized as being inadequate of measuring reading comprehension (Goodman, 2006; Manzo, 2005) and of focusing more on speed than comprehension (Rasinski, 2006; Riedel, 2007; Samuels, 2007). Further criticism arises in the predictive ability of the ORF measure as McGlinchey and Hixon (2004) and Stage and Jacobsen (2001) found that about a quarter of fourth-grade-level students were classified incorrectly based on their ORF measure for

their ability to pass a high-stakes assessment.

These concerns lead one to wonder if measures of silent reading fluency might be a better indicator of reading proficiency for upper elementary students. Correlations between ORF and reading comprehension decrease for maturing readers (Wagner, 2011). This is likely due to students' transition to silent reading. The transition from oral reading to silent reading begins in late second or third grade and is more firmly established in fourth and fifth grade (Johnson, Pool, & Carter, 2011; Kim, Wagner, & Lopez, 2012; Wagner, 2011)**.** This places an interesting paradox between practice and assessment, which has left measures of silent reading fluency (SRF) for upper elementary students overlooked and understudied (Share, 2008). For older students, valid assessing of reading proficiency, growth monitoring, and identifying of students with reading difficulties seems to necessitate the inclusion of a silent reading fluency measure (Biancarosa & Snow, 2004).

Purpose and Research Questions

Considering the developmental nature of reading and the implementation of appropriate assessments to determine reading proficiency levels, measure student growth, and predict achievement on high-stakes tests, a comparison of the DIBELS oral reading fluency measure with a measure of silent reading fluency for students in fourth and fifth grade is warranted. Specifically, this study addressed the following questions.

1. How do oral reading fluency measures and silent reading fluency measures compare as predictors of reading comprehension for fourth and fifth grade students? This information could prove valuable for educators seeking to identify and administer appropriate assessments for older students.

2. How does the relationship of oral reading fluency measures and silent reading fluency measures as predictors of reading comprehension vary as a function of reading proficiency level? If ORF and SRF measures vary by reading proficiency levels as predictors for reading comprehension, teachers could be provided with the stronger measure for guiding reading instruction.

3. What is the relationship of oral reading fluency measures, silent reading fluency measures, and the high-stakes measure for Colorado students (the Transitional Colorado Assessment Program [TCAP]) for fourth and fifth grade students? Since state assessment results determine status and growth percentiles and performance levels on state accreditation frameworks, this study could potentially inform educations regarding measures of ORF and SRF in relation to TCAP. An overview of these questions and the hypotheses are presented in Table 1.

Significance

This study sought to expand the research base by focusing on measures of ORF and SRF as predictors of reading comprehension at the fourth- and fifth-grade level. This study will help inform educators and assessment developers by examining the measurement predictors in evaluating reading proficiency levels. The results may help educators focus on which assessment type has the strongest relationship to reading proficiency. Reading assessments could then be selected based on the predictive relationship to the desired goals and outcomes of reading instruction.

Definition of Key Terms

For the purpose of this study, the following terms and definitions were used.

Oral reading fluency (ORF) is the ability to orally read "accurately, quickly, expressively, with good phrasing, and with good comprehension" (Rasinski, 2009, p. 4).

Silent reading fluency (SRF) requires "fluent recognition of printed words, ability

Table 1

Research Questions and Hypotheses

Question	Hypothesis/null hypothesis
1. How do oral and silent reading fluency measures compare as predictors of reading comprehension for fourth and fifth grade?	*Null hypothesis:* There will be no difference between measures of oral and silent reading fluency as predictors of reading comprehension. *Alternative Hypothesis 1:* Silent reading fluency measure will be a stronger predictor of reading comprehension than the oral reading fluency measure. *Alternative Hypothesis 2:* Oral reading fluency measure will be a stronger predictor of reading comprehension than the silent reading fluency measure.
2. Does the relationship of oral reading fluency measures and silent reading fluency measures as predictors of reading comprehension vary as a function of reading proficiency level?	*Null hypothesis:* There will be no difference between measures of silent reading fluency and oral reading fluency as a predictor for reading comprehension based on reading proficiency level. *Alternative Hypothesis 1:* Silent reading fluency measure will be a stronger predictor of comprehension than the oral reading fluency measure for students with a higher reading proficiency level. *Alternative Hypothesis 2:* Oral reading fluency measure will be a stronger predictor of reading comprehension than the silent reading fluency measure for students with a higher reading proficiency level.
3. What is the relationship of oral reading fluency measures, silent reading fluency measures, reading comprehension and the Transitional Colorado Assessment Program reading proficiency level for fourth- and fifth-grade students?	*Null hypothesis:* There will be no relationship between oral reading fluency measures, silent reading fluency measures, reading comprehension and the Transitional Colorado Assessment Program. *Alternative Hypothesis 1:* Silent reading fluency measure will have a stronger relationship to the Transitional Colorado Assessment Program reading proficiency than the oral reading fluency measure and reading comprehension. *Alternative Hypothesis 2:* Oral reading fluency measure will have a stronger relationship to the Transitional Colorado Assessment Program reading proficiency than the silent reading fluency measure and reading comprehension.

to process grade-level appropriate sentence structure, knowledge of grade-level-appropriate vocabulary, adequate working memory capacity to process realistic sentences, the ability to make appropriate inferences, and possession of relevant background knowledge" (Wagner, Torgesen, Rashotte, & Pearson, 2010, p. 3) while reading silently.

Reading comprehension is "the process of simultaneously extracting and constructing meaning through interaction and involvement with written language" (Snow, 2002, p. 11).

State assessment will refer to a state mandated assessment that is given to all students at the designated grade levels that measure the students reading ability based on state expectations (CDE, 2014b).

Summary

Many screening measures and reading assessments implemented are insufficient in capturing the varying components of reading (Goffreda & DiPerna, 2010; Samuels, 2007). Buly and Valencia (2002) brought forth the fact that many decisions are being made off of scant data and it is imperative that we consider broadening the scope of assessments used to determine student reading proficiency. It is important for schools to use appropriate assessments for monitoring progress, identifying students at risk, and predicting achievement on high-stakes tests. To help educators make informed decisions about fourth- and fifth-grade students, perhaps we need to rethink the use of measures of ORF as the primary measure as SRF assessments may be a better measure for older

students because ORF has been shown to have a weak correlation with reading success for older students.

CHAPTER II

REVIEW OF THE LITERATURE

Currently, there is a heavy reliance on measures of ORF for assessing and monitoring the reading proficiency of fourth- and fifth-grade students. However, research has shown correlations between ORF and reading comprehension decrease for maturing readers (Denton et al., 2011; Jenkins & Jewell, 1993; Silberglitt, Burns, Madyun, & Lail, 2006; Valencia et al., 2010; Yovanoff, Duesbery, Alonzo, & Tindal, 2005). Assessment of silent reading fluency might help educators effectively monitor students' progress in a manner that more closely resembles the expectations of the development of silent reading proficiency by fourth- and fifth-grade students. A better understanding of the relationships among measures of oral and silent reading fluency and reading comprehension for grades 4-5 students and the use of fluency scores to identify students at risk for failure on a high-stakes reading test is needed.

Therefore, the purpose of this literature review was to analyze and synthesize previous research related to the use of appropriate assessments to identify students at risk and to predict achievement on high-stakes outcome tests for fourth- and fifth-grade students. Objectives for this literature review were as follows.

1. To describe the current state of research regarding the assessment-intervention connection and the importance of appropriate measures to guide instruction.
2. To describe the use of oral and silent reading fluency measures to identify students at risk and to predict achievement on high-stakes outcome tests.
3. To compare the use of oral reading fluency measures with silent reading fluency measures for the identification of students at risk and prediction of achievement on high-stakes outcome tests.

4. To formulate conclusions based on the current research to guide the focus and design of this study.

Locating the Studies

This review of the literature included a search of the following data bases: Academic Search Premier, CQ Researcher, EBSCO Host, Education Full Test, ERIC, Professional Development Collection, PsychINFO, Psychology and Behavioral Sciences Collection, and Web of Science. The following descriptors were used for these searches: assessment-intervention connection, selecting appropriate measures to identify fourth- and fifth-grade students at risk in reading, oral fluency + reading comprehension, silent reading fluency + reading comprehension, oral fluency + state reading assessment, and silent reading fluency + state reading assessment. As articles were retrieved, reference lists were searched for additional sources.

Research included in the review of the literature meet the following criteria.

1. published in peer-reviewed journals
2. conducted after 1980
3. conducted in the United States with students in grades one through eight
4. focused on general education students
5. minimum number of participants no less than 20

Overview of Research

Importance of Assessment to Guide Instruction

Assessment has two main purposes in schools: a legal aspect and instructional

decision-making (Coburn, Pearson, & Woulfin, 2011; Perie, Marion, Gong, & Wurtzel, 2009). Laws and mandates require students to meet state proficiency levels in reading as determined through high-stakes outcome assessments to meet accountability policies (O'Reilly, Sabatini, Bruce, Pillarisetti, & McCormick, 2012**).** While some of these laws and mandates are not new, they have been updated and reintroduced to help assure the academic reading success of all children. NCLB was introduced in 2001 as an update of the Elementary and Secondary Education Act (ESEA) that was signed into law in 1965 (U.S. Department of Education, 2015). NCLB was signed into law with the intent that "all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments" (No Child Left Behind (NCLB), 2002, sect. 1001). NCLB required large-scale summative assessments in reading and math for every grade from third to eighth and once in high school (NCLB, 2002; Perie et al., 2009). The reading and math assessments were aligned to state standards and assessed students at a given point in time on the content knowledge and skill as defined by the proficiency levels. The assessments were disaggregated by recognized subgroups such as ethnicity and socioeconomic status (SES) to identify possible achievement gaps (Ardoin et al., 2004; Duncan, 2009; Kamil et al., 2011). One of the downfalls of NCLB was that every state was allowed to set their own bar for proficiency with the result that all states were measuring their outcomes differently (Duncan, 2010). Additionally, with the difference in content and performance standards in individual state tests, screeners used for one state test may not apply to another state test (Jenkins, Hudson, Johnson, 2007).

ESEA is being reauthorized and will be known as Every Student Succeeds Act after it is signed by the president (Association for Supervision and Curriculum Development [ASCD], 2015). The Act will start in the 2017-18 school year. With this reauthorization, states will still be required to conduct large-scale summative assessments but will have leeway with the accountability system. Identifying low performing schools and focusing on students not meeting proficiency will still be a priority to assure that students who are not meeting expectations are provided additional support.

The Individual with Disabilities Education Implementation Act (IDEA) was a second mandate that focused on student outcomes. IDEA mainly focused on intervention and students with disabilities. The updated act of IDEA in 2004 and 2008 from the origin in 1975 supported the implementation of NCLB and early intervention of at-risk students (U.S. Department of Education, Office of Planning, Evaluation and Policy Development 2011). This law was closely aligned to the NCLB requirements (National Assessment of IDEA Overview, 2011; U.S. Department of Education, 2007). IDEA also included the assessment of all students either to the state assessment or an alternative assessment for grades 3 through 8 and once in high school to assure that no students were being excluded from testing (U. S. Department of Education, Office of Planning, Evaluation and Policy Development, 2011).

As a result of the need for accountability that is in place with ESEA and IDEA laws, high-stakes assessment has become a major part of the education system in the United States (Chappuis & Chappuis, 2002; Deno, 1985; Hall, 2006; Stage & Jacobsen, 2001). State assessment results are used to determine where students are in reaching the

set levels of proficiency as outlined by ESEA and IDEA and to help identify students who may be at risk of falling behind in reading or lacking skills to advance a grade level. State assessments are used to monitor growth and improvement of all students (Colorado Department of Education, 2008; Hardcastle & Justice, 2006).

While laws and mandates require assessments for accountability, teachers and principals use assessments to guide instruction and differentiate lessons based on the results (Hamilton et al., 2009; Kerr, Garvin, Heaton, & Boyle, 2006; Kim et al., 2010). Educators need an accurate representation of student progress and achievement (Buly & Valencia, 2002; Hall, 2006). Recognizing the importance of assessment to guide instruction, professional educational associations collaborated and created assessment standards to ensure that assessment results would benefit students. These seven standards were created in 1990 and prior to the reauthorization of ESEA or IDEA by the American Federation of Teachers, National Council on Measurement in Education, and the National Education Association (AFT, NCME, & NEA, 1990). The seven standards are:

1. Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2. Teachers should be skilled in developing assessment methods appropriate for instructional decision.
3. Teachers should be skilled in administering, scoring, interpreting the results of both externally produced and teacher-produced assessment methods.
4. Teachers should be skilled in using assessment results when making decision about individual students, planning, teaching, developing curriculum, and school improvement.
5. Teachers should be skilled in developing valid pupil grading procedures that use pupil assessments.
6. Teachers should be skilled in communicating assessment results to students,

parents, other lay audiences, and other educators.

7. Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.

The standards were intended to help guide teachers in the classroom and provide a foundation as a basis for their classroom practices.

There are several types of assessments used to help outline the learning of students and provide feedback on what the student has learned. Some of the common types are summative, interim, and formative (Bulkey, Olah, & Blanc, 2010; Perie et al., 2007. Summative assessments capture the learning over a certain period of time which can be a quarter, unit, midterm, final, or year and measure the learning at a point in time (Popham, 1999; Stiggens, 2004). This type of assessment is considered an "assessment *of* learning" (Stiggens, 2004) and is often used for literacy accountability such as high-stakes state reading assessments and is the roadmap of academic needs (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2014). Summative assessments help identify students who are proficient at the skills assessed but provide little information on sub-skills. Interim assessments are able to provide critical data on sub-skills that the student has not met to the proficiency level set (Bennett, 2011; O'Reilly et al., 2012). Interim assessments are done occasionally throughout the school year over a long period of time and resemble summative assessments. Defined by Perie et al. (2009), interim assessments:

> (1) are to evaluate students' knowledge and skills relative to a specific set of academic goals, typically within a limited time frame; and, (2) are designed to inform decisions at both the classroom and beyond the classroom level, such as the school or district level. (p. 1)

Examples of interim assessments include benchmarks, chapter tests, essays and unit

projects. These assessments evaluate what a student has learned during the allotted time. Interim assessments are used to help predict the outcome of summative, high stakes assessments (Bulkey, Olah, & Blanc, 2010; Datnow & Hubbard, 2015; Jenkins et al., 2007; Perie et al., 2009).

Formative assessments are shorter in length and provide frequent information on how students are performing on a frequent basis and promote learning along the way (Chappuis, 2009; Good, 2011; Heritage, Kim, Vendlinski, & Herman, 2009; Stiggens & Knight, 1997). Through the use of formative assessments "...evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers, to make decisions about next steps in instruction..." (Wiliam, 2011 p. 24). Formative assessments are intended to help teachers and students understand what learning targets have been met (Brookhart, 2011; Bulkey et al., 2010). This type of assessment focuses on improving learning and the process of learning (Ardoin et al., 2004; Buffum, Mattos, & Weber, 2009; Jenkins & Jewell, 1993; Ruiz-Primo, 2011; Wiliam, 2011) and has a narrow scope of learning targets (Perie et al., 2009). These assessments may include class discussions, homework, demonstrations of learning, exit tickets, quizzes, progress monitoring, ongoing observations, and presentations. Implementing formative assessment is a powerful tool for guiding literacy instruction and for identifying students at risk of not meeting set reading proficiency levels (Bennett, 2011; Shepard, 2009).

A variety of assessments are used to help identify students who are at risk of not meeting high-stakes assessment targets (Jenkins et al., 2007). But, the underlying purpose of assessment for teachers is to help guide instruction (Hamilton et al., 2009; N. R.

Hoover & Abrams, 2013; Keller-Margulis, 2012; National Center on Response to Intervention [NCRTI], 2010; U.S. Department of Education, Office of Planning, Evaluation and Policy Development, 2011; Ysseldyke, Burns, Scholin, & Parker, 2010). The data obtained from the classroom and state assessments are used to judge or evaluate the progression of instruction in the classroom and provide specific support for each student that will result in the student reaching the proficiency level set by the state (Datnow & Hubbard, 2014; Heritage, 2010). However, some teachers do not believe that state assessments results provide adequate information to inform instruction (Shepard, 2000) and this data alone is not useful (Datnow & Hubbard, 2014).

Students are provided interventions if they are not meeting grade level expectations or acceptable levels of reading progress and are considered to be at risk. However, one measure alone cannot meet all the expectations outlined (Lipson, Chompsky-Higgins, & Kanfer, 2011). Combining assessment types such as summative and formative that both link to instructional goals is beneficial (Brookhart, 2010). Multiple measures are needed to identify areas of concern (Lipson et al., 2011; Mandinach, Gummer, & Muller, 2011). A variety of assessment types and framework are needed to provide a clear picture of a student's reading performance and to help guide instruction.

Shift to Response to Intervention/Multi-Tiered System of Supports

A component of IDEA helped schools transition to RtI that was a method of monitoring student's response to instruction according to Fuchs, Fuchs, and Compton

(2012). The National Center on Response to Intervention (2010) created the following definition of RtI:

> Response to intervention integrates assessment and intervention within a multi-level prevention system to maximize student achievement and to reduce behavioral problems. With RtI, schools use data to identify students at risk for poor learning outcomes, monitor student progress, provide evidence-based interventions and adjust the intensity and nature of those interventions depending on a student's responsiveness, and identify students with learning disabilities or other disabilities. (p. 2)

RtI provided the ability to intervene early based on assessments results and included support along the way as the central component of the framework and helped prevent reading difficulties (Lipson et al., 2011). The prior method of waiting until a student failed has been changed to an "all students can succeed" model (Gersten & Dimino, 2006; Hardcastle & Justice, 2006; O'Reilly et al., 2012; Zvoch & Stevens, 2011). "RtI is a process that incorporated both assessment and intervention so that immediate benefits come to students" (Mesmer & Mesmer, 2008, p. 287).

RtI has been implemented differently around the country. Although three tiers of intervention are the most common model, as few as one and many as nine tiers have been implemented, with tiers meaning different services (Barnes & Harlacher, 2008; Fuchs et al., 2012). RtI provides layers of support through the tiers but the intervention and monitoring of students increase at each level (Al Otaiba et al., 2014; Hughes & Dexter, 2011; Vaughn, Wanzek, Murray, Linan-Thompson, & Woodruff, 2009). The three most common tiers include general core instruction with universal screening (Tier 1), intervention instruction with progress monitoring (Tier 2), and intense, small group interventions (Tier 3) prior to placement in special education (Barnes & Harlacher, 2008;

Nelson & Machek, 2007; Vaughn, Wanzek, Woodruff, & Linan-Thompson, 2007). Student instruction and intervention becomes more individualized, intensive, and frequent as students move through the RtI framework (J. J. Hoover & Lover, 2011; Keller-Margulis, 2012).

While RtI provided academic support for all students to reach set levels of proficiency, additional components were needed to meet the social/emotional needs of students. As a result, a new framework identified as Multi-Tiered System of Supports (MTSS) combined the RtI framework and the Positive Behavioral Interventions and Supports along with a variety of other support systems (Averill & Rinaldi, 2011; Florida Department of Education, 2015 Metcalf, 2012). By combining these frameworks, academic as well as social/emotional needs can be identified. MTSS is "a whole school, prevention-based framework for improving learning outcomes for every student through a layered continuum of evidence-based practices and systems" (CDE, 2014a). MTSS includes essential elements of shared leadership, data-based problem solving and decision making, layered continuum of supports, evidence-based instruction, instruction and intervention, assessment practices, universal screening and progress monitoring, and family engagement.

The MTSS framework includes monitoring all students through an interim measure to determine an initial level of reading proficiency with a focus on learning needs. The measure connects students to the MTSS framework known as Universal Support (Tier 1) and identifies students before they start to fail (O'Reilly et al., 2012). This interim measure is considered a screening for all students and is generally conducted

three times a year (Ysseldyke et al., 2010) to identify students potentially at risk for not meeting set expectations or grade level reading skills (Jenkins et al., 2007). Students who do not meet established cut points are assessed with additional measure to determine specific areas of need and help guide next steps for instruction. Depending on the need and outcome of the additional measures, students may continue to be provided instruction by the classroom teacher or may move to Targeted Support (Tier 2) based on lack of progress and low rate of growth (Hughes & Dexter, 2011). The bottom 20% of Universal students are considered candidates for Targeted Support (Metcalf, 2012).

At Targeted Support, intervention based on assessment results and the student's needs are implemented. With Targeted Support comes progress monitoring, a type of formative assessment, of specific skills where the student is assessed frequently on a dynamic measure that will help determine if the student is responding to the instruction being provided (Keller-Margulis, 2012; Mesmer & Mesmer, 2008; Vaughn & Fletcher, 2012; Ysseldyke et al., 2010). The frequency of progress monitoring varies from one time a month (Gersten et al., 2008) to weekly or biweekly (Fuchs & Fuchs, 2006; Mesmer & Mesmer, 2008; Vaughn et al., 2009). If the intervention is effective, student scores will quickly increase and the student will continue to receive the services that are provided until assessment results determine that the student no longer needs additional support. If the assessment results do not show adequate growth in the determined amount of time, intervention services may change to a different learning outcome. Progress monitoring will continue to assure the student is making adequate growth based on the new intervention. With the focus on prevention and intervention support, a majority (90-95%)

of students should meet grade level expectation before reaching Intensive (Yell, 2004).

If the student is still not making progress, the student is moved to Intensive Support (Tier 3) support services where the student receives small group, intensive intervention that is focused on specific learning concepts based on assessment results (Vaughn et al., 2009). At this point, multiple measures are used to determine appropriate placement and support for the student and progress monitoring can be weekly to twice weekly (Ysseldyke, Burns, Scholin, & Parker, 2010). The purpose of intensive varies in some frameworks. Continual progress monitoring based on the specific reading need and data analysis is done through the MTSS framework based on a determined schedule and analyzed regularly to monitor the growth of the student and predict outcomes on state reading assessments.

An important component of MTSS may be the wealth of data obtained through assessments and progress monitoring (Hardcastle & Justice, 2006; Zvoch & Stevens, 2011), but educators must know what to do with the data or it is meaningless (Goodwin, 2013/2014). According to Blanc et al. (2010), there are four basic steps in using data to inform instruction: (1) data needs to be organized and assessed; (2) problems and solutions need to be identified; (3) the intervention needs to be determined and implemented; and (4) assess the intervention and modification needs to be identified. These steps are cyclical and continue as needed until the student reaches set expectations. The data gathered needs to focus on what is learned by the student, not what the teacher taught, and is essential for the reading progress of students (CDE, 2014a; Duncan, 2010; U.S. Department of Education, 2011). The use of the data obtained is part of the

instructional improvement cycle which includes collaborative conversations and multiple data sources (Bocala & Boudett, 2015; Mandinach et al., 2011; U.S. Department of Education, Office of Planning, Evaluation and Policy Development, 2011). “To educators, the wrong data can often be seductively appealing. But the right data will, in fact, help teachers do a better job with students. Those are the data we need” (Popham, 2003, p. 49). Educators take the data gathered from summative, interim, and formative assessments, and modify instruction based on the interpretation of the growth the student is making (Hamilton et al., 2009). Given the importance of valid assessments to monitor student progress, and predict outcomes on high stakes tests, identification of appropriate assessments is essential.

Oral Reading Fluency

What is Oral Reading Fluency?

According to the National Reading Panel (2000), as text becomes more complex, fluency is influenced by students’ knowledge about how sentences are constructed and put together to make meaning. But what exactly is fluency? Definitions in the literature vary. Teachers use the terms “automaticity” and “fluency” to indicate the same phenomenon but automaticity is a component of fluency. Armbruster et al. (2001) extended the definition of fluency to include automaticity. They defined automaticity as fast, effortless word recognition without reading with expression, but asserted that fluency is the bridge between word recognition and comprehension. Applegate, Applegate, and Modla (2009) defined fluency as “an indicator of speed, accuracy, and

prosody in oral reading" (p. 513). Rasinski (2003) defined fluency as "the ability to read quickly and accurately with appropriate and meaningful expression" (p. 16). He updated the definition (Rasinski, 2009) to "the ability to read accurately, quickly, expressively, with good phrasing, and with good comprehension" (p. 4). For this study, oral reading fluency will be defined as the ability to read with accuracy and automaticity (reading rate) as indicators of reading comprehension.

In the 1980s, fluency was described as one of the "most neglected" reading skills (Allington, 1983). LaBerge and Samuels (1974) hypothesized that if readers did not have automaticity in word recognition then comprehension would be affected. Numerous studies have concluded that an increase in fluency results in increased reading comprehension (Schwanenflugel et al., 2006; Wayman, Wallace, Wiley, Ticha, & Espin, 2007). Comprehension is the intended outcome of reading and is considered a critical academic skill (Hosp & Fuchs, 2005; Shinn & Good, 1992). In the following paragraphs, research about the relationship of oral reading fluency and comprehension is examined and gaps in the literature are identified.

Oral Reading Fluency and Comprehension

Research suggests that a relationship exists between fluency measures and comprehension. Fluent reading places a cognitive demand on students for both decoding and comprehension (LaBerge & Samuels, 1974). If students are focusing on decoding, there is limited ability to focus on comprehension. After a student has mastered the task of decoding, comprehension becomes the task at hand for readers. Shinn and Good (1992) noted that "decoding affects comprehension; comprehension does not affect

decoding" (p. 2). Decoding becomes the bridge to comprehension (Rasinski, 2003).

While research has provided mixed results about the relationships between measures of ORF and comprehension, some studies have indicated a positive relationship between ORF measures and comprehension. Research has shown variance in reading comprehension as a result of ORF rates in the primary grades (Ardoin et al., 2004; Jenkins & Jewell, 1993). DIBELS subcomponents and GRADE were used to determine the relationship of ORF measures and reading comprehension with primary grade students in a study conducted by Riedel (2007). The end of first grade ORF measure predicted reading comprehension status with 80% accuracy for first grade and 71% accuracy at the end of second grade. Riedel's results fell within the range of 0.67 for first grade and 0.54 for second grade, which indicates that the ORF measure is an accurate predictor of reading comprehension for first and second grade. Correlations for ORF measures and comprehension reported by Cook (2003) were slightly higher for first graders with 0.73. Salvador, Schoeneberger, and Tingle (2009) supported these findings for third-grade students with a moderate relationship of 0.66 between ORF measures and reading comprehension.

