TEACHER ACCEPTABILITY OF ORAL READING FLUENCY

By

Sarah Stebbe Rowe

A DISSERTATION

Submitted to Michigan State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

School Psychology

ABSTRACT

TEACHER ACCEPTABILITY OF ORAL READING FLUENCY

By

Sarah Stebbe Rowe

Many schools are adopting a Response to Intervention (RTI) model to support and evaluate learning (Fuchs & Fuchs, 2006). Universal screening and progress monitoring are two essential components of RTI that generally support improved student outcomes (Shinn, 2007). In many schools, teachers collect and use a tool called oral reading fluency for these purposes. Some teachers are resistant to using oral reading fluency due to its disputed validity and their lack of time, resources, and knowledge about how to use these data (Foegen, Espin, Allinder, & Markell, 2001; Roehrig, Duggar, Moats, Glover, & Mincey, 2008; Samuels, 2007; Yell, Deno, & Marston, 1992). With a greater understanding of teacher resistance, school psychologists can address the critical factors that relate to the acceptability, use, and effectiveness of oral reading fluency (National Association of School Psychologists, 2008, p. exix).

The purpose of this study was to explore teacher acceptability of oral reading fluency for universal screening and progress monitoring. First to sixth grade teachers from mid-