There is some evidence supporting the importance of ORF measures and reading comprehension with older students. Hosp and Fuchs (2005) conducted a study with 310 students in grades one through four from four different schools in which they administered ORF passages (Fuchs & Deno, 1992) and the Woodcock Johnson Mastery (Woodcock, 1987) subtests of word attack, word identification, passage comprehension, basic skills and total reading-short. For students in grades 1-3, ORF measures had a

stronger relationship with word reading than with comprehension. For students in fourth grade, a stronger relationship between ORF measures and comprehension emerged over word reading. Hosp and Fuchs (2005) indicated a need for replication studies with norm-referenced reading tests and inclusion of higher grade levels.

Third- and fifth-grade students were given a battery of assessments in a study by Shinn and Good (1992). In the study, 364 students received 96% of their instruction in the general education classroom. Measures of ORF were examined to determine if it was a good indicator of reading proficiency in relation to phonetically regular and regular nonsense words, literal comprehension, inferential comprehension, cloze items, and written retell. It was determined that ORF measures provided an estimate of reading comprehension for third and fifth grade levels.

At the fourth-grade level, and as an extension of the National Assessment of Education Progress (NAEP), 1,136 fourth-grade students read an ORF passage and completed a reading comprehension measure. Fifty-five percent of the students were considered to be fluent as rated on a four-point scale as being a three or four (Pinnell et al., 1995). As the fluency rate increased, the reading proficiency also increased reinforcing the link between reading fluency and reading comprehension.

In a small scale study of 51 elementary students (grades 4-5) and 42 secondary students (grades 10-12) identified as skilled readers, Hale et al. (2007) reported students at both levels answered more comprehension questions correctly after reading a passage aloud than after reading a passage silently. These results indicated that comprehension was enhanced through oral reading. Unfortunately, this study failed to monitor student

completion of the silent reading passage and thus, the researchers "had no way of knowing if the student was actually reading and/or read the entire passage" (p. 17).

Over the past two decades, the result of some research has indicated a positive relationship between measures of ORF and reading comprehension for elementary students. Typical coefficients from these studies for ORF measures and comprehension have ranged from 0.60 to 0.90 for kindergarten through 3rd grade students. However, it has not been determined how this relationship varies as a function of grade levels as other studies have suggested negative or weak relationships between ORF measures and reading comprehension for maturing readers. Valencia et al. (2010) found that as the proficiency level of readers increased, the correlation between words read correctly per minute and comprehension decreased in the study that included second, fourth, and sixth graders. The study included the Iowa Test of Basic Skills for reading comprehension and oral reading passages with comprehension questions and a modified NAEP prosody rubric. The variance accounted for 23%-30% of the correlation. In a study of oral reading fluency measures conducted with students in grades 2-6, Jenkins and Jewell (1993) administered the Gates-MacGinitie Reading Test, the Metropolitan Achievement Test (MAT), and a maze passage task. The study revealed a negative trend across grade levels for correlations between measures of ORF and reading comprehension as measured by the two achievement tests. Correlations declined from 0.86 in fourth grade to 0.67 in sixth grade for ORF measures and Group Reading Assessment and Diagnostic Evaluation (GRADE) and from 0.87 in second grade to 0.60 in sixth grade for ORF measures and MAT. Jenkins and Jewell noted that starting at the intermediate grades, oral reading

fluency measures no longer reflects growth in reading proficiency. Expanding the grade levels and looking at students in grades 2-10, Applegate et al. (2009) had students read two passages. Students read one passage orally and one passage silently. After reading each passage, the students were given 10 open-ended comprehension questions and did a retell. One third of the students who were perceived as fluent, high readers, as identified by parents or teachers as strong readers, were not able to answer text based comprehension questions correctly as indicated by the Critical Reading Inventory and high-level comprehension questions. This indicates that ORF measures alone cannot determine a student's comprehension proficiency level.

Torgesen, Nettles, Howard, and Winterbottom (2003) revealed that the importance of ORF measures as an aid to reading comprehension varies by grade for 4th, 6th, 8th, and 10th-grade students. The results of this study found different indicators had the strongest relationship between ORF measures and reading comprehension. In fourth grade, the relationship between ORF measures and maze on the Florida Comprehensive Assessment Test (FCAT) was equal; sixth-grade results indicated that maze had the strongest relationship with FCAT; eighth-grade had maze and the Test of Sentence Reading Efficiency (TOSRE) equal on the relationship with FCAT; 10th-grade had TOSRE with the strongest relationship with FCAT. Yovanoff et al., (2005) emphasized the cliché of when children are learning to read then reading to learn. From their study that included 5,973 students in fourth to eighth grade measuring ORF, vocabulary, and reading comprehension, fluency was found to be less important for grades five through eight than grade four. They characterized grade four as "a pivotal grade, where we

anticipate different regression coefficients than for later grades" (Yovanoff et al., 2005, p. 9) with ORF measures, vocabulary, and reading comprehension. From their research, fourth grade is when teachers transition from oral reading competency to independent reading. When a student transitions from fourth to fifth grade, does the ORF measure become less important due to the student relying on a different form of fluency?

Denton et al. (2011) noted a weaker relationship between ORF measures and reading comprehension for sixth- to eighth-grade students than for primary grade students. Silberglitt et al. (2006) and Torgesen el al. (2003) reported correlations similar to Denton's for 6th to 8th grade students of 0.50 and 0.60 for ORF measures and reading comprehension. Silberglitt et al. emphasized that ORF measures accounted for 50.4% of the variance in comprehension scores at the third-grade level but only 26% at the eighth-grade level. The researchers indicated that additional studies, which include ORF measures, were needed and generalizations from findings in younger grades are not always appropriate.

Studies have examined the relationship between measures of ORF using a variety of probes and reading comprehension. The results are mixed and do not clearly identify the relationship of ORF measures with reading comprehension for maturing readers. It is unclear how ORF measures and reading comprehension correlate at the fourth- and fifth-grade level. Researchers have emphasized the need for future studies of measures of oral reading fluency in relationship to other assessments to more clearly define the relationship of ORF and reading comprehension at various grade levels (Denton et al., 2011; Hale et al., 2007; Salvador et al., 2009).

How is Oral Reading Fluency Measured?

Oral reading fluency is generally assessed individually with a student reading a grade level passage for one minute. As a student reads, the educator documents "words pronounced incorrectly, substitutions, and omissions as these are all considered errors" (Hall, 2006, p. 251). The number of errors is totaled and this is subtracted from the total words read. To assure an accurate reflection of a student's reading fluency, the child reads three one-minute timed passages and the median score from the three passages is recorded. Accuracy is noted by the percentage of words read correctly. Automaticity is measured by the number of words read per minute. Reading proficiency level is based on set cut scores of accuracy and automaticity to help identify the level of instruction that aligns with the student's reading level. Oral reading fluency passages are seen as a quick and easy way to screen and monitor student growth in reading proficiency.

The most widely used ORF measure is DIBELS and is part of the Reading First initiative which some feel has helped DIBELS gain its popularity (Riedel, 2007). However, many teachers have concerns about the fluency measures (Riedel, 2007; Shinn & Good, 1992). Teachers question if the measure still needs administered once a student reaches a set level of fluency (Salvador et al., 2009). Shinn and Good noted that measures of ORF have a face validity concern with teachers.

With the DIBELS ORF assessment, comprehension is assessed through a retell of the passage read. Retell is generally given if a student has read at least 40 words in the 1 minute allotted time frame (Good et al., 2011). After reading the passage, the student is asked to tell about what was read. The administrator marks the number of words that

relate to the passage as the student retells what was read and marks the final score after one minute. The quality of the response is rated using a rubric**.** The rubric has four levels which all require the reader to include details from the passage in the retell. Level one is providing two or fewer details and level two is providing three or more details. Level three needs three or more details in a meaningful sequence and level four requires three or more details in a meaningful sequence and captures the main idea (Good et al., 2011). The number of words retold and the quality of response are generally recorded and used for comprehension instruction.

However, DIBELS retell is often criticized for not being an accurate indicator of reading comprehension (Goodman, 2006; Manzo, 2005). With both proponents and critics agreeing that comprehension is the intended outcome of reading (Good et al., 2011; Goodman, 2006), the subtests within DIBELS need to have a strong relationship with comprehension. If the relationship does not exist and students are misidentified as needing support in reading comprehension support, valuable instructional time is wasted; but if there is a relationship, students can be identified when additional support is warranted. Critics also believe that the DIBELS ORF focuses on speed reading and not comprehension of what is being read (Pressley, Hilden, & Shankland, 2005). Samuels (2006) argued that decoding and comprehension occur at the same time which is not done simultaneously with DIBELS ORF. It is unclear how closely reading comprehension is related in DIBELS with ORF and the subtests.

Theoretical Foundation

Over the past several years, there has been widespread use of ORF assessment in

elementary grades (Riedel, 2007). The theoretical framework behind the use of fluency measures stem from the work of LaBerge and Samuels (1974), which stated that readers have limited cognitive resources available for any given task at one time; building on this theory, measures of ORF are viewed as the bridge to comprehension (Rasinski, 2003). As students become more proficient with word recognition (as measured by ORF), cognitive resources are freed for comprehension and higher level processing of text. Therefore, ORF measures for students in grades 1-3 tend to make sense as beginning readers focus heavily on word recognition but, as students become proficient readers, their cognitive resources switch to comprehension (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Schilling et al., 2007; Wagner, 2011). This would suggest a drop in oral reading fluency rate as comprehension becomes the focus over speed (Salvador et al., 2009; Silberglitt et al., 2006). However, ORF measures generally promote speed as the goal with comprehension as an afterthought (Goodman, 2006; Pressley et al., 2005; Samuels, 2007). The dependence on ORF provides an incomplete picture of reading proficiency.

Theoretically, SRF measures might be a better predictor of reading growth and proficiency for maturing readers because "Everyday academic tasks require proficient silent reading skills" (Price, Meisinger, & Louwerse, 2012, p.10). Additionally, measures of SRF better reflect instructional practice because as students progress through the grade levels, silent reading becomes the prominent form of reading in the classroom. Students have natural opportunities in the classroom to read silently (Hiebert, Samuels, & Rasinski, 2012). "Skilled children and adult readers rarely read aloud" (Price et al., 2012, p. 1), but emphasis is still placed on oral reading measures even though this is not what

happens daily with literacy activities. In fact, SRF measures may better serve as an indicator for reading comprehension than ORF measures.

Proficient readers read faster silently and use it as their primary mode of reading (Wagner, 2011). As students read silently, reading rates exceed that of oral reading rates by 30% (Hasbrouck & Tindal, 2006) which could partially account for why many state reading assessments administered to students in the upper grades have the students reading silently. State reading outcomes for proficiency levels are considered evidence that a student has learned the standards and expectations set before them. However, there is concern over the lack of compatibility between the wide use of ORF measures as a predictive assessment for all grade levels and state reading assessments which require students to read silently (Schilling et al., 2007; Salvador et al., 2009). A theory noted by Juel and Holmes (1981) indicated that comprehension is impacted differently when reading orally and reading silently. Kragler (1995) indicated that the mode of reading may have differential effects on comprehension. Therefore, a SRF measure could potentially be a more accurate predictor of a state reading assessment proficiency level as the two measures use the same reading mode. Measures that are more compatible to the high stakes assessment would provide information for instructional decisions. As noted by Wagner (2011), "the lack of attention to silent reading fluency may reflect the assumption that silent reading fluency may develop naturally from oral reading fluency, and are manifestations of the same underlying reading skill" (p. 2).

The purpose of the present study is to investigate the extent to which measures of oral and silent reading fluency compare as predictors for students in grades 4 and 5 and

determine if this varies as a function of grade level or reading proficiency level. Fourth and fifth grade are important transition grades for silent reading and comprehension as there is a greater demand on high-level reasoning and inferencing (Denton et al., 2011).

Silent Reading Fluency

What is Silent Reading Fluency?

Johnson et al. (2011) defined SRF as "the ability to simultaneously decode and comprehend" (p. 51). SRF goes beyond the ability to read words in one's head. It includes the ability to decode and comprehend what is being read (Griffith & Rasinski, 2004; Samuels, 2006). SRF requires students to monitor the meaning of the passage (Torgesen et al., 2003) and requires students to use their foundational knowledge as they read (Applegate et al., 2009; Fuchs et al., 2001; Hiebert et al., 2012). The definition of SRF is not differentiated from the ORF definition with the exception that prosody can't be measured with SRF (Rasinski, Samuels, Hiebert, Petscher, & Feller, 2011). However, ORF is related to speed of speech production while silent reading fluency is related to capacities of eye movement (Hiebert et al., 2012; Price et al., 2012). Silent reading rates tend to surpass oral reading rates once a student's reading proficiency is established (Hiebert et al., 2012). SRF requires "fluent recognition of printed words, ability to process grade-level appropriate sentence structure, knowledge of grade-level-appropriate vocabulary, adequate working memory capacity to process realistic sentences, the ability to make appropriate inferences, and possession of relevant background knowledge" (Wagner et al., 2010). For this study, SRF will be defined as the ability to simultaneously

decode and comprehend.

Silent Reading Fluency and Comprehension.

The relationship between measures of SRF and reading comprehension is not clearly understood. Some studies have been conducted that examine SRF measures at varying grade levels and reader ability levels but lack consistent outcomes. Additionally, the definition used to identify students' reading ability has been based on different indicators. Wagner (2011) classified skilled readers based on their word identification score. Applegate et al. (2009) identified strong readers as determined by reading group with parent and teacher identification. Miller and Smith (1990) and Hale et al. (2007) identified students as average and poor readers based on oral reading scores. With varying indicators of reading ability, it is difficult to determine the relationship between measures of SRF and proficiency level from the studies that have been conducted. Hale et al. noted "student's reading proficiency may affect the reading mode that best facilitates comprehension" (p. 10). Clear proficiency levels need to be identified so the relationship between measures of SRF and comprehension can be better understood.

Two recent studies have sought to examine the relationship between SRF measures and reading comprehension. Wagner (2011) studied 316 first grade students and included three Woodcock Johnson III measures: passage comprehension, word identification, oral comprehension, the Test of Word Reading Efficiency (TOWRE-2), DIBELS ORF, and the TOSREC. To identify reading ability, students were divided into subgroups based on the word identification score. Skilled readers were the top third, while average readers comprised the bottom third. The findings indicated SRF measures

were strongly related to reading comprehension for first grade skilled students. However, ORF measures were identified as a better predictor of reading comprehension for all first grade students. Considering that most reading in first grade is done orally and students are just learning to read, SRF measures being a stronger predictor aligns for skilled students hints at the notion that as reading proficiency increases, measures of SRF may be a better predictor of comprehension.

Johnson et al. (2011) had 226 students in grades one through five complete measures of AIMSweb, TOSREC, Measures of Academic Progress (MAP), and the state reading assessment for students in grades three through five. Correlations between measures of ORF and SRF were high for all grades except fourth grade, which were low and not significant. Yovanoff et al. (2005) noted fourth grade was the pivotal grade which could account for the variation at this grade level. With their colleagues, both Johnson and Yovanoff found SRF measures and reading comprehension to be correlated, but ORF measures were more predictive for first grade and fourth grade correlations were low. With the limited research that includes SRF as a predictive measure, additional studies need conducted at sequential grade levels to verify the findings and address gaps in the research that exist.

Price et al. (2012) used the Gates-MacGinitie and the AIMSweb reading maze to measure reading comprehension with 59 fourth and sixth graders. The researchers also used underlining as the mode to track silent reading fluency. The maze task did not have a significant correlation with the comprehension measures, but the correlation with SRF measures and the reading comprehension measures was strong. This study indicated that

measures of SRF can be an accurate predictor of reading comprehension.

Miller and Smith (1990) found when reading silently 'average readers' had higher comprehension which contradicts the results from Hale et al. (2007), that indicated students who read passages orally had higher comprehension than students who read silently. In a study by Fuchs et al. (2001), SRF comprehension scores were substantially and statistically significantly lower than those of ORF scores for the 265 fourth-grade students. A mixed result of outcomes from these studies helps identify the need for further research in predictors of reading comprehension and if it varies as a function of grade level. Hale et al. indicated that future research is needed to examine the relationship between ORF and SRF with additional measures.

Research has supported the need for additional studies with measures of SRF and reading comprehension in order to clearly define the relationship (Denton et al., 2011; Fuchs, et al., 2001; Johnson et al., 2011; Torgesen et al., 2003). Even though the use of silent reading increases throughout the school years and becomes the dominant method for reading, measures of SRF are understudied, overlooked, and limited research has been conducted (Share, 2008). Fuchs et al. indicated that this may be in part to the fact that SRF is not easy to measure.

How is Silent Reading Fluency Measured?

SRF is a challenge to measure in the classroom since it is not an observable action (Denton et al., 2011; Fuchs et al., 2009; Price et al., 2012). In SRF, monitoring of where the student is in the passage and what words are read correctly is an unknown. This information has led to concerns with SRF over what have been termed as 'fake' readers

(Griffith & Rasinski, 2004). Thus, researchers have devised several methods to assess this important skill. Students can read a passage silently then circle the last word read in the given time (Fuchs et al., 2001). This method of measuring SRF has limited use as it is unknown if the circled word truly represents actual reading of all words to that point or if comprehension occurred. Another way to measure SRF is to have students answer comprehension questions or retell the passage after they complete their silent reading passage (Denton et al., 2011; Fuchs et al., 2001). This is referred to as question answering or passage recall (Johnson et al., 2011). Question answering does not document the number of words read. This measure can provide information on comprehension but does not account for the number of words read and scoring on retell provides only an indication as the retell score is based on a rubric. A different type of silent reading fluency measure is a cloze or maze activity where typically every Nth word is eliminated and the student chooses the best word to complete the blank (Ardoin et al., 2004; Price et al., 2012; Wayman et al., 2007). The student's score is the number of correct blanks completed in the allotted time. This assessment is intended to measure SRF and comprehension. The assessment allows for whole group administration and takes limited time to administer while providing an indicator of comprehension. Sentence verification and strings of words with no spaces are additional types of SRF measures (Denton et al., 2011; Hammill, Wiederholt, & Allen, 2006). Sentence verification requires the student to read a sentence then answer yes or no questions based on the sentence read and provides comprehension information (Denton et al., 2011; Wagner, 2011). A string of words with no spaces requires the student to draw lines between the words. The student's scores is

the number of words that were correctly identified in the set time period. This assessment does not provide a comprehension measure.

With the variety of ways to measure SRF and the little attention focused on this skill, a clear preference based on age or ability has not been established as to which SRF measure is most reflective of a student's reading proficiency level. However, as students' progress through school, silent reading becomes the dominant means of reading in the classroom and SRF measures would be useful to use with older students (Denton et al., 2011). With the focus on silent reading in the classroom, a measure that can be used to determine SRF and comprehension is warranted that does not consume student instructional time. A measure that not only documents SRF rate but also comprehension can provide needed information to help determine if a student is comprehending which is the ultimate goal of reading. With the intended goal of reading being comprehension and the high stakes assessments that many states administer, research about the relationship of ORF and SRF measures with state assessments is examined next.

Correlations of Oral Reading Fluency and Silent Reading Fluency with State Reading Assessments

What are State Reading Assessments?

State reading assessments are measures of academic accountability that states require students to take at varying grade levels. State reading assessments are considered "high stakes" assessments (Good et al., 2001; McGlinchey & Hixson, 2004) that focus on comprehension of main idea, cause/effect, and comparison. The purpose of the state

reading assessment is to identifying a reading proficiency level on outcomes that indicate the student's level of achievement in meeting the standards in reading and that the teacher can use to help guide instruction. This will assure that students are meeting the grade level standards in order to reach the goals and objectives set by the state board of education (Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Shaw & Shaw, 2002; Stage & Jacobsen, 2001; Vander Meer, Lentz, & Stollar, 2005).

State reading assessments are generally criterion referenced and administered to the whole class under standardized procedures. State reading assessments may contain multiple choice, short answer, and extended response questions based on fiction, poetry, and nonfiction passages that assess the students' knowledge and skills through interpretation, analyzing and critical thinking (Roehig et al., 2008; Vander Meer et al., 2005). Answers are generally machine scored and a raw score is derived (McGlinchey & Hixon, 2004). The raw scores are converted to scale scores and schools/districts receive a score for each student who completed the assessment.

Importance of Correlation to State Reading Assessments

School districts throughout the U.S. need indicators to determine how students are performing in reading prior to the state reading assessment. Both ORF and SRF can be predictive measures to assist states in determining a student's proficiency level (Johnson et al., 2011; Riedel, 2007; Stage & Jacobsen, 2001). Several studies have correlated ORF measures to state reading assessments that show a strong correlation at varying, but not consecutive, grade levels with results. Correlations of 0.65 - 0.80 have been found

between DIBELS ORF and several state assessments (Good et al., 2001; Roehrig et al., 2008; Shaw & Shaw, 2002; Vander Meer et al., 2005).

With the focus on third grade, Salvadro et al. (2009) included 9,562 students to predict third-grade students end-of-year proficiency results on the North Carolina Reading Assessment with DIBELS. The study found the results between DIBELS ORF and the state reading assessment were moderately correlated. Third grade is a pivotal point since mastery of reading should be established at this level that indicates success and future comprehension reading mastery (CDE, 2013c; Good et al., 2001; Hosp & Fuchs, 2005). Salvado et al. (2009) disaggregated the results by subpopulations to include ethnicity, special accommodations, and economic status and found the correlation remained stable for the ORF measure and state reading outcomes. Roehrig et al. (2008) included 35,207 students who completed the Stanford Achievement Test (SAT, 10th edition) as a standardized measure for determining the relationship between measures of ORF, reading comprehension, and the Florida state reading assessment. The results indicated that DIBELS has a high correlation to the reading comprehension measures of the SAT and Florida reading assessment. Predictive placement for 58 grade 3 Colorado students was conducted by Shaw and Shaw (2002). The finding for the TCAP for third grade in correlation to DIBELS spring benchmark indicated that 91% of the students who scored above the DIBELS national cut point scored at or above grade level on the state assessment. Overall, when measures of ORF for third-grade students were compared to state reading assessments, a moderate to strong correlation was indicated. But, when students move beyond third grade, what is the correlation between measures of ORF and

SRF as predictors of reading comprehension on state assessments?

Three research studies analyzed the predictive relationship of ORF measures and state performance level outcomes at the fourth-grade level. Vander Meer et al. (2005) conducted a study with 364 students using three different ORF benchmark results (DIBELS spring score in third- and fourth-grade CBM from Houghton Mifflin Reading Series from fall and spring) with a correlation to the fourth-grade Ohio Proficiency Test. All coefficients had a significant correlation with ORF measures and the Ohio Proficiency Test. Similarly, Stage and Jacobsen (2001) used ORF measures from reading passages in the Silver Burdette and Ginn curriculum with 173 fourth-grade students to predict the Washington Assessment of Student Learning (WASL). Their findings indicated that ORF measures had a 0.51 correlation to the WASL, which is lower than some standardized assessment results as a medium effect size was obtained. McGlinchey and Hixon (2004) found a positive correlation between ORF measures from the Macmillan Connections Reading Program and the Michigan Education Assessment Program (MEAP). The study included 1,362 fourth-grade students over an 8-year period. These fourth grade results were similar to results that Yovanoff et al. (2005) found when ORF measures and reading comprehension were investigated with 6th grade state assessment results and a moderate correlation with the reading comprehension measure was found. These studies have indicated a relationship between measures of ORF and state reading assessments at varying grade levels but the relationship appears to be stronger at the third grade level. A research gap has been identified from not having research using the same ORF measure and the same state assessment at two consecutive

grade levels.

While several studies have investigated the correlations of ORF measures to state assessments, few SRF measure correlations to state assessments have been conducted. As previously described, on the Idaho State Assessment Test (ISAT), Johnson et al. (2011) found different measures at the different grade levels indicate some assessments may be better indicators of reading proficiency at certain ages than others. Similarly, Torgesen et al. (2003) found that at 10th grade that ORF measures and TOSREC were the strongest predictors on the Florida state reading assessment. These studies all indicate that SRF measures could be a predictor of the reading proficiency level as determined by state reading assessments. With the time needed to administer predictive indicators and the need for compatibility of how students are instructed in the classroom with how students are being assessed, SRF may be a predictive measure that can save valued classroom instruction time and be a predictor on high stakes reading assessments.

Summary of Findings from Review of Literature

To date, there are a handful of studies that have focused on the relationship between measures of ORF, SRF, and state reading assessments, but gaps in research exist that were addressed in this study. One identified gap is that the definition for ORF is not consistent throughout the research. Many studies do not include a measure of comprehension with ORF, which causes many to believe that ORF is a measure of speed and does not assure comprehension. The definition in this study includes accuracy, automaticity, and reading comprehension, which is not consistently used throughout the

studies.

Second, ORF measures correlate with reading comprehension measures at the primary grades but as grade level increases the outcomes vary. It is unclear if measures of ORF are a stronger predictor than other measures of reading comprehension at the fourth- and fifth-grade level or if this varies as a function of grade level or proficiency level. Third, SRF measures have received limited attention when compared to ORF measures and the relationship with SRF and reading comprehension is even less clear. As students increase in grade, classroom expectations of silent reading also increase which would lead one to believe that SRF measures would be a strong predictor for reading comprehension.

Finally, most research studies of state reading assessments have not included consecutive grade levels. As state reading assessments vary in content and expectations each state should conduct studies that align with their intended outcomes. With many states putting emphasis and expectations on high outcomes, accurate predictors are needed to assure the students achieve the goals set. As indicated by Yovanoff et al. (2006), with assessments and instruction going hand in hand, it is critical that the measures used are effective at predicting reading comprehension. The Appendix outlines the studies, grade level(s), participants, and findings from the literature review. Additional studies to more clearly define the relationship between measures of ORF, reading comprehension measures, and state assessments are warranted.

CHAPTER III

METHODOLOGY

The purpose of this study was to analyze specific measures ORF and SRF as predictors of reading comprehension and as predictors of student achievement on a high-stakes reading test at the fourth- and fifth-grade level. The majority of the studies conducted thus far have not focused on analysis of both measures of ORF and SRF as predictors for reading comprehension and have not done so with two consecutive elementary grade levels. This study can help inform educators and assessment developers by examining the accuracy of the predictors in determining reading proficiency levels. The results may help educators select reading assessments based on the predictive relationship to the desired goals and outcomes of reading instruction. The following questions were used to guide this study:

Questions and Hypotheses

Specifically, this study addressed the following questions and hypotheses.

1. How do oral reading fluency measures and silent reading fluency measures compare as predictors of reading comprehension for fourth- and fifth-grade students?

The null hypothesis for this question is that there will be no difference between measures of oral reading fluency and silent reading fluency as predictors of reading comprehension. However, it is expected that silent reading fluency measures will be a stronger predictor of reading comprehension because research suggests correlations between ORF measures and reading proficiency decreases for maturing readers (Wagner,

2011).

2. Does the relationship of oral reading fluency measures and silent reading fluency measures as predictors of reading comprehension vary as a function of reading level?

The null hypothesis for this question is there will be no difference between measures of oral reading fluency and measures of silent reading fluency as a predictor for reading comprehension as a function of reading level. But, it is expected that silent reading fluency measures will be a stronger predictor of comprehension than oral reading fluency measures for students with a higher reading proficiency level because as students develop stronger reading proficiency, they are more likely to read silently (Price, et al., 2012). Conversely, oral reading fluency is likely to be a better predictor for students with lower reading proficiency (Cook, 2003; Riedel, 2007).

3. What is the relationship of oral reading fluency measures, silent reading fluency measures, reading comprehension and the Transitional Colorado Assessment Program reading proficiency level for fourth and fifth grade students?

The null hypothesis for this question is there will be no relationship between oral reading fluency measures, silent reading fluency measures, reading comprehension, and the Transitional Colorado Assessment Program. However, it is expected that silent reading fluency measures will have a stronger relationship to the Transitional Colorado Assessment Program reading proficiency than oral reading fluency measures and reading comprehension as it aligns to classroom practices.

Design

This research was a correlational prediction design study. As defined by Creswell

(2008), a correlational study is a "quantitative design in which investigators use a correlation statistical technique to describe and measure the degree of association (or relationship) between two or more variables or sets of scores" (p. 638). The word correlation was generically used sometimes "to refer to any statistical association between a pair of variables" (Warner, 2013, p. 1080). Prediction design included variables that served as predictors of an outcome and included several possible statistical procedures. Correlational design can help explain the relationship among variables and is not causation for the relationship or prediction revealed.

Variables in a correlation prediction design are referred to as predictor and criterion variables. For this study there were multiple predictor variables and one criterion variable. For the first question, "How do oral reading fluency measures and silent reading fluency measures compare as predictors of reading comprehension for fourth- and fifth-grade students?", the predictor variables are ORF as measured by DIBELS Next ORF with Retell and SRF as measured by the Test of Silent Reading Efficiency and Comprehension and DIBELS Daze. The criterion variable is reading comprehension as measured by the Group Reading Assessment and Diagnostic Evaluation. The second question, "Does the relationship of oral reading fluency measures and silent reading fluency measures as predictors of reading comprehension vary as a function of reading level?", examines the influence of the covariate of proficiency of measures of ORF and SRF on the relationship to reading comprehension. The third question, "What is the relationship of oral reading fluency measures, silent reading fluency measures, reading comprehension and the Transitional Colorado Assessment

Program reading proficiency level for fourth and fifth grade students?", analyzes the relationship of the predictor variables of ORF measures, SRF measures, and reading comprehension to the criterion variable of reading proficiency identified by a high-stakes state reading assessment as measured by the TCAP.

Setting

This study was conducted in a rural school district in Colorado. The district was selected based on convenience for the researcher. According to national statistics, the United States had 52.4% students in kindergarten through 12th-grade documented as nonminority and 47.6% documented minority (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2014). In comparison, the district selected for this study had 72.2% nonminority and 27.8% minority (CDE, 2013b). This indicates that the selected district has almost a 20% higher nonminority population and almost 20% lower minority population than the national percentages. Free and reduced lunch rate for the selected district was 38.9% as compared to the Colorado rate of 45% and the national rate of 48.1% (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2014). The district has a 6% lower free and reduced lunch rate than the state and 9% lower than the national rate.

The selected district spreads over a large geographic area and consists of seven elementary schools, two middle schools, and two high schools with approximately 4,500 students in grades K-12. The district contains a mix of schools sizes. Three elementary schools have student populations of over 400; four elementary schools have populations

that range from 135 to 300. All class sizes at the elementary school are held to district guidelines to assure similar class size throughout the district regardless of the school size. District guidelines set a teacher/student ratio of 1:20 for kindergarten and 1st grade, 1:22 for second and third grade, and 1:24 for fourth and fifth grade. All district elementary schools have participated in the Colorado Read to Achieve project and in state-sponsored K-5 reading initiatives with teachers receiving training in scientifically based reading research strategies.

Participants

Two of the seven elementary schools were randomly selected through simple random sampling to participate in this study. Each of the seven elementary schools was assigned a number; the two numbers that were randomly drawn from corresponding cards were selected to participate in the study. The two schools had similar student enrollments of 477 and 445. The two randomly selected schools had similar demographics and economic percentages as the district. Table 2 shows the demographics and economic details.

Free/Reduced lunch percentages ranged from 39% to 42% for the two schools, which is slightly higher than the district, but more comparable to the State. Diversity numbers ranged from 26% to 35%, with School 1 having a higher minority population when compared to district statistics of 27%. In comparison, the number of minority students at the State level is significantly higher at 46%.

The six fourth-grade teachers and seven fifth-grade teachers at the two selected

Table 2

Population and Sample Demographics

		Free reduced lunch		Minority		Nonminority		Male		Female	
Population	Total students	*N*	% of total	*n*	% of total	*n*	% of total	*n*	% of total	*n*	% of total
State	876,199	367,800	42	404,701	46	484,305	54	449,636	51	427,363	49
District	4,324	1,621	37	3,146	27	1,178	73	2,177	50	2,147	50
School 1	477	202	42	169	35	308	64	247	52	230	48
School 2	445	174	39	116	26	329	74	223	50	222	50

schools were invited to participate in the study. All teachers agreed to participate in the study. This resulted in three fourth-grade teachers and three fifth-grade teachers from School 1 and three fourth-grade teachers and four fifth-grade teachers from School 2. There was a total of 117 students in fourth grade and 155 students in fifth grade in the participating schools. An information letter was sent home to all grade 4 and 5 students in the two schools explaining the purpose of the study and providing students a non-participating option. Of the 272 students invited to participate in the study, 219 agree to participate (80.5% of the sample population); 53 students declined participation. By grade level, 98 fourth-grade students agreed to participate (83.8% of the sample population), and 121 fifth-grade students agreed to participate (78.1% of the sample population). Unfortunately, additional students opted out of assessments during the data collection phase of this study. Forty-two students who had agreed to participate did not complete one or both of the two measures specific to this study (TOSREC or GRADE). This resulted in an overall sample size of 177 participants (65% of the total sample population), with 75 fourth-grade participants (27.6% of the sample population), and 102 fifth-grade participants (37.5% of the sample population).

However, all fourth- and fifth-grade students are required to take the district assessments, three of which were used in this study (DIBLES ORF a.k.a. DORF, Daze, TCAP). This data is publicly available, which allowed for comparison of study participants with non-participants in regard to these selected reading competencies.

Participating student demographic data was gathered from the student registration completed each year by the parent or guardian. This information included a free and

reduced lunch application. An analysis was conducted to determine if sample bias occurred due to the number of students (19.5%) who opted out of the study and the number of students who did not complete one or both of the additional measure for this study. Comparison of the participating and nonparticipating students showed no statistically significant difference between the groups for student characteristics of socioeconomic status, minority classification, and gender (Table 3). Tables 4, 5, and 6 represent the student populations by school and by grade level and teacher. In regard to the reading measures, there was no statistically significant differences between nonparticipants and participants on the TCAP, $t(268) = 0.57$, $p = 0.057$. However, there was a statically significant difference between nonparticipants and participants on the DORF, $t(270) = 2.205$, $p = 0.03$, with the non-participants scoring higher ($M = 142.47$, $SD = 43.07$) than the participants ($M = 132.15$, $SD = 33.55$). There was also a statically significant difference between nonparticipants and participants on the Daze, $t(270) = 2.285$, $p = .02$, again with the nonparticipants scoring higher ($M = 28.45$, $SD = 10.68$) than the participants ($M = 25.71$, $SD = 8.72$). Given these results, in regard to student demographics the participant sample resembled the sample population in regard to demographic data and overall reading proficiency as measured by the TCAP. There was a slight difference between groups on the scores of the other two district-mandated assessments of reading subskills (DORF and Daze) favoring the nonparticipants. Nevertheless, the comparisons of this study should serve to provide useful information about the relationship of oral and silent reading fluency measures to fourth and fifth grade students' reading comprehension.

Table 3

Comparisons for Participating versus Nonparticipating

	Participants ($N = 177$)		Nonparticipants ($N = 95$)				
Comparisons	*N*	%	*N*	%	χ^2	*T*	*P*
Free reduced lunch	76	43	33	35	1.90		0.17
Minority	57	32	25	26	1.02		0.31
Nonminority	120	68	70	74			
Male	86	49	45	47	0.04		0.85
Female	91	51	50	53			
DORF						0.03	
Daze						0.02	
TCAP						0.57	

Table 4

Participating Student Demographics by School

	School 1 ($N = 102$)		School 2 ($N = 75$)	
Demographics	*n*	%	*n*	%
Free/reduced lunch	46	45	30	40
Minority	39	38	18	24
Nonminority	63	62	57	76
Male	50	49	36	48
Female	52	51	39	52

Table 5

Participating Student Demographics by Grades

	4th grade ($n = 75$)		5th grade ($n = 102$)	
Demographics	*n*	%	*n*	%
Free/reduced lunch	24	32	52	51
Minority	21	28	36	35
Nonminority	54	72	66	65
Male	39	52	52	51
Female	36	48	50	49

Table 6

Student Demographics by Teacher for Participating Students

	School 1								School 2					
	Teacher A (N = 17)		Teacher B (N = 14)		Teacher C (N = 17)				Teacher D (N = 12)		Teacher E (N = 8)		Teacher F (N = 7)	
Demographics	*n*	%	*n*	%	*n*	%			*n*	%	*n*	%	*n*	%
4th Grade														
Free/reduced lunch	7	41	7	50	3	25			2	17	4	50	1	14
Minority	5	29	4	29	5	35			3	25	2	25	2	29
Nonminority	12	71	10	71	12	65			9	75	6	75	5	71
Male	8	46	7	50	8	55			4	33	5	63	4	57
Female	9	53	7	50	9	45			8	67	37	37	3	43
	Teacher G		Teacher H		Teacher I		Teacher J		Teacher K		Teacher L		Teacher M	
5th Grade	(N = 19)		(N = 15)		(N = 20)		(N = 12)		(N = 10)		(N = 12)		(N = 14)	
Free/reduced lunch	5	26	9	60	15	75	7	58	6	60	5	42	5	36
Minority	4	21	8	53	13	65	3	25	3	30	2	17	3	21
Nonminority	15	79	7	47	7	35	9	75	7	70	10	83	11	79
Male	11	58	10	67	6	30	5	42	5	50	6	50	7	50
Female	8	42	5	33	14	70	7	58	5	50	6	50	7	50

Instrumentation

The instrumentation for this study included reading assessments that have been adopted by districts and states across the nation and measures that align with classroom practice at the fourth- and fifth-grade level. With the state assessment being the driving force behind school performance, schools need predictive indicators that align with reading outcomes that can be used to predict student results. Assessments that measure ORF, SRF, and reading comprehension are needed to support and guide reading instruction in the classroom to reach proficiency levels set by the state. Currently, teachers rely heavily on ORF measures as the main data source but ORF decreases as students grow older and assessments that can be predictors of reading comprehension are needed.

Oral Reading Fluency

ORF was measured with DIBELS Next. DIBELS Next ORF was selected as it was one of the state approved assessments for reading proficiency and the district adopted reading component. DIBELS was originally created to assess students in kindergarten through third grade but was expanded to included students in fourth through sixth grade. DIBELS is now in its 7th edition and is commonly known as DIBELS Next (Good et al., 2013). DIBELS Next assessments passages are criterion-referenced reading measures. These assessments are used to compare growth over time to determine each student's reading progress. The criterion-referenced target scores and cut points from DIBELS Next help identify how students should be grouped for instruction based on their

identified levels of risk. The DIBELS Next ORF scores are reported to predict future reading achievement and academic success (Good et al., 2011).

DIBELS Next ORF consists of three reading passages at each grade level. The students are asked to read the passages aloud and are timed for one minute. The number of words accurately read during the time is calculated. Word omissions, substitutions, and hesitations of more than three seconds are recorded as errors. Word self-corrections within three seconds are scored as accurate. The median correct words per minute read from the three passages are recorded as the ORF score (Good et al., 2011; Wagner, 2011). One third of the passages in fourth through sixth grades are narrative and two-thirds are expository (Good et al., 2013). To prevent ceiling effects, the passage lengths are designed so that most students do not finish in one minute (Good et al., 2011).

DIBELS Next also includes a measure of comprehension after oral reading. Measures of ORF that do not include a comprehension component neglect the importance of the transfer of decoding skills to comprehension (Rasinski, 2009). After reading each ORF passage, the comprehension check for this study was a Retell to provide an indication if the student was reading for meaning (Good et al., 2013). The students were asked to tell "as much as you can about the story you just read" (Good et al., 2011, p. 81). As the students retold the story, the administrator analyzed how many words the student used to retell the story. Each word that relates to the story was counted; those that were off topic were ignored. If a student hesitated for 3 seconds during a Retell, the administration provided a probe to encourage the student to continue by repeating what was originally asked or saying "Can you tell me anything more about the story?" (Good,

et al., 2011, p. 81). After 5 seconds of hesitation, the administrator discontinued the Retell. After the Retell, the response was rated using the DIBELS Retell quality of response rubric that has four levels. The quality of response is scored by the number of details provided. Level 1 has 2 or few details; level 2 has 3 or more details; level 3 has 3 or more details that are in a meaningful sequence; level 4 has 3 or more details in a meaningful sequence and captures the main idea of the passage read. By administering the Retell component of DIBELS in addition to ORF, the reading definition for this study was met.

DIBELS ORF has been established as a valid and reliable measure. Validity of the DIBELS ORF is reported in the DIBELS Next Technical Manual (Good et al., 2013). Criterion-related validity ranges from 0.44 to 0.61 with coefficients of 0.52 and 0.45 for fourth and fifth grade respectively on end of year results with Retell coefficients of 0.78 and 0.77 for fourth and fifth grade with Daze (Good et al., 2013). The predictive validity for end of year DIBELS ORF with the Stanford Achievement Test–10th edition (SAT10) was 0.81 for fourth grade and 0.83 for fifth grade. Concurrent validity of ORF with NAEP Oral Reading Study was 0.89 and 0.96 for the fourth and fifth grade levels (University of Oregon Center on Teaching and Learning, 2012). Validity above 0.70 is considered strong and 0.50 - 0.69 is considered moderate (Good et al., 2013). Predictive validity of ORF was tested with alternate-form, test-retest, and inter-rater. As reported in the DIBELS Next Technical Manual (Good et al., 2013), the alternative-form reliability for grade four was 0.96 and 0.96 for grade five. The test-retest reliability was 0.97 for both grades four and five. The interrater reliability was 0.99 for both grades.

Validity of the DIBELS Retell is reported in the DIBELS Next Technical Manual (Good et al., 2013). Criterion-related validity of DIBELS Retell is reported as ranging from 0.44 to 0.61 with coefficients of 0.52 and 0.45 for fourth and fifth grade respectively on end of year results with ORF and coefficients of 0.78 and 0.77 for fourth and fifth grade with Daze (Good et al., 2013). The predictive validity for end of year DIBELS Retell with the SAT10 was 0.65 for fourth grade and 0.69 for fifth grade (University of Oregon Center on Teaching and Learning, 2012). Concurrent validity of Retell with NAEP Oral Reading Study was 0.62 and 0.65 for the respective grade levels. Predictive validity of Retell was tested with alternate-form, test-retest, and inter-rater. The alternative-form reliability for grade four was 0.80 and 0.65 for grade five and the test-retest reliability was 0.36 for grade four and 0.58 for grade five. The inter-rater reliability was 0.98 and 0.96 for fourth and fifth grade.

Silent Reading Fluency

SRF was evaluated with two assessments, the Test of Silent Reading Efficiency and Comprehension and DIBELS Daze. The Test of Silent Reading Efficiency and Comprehension (TOSREC) is an assessment of silent reading of connected text for speed, accuracy, and comprehension (Wagner et al., 2010). A strength of this assessment is that it measures comprehension during silent reading. Measures of SRF that do not include a comprehension component (such as those in which the child circles the last word read) do not adequately measure SRF; thus, leading to questionable results (Miller & Smith, 1990; Wagner, 2011). The TOSREC is a SRF measure that can be used as a predictor for reading comprehension that aligns with classroom reading practices.

Another strength of TOSREC, leading to its selection for this study, is the efficiency of administering the assessment whole group and the research conducted with the correlation to multiple measures including the Test of Word Reading Efficiency (TOWRE), Woodcock Johnson III, and GRADE (Wagner et al., 2010). Johnson et al. (2011) noted that the TOSREC takes minimal time away from classroom instruction. Administering the three-minute measure and scoring take about 30 minutes per class to complete (Johnson et al., 2011). The measure can be group or individually administered and can be used for screening and progress monitoring. SRF comprehension is explicitly assessed through the questions that students are asked and affects students' scores since each incorrect answers is scored as a minus one (Johnson, et al., 2011).

There are four alternative forms available that can be used for screening and progress monitoring. The forms are normed for fall, winter, spring, or any time of year at each grade level. Form O was used for this study as it is normed for spring. This allows the TOSREC to be used as a progress monitoring measure that will show student growth. The TOSREC provides raw scores, indexes, and percentiles that allow comparability of a student's individual score to national norms.

For this assessment, students read sentences silently and verify their comprehension by answering true/false questions immediately following the sentence reading by circling yes or no. Students have two sample items to explain the task, five practice items, and can complete up to 50 test items. The students have three minutes to complete as many questions as possible. The scores were computed by counting the correct number of responses and subtracting the incorrect responses. Incorrect items are

scored as a -1 to account for guessing. The raw scores can range from 0–60.

TOSREC has been shown to be a valid and reliable measure of SRF. The alternate-form reliability coefficients for all forms and grade levels exceeded 0.85. The alternate-form reliability for fourth and fifth grade was 0.86 and 0.89, respectively. In a study by Wagner (2011), reading comprehension levels and the TOSREC exceeded 0.70 correlation coefficient; and a sensitivity of 96% and specificity of 84%.

The second SRF assessment administered, a component of DIBELS Next, is a three-minute, whole class administered measure called the Daze. (Good et al., 2011). Daze assesses a student's ability to construct meaning from text using word recognition skills, background information and prior knowledge, familiarity with linguistic properties such as syntax and morphology, and cause and effect reasoning skills (Good et al., 2011).

Daze is a cloze measure which replaces approximately every seventh word in the passages with a box containing the correct word and two distracter words. Standardized directions require a student to read a passage silently and circle the word that best completed the sentences. Credit is given if the student selects the words that best fit the omitted words in the reading passage. The number of correct and incorrect responses is recorded. The score is adjusted by subtracting half the number of errors made from the number correct to compensate for guessing.

Validity of the DIBELS Daze is reported in the DIBELS Next Technical Manual (Good et al., 2013). Criterion-related validity of DIBELS Daze is reported as ranging from 0.44 to 0.61 with coefficients of 0.52 and 0.45 for fourth and fifth grade respectively on end of year results with Retell and coefficients of 0.78 and 0.77 for fourth

and fifth grade with ORF (Good et al., 2013). The predictive validity for end of year DIBELS Daze with the SAT10 was 0.78 for fourth grade and 0.77 for fifth grade (University of Oregon Center on Teaching and Learning, 2012). Concurrent validity of Daze with GRADE was 0.78 and 0.77 for the respective grade levels. As reported in the DIBELS Next Technical Manual (Good et al., 2013), the alternative-form reliability for grade four was 0.93 and 0.94 for grade five. The inter-rater reliability was 0.98 and 0.99 for fourth and fifth grade.

Reading Comprehension

The Group Reading Assessment and Diagnostic Evaluation (GRADE) was used in this study as a measure of reading comprehension. GRADE is a norm-referenced diagnostic tool with two forms parallel in content and difficulty. Form A was used for this study. GRADE requires students to read passages then answer multiple-choice questions. GRADE provides raw scores which can be converted to stanines, standard scores, percentiles, normal curve equivalences, and grade equivalences.

GRADE reading comprehension allows for whole group administration and can be administered in two shorter sessions for students. GRADE is untimed, allowing students to complete the assessment at their own pace for a more accurate reflection of their reading ability. However, the assessment generally takes students 45-90 minutes to complete. The reading comprehension components are broken into two parts: sentence comprehension and passage comprehension. There are 19 questions in the sentence comprehension and 28 in the passage comprehension. Questions are ordered randomly between easier and harder to allow for all students to be encouraged to continue to give

their best effort. Each item is scored as correct or incorrect; therefore, the range of scores is 0-19 for the sentence comprehension subtest and 0-28 for the passage comprehension subtest. From the total of correct answers, raw scores are obtained and converted into normative scores.

There is evidence for the validity and reliability of GRADE. Criterion-related validity for grades 1-6 with ITBS, CAT, Gates-MacGinitie has a coefficient range of 0.69 to 0.90 with a median of 0.83. The predictive validity with TerraNova at grades 2, 4, and 6 ranges from of 0.76 to 0.86 with a median of 0.77. Alternate form reliability ranges from of 0.81 to 0.94 with a median of 0.89. Reliability coefficient for students for test-retest is in the 0.77 to 0.96 range with 0.90 as the median. Internal reliability ranges from of 0.91 to 0.99 with a median of 0.96.

State Assessment

State reading assessments measure a student's progress to set standards or performance frameworks for each given state. The results provide a picture of the student's performance at a given time to educators and the community. State assessments are used to ensure that students are meeting the same expectations throughout the state. State assessments are administered with standardized procedures, allow student accommodations as verified in individual education plans, and can be timed or untimed depending on the state.

The state assessment of interest in this study is the TCAP. In 2009, the State adopted new Colorado Academic Standards and integrated the Common Core Standards when they were finalized. When the items were reviewed and a content analysis was

conducted between the two sets of standards, the Colorado Standards had a significantly high alignment to the Common Core and in some cases exceeded the expectations identified in the Common Core. With the strong alignment, TCAP is considered to be aligned with the new national assessments that began in 2015. TCAP is designed to help Colorado schools transition from the content model standards to the new Colorado Academic Standards.

TCAP is used in Colorado to determine the reading proficiency level for each student. All students in grades 3-10 enrolled in a public school in Colorado are required by law to take this paper/pencil assessment. TCAP is a standardized assessment with written protocol and procedures. Fiction, nonfiction, poetry, and vocabulary are included in the assessment and are assessed through multiple-choice and short answer items that are computer scored. TCAP also includes constructed response items which are scored by trained readers with continued checks for inter rater reliability (Colorado Department of Education, 2013a). The overall score for the items is computed to obtain a raw score. The raw score is converted into a scale score that is used to identify what level of reading proficiency the student had reached. The scale score ranges are 180-940 fourth grade and 220-955 for fifth grade.

Criterion-related validity for 2013 TCAP has coefficient ranging from 0.86 to 0.94 with 0.93 for fourth and fifth grade (CTB McGraw Hill, 2013) with the 2012 TCAP. The 2013 predictive validity with the 2012 TCAP results had correlation coefficients of 0.91 for fourth grade and 0.92 for fifth grade. Concurrent validity is measured by the number of items flagged and reviewed for each grade level. Items that are flagged are

broken down by subgroup and removed if necessary. As indicated in the technical manual (CTB McGraw Hill, 2013), the alternative-form reliability with the 2013 form was 0.99 for grade four and 0.98 for grade five. The interrater scoring reliability for the constructed responses was between 0.97 to 0.99 for fourth grade and 0.90 to 0.99 for fifth grade.

Assessment Procedures and Data Collection Fidelity

All assessments were administered during a set assessment window over nine consecutive weeks as outlined in Table 7. Three of the measures, DIBELS Next ORF, Daze, and TCAP were District or State mandated assessments for students in grades 4 and 5. These assessments required no change in teacher practices or different expectations. The DIBELS Next assessments were administered during the district-mandated end of year two-week window by the classroom teacher. The other two measures, TOSREC and GRADE, were assessments administered as part of the research study.

As the school district's certified DIBELS Next trainer, the researcher trained all reading interventionists, instructional coaches, and school principals at the beginning of every school year. The interventionist, instructional coach, and principal then trained their school staff using the same materials and procedures. The interventionists, instructional coach, principal, and the researcher observed teachers administering DIBELS ORF and Daze benchmark measures to assure quality and reliability of results by using an administration checklist and shadow scoring. Two random observations were conducted per teacher during the benchmarking period by the researcher at each site to

Table 7

Assessment Procedure Outline

Assessment	When given	By whom	How administered	How scored
TCAP	March 10th – 21st	Classroom teacher	Group	State and machine
TOSREC	April 14th – 25th	Researcher	Group	Researcher
GRADE	April 14th – 25th	Researcher	Group	Researcher
DIBELS Next: ORF and Daze	April 24th – May 9th	Classroom teacher	Individual	Classroom teacher

maintain fidelity and integrity of the results. Scoring was completed by the classroom teacher who received and completed training at their school site from their interventionist, instructional coach, or school principal. During the random observations, accuracy was checked using the observational checklist and shadow scoring the teacher during the administration of DIBELS ORF benchmark.

State assessments require standardized procedures and protocols be followed. As the District Assessment Coordinator (DAC), the researcher trained all School Assessment Coordinators (SACs). Each school SAC attends the required training and trains all participating teachers at the school site. The SAC assured that all standardized procedures are followed throughout the assessment process. There are several trainings prior to the actual assessment window and administration of TCAP to ensure all questions are answered and teachers are prepared for the assessment. All teachers who administer the assessment are required to agree and sign that all standardized procedures will be followed and if a breach occurs, it is reported immediately. TCAP is administered to fourth and fifth grade students by the classroom teacher in March. Completed booklets

were packed and sent to McGraw Hill Education/CTB for scoring.

To assure consistent procedures were followed, the researcher administered the TOSREC and GRADE to participants in grades 4 and 5. Each classroom teacher observed the assessment administration and followed along in the administration manual to confirm it was being administered in a standardized format.

The TOSREC and GRADE were administered to study participants in each classroom during a 2-week testing window from April 14th to 25th. Nonparticipating students went to the school library with the assistant principal to complete a book search that aligned with the International Baccalaureate unit they were working on in the classrooms. The TOSREC was administered in a group setting 3 weeks after the State-required grade level assessment had been completed. Unlike the TOSREC, which is timed, the GRADE assessment was completed in two untimed sections. The first section was sentence comprehension. The second section, completed the next day, was passage comprehension with the same procedures and directions followed. The researcher sealed with tape the section that had previously been completed to assure that no answers were changed. The researcher scored the TOSREC and GRADE by using the scoring key provided by the publisher. All answer sheets were rechecked by the researcher to verify accuracy of the recorded score. Teachers verified scoring accuracy through a random check of 10% of the participants' books from their class.

Threats to Internal and External Validity

There are some possible threats to the internal and external validity of this study. To help control for threats, additional training, monitoring, and observations were

conducted throughout the study.

Internal Validity

As defined by Warner (2013), internal validity is "the degree to which results from a study can be used as evidence of a causal connection between variable" (p. 1,093). Identified potential threats to internal validity for this study included testing, instrumentation and experimenter bias.

Testing. How a student performs on a measure at the end of the study may differ from the initial testing if participants become familiar with a repeated measure (Freedman, Pisani, Purves, 2007). DIBELS probes for ORF and Daze were used as progress monitoring assessments by the school district in which this study was conducted. DIBELS probes were given weekly or bi-weekly based on the child's risk indicator. However, study participants had not previously read the end of year ORF or Daze passages used in this study. Additionally, participating teachers encouraged students to do their best prior to starting the ORF and Daze assessments. GRADE and TOSREC were new assessments to the students and TCAP was only administered once, therefore, testing validity was not a concern with these measures.

Instrumentation. Potential threats to instrumentation included lack of standard administration procedures and conditions of testing. However, if standardized procedures have been followed as outlined, the results will portray the intended outcome (Creswell, 2008). Instrumentation validity was a concern as ORF probes were given frequently to students. This may lead to the classroom teacher paying less attention to detail when administering the ORF and Daze.

To help minimize this threat, training sessions were conducted with participating teachers for each of the five measures used in the study. The researcher provided a 2-hour refresher training held at the district office to review the administration procedures for the ORF and Daze measures. This training was held on April 14th, 2 weeks prior to the start of the test administration on April 28th. All participating fourth- and fifth-grade teachers were in attendance. The training consisted of a PowerPoint presentation that outlined the proper standardized administration procedures. Teachers were free to ask questions throughout the training. After the training was completed, direction packets were provided to each test administrator who practiced giving the assessment to a partner during the training. All teachers were observed by the researcher during the practice session to confirm accurate procedures and scoring of the ORF and Daze measures. Teachers continued to practice until each felt comfortable administering the assessments and were following outlined administration procedures.

During the assessment window, two random 20 minute observations of each teacher were conducted on the teacher's scheduled assessment date by the researcher during administration of the assessments to assure the teacher followed the set procedures as outlined in the directions packet. This resulted in 26 observations over the 2-week period. The number of observations per day varied as determined by the teacher's assessment date during the window. The researcher had the same materials as the teacher and followed along during the administration of the assessment and completed scoring of the student alongside the teacher. Each observation included the entire ORF measure with Retell assessment which takes approximately 10 minutes to complete per student. The

Daze took approximately 4 minutes. The teacher's total assessment time depended on the number of students in the class.

Training for the TCAP measure was conducted at each school site by the School Assessment Coordinators (SAC) who attended the researcher's district training. The district training was presented to all SACs during the last week of January. The training was conducted by PowerPoint that was provided by the Colorado Department of Education and McGraw Hill Education/CTB. The training lasted 2 hours; 1 hour for the presentation and 1 hour for questions and answers. All SACs attended the training. There is one SAC at each school. All teachers who participated in the study attended the training at their site that was conducted by the SAC-which included a power point and review of the procedures manual. Test procedures manuals were provided to each teacher to follow for standardized procedures. The researcher observed the training to assure all procedures were followed. During the administration of the TCAP, the researcher observed each teacher one time for 25 minutes during administrations to check to see if procedures were being followed. The TCAP took approximately 180 minutes.

The publisher-provided procedures manuals for the GRADE and TOSREC measures provide a written script for test administration. These two measures were administered by the researcher, in each participating classroom. To minimize instrumentation threats for the GRADE and TOSREC assessments, the researcher carefully followed the procedures as outlined in the publishers' administration manual, and the participating classroom teachers followed along to assure that standardized procedures were followed. The GRADE took approximately 90 minutes per class, and the

TOSREC took approximately 10 minutes per class.

Testing conditions were standardized for all assessments. For the individually administered ORF measure, each student walked from his/her classroom to a quiet, well-lit, familiar room in the same hallway for test administration with his/her teacher. For the Daze, TCAP, GRADE, and TOSREC, students completed the assessment in their classroom at their usual desk. All testing environments were free from additional distractions and excessive noise in the hallways.

Experimenter bias. Experimenter bias happens when the administrator conducting the study affects the outcome by behaving differently when obtaining results (Krathwohl & Smith, 2005). Classroom teacher bias in administering the ORF measure, Daze or TCAP could be impacted from preconceived notions such as reading ability due to grade level, past progress monitoring scores, or previous assessment results. Random checks were conducted by the researcher to assure the teacher administrators followed standardized protocol consistently when administering the assessments. The random checks were part of the observations that were completed to control for potential threats to instrumentation. Training of the importance of following procedures exactly as written and not adjusting wording with different participants was completed two weeks prior to the start of the assessments during the 2-hour session on April 14 that included standardized administration procedures. For the researcher who administered the TOSREC and GRADE, an acceptance of the obtained results was needed to help minimize experiment bias.

External Validity

External validity is the "degree to which research results were generalized to participants, settings, and materials beyond those actually included in the study" (Warner, 2013, p. 1,086). Threats to external validity for this study included sample size, geographic location, and order effect.

Sample size and location. The combined factors of sample size and location make replication results of the study unknown. If the sample size of a correlational study has a low N size, the results can be skewed, by one or two scores. Warner (2013) suggested that researchers have a sample size of at least 100 participants so there is a wider range of scores for the predictor and outcome variables. "Larger sizes contribute to less error variance and better claims of representation" (Creswell, 2008, p. 370). Data analysis examined the grade levels independently with statistical control for the classroom grouping variable. A biased sample could also be obtained even though random selection of the schools was conducted. The location of a rural community with populations of students with similar demographics could be hard to replicate in urban settings.

Order effect. Order effect refers to the order in which something happened or was administered (Warner, 2013). In this case, the assessments were in a specific order which could affect the outcome, motivation, and effort that students put into it. The TCAP assessment was given first and was a high-stakes, multi-day, high-stress measure. This assessment window is from March 10th to March 21st and is set to assure that all students complete the measure as close to the same time as possible throughout the State

so results can be compared. The State assessment was followed by two measures that had not been previously administered to students. These two measures were given between the State and district assessments to provide a closer alignment of results with limited impact on learning that could occur between assessments. The final assessments were the third, and end of year, administration of the measure of DIBELS ORF, Retell, and Daze. This assessment window, from April 28th to May 9th, is also set by the State and has limited flexibility to alter. For the five measures of this study not to overlap and be administered within a 2-month time span, the order of the assessments had to follow the sequence outlined. As a student proceeded through the battery of assessments, order effect could potentially skew the results. However, internal and external validity were considered in conducting this study and care was taken to address the identified threats to the greatest extent possible.

Data Analysis

Data was collected over a 9-week period from 75 fourth- and 102 fifth-grade students to investigate the relationship of measures of ORF and SRF and reading comprehension. Furthermore, the study sought to investigate if the relationship varied as a function of reading proficiency level. This study analyzed specific measures ORF and SRF as predictors of reading comprehension and as predictors of student achievement on a high-stakes reading test at the fourth- and fifth-grade level.

Data Entry

The DIBELS ORF with Retell and Daze assessment results were recorded by the

classroom teacher and entered into the district data base system within 3 days after completion. The system allowed teachers to enter individual students' results that record the date assessed, proficiency level, and individual measures assessed. The researcher randomly selected 10% of the student booklets and verified the entries in the data base. TCAP was shipped to McGraw Hill Education/CBT company for scoring. McGraw Hill Education/CBT provided computed scores in data files that were uploaded into the district data base for analysis as this was a state mandated assessment. TOSREC scores were entered into a spreadsheet by the researcher within two days of completion of the assessment. GRADE scores were entered into a spreadsheet by the researcher upon completion of both sections. Ten percent of the student measures were randomly selected and checked for accuracy of data entry by the researcher's administrative assistant. No discrepancies were found.

Preliminary Analysis and Descriptive Statistics

Preliminary analysis and descriptive statistics were used to evaluate the means, standard deviation, minimum and maximum scores. Additionally, reliability coefficients and correlations between variables were reviewed. Reading proficiency levels were based on the cut scores set by DIBELS Next. The commercial software package SPSS 23 was used for preliminary analysis and to calculate descriptive statistics and correlation analysis. Secondary analysis included structural equation modeling (SEM) to examine the relationships between ORF measures, SRF measures, reading comprehension and the Colorado State reading assessment. SEM allows multiple latent variables to be measured and causation in both directions of variable pairs (Cohen, 2008). With SEM, the variables

can be analyzed to estimate the variance within each assessment. The secondary analysis also used the SPSS 23 software with an add-on regression component, PROCESS (Hayes, 2013) to examine the relationships between ORF measures, SRF measures, reading comprehension and the Colorado State reading assessment.

Structural Equation Models

Regression analysis and structural equation modeling was used to examine the relationship of ORF measures, SRF measures, reading comprehension, and TCAP for fourth and fifth graders, and how the relationship may vary based on grade level and reading proficiency level.

Analysis began with models to analyze the association between reading comprehension (as measured by the GRADE) and measures of ORF or measures of SRF as moderated by grade level. These models are represented by the equation: $Y = i + b_1X + b_2M + b_3XM + e$. For this study, this equation can be read as:

- Comprehension = intercept + regression coefficient*ORF + regression coefficient*grade level + regression coefficient*ORF*grade level + error or
- Comprehension = intercept + regression coefficient*SRF + regression coefficient*grade level + regression coefficient*SRF*grade level + error.

These equations can be represented by the conceptual models (Hayes, 2013) as shown in Figure 1. Analysis was also conducted to evaluate the association between reading comprehension and SRF as moderated by proficiency level: Comprehension = intercept + regression coefficient*ORF + regression coefficient*proficiency level + regression coefficient*SRF*proficiency level + error.

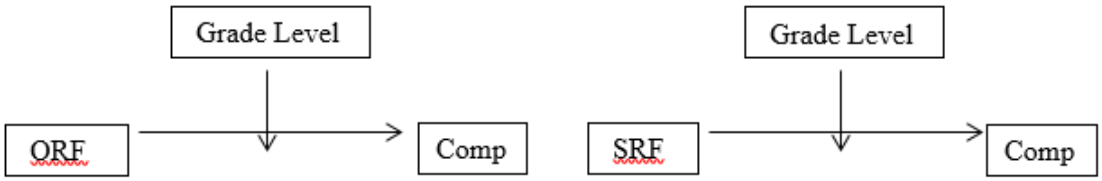


Figure 1. Reading comprehension against measures of oral reading fluency or silent reading fluency as moderated by grade level.

This equation can be represented by the conceptual model (Hayes, 2013) as shown in Figure 2. Finally, the relationship of ORF measures, SRF measures, reading comprehension and the Transitional Colorado Assessment Program reading proficiency was analyzed using a mediation model which showed the direct and indirect effects with TCAP. The indirect effect was from measures of ORF or SRF to comprehension on TCAP. The direct effect was measures of ORF or SRF on TCAP. Figures 3 and 4 show theSE mediation models.

Summary

In summary, this study analyzed specific measure of ORF and SRF as predictors of reading comprehension and as predictors of fourth and fifth grade student achievement on a high-stakes reading tests. DIBELS ORF and Retell were used to measure ORF. Daze and TOSREC measured SRF. Reading comprehension was measured with GRADE. The TCAP was the state reading assessment. Descriptive statistics, reliability coefficients, and correlations between variables were analyzed to compare how measures of ORF and SRF predict reading comprehension and if this varied as a function of grade level or reading proficiency level. Results from these assessments were used in the structural equation models to examine how measures of ORF and SRF compare as predictors of reading

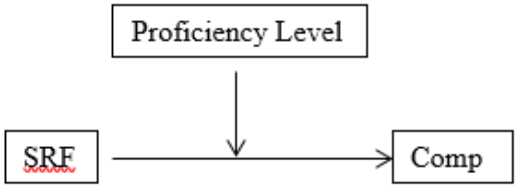


Figure 2. Regression model for measures of silent reading fluency and reading comprehension by proficiency level.

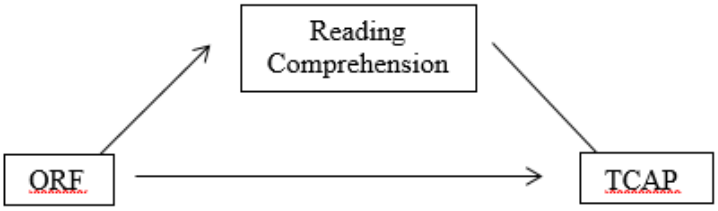


Figure 3. Meditation model for oral reading fluency measures, reading comprehension, and Transitional Colorado Assessment Program.

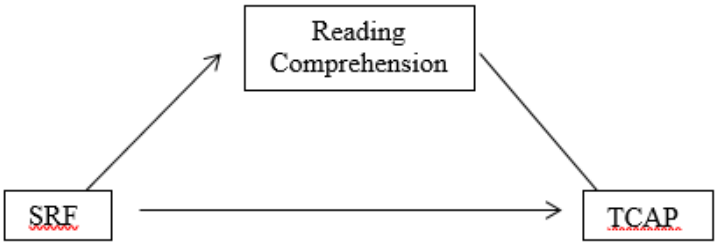


Figure 4. Meditation model for silent reading fluency measures, reading comprehension and Transitional Colorado Assessment Program.

comprehension. With more silent reading occurring in the upper elementary classrooms, silent reading might be a strong predictor of reading comprehension and of student achievement on high-stakes assessments for older students.

CHAPTER IV

RESULTS

The purpose of this study was to evaluate how measures of oral and silent reading fluency compare as predictors of reading comprehension and how these vary as a function of proficiency level for fourth- and fifth-grade students. Additionally, the study sought to examine the relationship between ORF measures, SRF measures, reading comprehension, and the TCAP with these students. This study used a correlational prediction design with measures for the variables of ORF, SRF, reading comprehension, and TCAP. Participants were 177 fourth- and fifth-grade students from two randomly selected schools in a school district in Colorado. Because of a testing error revealed during the analysis of assumptions of normality, two participants were removed from the study analysis. This yielded a final sample size of 175 participating students.

Descriptive Statistic Results

For each of the measures, descriptive statistics were analyzed for distributions and central tendency.

Oral Reading Fluency

Assumptions of normality. ORF was assessed using DIBELS Next with retell. DIBELS is an individually administered standardized measure in which the student reads a grade level passage for 1 minute and then retells what was read. The measure has three 1-minute reads and the median score of words correct per minute is recorded. Student

scores are reported as the number of words read correctly in one minute. The end year grade level score for the 50th percentile is 123-word count per minute (WCPM) for fourth grade and 167 WCPM for fifth grade (Good et al., 2013). Examination of the score distributions for this study revealed that scores were approximately normally distributed with the skew and kurtosis values less than 1. The skewness was 0.122 (*SE* = 0.184) and Kurtosis was 0.500 (*SE* = 0.365). Figure 5 shows the distribution of the ORF measure. This variable was accepted as normally distributed. There were seven outliers on the measure. Three scores were three standard deviations above the mean and four scores were three standard deviations below the mean. Figure 6 shows the outliers in ORF scores.

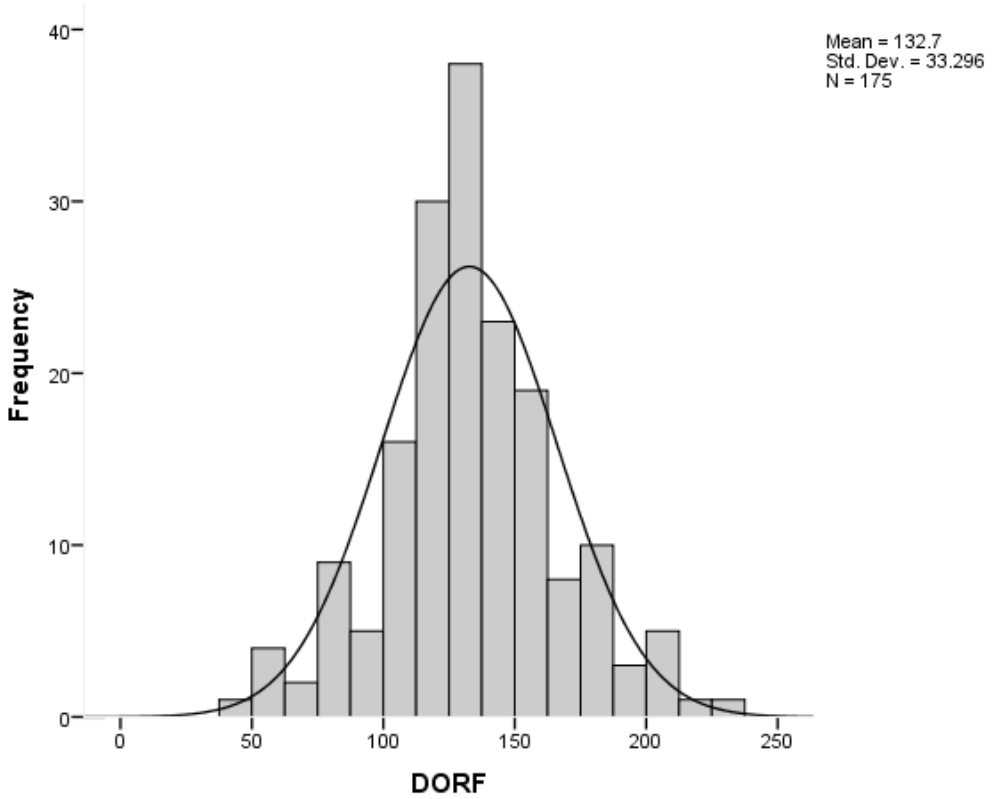


Figure 5. Distribution of oral reading fluency measure.

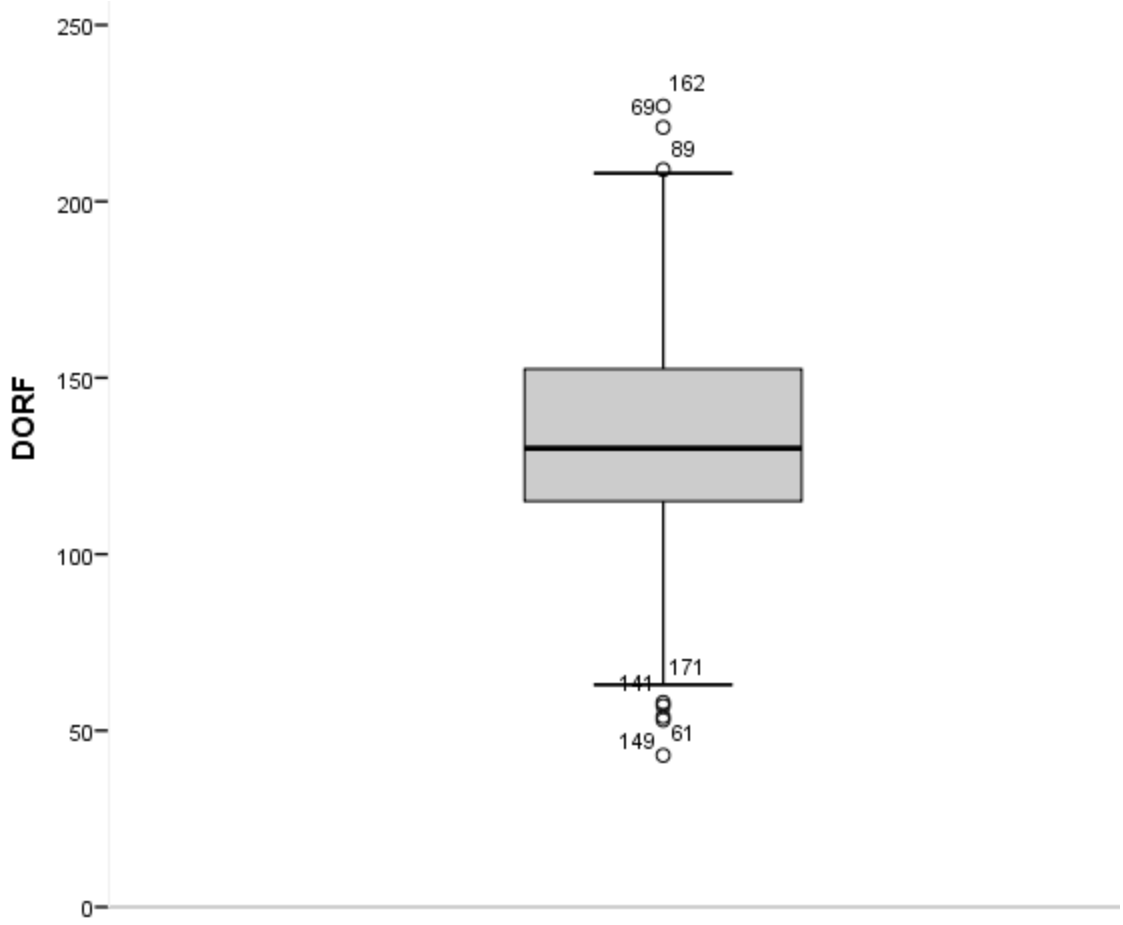


Figure 6. Boxplots of oral reading fluency scores.

Descriptive statistics. DIBELS Next ORF scores had a range in words read of 184 with 43 being the minimum and 227 being the maximum. The mean was 132.70 (*SD* = 2.517), with a median of 130, and a mode of 116 for the 175 participants. For DIBELS Next ORF by grade level, fourth-grade students had a range in words read of 141 with 67 being the minimum and 208 being the maximum. The mean was 136.85 (*SD* = 3.609), median of 134, and a mode of 140 for the 75 participants. For fifth-grade students, DIBELS Next ORF had a range in words read of 184 with 43 being the minimum and 227 being the maximum. The mean was 129.58 (*SD* = 3.457), with a median of 129, and

a mode of 127 for the 100 participants. Table 8 shows the descriptive statistics for the total sample and for fourth- and fifth-grade levels.

Reading proficiency levels were classified using the Hasbrouck and Tindal (2006) percentile rank based on spring scores for WCPM. Participants' DIBELS Next ORF scores were used to create three classifications representing participants scoring in the bottom quartile (with a 25th percentile or lower), participants scoring in the middle quartiles (26th-74th percentile), and participants scoring in the top quartile (75th percentile or above). Table 9 shows the three classifications and the descriptive statistics in relation to reading proficiency.

Table 8

DIBELS Next Oral Reading Fluency Descriptive Statistics

	Mean						
Sample	Statistic	*SE*	Median statistic	*SD* statistic	Variance statistic	Minimum statistic	Maximum statistic
Total sample (N = 175)	132.70	2.52	130	33.30	1108.62	43	227
4th grade (n = 75)	136.85	3.61	134	31.26	977.13	67	208
5th grade (n = 100)	129.85	3.46	129	34.57	1195.20	43	227

Table 9

DIBELS Next Oral Reading Fluency Descriptive Statistics by Reading Proficiency Level

		Mean						
Quartile	*n*	Statistic	*SE*	Median statistic	*SD* statistic	Variance statistic	Minimum statistic	Maximum statistic
Bottom quartile	27	82.37	3.54	83.00	18.41	338.78	43	109
Middle quartile	109	128.93	1.28	129.00	13.39	179.29	101	162
Upper quartile	39	178.08	3.13	175.00	19.52	381.02	153	227

Silent Reading Fluency

Assumptions of normality. SRF was evaluated with two assessments, the Test of Silent Reading Efficiency and Comprehension and Daze (TOSREC). The TOSREC is an assessment of silent reading of connected text for speed, accuracy, and comprehension (Wagner et al., 2010). The score on the 60-item test are computed by counting the correct number of responses and subtracting the incorrect responses (scored as a -1 to account for guessing). The raw scores can range from 0-60. There is one point per item and a student's score cannot go below 0.

Examination of the score distributions for TOSREC revealed that scores were approximately normally distributed with the skew and kurtosis values less than 1. The skewness was 0.035 (*SD = 0.184)* and Kurtosis was 0.283 (*SD = 0.365)*. Figure 7 shows the distribution of the TOSREC scores. This variable was accepted as normally distributed. There was one outlier on this measure that was three standard deviations above the mean (Figure 8).

Descriptive statistics. TOSREC scores had a range of 50 with 6 being the minimum and 56 being the maximum. The mean was 28.90 (*SD* = 0.680), median of 29, and a mode of 28 for the 175 participants. By grade level, fourth-grade students had a range of 44 with 12 being the minimum and 56 being the maximum. The mean was 27.16 (*SD = 0.988)*, median of 27, and a mode of 23 for the 75 participants. For fifth-grade students, the range was 43 with 6 being the minimum and 49 being the maximum. The mean was 30.20 (*SD = 0.913)*, median of 31, and a mode of 28 for the 100 participants. Table 10 shows descriptive statistics for the total sample, fourth- and fifth-grade levels.

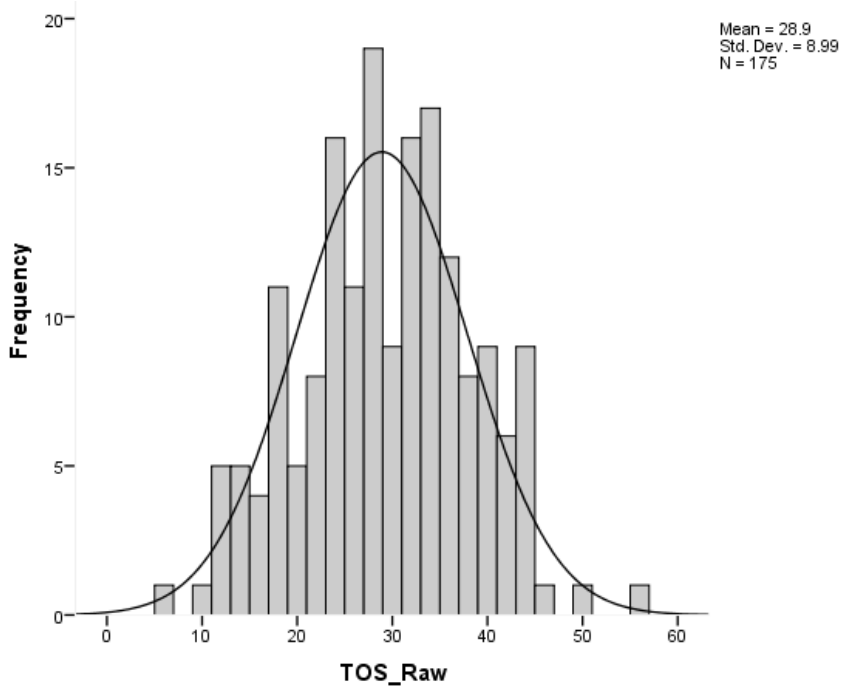


Figure 7. Distribution of Test of Silent Reading Efficiency and Comprehension measure.

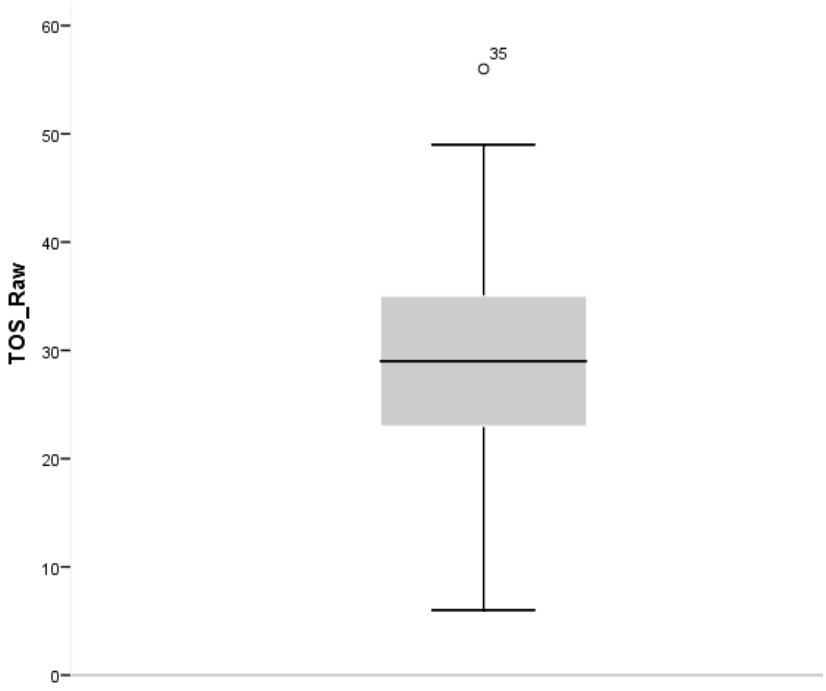


Figure 8. Boxplots of Test of Silent Reading Efficiency and Comprehension scores.

Table 10

Test of Silent Reading Efficiency and Comprehension Descriptive Statistics for All, Fourth- and Fifth-Grade Participants

	Mean						
Sample	Statistic	*SE*	Median statistic	*SD* statistic	Variance statistic	Minimum statistic	Maximum statistic
Total sample	28.90	0.68	29	8.99	80.89	6	56
4th grade	27.16	0.99	27	8.56	73.19	12	56
5th grade	30.20	0.91	31	28.00	83.35	6	49

Daze

Assumptions of normality. Daze, the second measure of SRF assessment, was administered as a component of DIBELS Next. Daze is a cloze measure that replaced approximately every seventh word in the passages with a box containing the correct word and two distracter words. Credit is given if the student selects the words that best fit the omitted words in the reading passage. The number of correct and incorrect responses is recorded. The score is adjusted by subtracting half the number of errors made from the number correct to compensate for guessing. The possible range for the fourth-grade score is 0-57 and 0-63 for fifth grade with the grade level expectation for both grades being a minimum of 24.

Examination of Daze score distributions revealed that scores were approximately normally distributed with the skew and kurtosis values less than 1. Skewness was 0.110 (*SD = 0.184)* and Kurtosis was -0.142 (*SD = 0.365)*. Figure 9 shows the distribution of the Daze scores. This variable was accepted as normally distributed. There were three outliers on the measure. One score was three standard deviations below the mean and two scores were three standard deviations above the mean as presented in Figure 10.

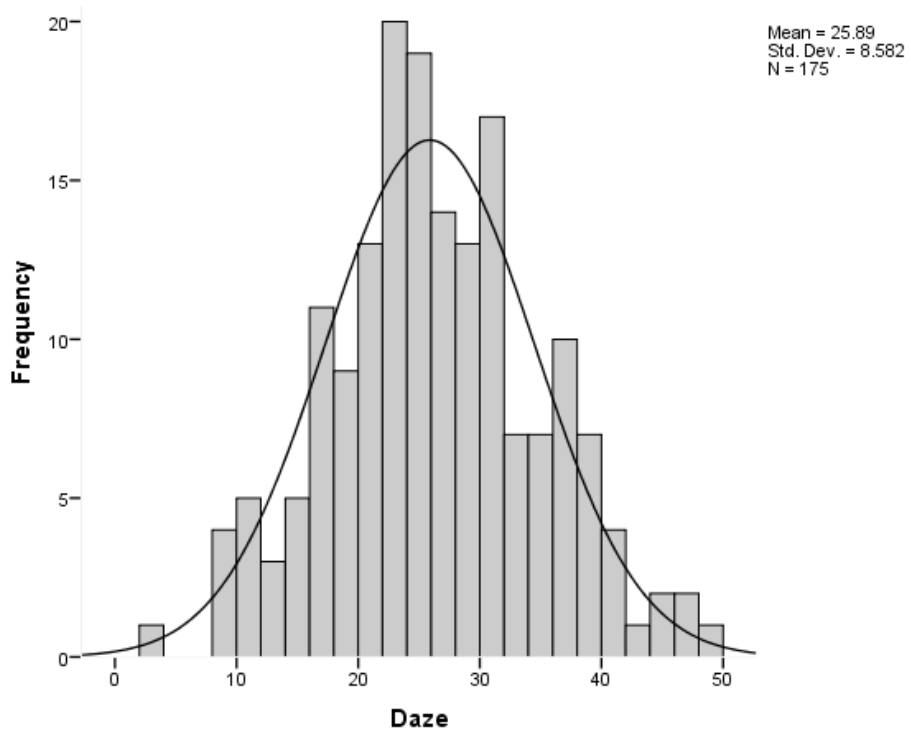


Figure 9. Distribution of Daze measure.

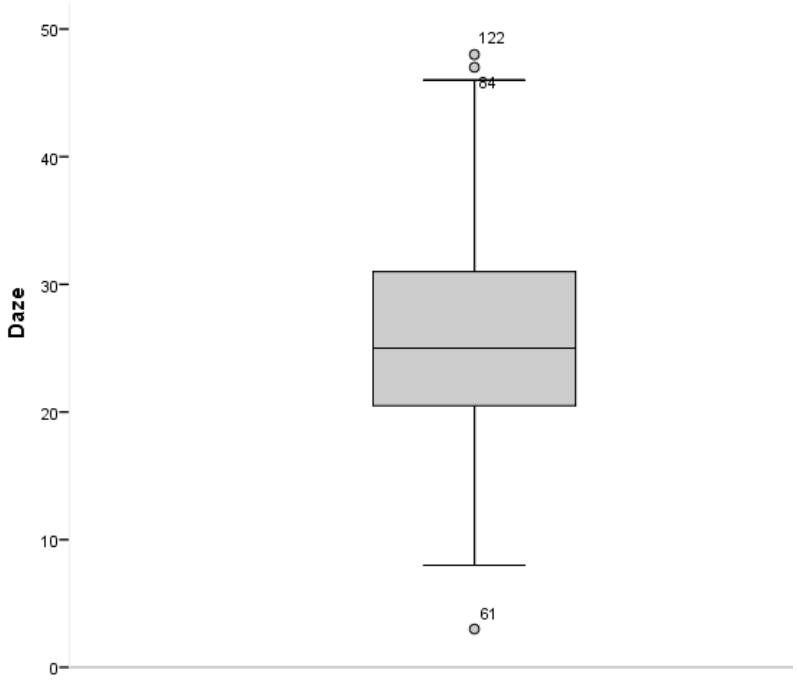


Figure 10. Boxplots of Daze scores.

Descriptive statistics. Daze scores had a range of 45 with 3 being the minimum and 48 being the maximum. The mean was 25.89 (*SD = 0.649)*, median of 25, and a mode of 22 for the 175 participants. For Daze by grade level, fourth-grade students had a range of 39 with 9 being the minimum and 48 being the maximum. The mean was 27.29 (*SD = 1.002)*, median of 26, and a mode of 22 for the 75 participants. For fifth-grade students, Daze had a range of 44 with 3 being the minimum and 47 being the maximum. The mean was 24.83 (*SD = 0.840)*, median of 24, and a mode of 23 for the 100 participants. Table 11 shows the descriptive statistics for the whole group, fourth- and fifth-grade students.

Group Reading Assessment and Diagnostic Evaluation

Assumptions of normality. The Group Reading Assessment and Diagnostic Evaluation (GRADE) was used in this study as a measure of reading comprehension. Students read passages then answered multiple-choice questions. The reading comprehension component had two subtests for sentence and passage comprehension. There are 19 questions in the sentence comprehension and 28 in the passage

Table 11

Daze Descriptive Statistics for All, Fourth- and Fifth-Grade Participants

	Mean						
Sample	Statistic	*SE*	Median statistic	*SD* statistic	Variance statistic	Minimum statistic	Maximum statistic
Total sample	25.89	0.65	25	8.58	73.65	3	48
4th grade	27.29	1.00	26	8.68	75.35	8	48
5th grade	24.83	0.84	24	8.40	70.51	3	47

comprehension. Each item is scored as correct or incorrect; therefore, the range of scores is 0-19 for the sentence comprehension subtest and 0-28 for the passage comprehension subtest. From the total of correct answers, raw scores were obtained.

Examination of the score distributions revealed that scores were approximately normally distributed with the skew and kurtosis values less than 1. The skewness was -0.422 (*SD* = 0.*184)* and Kurtosis was -0.764 (*SD* = 0.*365)*. Figure 11 shows the distribution of the GRADE scores. This variable was accepted as normally distributed. There were no outliers for GRADE scores. Figure 12 shows the boxplot scores.

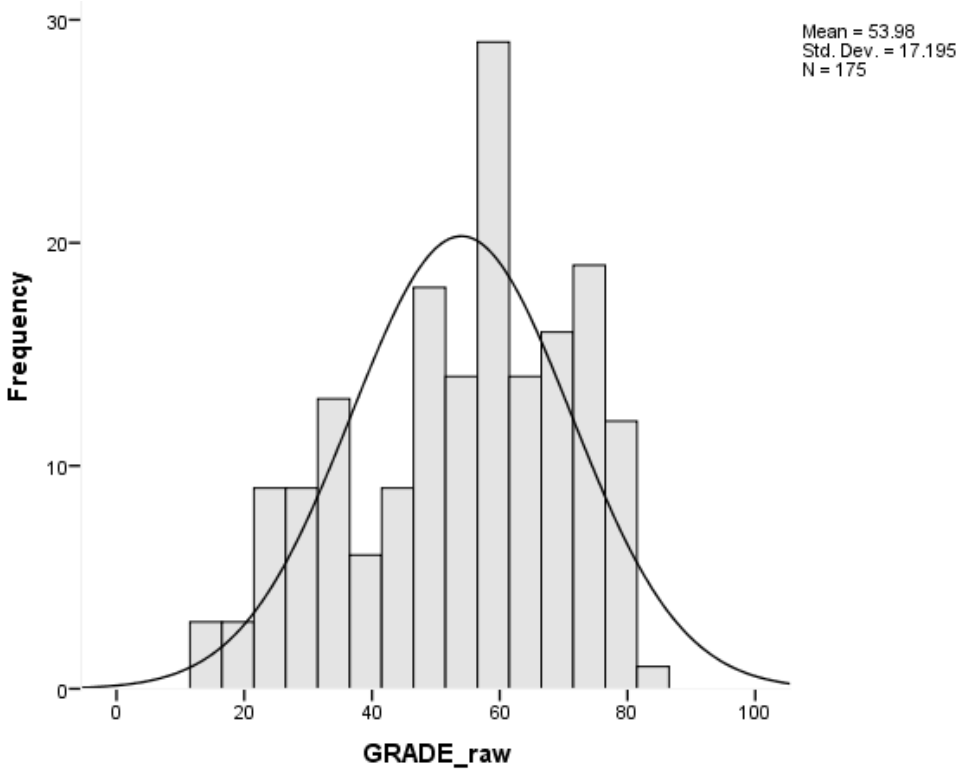


Figure 11. Distribution of Group Reading Assessment and Diagnostic Evaluation measure.

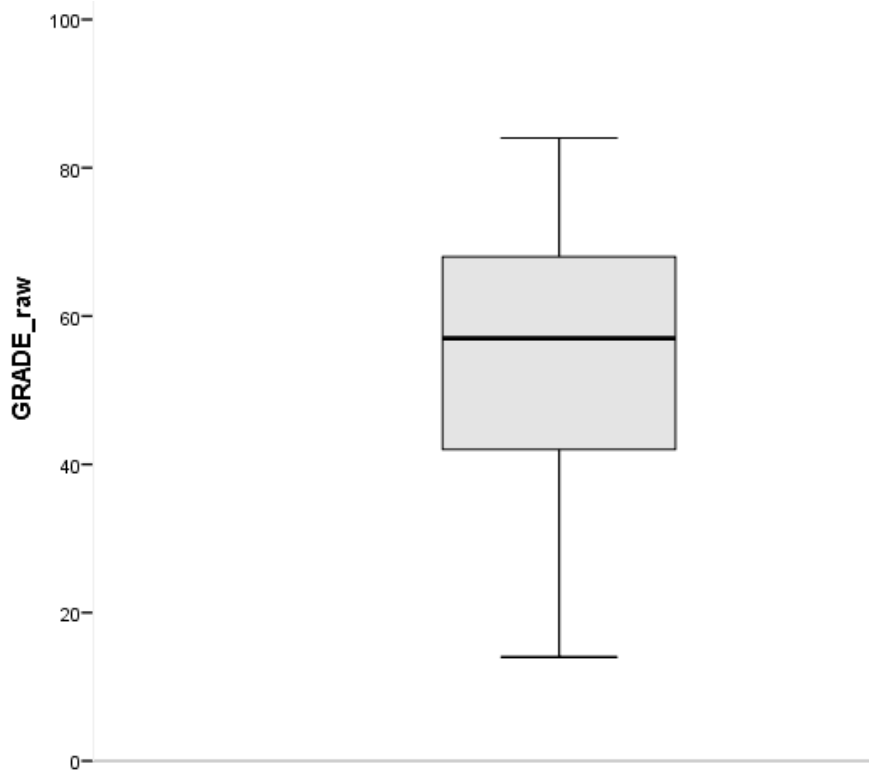


Figure 12. Boxplot of Group Reading Assessment and Diagnostic Evaluation scores.

Descriptive statistics. GRADE had a range of 70 with 14 being the minimum and 84 being the maximum. The mean was 53.98 (*SD = 1.300)*, median of 57, and a mode of 60 for the 175 participants. For GRADE by grade level, fourth-grade students had a range of 57 with 21 being the minimum and 78 being the maximum. The mean was 59.20 (*SD = 1.769)*, median of 62, and a mode of 76 for the 75 participants. For fifth-grade students, GRADE had a range of 70 with 14 being the minimum and 84 being the maximum. The mean was 50.07 (*SD = 1.756)*, median of 54, and a mode of 34 for the 100 participants. Table 12 shows the descriptive statistics for the whole group, fourth- and fifth-grade students.

Table 12

Group Reading Assessment and Diagnostic Evaluation Descriptive Statistics for All, Fourth- and Fifth-Grade Participants

	Mean						
Sample	Statistic	*SE*	Median statistic	*SD* statistic	Variance statistic	Minimum statistic	Maximum statistic
Total sample	53.98	1.30	57	17.20	295.66	14	84
4th grade	59.20	1.77	62	15.32	234.62	21	78
5th grade	50.07	1.76	54	17.56	308.19	14	84

Transitional Colorado Assessment Program

Assumptions of normality. TCAP is Colorado's assessment program that was designed to help schools transition from the content model standards to the new Colorado Academic Standards and was used for this study. The assessment provided a snapshot of the students reading performance with alignment to the expectations set by the state. TCAP was a standardized assessment that required all students to follow the same written protocol and procedures.

Student scores are reported as a scaled score indicating the reading proficiency level. The possible scaled score ranges for this measure are 180-940 fourth grade and 220-955 for fifth grade. Examination of the score distributions revealed that scores were approximately normally distributed with the skew and kurtosis values less than 1. The skewness was -0.574 (*SD = 0.184)* and Kurtosis was .653 (*SD = 0.365)*. Figure 13 shows the distribution of the TCAP scores. This variable was accepted as normally distributed. There were six outliers on the measure. Five scores were three standard deviations below the mean and one score was three standard deviations below the mean as presented in Figure 14.

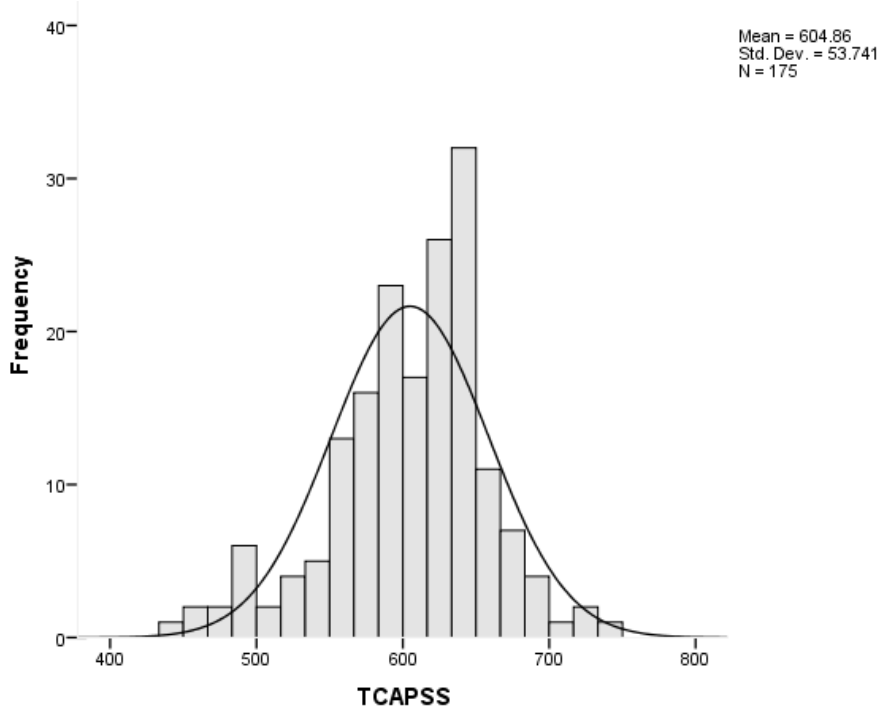


Figure 13. Distribution of Transitional Colorado Assessment Program measure.

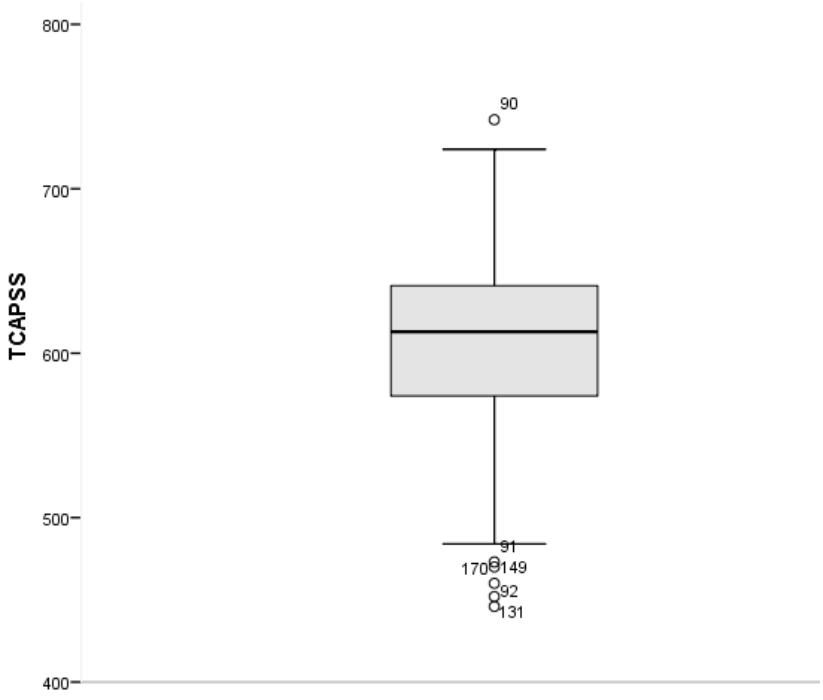


Figure 14. Boxplots of Transitional Colorado Assessment Program scores.

Descriptive statistics. TCAP had a range in scores of 296 with 446 being the minimum and 742 being the maximum. The mean was 604.86 (*SD = 4.062)*, median of 613, and a mode of 627 for the 175 participants. When looking at TCAP by grade level, fourth-grade students had a range of 209 with 491 being the minimum and 700 being the maximum. The mean was 596.95 (*SD = 4.683)*, median of 597, and a mode of 570 for the 75 participants. For fifth-grade students, TCAP had a range of 296 with 446 being the minimum and 742 being the maximum. The mean was 610.80 (*SD = 6.135)*, median of 623.50, and a mode of 636 for the 100 participants. Table 13 shows the descriptive statistics for the whole group, fourth- and fifth-grade students.

Correlations

Correlation coefficients were computed for the five measures. All measures are significantly correlated at the $p < .01$ level (Table 14).

Table 13

Transitional Colorado Assessment Program Descriptive Statistics for All, Fourth- and Fifth-Grade Participants

	Mean						
Sample	Statistic	*SE*	Median statistic	*SD* statistic	Variance statistic	Minimum statistic	Maximum statistic
Total sample	604.86	4.06	613	53.74	2888.06	446	742
4th grade	596.95	4.68	597	40.56	1644.67	491	700
5th grade	610.80	6.14	623.50	61.35	3763.56	446	742

Table 14

Correlations Among Measures

Measure	DORF	TOSREC	Daze	GRADE
DORF				
TOSREC	0.680*			
Daze	0.743*	0.713*		
GRADE	0.673*	0.613*	0.730*	
TCAP	0.690*	0.776*	0.729*	0.777*

* $p < .01$.

Regression and Structural Equation Model Results

This study sought to investigate the effect of measures of ORF and SRF as predictors of reading comprehension and as predictors of student achievement on high stakes reading test at the fourth- and fifth-grade level.

Comparison of Oral Reading Fluency and Silent Reading Fluency as Predictors of Reading Comprehension

The first research question focused on comparison of an ORF measure and SRF measure as predictors of reading comprehension for fourth- and fifth-grade students. It was hypothesized that the SRF measure would be a stronger predictor of reading comprehension because research suggests correlations between ORF measures and reading proficiency decreases for maturing readers (Wagner, 2011). This analysis included three models for ORF measures and three models for SRF measures.

Analysis began with models to analyze the association between reading comprehension (as measured by the GRADE) and the ORF measure as moderated by

grade level. The linear regression model was used to analyze the variance in reading comprehension as accounted for in the ORF measure for the total sample and by grade level. The first model with reading comprehension regressed on ORF for the total sample accounted for 45.0% of the variance $F(1,173) = 143.16$, $p < .001$. Reading comprehension regressed on ORF by grade level accounted for 49.2% of the variance $F(1,73) = 72.69$, $p < .001$ for fourth grade and 41.9% of the variance $F(1,98) = 72.37$, $p < .001$ for fifth grade. Results of the model show that the ORF measure is a significant predictor of reading comprehension ($t = 11.965$, $p = <.001$). ORF is also a significant predictor of reading comprehension at the fourth-grade level ($t = 8.526$, $p = <.001$) and fifth-grade level ($t = 8.507$, $p = <.001$).

Table 15 and Figure 15 present information for the reading comprehension and ORF models. Table 16 shows the impact of the ORF measure on reading comprehension for fourth to fifth grade.

Three additional models were used to analyze the association between reading comprehension (as measured by the GRADE) and the SRF measure as moderated by grade level. The study included two measures for SRF: TOSREC and Daze. While both measures were completed by the participants in the study based on the reliability of each

Table 15

Analysis of Oral Reading Fluency Measure on Reading Comprehension for Participants

Oral reading proficiency	*B*	*SE B*	*β*	*t*	*p*
Total sample	0.348	0.029	0.673	11.965	<.001
4th grade	0.346	0.041	0.706	8.526	<.001
5th grade	0.331	0.039	0.652	8.507	<.001

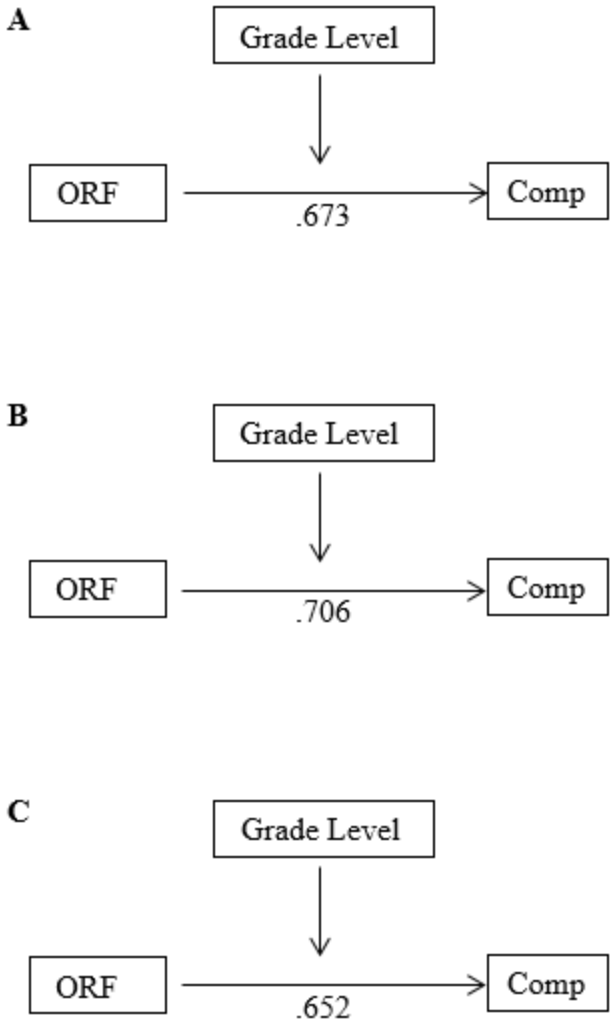


Figure 15. Regression of oral reading fluency measure on reading comprehension for total participants (A); fourth-grade participants (B); fifth-grade participants (C).

Table 16

Impact of Oral Reading Fluency Measure on Reading Comprehension for Participants

Oral reading fluency	*R*	R^2	Adjusted R^2	*F*	*p*
Total sample	0.673	0.453	0.450	143.155	<.001
4th grade	0.705	0.499	0.492	72.689	<.001
5th grade	0.652	0.425	0.419	72.369	<.001

assessment, only one assessment was included in the final analysis. As the two measures were highly correlated ($r = 0.713$), Daze was selected for inclusion in the regression model as it is accessible and frequently used by teachers (Riedel, 2007). Reading comprehension regressed on the SRF measure accounted for 53.0% of the variance $F(1,173) = 197.39$, $p < .001$. Reading comprehension regressed on SRF by grade level account for 43.6% of the variance $F(1,73) = 58.32$, $p < .001$ for fourth grade and 58.7% of the variance $F(1,98) = 141.94$, $p < .001$ for fifth grade. Results of the model show that for the SRF measure is a significant predictor of reading comprehension ($t = 14.050$, $p = <.001$). SRF is also a significant predictor of reading comprehension at the fourth-grade level ($t = 7.636$, $p = <.001$) and fifth-grade level ($t = 11.914$, $p = <.001$). Thus, as the grade level increased, the SRF measure accounted for more variance as predicted indicating that measures of SRF are potentially a better measure for predicting reading comprehension.

Table 17 and Figure 16 present information on reading comprehension and the SRF measure models. Table 18 shows the impact of the SRF measure on reading comprehension from fourth to fifth grade.

Table 17

Analysis of Daze on Reading Comprehension for Participants

Daze	*B*	*SE B*	*β*	*t*	*P*
Total sample	1.463	0.104	0.730	14.050	<.001
4th grade	1.176	0.154	0.666	7.636	<.001
5th grade	1.608	0.135	0.769	11.914	<.001

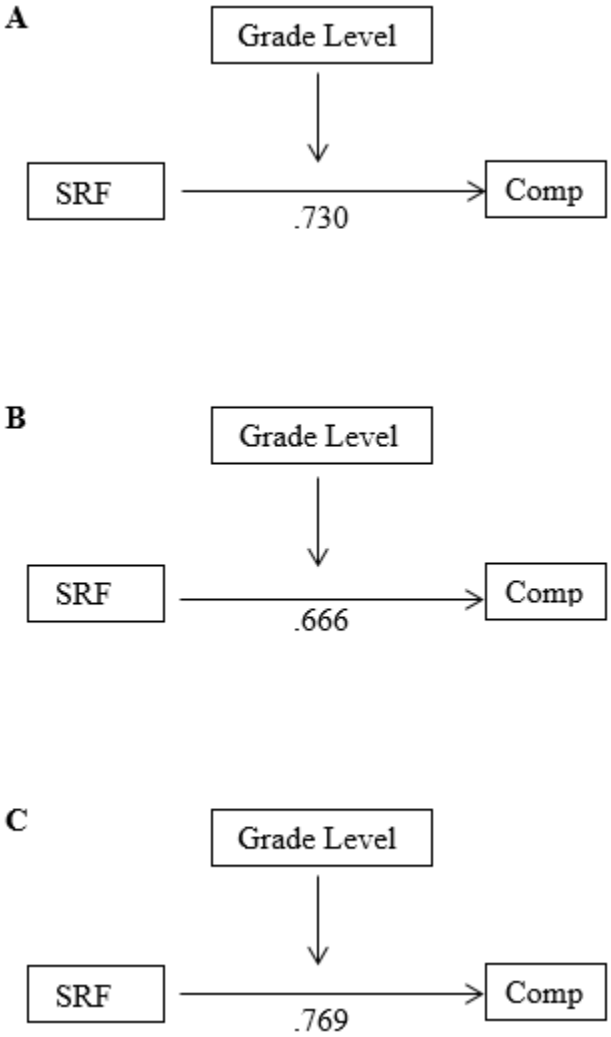


Figure 16. Regression of silent reading fluency measure on reading comprehension for total participants (A); fourth-grade participants (B); fifth-grade participants (C).

Table 18

Impact of Daze on Reading Comprehension for Participants

Daze	*R*	R^2	Adjusted R^2	*F*	*P*
Total sample	0.730	0.533	0.530	197.393	<.001
4th grade	0.666	0.444	0.436	58.316	<.001
5th grade	0.769	0.592	0.587	141.937	<.001

Oral Reading Fluency and Silent Reading Fluency as Predictors as a Function of Reading Proficiency Level

The second question addressed how the relationship of measures of ORF and SRF as predictors of reading comprehension varies as a function of reading proficiency level. It was expected that the SRF measure would be a stronger predictor of comprehension for students with a higher reading proficiency level because as students develop stronger reading proficiency, they are more likely to read silently (Price et al., 2012). Conversely, oral reading fluency is likely to be a better predictor for students with lower reading proficiency (Cook, 2003; Riedel, 2007).

Reading Proficiency was examined as a moderator of the relation between a measure of SRF and reading comprehension. Reading proficiency levels were classified using the Hasbrouck and Tindal (2006) percentile rank based on spring scores for WCPM. Three classifications representing participants scoring were created: bottom quartile (with a 25th percentile or lower), middle quartiles (26th-74th percentile), and top quartile (75th percentile or above). Proficiency level was analyzed as a whole-group sample as the sample size did not allow for analysis at the grade level. The variables for the SRF measure and Reading Proficiency Level were centered prior to analysis. This model accounted for a significant amount of the variance, R^2= 0.59, $F(3,171) = 140.95$, $p < .001$. Results indicated that reading proficiency level, $t(171) = 4.12$, $p = .0001$, and SRF, $t(171) = 8.78$, $p < .000$, were both associated with reading comprehension. The interaction between the SRF measure and reading proficiency level was also significant $t(171) = 4.67$, $p < .000$. The interaction between the SRF measure and reading

proficiency accounted for an *R*-squared increase of 0.031, $F(1, 171) = 21.7614$, $p < .000$. Thus, Reading Proficiency Level is a significant moderator of the relationship between the SRF measure and reading comprehension. Table 19 shows the model results. Figure 17 presents the slopes for the moderating effect of reading proficiency. Examination of the slopes shows that as reading proficiency increases, the correlation between the SRF measure and reading comprehension increases. Figure 18 shows the model effect for the SRF measure and reading comprehension by proficiency level.

Table 19

Model Results

Results	Coefficient	*SE*	*T*	*P*
Constant	55.794	1.055	52.884	0.000
DORF PR	5.028	1.220	4.121	0.0001
Daze	1.123	0.128	8.776	0.000
Int-1	-0.331	0.071	-4.665	0.000

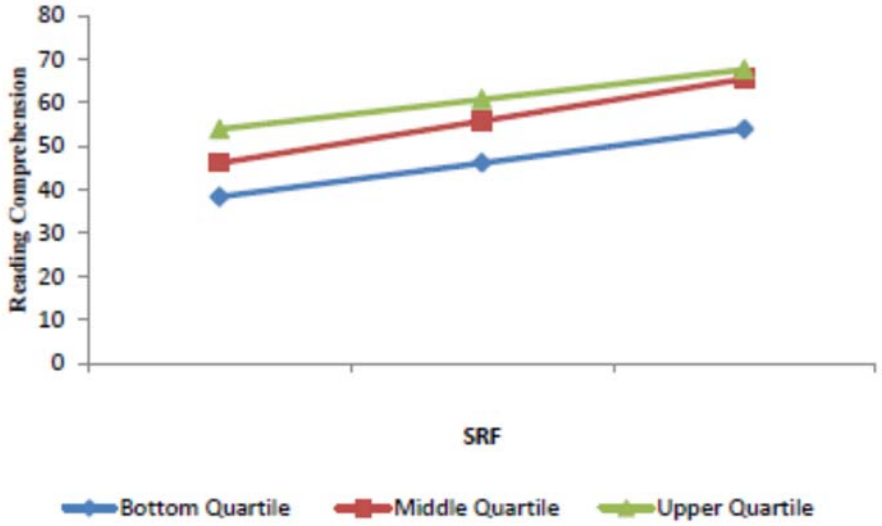
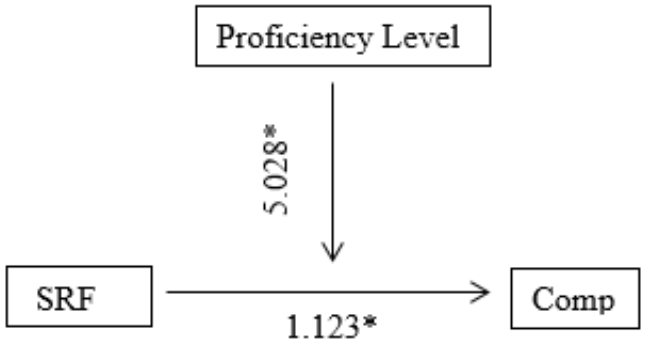


Figure 17. Slopes for the moderating effect of proficiency level.



*$p < .001$

Figure 18. Model effect for the SRF measure and reading comprehension by proficiency level.

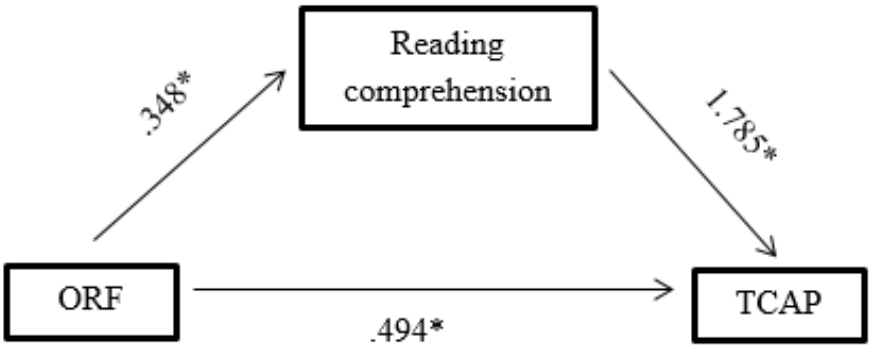
Relationship of Oral Reading Fluency, Silent Reading Fluency, Reading Comprehension, and Transitional Colorado Assessment Program

The third question analyzed the relationship of ORF measures, SRF measures, reading comprehension and the TCAP reading proficiency level for fourth- and fifth-grade students. It is expected that measures of SRF will have a stronger relationship to the TCAP reading proficiency than oral reading fluency measures and reading comprehension as it aligns to classroom practices.

An analysis of the relationship of measures of ORF or SRF and TCAP as mediated by reading comprehension was conducted. The first mediation model analyzed the relationship of the ORF measure, reading comprehension, and TCAP. This model accounted for 45.3% of the variance $F(1,173) = 143.15$, $p < .001$ with a coefficient of 0.3475 for the indirect effect of ORF on reading comprehension. The impact of the ORF measure on reading comprehension was $t = 11.965$, $p = <.001$. The significance was tested using bootstrapping procedures using 1,000 bootstrapped samples with a 95% confidence interval. The indirect effect was 0.6201and the 95% confidence interval range

from 0.4725, 0.8105. The indirect effect was statistically significant. The direct effect of the ORF measure on TCAP accounted for 65.5% of the mediation $F(2,172) = 163.40$, $p < .001$. The impact of the ORF measure on TCAP was $t = 5.060$, $p < .001$. Figure 19 presents the standardized regression of the ORF measure, reading comprehension and TCAP. Table 20 presents the outcome on TCAP.

The second model analyzed the relationship of the SRF measure and TCAP as mediated by reading comprehension. The second model accounted for 53.3% of the variance $F(1,173) = 197.39$, $p < .001$ with a coefficient of 1.4626 of the SRF measure mediated by reading comprehension. The impact of the SRF measure on reading



*$p < .001$.

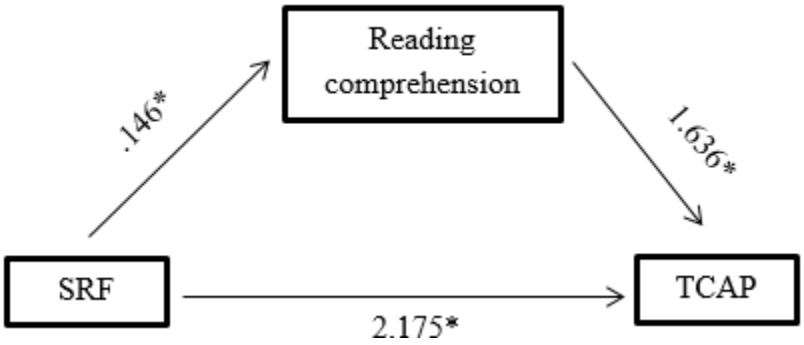
Figure 19. Mediation model for the oral reading fluency measure, reading comprehension, and Transitional Colorado Assessment Program.

Table 20

Transitional Colorado Assessment Program Outcome for DIBELS Oral Reading Fluency (DORF) and Group Reading Assessment and Diagnostic Evaluation

Variable	Coefficient	*SE*	*t*	*P*
DORF	0.4943	0.0977	5.0599	< .001
GRADE	1.7846	0.1892	9.4334	< .001

comprehension was $t = 14.050$, $p = <.001$. The indirect effect of reading comprehension on TCAP from the SRF measure had a coefficient of 1.6363. The significance was tested using bootstrapping procedures using 1,000 bootstrapped samples with a 95% confidence interval. The indirect effect was 2.3933 the 95% confidence interval range from 1.7438, 3.2007. The indirect effect was statistically significant. The direct effect of the SRF measure on TCAP accounted for 66.0% of the mediation $F(2,172) = 167.08$, $p < .001$. The impact of the SRF measure on TCAP was $t = 5.340$, $p <.001$. Figure 20 presents the standardized regression of the SRF measure, reading comprehension and TCAP. Table 21 present the outcome on TCAP.



*$p < .001$.

Figure 20. Meditation model for the silent reading fluency measure, reading comprehension, and Transitional Colorado Assessment Program.

Table 21

Transitional Colorado Assessment Program Outcome for Daze and Group Reading Assessment and Diagnostic Evaluation

Variable	Coefficient	*SE*	*t*	*P*
Daze	2.1747	0.4073	5.3398	< .001
GRADE	1.6363	0.2033	8.0499	< .001

Summary

This study evaluated how measures of oral and silent reading fluency compare as predictors of reading comprehension and how these vary as a function of proficiency level for fourth- and fifth-grade students. Additionally, this study sought to examine the relationship between ORF measures, SRF measures, reading comprehension, and the TCAP with these students. The study used a correlational prediction design with measure for the variables of measures of ORF, SRF, reading comprehension, and TCAP. The participants were 175 fourth- and fifth-grade students from two randomly selected schools in a school district in Colorado.

Descriptive statistics show that all measures are highly correlated and approximately normally distributed. There was a high correlation for reading comprehension with the SRF measure ($r = 0.730$) over the ORF measure ($r = 0.673$) overall, although both measures were related to reading comprehension. When analyzed by grade level, fourth grade has a higher correlation for the ORF measure ($r = 0.706$) than the SRF measure ($r = 0.666$) and fifth grade has a higher correlation for the SRF measure ($r = 0.769$) than the ORF measure ($r = 0.652$).

Regression analysis was conducted to evaluate the impact of measures of ORF and SRF on reading comprehension. Both ORF and SRF measures are significant predictors of reading comprehension. However, the SRF measure accounted for 53.3% of the variance and the ORF measure accounted for 45.0% of the variance for the total sample. Results of this analysis for ORF and SRF measures show that SRF has as stronger direct effect on reading comprehension. Additionally, results of the study

indicate that as grade level increased from fourth to fifth grade, the SRF measure was a stronger predictor of reading comprehension.

Participants' reading proficiency level of ORF based on Hasbrouck and Tindal (2006) percentile ranks for SRF on reading comprehension is strongly related. Using the three classifications of bottom, middle, and top quartile, the SRF measure accounted for 59.01% of the moderation. The proficiency level impact of the SRF measure on reading comprehension was $r = 0.7682$. Analysis of the relationship of ORF measures, SRF measures, reading comprehension, and TCAP, showed the ORF measure and the SRF measure had similar relationships with TCAP with ORF accounting for 65.5% of the variance and SRF accounting for 66.0% of the variance. The direct effect for measures of ORF and SRF on TCAP show minimal difference in effect.

CHAPTER V

DISCUSSION

The purpose of this study was to evaluate how measures of oral and silent reading fluency compare as predictors of reading comprehension and how these vary as a function of reading proficiency level for fourth- and fifth-grade students. Additionally, the study sought to examine the relationship between ORF) measures, SRF measures, reading comprehension, and the TCAP with these students. This study used a correlational prediction design with measures for the variables of ORF, SRF, reading comprehension, and TCAP. Participants were 175 fourth- and fifth-grade students from two randomly selected schools in a school district in Colorado.

The instrumentation for this study included reading assessments that have been adopted by districts and states across the nation and measures that align with classroom practice at the fourth- and fifth-grade level. With the state assessment being the driving force behind school performance, schools need predictive indicators that align with reading outcomes that can be used to predict student results. Assessments that measure ORF, SRF, and reading comprehension are needed to support and guide reading instruction in the classroom to reach proficiency levels set by the state.

DIBELS Next ORF was selected as it was one of the state approved assessments for reading proficiency and the district adopted reading component. DIBELS is now in its 7th edition and is commonly known as DIBELS Next (Good et al., 2013). SRF was evaluated with two assessments, the Test of Silent Reading Efficiency and Comprehension and DIBELS Daze. TOSREC is an assessment of silent reading of

connected text for speed, accuracy, and comprehension (Wagner et al., 2010). The second SRF assessment, Daze, is a component of DIBELS Next, is a 3-minute, whole class administered measure (Good et al., 2011). The GRADE was used in this study as a measure of reading comprehension. The state assessment of interest in this study is the TCAP that aligns with the State adopted Colorado Academic Standards and integrated the Common Core Standards.

Predictors of Reading

Literacy is the foundation of learning and considered a discipline rather than a subject (Dole, 2003; Paris & Hamilton, 2009). Educators are relying on accurate measures to aid identification of students at risk, monitor student progress, and guide instructional practices (Buly & Valencia, 2002, 2003; Goffreda & DiPerna, 2010; Kamil, Afflerbach, Pearson, & Birr Moge, 2011; Pyle & Vaughn, 2012). This study sought a comparison of the ORF measure with a measure of SRF for students in fourth and fifth grade to predict reading comprehension, determine if there is a relationship as a function of proficiency, and the relationship of ORF measures, SRF measures, and TCAP. As an operating hypothesis, it was predicted that measures of SRF would be a stronger predictor of reading comprehension than measures of ORF and also a stronger predictor for students with a high reading proficiency level. SRF measures were also predicted to have a stronger relationship to the TCAP reading proficiency than oral reading fluency measures and reading comprehension. Currently, teachers rely heavily on measures of ORF as the main data source but ORF decreases as students grow older and assessments

that can be predictors of reading comprehension are needed. Correlations between measures of ORF and reading comprehension decrease for maturing readers (Wagner, 2011). This is likely due to students' transition to silent reading. The transition from oral reading to silent reading begins in late second or third grade and is more firmly established in fourth and fifth grade (Johnson et al., 2011; Kim et al., 2012; Wagner, 2011).

Preliminary analysis and descriptive statistics were used to evaluate the means, standard deviation, minimum and maximum scores. Additionally, reliability coefficients and correlations between variables were reviewed. Secondary analysis included structural equation modeling (SEM) to examine the relationships between ORF measures, SRF measures, reading comprehension and the Colorado State reading assessment. Regression analysis and structural equation modeling was used to examine the relationship of ORF measures, SRF measures, reading comprehension, and TCAP for fourth and fifth graders, and how the relationship may vary based on grade level and reading proficiency level.

The discussion of the results are organized around the three research questions.

1. How do oral reading fluency measures and silent reading fluency measures compare as predictors of reading comprehension for fourth and fifth grade students?

2. Does the relationship of oral reading fluency measures and silent reading fluency measures as predictors of reading comprehension vary as a function of reading proficiency level?

3. What is the relationship of oral reading fluency measures, silent reading fluency measures, and the high-stakes measure for Colorado students (the Transitional Colorado Assessment Program) for fourth and fifth grade students?

Comparison of Oral Reading Fluency and Silent Reading Fluency as Predictors of Reading Comprehension

The first research question focused on comparison of measures of ORF and SRF as predictors of reading comprehension for fourth- and fifth-grade students. The null hypothesis was that there would be no difference between oral and silent reading fluency measures as predictors of reading comprehension. It was hypothesized that measures of SRF would be a stronger predictor of reading comprehension than oral reading fluency measures because research suggests correlations between ORF and reading proficiency decrease for maturing readers (Wagner, 2011).

ORF was measured with DIBELS Next (Good et al., 2013). Participants' WCPM scores ranged from 67 to 208 for fourth grade and 43 to 227 for fifth grade. SRF was evaluated with two assessments, the TOSREC and DIBELS Daze. In the study, the TOSREC and Daze assessments were closely correlated, $r(173) = +0.713$, $p < .001$; therefore, Daze was selected for use in the analysis on the basis that it is a free measure and accessible to teachers (Riedel, 2007).

Results of the regression analysis indicated that 45.0% of the variance in reading comprehension was accounted for by the ORF measure for the sample population, $F(1, 173) = 143.15$, $p < .001$ as compared to 53.0% of the variance accounted for by the SRF measure, $F(1, 173) = 197.39$, $p < .001$. When analyzed by grade level 49.2% of the variance in reading comprehension was accounted for by the ORF measure for fourth grade, $F(1, 73) = 72.69$, $p < .001$ and 43.6% of the variance was accounted for by the SRF measure, $F(1, 73) = 58.32$, $p < .001$. For fifth grade, 41.9% of the variance in reading

comprehension was accounted for by the ORF measure, $F(1, 98) = 72.37$, $p < .001$ and 58.7% of the variance was accounted for by the SRF measure, $F(1, 98) = 141.94$, $p < .001$. For fifth-grade participating students, the SRF measure had a higher correlation with reading comprehension than the ORF measure. However, the ORF measure had a higher correlation with reading comprehension than the SRF measure for fourth-grade participating students. Table 22 shows the correlations for the measures of ORF and SRF with reading comprehension

Results show that as grade level increases, the relationship between ORF measures, SRF measures, and reading comprehension changes. The ORF measure has a higher correlation with reading comp for fourth-grade students, while the SRF measure has a higher correlation than the ORF measure with reading comprehension for fifth-grade students. The null hypothesis was rejected as results indicate there was a difference between measures of ORF and SRF as predictors of reading comprehension. The hypothesized result that the SRF measure would be a stronger predictor of reading comprehension was accepted for fifth-grade students. This finding aligns with the results of Riedel (2007) who reported ORF prediction of reading comprehension with 80%

Table 22

Oral Reading Fluency and Silent Reading Fluency Measures Reading Comprehension Comparison

	Oral reading fluency			Silent reading fluency		
Correlation	Total	4th	5th	Total	4th	5th
Adjusted R^2	0.450	0.492	0.419	0.530	0.436	0.587
F	143.155	72.689	72.369	197.393	58.316	141.937
P	< .001	< .001	< .001	< .001	< .001	< .001

variance for first grade and 71% variance for second grade and Salvador et al. (2009) with 66% variance for third grade. The results from this study help to complete this scale with 49% variance for fourth grade and 42% variance for fifth grade. Thus, we see the declining correlation of ORF with reading comprehension as students progress throughout the elementary grades. These results also support the findings of Denton et al. (2011) who reported a lower relationship between ORF and reading comprehension for 6th to 8th grade students (r = 0.50 - 0.51) than was found with younger elementary students. The results of this study align with Miller and Smith (1990), who noted that silent reading measures may be more conducive to reading comprehension for "good" readers for third- through fifth-grade students than oral reading. Accurate measures to reflect reading comprehension have been sought by teachers to support learning that aligns to classroom practice. With data driven instruction and practices becoming a driving force for student progress, assessment results need to accurately predict student outcomes. Aligning classroom reading practices with predictive outcome measures is a needed component to support student success. Thus, measures of SRF might be a better predictor for maturing readers to determine reading proficiency, monitor student progress, and guide instructional practices.

Oral Reading Fluency and Silent Reading Fluency as Predictors as a Function of Reading Proficiency Level

The second question addressed how the relationship of ORF measures and SRF measures as predictors of reading comprehension varied as a function of reading proficiency level. The null hypothesis was that reading proficiency level would not

influence the relationship of the SRF measure as a predictor of reading comprehension. It was hypothesized that silent reading fluency would be a stronger predictor of reading comprehension for students with a higher reading proficiency level because as students develop stronger reading proficiency, they are more likely to read silently (Price et al., 2012). Conversely, oral reading fluency is likely to be a better predictor for students with lower reading proficiency (Cook, 2003; Riedel, 2007).

Reading proficiency levels were classified using the Hasbrouck and Tindal (2006) percentile rank based on spring scores for WCPM. Participants' DIBELS Next ORF scores were used to create three classifications representing participants scoring in the bottom quartile (with a 25th percentile or lower), participants scoring in the middle quartiles (26th-74th percentile), and participants scoring in the top quartile (75th percentile or above).

The analysis for the SRF measure by reading proficiency was conducted at the whole group level as the number of participants at the fourth- and fifth-grade level did not allow for grade-level analysis. The model accounted for 59.01% of the moderation $F(3, 171) = 140.95$, $p < .001$. Results indicated that reading proficiency level, $t = 4.12$, $p = .0001$ and the SRF measure, $t = 8.78$, $p = .0000$ were both associated with reading comprehension. The interaction between reading proficiency level and the SRF measure was also significant $t(171) = 4.67$, $p = .000$ and the interaction between reading proficiency and the SRF measure accounted for an *R*-squared increase of 0.031. Reading proficiency level is a significant moderator of the relationship between reading comprehension and the SRF measure.

In this study, results indicated a moderate-high impact on reading comprehension for the SRF measure when moderated by proficiency level for the sample population. The null hypothesis was rejected reading proficiency was a significant moderator of the relationship between the SRF measure and reading comprehension. For students in all quartiles, as reading proficiency level increased, reading comprehension also increased. Overall, this finding supports Wagner (2011) who divided students into three subgroups and found skilled readers were the top third in reading comprehension based on SRF. The results support Price et al. (2012) findings that SRF and reading comprehension measures had a strong correlation for fourth and sixth graders. The SRF measure is a strong predictor of reading comprehension for older students and proficiency level is a strong component for predicting reading comprehension.

Relationship of Oral Reading Fluency, Silent Reading Fluency, Reading Comprehension, and Transitional Colorado Assessment Program

The relationship of ORF measures, SRF measures, reading comprehension and the TCAP reading proficiency level for fourth- and fifth-grade students was also examined in this study. The null hypothesis was that there would be no relationship between ORF measures, SRF measures, reading comprehension and the TCAP. It was hypothesized that silent reading fluency would have a stronger relationship to the TCAP reading proficiency than ORF as it aligns to classroom practices.

TCAP is Colorado's assessment program that provided a snapshot of the students reading performance with alignment to the expectations set by the state. Student scores are reported as a scaled score indicating the reading proficiency level. The possible scaled

score ranges for this measure are 180-940 fourth grade and 220-955 for fifth grade.

Results of the structural equation model indicated that the SRF measure has a higher indirect effect on reading comprehension, $F(1, 173) = 197.39$, $p < .001$ than the ORF measure, $F(1, 173) = 143.15$, $p < .001$. This accounted for 53.3% of the variance on reading comprehension from the SRF measure and 45.3% of the variance on reading comprehension from the ORF measure. The direct effect on TCAP from the SRF measure accounted for 66.0% of the variance $F(2, 172) = 167.078$, $p < .001$ while the ORF measure accounted for 65.5% of the variance $F(2, 172) = 163.404$, $p < .001$. The indirect effect of reading comprehension on TCAP was tested using a bootstrap estimation approach with 1,000 samples and 95% confidence interval. These results indicated the indirect coefficient was from the SRF measures 2.39 with bootstrap $SE = 0.366$ while the ORF measure was 0.62 with bootstrap $SE = 0.083$. Table 23 shows the comparison of models for measures of ORF and SRF with TCAP.

Results show that both measures of ORF and SRF had a significant correlation with TCAP. The null hypothesis was rejected as there is a relationship between the ORF measure, the SRF measure, reading comprehension, and TCAP. The hypothesized result was accepted as the SRF measure has a stronger relationship to TCAP than the ORF

Table 23

Comparison of Models for Measures of Oral Reading Fluency and Silent Reading Fluency on Transitional Colorado Assessment Program

Measure	Adjusted R^2	*F*	*P*
Oral reading fluency	0.655	163.404	< .001
Silent reading fluency	0.660	167.078	< .001

measure. Additionally, the ORF measure had a strong relationship with TCAP. Results of this study align with previous research showing that DIBELS has correlation coefficients results between 0.65 - 0.80 with selected state assessments (Good et al., 2001; Johnson et al., 2011; Roehrig et al., 2008; Shaw & Shaw, 2002; Stage & Jacobsen, 2001; Vander Meer et al., 2005). The direct effect on the different measures showed that measures of SRF and ORF had a statistical significance on the TCAP.

Results from this study indicated the direct effect on TCAP from measures of ORF or SRF was statistical significance and the indirect effect of the SRF measure had a higher correlation on reading comprehension than the ORF measure. Therefore, teachers can rely on SRF measures for upper elementary students as predictive measures on high-stakes assessments.

Limitations

One limitation of this study is the number of nonparticipating students from the total possible sample population. The total sample population had 272 students, of which 219 originally agreed to participate in the study. However, 95 students opted out of district-mandated tests prior to the start of the study, as allowed in the state of Colorado (CDE, 2015) or did not have sufficient data for inclusion in the study. Opting out of State assessments, such as TCAP, can be either by parent or student choice.

These 95 students reduced the sample size by 35% of the total sample population. To evaluate if a representative sample population completed all measures, analysis was conducted to compare participating students with non-participating students who had

agreed to be part of the study, but later opted out of state testing. The students who opted out were nonminority (70%) and did not qualify for free/reduced lunch status (67%). This aligns with Colorado opt out students who were more likely to be nonminority and less likely to qualify for free/reduced lunch status **(**Bennett, 2016). Results from the *t* test indicated that there were differences between the two groups on the DORF $t(270) = 2.21$, $p < .03$, and Daze $t(270) = 2.29$, $p < .02$. Therefore, the findings of this study can be used to can be used to determine predictability and relationships between ORF measures, SRF measures, reading comprehension, and TCAP. A second limitation of the study was that this study used a convenience sample from one district in Colorado with about 27% diversity. This limits the generalizability to similar populations.

Future research could address these limitations by providing a longer window that would allow students to make up missed assessments. Additionally, parent meetings to answer questions and help parents understand how the assessment results could be used to support instruction in the classroom and student achievement could be scheduled. This would potentially increase student participation in all measures. Larger sample sizes could also allow for data analysis of the moderating effect of reading proficiency on the relationship of SRF measures to reading comprehension by grade level. Extending the research to include additional grade levels and more diverse populations could also be addressed in future research.

As indicated in the study, ORF measures are not the only assessment for determining a student's reading level. Research could be extended by including additional SRF measures that are available to teachers to provide information about

fluency and how these are correlated to reading comprehension and state assessments utilizing the same model from this study. Additional study designs could also be conducted to further the information available on fluency and the relationship to comprehension. In spite of these limitations, this study provide evidence that the SRF measure is a strong predictor of reading comprehension for fifth-grade students, that the relationship of ORF measures, SRF measures, and reading comprehension changes over time, the moderating effect between SRF measures and reading comprehension is significant and that SRF is a predictive measure with high-stakes State assessments.

Implications for Practice

The transition from oral reading to silent reading begins in late second or third grade and is more firmly established in fourth and fifth grade (Johnson et al., 2011; Kim et al., 2012; Wagner, 2011)**.** However, many districts continue to use ORF measures to assess and progress monitor students in upper elementary grades. This has placed an interesting paradox between practice and assessment which has left measures of SRF for upper elementary students overlooked and understudied (Share, 2008). Generally, oral reading happened when students were called upon to read a section out loud or when being assessed for benchmark measure or progress monitoring. Many teachers felt the frustration of the practices between assessment and classroom not being aligned (Manzo, 2005; Riedel, 2007; Salvador et al., 2009; Shinn & Good, 1992).

The goal of the study was to identify measures of reading fluency predictive for reading comprehension and high-stakes state assessments. Results of the study, which

support the findings of previous research, indicated that as a student progresses through the grade levels and silent reading becomes more dominant in classroom practice, silent reading fluency had a higher correlation with reading comprehension than oral reading fluency measures for fifth-grade students. Therefore, aligning classroom assessment measures that correlate with reading comprehension of students can be implemented. Classroom teachers can start to align practices and assessment results. By aligning assessment outcomes and classroom strategies, teachers can provide instructional support to help students reach the goal of meeting or exceeding grade level standards and moving towards being career and college ready or providing support through the MTSS process to close the reading gap that exists. Reading proficiency level also had a moderating effect on the relationship of SRF with reading comprehension as hypothesized. With the alignment of classroom practice and reading assessments that are accurate predictors of reading comprehension, teachers can analyze the SRF measures data to monitor growth and support students who are not meeting grade level expectation through the MTSS process.

With supporting students through the MTSS process, additional training and conversations also need to take place. Currently, teachers know how to analyze data from the assessments currently administered in their schools. But, when new assessments are administered, teachers, interventionists, and instructional coaches need training in administering the assessment as well as interpreting the results. Too often, it is assumed that teachers know what the results indicate and are expected to change instructional practices based on these results without an understanding of where a student's abilities

are for literacy. Additionally, conversations amongst teachers to discuss the abilities of the students as identified through the assessments need to take place. Teachers need to not just use a score to group students but should understand a student's ability and group students by academic need to provide instructional support. An additional component needed to allow time for teachers to explore additional measures that are available. Too often, districts have required assessments that teachers are required to administer but aren't able to explore other options available. By being able to explore additional measures, teachers might find an assessment that aligns to classroom practices, meets the needs of the district, and is reliable and valid. Based on the results, teachers might need to administer an ORF measure to an older student who needs support in skills that align with learning to read and is indicated in the practices in the classroom or a SRF to a younger student who is demonstrating abilities that align with upper grade level reading standards. Teachers need to be able to meet the needs of their students academically based on the outcome measures.

Predictive measures of reading comprehension can also have implications on classroom practice. Students with less-developed ORF or SRF skills will most likely struggle with reading comprehension. Knowing the relationship between these measures, classroom teachers can structure differentiated lessons for small group instruction and provide instructional support in areas of need through analysis of the results for students who are performing below grade level expectations. The relationship of ORF and SRF measures changes over time due to increased proficiency associated with grade level development. In the primary grades, ORF is dominant in classroom practice and has a

stronger correlation with reading comprehension for younger students. As students mature and SRF becomes dominant in classroom practice, the SRF measure has a higher correlation with reading comprehension. Fortunately, the change in relationship of ORF and SRF measures aligns with typical instructional practices. Additionally, the SRF measure is a stronger predictor of reading comprehension and high-stakes assessments as students mature. SRF measures can be a viable alternative to ORF measures for upper elementary students as predictors or reading comprehension and high-stakes assessments. Classroom practices and assessment could be aligned to help guide instruction and support individual needs based on the analysis of the assessment. This would allow teachers to use what is being practiced in the classroom with assessment that accurately predicts reading comprehension.

Conclusion

In summary, results of this study indicated that as grade level increases, the relationship between ORF measures, SRF measures, and reading comprehension changes. As students progress from fourth to fifth grade, the ORF measure has a higher correlation with reading comprehension for fourth-grade students, while the SRF measure has a higher correlation than ORF measure with reading comprehension for fifth-grade students. Proficiency level as a moderator shows a significant relationship between the SRF measure and reading comprehension. Finally, determining the relationship between ORF measures, SRF measures, reading comprehension, and TCAP, indicated that both measures of ORF and SRF have a significant correlation with TCAP.

REFERENCES

Al Otaiba, S., Conner, C. M., Folsom, J. S., Wanzek, J., Greulich, L., Schatschneider, C., & Wagner, R. K., (2014). To wait in Tier 1 or intervene immediately: A randomized experiment examining first-grade response to intervention in reading. *Exceptional Children, 81*(1), 11-27.

American Federation of Teachers, National Council of Measures in Education, & National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Retrieved from: http://buros.org/standards-teacher-competenceeducational-assessment-students

Applegate, M. D., Applegate, A. J., & Modla V. B. (2009). "She's my best reader; she just can't comprehend": Studying the relationship between fluency and comprehension. *The Reading Teacher, 62*, 512-521.

Ardoin, S. P., Witt, J. C., & Suido, S. M. (2004). Examining the incremental benefits administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review,* 33, 218-233.

Allington, R. (1983). Fluency: The neglected reading goal. *Reading Teacher, 36*, 556-561.

Armbruster, B. B., Lehr, F., & Osborn, J. (2001). *Put reading first: The research building blocks for teaching children to read.* Jessup, MD: National Institute for Literacy.

Association for Supervision and Curriculum Development. (2015). *Association legislative priorities include ESEA reauthorization, overhauled accountability systems.* Retrieved from: http://www.ascd.org/news-media/Press-Room/News-Releases/2015-leg-agenda.aspx

Averill, O. H., & Rinaldi, C. (2011). *Multi-tier systems of support: The RTI and PBIS approaches involve targeting specific areas in which students are struggling.* Retrieved from https://www.districtadministration.com/article/multi-tier-system-supports

Barnes, A. C., & Harlacher, J. E. (2008). Clearing the confusion: Response-to-intervention as a set of principles. *Education and Treatment of Children, 31*(3), 417-431.

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.

Bennett, R. E. (2016). *Opt out: An examination of issues* (Research Report RR-16-13) Princeton, NJ: Educational Testing Service. http://dx.doi.org/10.1002/ets2.12101

Biancarosa, G., & Snow, C. E. (2004). *Reading next - A vision for action and research in middle and high school literacy: A report from Carnegie of New York.* Washington, DC: Alliance for Excellence in Education.

Blanc, S., Christman, J. B., Liu, R., Mitchell, C., Travers, E., & Bulkley, K. E. (2010). Learning to learn from data: Benchmarks and instructional communities. *Peabody Journal of Education, 85*(2), 205-225.

Bocala, C., & Boudett, K. P. (2015). Teaching educators habits of mind for using data wisely. *Teachers College Record, 117*, 1-20.

Brookhart, S. (2010). Mixing it up: Combing sources of classroom achievement information for formative and summative purposes. In H. Andrade & Y. G. Cizek (Eds). *Handbook of formative assessment* (pp. 279-296). New York, NY: Routledge.

Brookhart, S. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice 30*(1), 3-12.

Buffum, A., Mattos, M., & Weber, C. (2009). *Pyramid response to intervention: RTI, professional learning communities, and how to respond when kids don't learn*. Bloomington, IN: Solution Tree.

Bulkey, K. E., Olah, L. N., & Blanc, S. (2010). Introduction to the special issue on benchmarks for success? Interim assessment as a strategy for educational improvement. *Peabody Journal of Education, 85*(2), 115-124.

Buly, M. R., & Valencia, S. W. (2002). Below the bar: Profiles of students who fail state reading assessments. *Educational Evaluation and Policy Analysis, 24*(3), 219-239.

California State Board of Education. (2006). *Reading/language arts framework for California's public schools: Kindergarten through grade twelve.* Sacramento, CA: Author.

Chappuis, J. (2009). *Seven strategies of assessment for learning.* Portland, OR: Assessment Training Institute.

Chappuis J., & Chapppuis, S. (2002). *Understanding school assessment: A parents and community guide to helping students learn.* Portland, OR: Assessment Training Institute.

Coburn, C. E., Pearson, P. D., & Woulfin, S. (2011). *Handbook of reading research* (Vol. IV). New York, NY: Routledge.

Cohen, B. H. (2008). *Examining psychological statistics.* Hoboken, NJ: Wiley.

Colorado Department of Education. (2008). *Response to intervention: A practitioners guide to implementation.* Retrieved from http://www.cde.state.co.us/RtI

Colorado Department of Education. (2013a). *State standards.* Retrieved from http://www.cde.state.co.us/CoReadingWriting/StateStandards.asp

Colorado Department of Education. (2013b). *October count.* Retrieved from http://www.cde.state.co.us/CoReadingWriting/StateStandards.asp

Colorado Department of Education. (2013c, March). *Colorado READ Act: Supporting K-3rd grade literacy development.* Retrieved from http://www.cde.state.co.us/sites/default/files/documents/coloradoliteracy/readact/download/read%20act%20fact%20sheet%203-21-2013.pdf

Colorado Department of Education. (2014a, January). *Colorado MTSS essential components.* Retrieved from https://www.cde.state.co.us/mtss/essentialcomponentsjan2014

Colorado Department of Education. (2014b, December). *Supporting improved standards and assessments: Transitioning to 21st century standards and assessments.* Retrieved from http://www.cde.state.co.us/communications/improvedstandardsandassessmentsfactsheet

Colorado Department of Education. (2015). *Data collection opt out requests.* Retrieved from https://www.cde.state.co.us/cdereval/datacollectionoptoutrequests

Cook, R. G. (2003). *The utility of DIBELS as a curriculum based measurement in relation to reading proficiency on high states tests* (Unpublished master's thesis). Marshall University Graduate College, Huntington, WV.

Creswell, J. W. (2008). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (3rd ed.). Upper Saddle River, NJ: Pearson Education.

CTB McGraw Hill Education. (2013). *Transitional Colorado Assessment Program: Technical Report 2013*. Denver, CO: Author

Datnow, A., & Hubbard, L. (2015). Teachers' use of assessment data to inform instruction: Lessons from the past and prospects for the future. *Teachers College Record, 117*, 1-25.

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

Denton, C. A. (2012). Response to intervention for reading difficulties in the primary grades: Some answers and lingering questions. *Journal of Learning Disabilities, 45*(3), 232-243.

Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J, Vaughn, S., Cirino, P. T., … Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading, 15*(2), 109-135.

Dole, J. A. (2003). Professional development in reading comprehension. In C. E. Snow & A. P. Sweet (Eds.), *Rethinking reading comprehension* (pp. 176-191). New York, NY: Guilford.

Duncan, A. (2009). *Robust data gives us the roadmap to reform.* Retrieved from http://www.ed.gov/news/speeches/robust-data-gives-us-roadmap-reform

Duncan, A. (2010). *Unleashing the power of data for school reform: Secretary Arne Duncan's remarks at the STATS DC 2010 Data Conference*. Retrieved from https://www.ed.gov/news/speeches/unleashing-power-data-school-reform-secretary-arne-duncans-remarks-stats-dc-2010-data

Florida Department of Education. (2015). *MTSS Implementation Components: Ensuring common language and understanding.* Retrieved from http://floridarti.usf.edu/resources/format/pdf/mtss_q_and_a.pdf

Freedman, D., Pisani, R., & Purves. (2007). *Statistics* (4th ed.). New York, NY: Norton.

Fuchs, L. S., & Deno, S. L. (1994). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58*(3)*,* 232-243.

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

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2012). Smart RTI: A next-generation approach to multilevel prevention. *Exceptional Children, 78*(3), 263-279.

Fuchs, L. S., & Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5*(3), 239-256.

Gersten, R., & Dimino, J. (2006). RTI (response to intervention): Rethinking special education for students with reading difficulties (yet again). *Reading Research Quarterly 41*(1), 99-108.

Gersten, R., Compton, D., Connor, C. M., Dimino, J., Santoro, L., Linan-Thompson, S., & Tilly, W. D. (2008). *Assisting students struggling with reading: Response to intervention and multi-tier intervention for reading in the primary grades. A practice guide* (NCEE 2009-4045). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides/

Goffreda, C. T., & DiPerna, J. C. (2010). An empirical review of psychometric evidence for the dynamic indicators of basic early literacy skills. *School Psychology Review, 39*(3), 463-483.

Good, R. (2011). Formative use of assessment information: It's a process, so let's say what we mean. *Practical Assessment Research & Evaluation, 16*(3).

Good, R. H., Kaminski, R. A., Cummings, K., Dufour-Martel, C., Peterson, K., Powell-Smith, K., … Wallin, J. (2011). *DIBELS Next assessment manual.* Eugene, OR: Dynamic Measurement Group.

Good, R. H., Kaminski, R. A., Dewey, E. N., Wallin, J., Powell-Smith, K. A., & Latimer, R. J. (2013). *DIBELS Next technical manual.* Eugene, OR: Dynamic Measurement Group.

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

Goodman, K. S. (2006). A critical review of DIBELS. In K. S. Goodman (Ed.), *The truth about DIBELS: What it is, what it does* (pp. 1-39). Portsmouth, NH: Heinemann.

Goodwin, B. (2013/2014). Using data to create continuous improvement opportunities. *Changing Schools, 70,* 3-4.

Griffith, L. W., & Rasinski, T. V. (2004). A focus on fluency: How one teacher incorporated fluency with her reading curriculum. *The Reading Teacher, 58*(2), 126-137.

Hale, A. D., Skinner, C. H., Williams, J., Hawkins, R., Neddenriep, C. E., & Dizer, J. (2007). Comparing comprehension following silent and aloud reading across elementary and secondary students: Implications for curriculum-based measurement. *Behavior Analyst Today*, 8(1), 9-23.

Hale, A. D., Hawkings, R. O., Sheeley, W., Reynolds, J. R., Jenkins, S., Schmitt, A. J., & Martin, D. A. (2011). An investigation of silent versus aloud reading comprehension of elementary students using maze assessment procedures. *Psychology in the Schools,* 48(1), 4-13.

Hall, S. L. (2006). *I've DIBEL'd, now what?* Longmont, CO: Sopris West.

Hamilton, L., Halverson, R., Jackson, S., Mandinach, E. S., Supovitz, J., & Wayman, J. (2009). *Using student achievement data to support instructional decision making.* (NCEE 2009-4067). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/wwc/publications/practiceguides

Hammill, D. D., Wiederholt, J. L., & Allen, E. A. (2006). *Test of silent contextual reading fluency*. Austin, TX: Pro-ED.

Hardcastle, B., & Justice, K. (2006). *RTI and the classroom teacher: A guide for fostering teacher buy-in and supporting the intervention process.* Horsham, PA: LRP Publications.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher, 59*(7), 636-644.

Hayes, A. F. (2013). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach.* New York, NY: Guilford.

Heritage, M. (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?* Washington, DC: Chief Council of State School Officers.

Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practices, 28*(3), 24-31.

Hiebert, E. H., Samuels, S. J., & Rasinski, T. V. (2012). Comprehension based silent reading rates: What do we know? What do we need to know? *Literacy Research & Instruction, 51*(2), 110-124.

Hoover, J. J., & Lover, E. (2011). Supporting school-based response to intervention: A practioner's model. *TEACHING Exceptional Children, 43*(3), 40-48.

Hoover, N. R., & Abrams, L. M. (2013). Teachers' instructional use of summative student assessment data. *Applied Measurement in Education, 26*, 219-231.

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review*, *34*(1), 9-26.

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

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review, 36*(4), 582-600.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading Aloud and Maze. *Exceptional Children, 59,* 421-432.

Johnson, E. S., Pool, J. L., & Carter, D. R. (2011). Validity evidence for the Test of Silent Efficiency and Comprehension (TOSREC). *Assessment for Effective Intervention* 37(1), 50-57.

Juel, C., & Holmes, B. (1981). Oral and silent reading of sentences. *Reading Research Quarterly, 16,* 545-568.

Kamil, M. L., Afflerbach, P. P., Pearson, P. D., & Birr Moje, E. (2011). *Handbook of Reading Research* (Vol. IV). New York, NY: Routledge.

Kansas Department of Education. (2010). *Kansas guide to early reading assessments.* Retrieved from http://community.ksde.org/Portals/55/Documents/Assessment/Early%20Reading%20Assessment/Kansas_Guide_to_Early_Reading_Assessment_Aug_2010.pdf

Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools* 49(4), 342-352.

Kerr, R., Garvin, J., Heaton, N., Boyle, E. (2006). Emotional intelligence and leadership effectiveness. *Leadership & Organization Development Journal*, 27(4), 265-279.

Kim, Y., Wagner, R., Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from grade 1 to grade 2. *Journal of Experimental Child Psychology*, 113, 93-111.

Kragler, S. (1995). The transition from oral to silent reading. *Reading Psychology: An International Quarterly, 16,* 396-408.

Krathwohl, D. R., & Smith, N. L. (2005). *How to prepare a dissertation proposal: Suggestions for students in education & the social and behavioral sciences.* New York, NY: Syracuse University Press.

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

Lipson, M. Y., Chompsky-Higgins, P., & Kanfer, J. (2011). Diagnosis: The missing ingredient in RTI assessment. *The Reading Teacher, 65*(3), 204-208.

Mandinach, E. B., Gummer, E. S., & Muller, R. D. (2011). *The complexities of integrating data-driven decision making into professional preparation in schools of education:* It's harder than you think. Washington, DC: CAN Education, Education Northwest, and WestEd.

Manzo, K. K. (2005). National clout of DIBELS test draws scrutiny. *Education Week,* 25(5), 1-12.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33*, 193-203.

Mesmer, E. M., & Mesmer, H. A. E. (2008). Response to intervention (RTI): What teachers of reading need to know. *The Reading Teacher, 62*(4), 280-290.

Metcalf, T. (2012). *What's your plan? Accurate decision making within a multi-tier system of support: Critical Areas in Tier 1*. Retrieved from http://www.rtinetwork.org/essential/tieredinstruction/tier2/whats-your-plan-accurate-decision-making-within-a-multi-tier-system-of-supports-critical-areas-in-tier-2

Miller, S. D., & Smith, D. E. (1990). Relations among oral reading, silent reading, and listening of students at differing competency levels. *Reading Research and Instruction, 29,* 73-84.

National Assessment of IDEA Overview. (2011). *NCEE 2011-4026.* Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U. S. Department of Education.

National Center on Response to Intervention. (2010). *Essential components of RTI: A closer look at response to intervention.* Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Center on Response to Intervention.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literate on reading and its implications for reading instruction.* Washington, DC: National Institute of Child Health and Human Development.

National Research Council. (2000). *Starting out right: A guide to promoting children's reading success.* Washington, DC: National Academy Press.

Nelson, J. M., & Machek, G. R. (2007). A survey of training, practice, and competence in reading assessment and intervention. *School Psychology Review 36*(2), 311-327.

No Child Left Behind Act. (2002). *20 U.S.C. 6301 et. Seq. (Statute).*

O'Reilly, T., Sabatini, J., Bruce, K., Pillarisetti, S., & McCormick, C. (2012). Middle school reading assessment: Measuring what matters under a RTI framework. *Reading Psychology, 33,*162-189.

Paris, S. G., & Hamilton, E. E. (2009). The development of children's reading comprehension. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 32-53). New York, NY: Routledge.

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2009). *Measures that matter: The role of interim assessment in a comprehensive assessment system.* Washington, DC: Achieve, The Aspen Institute, The National Center for Improvement of Educational Assessment.

Pinnell, G. S., Pikulski. J. J., Wixson, K. K., Campbell, J. R. Gough, P. B., & Betty, A. S. (1995). *Listening to children read aloud.* Washington, DC: U.S. Government Printing Office.

Popham, W. J. (1999). *Classroom assessment: What teachers need to know.* Needham Heights, MA: Allyn and Bacon.

Popham, W. J. (2003). The seductive allure of data*. Educational Leadership, 60*(5), 48-51.

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

Price, K. W., Meisinger, E. B., & Louwerse, M. M. (2012). Silent reading fluency using underlining: Evidence for an alternative method of assessment. *Psychology in the Schools, 49*(6), 1-13.

Pyle, N., & Vaughn, S. (2012). Remediating reading difficulties in a response to intervention model with secondary students. *Psychology in the Schools, 49*(3), 273-284.

Rasinski, T. V. (2003). Fluency is fundamental. *Instructor, 113*(4), 16-20.

Rasinski, T. V. (2006). Reading fluency instruction: Moving beyond accuracy, automaticity, and prosody. *The Reading Teacher, 59*(7), 704-706.

Rasinski, T. V. (2009). *Effective teaching of reading: From phonics to fluency*. Retrieved from http://www.timrasinski.com/presentations/effective_teaching_of_reading-from_phonics_to_fluency_2009.pdf

Rasinski, T., Rikli, A., & Johnston, S. (2009). Reading fluency: More than automaticity? More than a concern for the primary grades? *Literacy Research and Instruction, 48*(4), 350-361.

Rasinski, T., Samuels, S. J., Hiebert, E., Petscher, Y., & Feller, K. (2011). The relationship between silent reading fluency instruction protocol on students' reading comprehension and achievement in an urban school setting. *Reading Psychology, 32*, 75-97.

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly*, 42(4), 546-567.

Roehrig, A., Petscher, Y., Nettles, S., Hudson, R., & Torgesen, J. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46*(3), 343-366.

Ruiz-Primo, M. A. (2011). Informal formative assessment: The role of instruction dialogues in assessing students' learning. *Studies in Educational Evaluation 38,* 15-24.

Salvador, S. K., Schoeneberger, J. A., & Tingle, L. (2009). *DIBELS oral reading fluency predictive effectiveness for 3rd-grade reading performance in Charlotte-Mecklenburg Schools* (Technical Report). Charlotte-Mecklenburg, NC: Center for Research & Evaluation.

Samuels, S. J. (2006). Toward a model of reading fluency. In S. J. Samuels & A. Farstrup (Eds.), *What research has to say about fluency instruction* (pp. 40-63). Neward, DE: International Reading Association.

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.

Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *Elementary School Journal, 107*(5), 429-448.

Schwanenflugel, P. J, Meisinger, E. B., Wisenbaker, J. M., Kuhn, M. R., Strauss, G. P., & Morris, R. D. (2006). Becoming a fluent and automatic reader in the early elementary years. *Reading Research Quarterly, 41*(4), 496-522.

Share, D. L. (2008). On the anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin*, 134, 584-615.

Shaw, R., & Shaw, D. (2002). *DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado state assessment* (CSAP; Technical Report). Eugene, OR: University of Oregon.

Shepard, L. A. (2000). The role of assessment in learning culture. *Educational Researcher 29*(7), 4-14.

Shepard, L. A. (2009). Commentary: Evaluating the validity of formative and interim assessment. *Educational Measurement: Issues and Practice Fall, 2009, 28*(3), 32-37.

Shinn, M., & Good, R., (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*, 459-479.

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

Snow, C. E., (2002). *Reading for understanding: Toward a research and development program in reading comprehension*. Retrieved from http://www.rand.org/content/dam/rand/pubs/monograph_reports/2005/MR146

Stage, S. A., & Jacobsen, M. D., (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review 30*(3), 407-419.

Stiggens, R. (2004). New assessment beliefs for a new school mission. *Phi Delta Kappan, 86*(1), 22-27.

Stiggens R., & Knight T. (1997). *But are they learning? A commonsense parents' guide to assessment and grading in schools.* Portland, OR: Assessment Training Institute.

Torgesen, J., Nettles, S., Howard, P., & Winterbottom, R. (2003). *Brief report of a study to investigate the relationship between several brief measures of reading fluency and performance on the Florida Comprehensive Assessment Test-Reading in 4th, 6th, 8th, and 10th grades* (FCRR Report No.6). Tallahassee, FL: Florida Center for Reading Research at Florida State University. Retrieved from http://www.fcrr.org/TechnicalReports/Progress_monitoring_report.pdf

U.S. Department of Education (2007). *Alignment with No Child Left Behind Act*. Retrieved from: http://idea.ed.gov/explore/view/p/,root,dynamic,TopicalBrief,3

U.S. Department of Education (2015). *Elementary and Secondary Education Act*. Retrieved from http://www.ed.gov/esea

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics. (2014). *Digest of statistics.* Retrieved from http://nces.ed.gov/programs/digest/2012menu_tables.asp

U.S. Department of Education, Office of Planning, Evaluation and Policy Development. (2011). *Teachers' ability to use data to inform instruction: Challenges and supports,* Washington, DC: Author.

University of Oregon Center on Teaching and Learning (2012). *2012-2013 DIBELS Data System Update Part II: DIBELS Next Benchmark Goals (Technical Brief No. 1203).* Eugene, OR: University of Oregon.

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*(3), 270-291.

Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). *The relationship between oral reading fluency and Ohio proficiency testing in reading* (Technical Report). Eugene, OR: University of Oregon.

Vaughn, S., & Fletcher, J. M. (2012). Response to intervention with secondary school students with reading difficulties. *Journal of Learning Disabilities, 45*(3), 244-256.

Vaughn, S., Wanzek, J., Murray, C. S., Linan-Thompson, S., & Woodruff, A. L. (2009). Response to early reading intervention: Examining higher and lower responders. *Exceptional Children, 75*(2), 165-183.

Vaughn, S., Wanzek, J., Woodruff, A. L., & Linan-Thompson, S. (2007). Prevention and early identification of students with reading disabilities. In D. Haager, J. Klinger, & S. Vaughn (Eds.) *Evidence-based reading practice for response to intervention* (11-27). Baltimore, MD: Brookes.

Wagner, R. K. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first grade readers. *Scientific Studies of Reading, 15*(4), 338-362.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2010). *Test of Silent Reading Efficiency and Comprehension (TOSREC) examiner's manual.* Austin, TX: Pro-Ed.

Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques.* Thousand Oaks, CA: Sage.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education, 4,* 85-120.

Wiliam, D. (2011). *Embedded formative assessment.* Bloomington, IN: Solution Tree.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Test.* Bloomington, MN: Pearson Assessments.

Yell, M. (2004, February). *Understanding the three-tiered model.* Presented at the Colorado State Directors of Special Education Meeting, Denver, CO

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*, 4-12.

Ysseldyke, J., Burns, M. K., Scholin, S. E., & Parker, D. C. (2010). Instructionally valid assessment within response to intervention. *Teaching Exceptional Children Cognitive Psychology 42*(4), 54-61.

Zvoch, K., & Stevens, J. (2011, September). Modeling students' response to intervention using and individualized piecewise growth model. Paper presented at the Society for Research on Educator Effectiveness Fall Conference, Evanston, IL.

APPENDIX

LITERATURE REVIEW FINDINGS

Table A1

Literature Review Findings

Study	Grade(s) included	Number of students no < 20	Assessment Instrument	Findings
Applegate, M. D., Applegate, A. J., & Modla V. B. (2009). "She's my best reader; she just can't comprehend": Studying the relationship between fluency and comprehension. *The Reading Teacher,* 62, (6), 512-521.	2nd - 10th	171	Critical Reading Inventory 2 with one passage read orally and one passage read silently with 10 open-ended comprehension questions	One third of the students who were perceived as fluent, high readers, as identified by parents or teachers as strong readers, were not able to answer text based comprehension questions correctly.
Ardoin, S. P., Witt, J. C., & Suido, S. M. (2004). Examining the incremental benefits administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review,* 33, 218-233.	3rd	77	Silver, Burdett, and Ginn (1991) Reading-Curriculum Based Measures (CBM) and Maze; Woodcock Johnson III Letter Word Identification, Reading Fluency, and Passage Comprehension; Iowa Test of Basic Skills Reading Comprehension and Vocabulary	CBM median and the Woodcock Johnson broad reading score had significantly higher correlations between the identified subtests and the maze. The CBM single passage score had the strongest predictive power.
Cook, R. G. (2003). The utility of DIBELS as a curriculum based measurement in relation to reading proficiency on high stakes tests. Unpublished master's thesis. Marshall University Graduate College: Huntington, WV.	1st- 3rd	unknown	DIBELS ORF; Stanford Achievement Test - 9th Edition (SAT9)	Correlations for ORF and comprehension for first graders was 0.73. A positive correlation was found between ORF and the SAT9
Denton, C. A., Barth, A. E., Fletcher, J. M., Wexler, J, Vaughn, S., Cirino, P. T., Romain, M., & Francis, D. J. (2011). The relations among oral and silent reading fluency and comprehension in middle school: Implications for identification and instruction of students with reading difficulties. *Scientific Studies of Reading,* 15(2), 109-135.	6th - 8th	6th -564 7th - 312 8th - 545	Texas Education Agency, 2004a, 2004b; Group Reading Assessment and Diagnostic Evaluation; Woodcock Johnson III Test of Achievement, Passage Comprehension; ORF Curriculum-Based Measurement; Test of Word Reading Efficiency; ORF CBM Word Fluency; AIMSweb Maze CBM; Test of Silent Reading Efficiency and Comprehension; Test of Silent Contextual Reading Fluency; Kaufman Brief Intelligence Test - 2, Verbal	Lower relationship between ORF and reading comprehension for 6th to 8th grade students than was found with younger elementary students

(*table continues*)

Study	Grade(s) included	Number of students no < 20	Assessment Instrument	Findings
Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indictors of foundational reading skills for third grade high stakes outcomes. *Scientific Studies of Reading,* 5, 257-288.	K - 3rd	K - 353 1st - 378 2nd - 342 3rd - 364	DIBELS ORF, Phonemic Segmentation Fluency (PSF), Nonsense Word Fluency (NWF); CBM ORF; OSA - Reading/Literature	Students who attained the benchmark goal had at least a 90% chance of reaching subsequent goals. PSF was less strong with 55% of reaching subsequent goals.
Hale, A. D., Skinner, C. H., Williams, J., Hawkins, R., Neddenriep, C. E. & Dizer, J. (2007). Comparing comprehension following silent and aloud reading across elementary and secondary students: Implications for curriculum-based measurement. *Behavior Analyst Today,* 8(1), 9-23.	4th & 5th 10th - 12th	42 10th-12th 51 4th - 5th	Timed Reading Series passages; Woodcock Johnson III Letter-Word Identification, Reading Fluency, and Passage Comprehension	Students answered more comprehension questions correctly after reading orally as opposed to silent reading.
Hosp, M. K. & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review,* 34(1), 9-26.	1st - 4th	310	CBM Passages; Woodcock Johnson Reading Mastery Test-Revised Word Attack, Word Identification, Passages Comprehension, Basic Skills, Reading-Short	Grades 1 - 3, ORF had a stronger relationship with word reading than with comprehension; fourth grade students, a stronger relationship between ORF and comprehension emerged over word reading
Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading Aloud and Maze. *Exceptional Children,* 59, 421–432.	2nd - 6th	2nd - 47 3rd - 50 4th - 66 5th - 47 6th - 125	Gates MacGinitie Reading Tests; Metropolitan Achievement Tests (MAT); Maze Passages; Group Reading Assessment and Diagnostic Evaluation (GRADE); Oral Reading Measures	A negative trend across grade levels for correlations between ORF and reading comprehension as measured by the two achievement tests. Correlations declined from 0.86 in 4th grade to 0.67 in 6th grade for ORF and GRADE and from 0.87 in 2nd grade to 0.60 in 6th grade for ORF and MAT.
Johnson, E. S., Pool, J. L., & Carter, D. R. (2011). Validity evidence for the Test of Silent Efficiency and Comprehension (TOSREC). *Assessment for Effective Intervention*, 37, (1), 50-57.	1st - 5th	226	AIMSweb Eduformation Standard Reading Assessment Passages; Test of Silent Reading Efficiency and Comprehension; Measures of Academic Progress; Idaho State Achievement Test (ISAT)	ISAT found measures at the different grade levels indicate some assessments may be better indicators of reading proficiency at certain ages. Correlations between ORF and SRF were high for all grades except fourth grade, which were low and not significant. SRF and comprehension are correlated, but ORF was more predictive for first grade and fourth grade correlations were low.

(*table continues*)

Study	Grade(s) included	Number of students no < 20	Assessment Instrument	Findings
Kim, Y., Wagner, R., Lopez, D. (2012). Developmental relations between reading fluency and reading comprehension: A longitudinal study from grade 1 to grade 2. *Journal of Experimental Child Psychology*, 113, 93-111.	1st - 2nd	270	Woodcock Johnson III Oral Comprehension and two experimental passages; Test of Word Reading Efficiency - 2nd Edition Sight Word Passages; DIBELS ORF - 6th Edition	In 1st grade, list reading fluency was related to reading comprehension. In 2nd grade, text reading fluency was related to reading comprehension. In 1st grade, oral reading fluency was related to reading comprehension and silent reading fluency was related to 2nd grade.
McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review,* 33, 193–203.	4th	1,362	MacMillan Connections Reading Program CBM probes; Michigan Educational Assessment Program (MEAP)	A positive correlation between ORF measures from the Macmillan Connections Reading Program and the MEAP.
Miller, S. D. & Smith, D. E. (1990). Relations among oral reading, silent reading, and listening of students at differing competency levels. *Reading Research and Instruction,* 29, 73-84.	2nd - 5th	94	Analytical Reading Inventory with comprehension questions based on J.M. Smith criteria; one passage read orally and one passage read silenty	Silent reading comprehension was higher for 'good' readers in grades 3 - 5 than oral reading.
Pinnell, G. S., Pikulski. J. J., Wixson, K. K., Campbell, J. R. Gough, P. B. & Betty, A. S. (1995). *Listening to children read aloud.* Washington, DC: U.S. Government Printing Office.	4th	1,136	National Assessment of Educational Progress (NAEP)	Fifty-five percent of the students were considered to be fluent as rated on a four-point scale as being a three or four. As the fluency rate increased, the reading proficiency also increased reinforcing the link between reading fluency and reading comprehension.
Price, K. W., Meisinger, E. B., & Louwerse, M. M. (2012). Silent reading fluency using underlining: Evidence for an alternative method of assessment. *Psychology in the Schools,* 0 (00), 1-13.	4th & 6th	4th - 37 6th - 22	DIBELS ORF - 6th Edition; Silent Reading fluency with underlining; Gates-MacGinitie Reading and Vocubulary - 4th Edition; AIMSweb maze	The maze task did not have a significant correlation with the comprehension measures, but the correlation with SRF and the reading comprehension measures was strong. This study indicated that SRF can be an accurate predictor of reading comprehension.
Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly,* 42(4), 546-567.	1st Grade	1,518	DIBELS Letter Naming Fluency, Phoneme Segmentation Fluency, Nonsense Word Fluency, Oral Reading Fluency, and Retell Fluency; Group Reading Assessment and Diagnostic Evaluation; TerraNova Reading subtest	80% accuracy for predicting reading comprehension for 1st grade participants; 71% accuracy for predicting reading comprehension for 2nd grade students

(*table continues*)

Study	Grade(s) included	Number of students no < 20	Assessment Instrument	Findings
Roehrig, A., Petscher, Y., Nettles, S., Hudson, R., & Torgesen, J. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology,* 46(3), 343-366.	3rd	35,207	DIBELS ORF; Florida Comprehensive Assessment Test - Sunshine State Standards; Stanford Achievement Test (SAT)	DIBELS had a high correlation to the reading comprehension measures of the SAT and Florida reading assessment.
Salvador, S. K., Schoeneberger, J. A., & Tingle, L. (2009). DIBELS oral reading fluency predictive effectiveness for 3rd-grade reading performance in Charlotte-Mecklenburg Schools *(Technical Report). Charlotte-Mecklenburg: NC; Center for Research & Evaluation.*	2nd - 3rd	9,562	DIBELS ORF; North Carolina Reading Assessment	Third grade students had a moderate relationship of 0.66 between ORF and reading comprehension. The results between DIBELS ORF and the state reading assessment were moderately correlated. The disaggregated results by subpopulations to include ethnicity, special accommodations, and economic status and found the correlation remained stable for ORF and state reading outcomes.
Shaw, R., & Shaw, D. (2002). *DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado state assessment (CSAP). Eugene, OR; University of Oregon.*	3rd Grade	58	DIBELS ORF; Colorado State Assessment Program (CSAP)	91% of the students who scored above the DIBELS national cut point scored at or above grade level on the state assessment.
Shinn, M. & Good, R., (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review,* 21, 459–479.	3rd & 5th	3rd - 114 5th - 124	Curriculum-based passages from Harcourt-Brace-Jovanovich; Phonetically regular words taken from Test of Written Spelling; Phonetically regular nonsense words taken from the Woodcock Johnson Mastery Tests; Stanford Diagnostic Reading Test; Written retell and close from a 400 word folktale; SDRT Comprehension subtest	It was determined that ORF provided an estimate of reading comprehension for third and fifth grade levels in relation to phonetically regular and regular nonsense words, literal comprehension, inferential comprehension, close items, and written retell.
Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools,* 43, 527-535.	3rd, 5th, 7th, & 8th	5,472	R-CBM; Maze, Minnesota Comprehensive Assessments - Reading; Basic Standards Test-Reading	ORF accounted for 50.4% of the variance in comprehension scores at the third grade level but only 26% at the eighth grade level.
Stage, S. A., & Jacobsen, M. D., (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review,* 30(3), 407-419.	4th	173	Curriculum-based oral reading passages from Silver Burdette and Ginn; Washington Assessment of Student Learning (WASL)	ORF reading passages in the Silver Burdette and Ginn Curriculum to predict the WASL. Findings indicated that ORF had a 0.51 correlation to the WASL, which is lower than some standardized assessment results as a medium effect size was obtained.

(*table continues*)

Study	Grade(s) included	Number of students no < 20	Assessment Instrument	Findings
Torgesen, J., Nettles, S., Howard, P., & Winterbottom, R. (2003). Brief report of a study to investigate the relationship between several brief measures of reading fluency and performance on the Florida Comprehensive Assessment Test–Reading in 4th, 6th, 8th, and 10th grades (FCRR Report No.6). Tallahassee, FL: Florida Center for Reading Research at Florida State University. Retrieved from http://www.fcrr.org/TechnicalReports/Progress_monitoring_report.pdf	4th, 6th, 8th, & 10th	4th - 88 6th - 252 8th - 161 10th - 98	Espin Maze Passages; Florida Comprehension Assessment Test (FCAT) - based maze passages; Test of Silent Contextual Reading Fluency; Test of Sentence Reading Efficiency; Oral reading fluency with FCAT passages; Test of Silent Reading Efficiency (TOSRE)	4th grade, the relationship between ORF and maze on the Florida Comprehensive Assessment Test (FCAT) was equal; 6th grade results indicated that maze had the strongest relationship with FCAT; 8th grade had maze and the (TOSRE) equal on the relationship with FCAT; 10th grade had TOSRE with the strongest relationship with FCAT.
Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). *The relationship between oral reading fluency and Ohio proficiency testing in reading (Technical Report). Eugene, OR: University of Oregon.*	3rd - 4th	364	CBM-ORF; DIBELS ORF; Ohio Fourth Grade Reading Proficiency Test	All coefficients had a significant correlation with ORF and the Ohio Proficiency Test.
Wagner, R. K. (2011). Relations among oral reading fluency, silent reading fluency, and reading comprehension: A latent variable study of first grade readers. *Scientific Studies of Reading,* 15(4), 338-362.	1st	316	Woodcock Johnson III Oral Comprehension, Passage Comprehension, Word Reading Accuracy; Test of Word Reading Efficiency; DIBELS ORF; Test of Sentence Reading Efficiency and Comprehension;	Students were divided into subgroups based on the word identification score. Skilled readers were the top third, while average readers comprised the bottom third. ORF was identified as a better predictor of reading comprehension for all first grade students. SRF was strongly related to reading comprehension for first grade skilled students.
Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of atheoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice,* 24, 4–12.	4th - 8th	4th - 981 5th - 1,248 6th - 1,248 7th - 1,248 8th - 1,287	Oral reading fluency measures based on Lexile Framework and Flesch-Kincaid index with 15 selected response questions; vocabulary list of words in the *World Book Encyclopedia* (2001)	Fluency measures in grades 5 through 8 appear to be less important than in grade 4.

CURRICULUM VITAE

CHRISTY LYN SINNER BLOOMQUIST

Office: Durango School District, 201 East 12th Street, Durango, CO 81301
Home: 102 Stevens Creek Lane, Durango, CO 81301

Telephone: Office (970) 247-5411
Home (970) 903-1363

EDUCATION

University/College	Fields of Study Major and Minor	Degree	Year
Utah State University	Education/ Curriculum & Instruction	Ph.D.	2017
Adams State University	Educational Leadership/ Principal Licensure		2008
Fort Lewis College	Linguistically Different: ESL Endorsement		2000
Adams State University	Special Education I - Moderate Needs	MA	1994
Fort Lewis College	Humanities/ History	BA	1992

Educational Experience

MTSS/Data Specialist, Durango School District, 2016 – present
Responsible for PK-12 MTSS, interventions, assessments, data analysis
Director of Curriculum, Instruction, & Assessment, Durango School District, 2013-2016
Responsible for K-12 curriculum, instruction, and assessment
Executive Director of Student Achievement, Durango School District, 2011-2013
Responsible for K-12 achievement and data analysis
Director of Federal Programs and Assessment, Durango School District, 2010-2011
Responsible for federal grants and programs and K-12 assessments
Teacher on Special Assignment, Durango School District, 2009-2010
Responsible for data analysis and training of assessments

Assistant Principal for Summer School, Durango School District, 2007-2008
Oversaw all levels of summer school and curriculum, instruction, and assessment
Acting Principal in Principal absences, Durango School District, 2007-2009
Administrator on duty, responsible for all school components
Title One Lead Teacher, Durango School District, 2005-2009
Responsible for leadership and Title I programs in district
Classroom Teacher and Intervention Specialist, Durango School District, 1994-2005
Responsible for all education components

Additional Training:
Standards Based on Outcomes
Positive Discipline/Classroom
Assessment Academy I
Six-Trait Writing/Assessment Instruction
Orton-Gillingham Multi-Sensory Education
Multi-Sensory Reading Strategies
Professional Learning Communities
Thoughtful Classroom – Five Highly Effective Practices
Becoming an Effective Instructional Coach
Cognitive Coaching
Data Team Training
Common Formative Assessment Certificate
Learning Focused Supervision

Adjunct Class Instructor

Decision Making for Results and Data Teams, Adams State University, Fall 2012.
SIOP: The SIOP Model in Classroom Instruction, Adams State University, Spring 2012.
Science/Social Studies Curricular Alignment to the Common Core, Adams State University, Summer 2011.
Language Arts/Math Common Performance Assessments, Adams State University, Spring 2011.
Creating Common Formative Assessments, Adams State University, Spring 2011.
Thoughtful Classroom: Moving Deeper with Case Studies, Adams State University, Fall 2010 and Spring 2011.
Intervention Strategies, Adams State University, Fall 2010.
Essentials of Writing, Adams State University, Spring 2010.
Better Answers: Written Performances, Adams State University, Summer 2009.
Bringing Words to Life, Adams State University, Summer 2009.

Professional Memberships

International Reading Association

National Council of Teachers of Mathematics
Teachers of English to Speakers of Other Languages International Association
Colorado Association of School Executives
Durango Education Association

Services

Board of Directors, Children's Museum of Durango
Board of Directors, Blue Sky Association

Professional Presentations

National

1. "Assessment Literacy: What teachers need to know and resources available." National Title One Conference, Nashville, TN, January 2013.

Regional

1 "Getting Assessment for Learning Right." Daniels, May 2015.
2 "Formative Assessments." San Luis Board of Cooperative Services, March 2015.
3 "Common Formative Assessments that Work." Sargent School District, February 2015.
4 "Formative Assessments . . . What's the plan and Why?" Colorado Association of School Boards, December 2014.
5 "Assessment for Learning!" Gates Foundation, November 2014.
6 "Common Formative Assessments." Colorado Education Association, August 2014.
7 "Utilizing an Assessment/Data Management System." Colorado Legacy Foundation, Integration Showcase, June 2014.
8 "SchoolVault." San Juan Board of Cooperative Services, May 2014.
9 "Common Formative Assessment Administration." Colorado Association of School Executives, February 2014.
10 "Changes In Education." Durango Education Foundation, January 2014.
11 "7 Keys to Comprehension." Title One Parent Conference, September – May 2012.
12 "Understanding ILPs, Title One, and IEPs." Young Child Conference, October 2011.
13 "Unified Improvement Plan: Data Analysis to identify Root Cause." San Juan Board of Cooperative Services Regional Training, October 2010.
14 "NWEA: Data Analysis to Guide Instruction." Connecting the Dots Conference, September 2010.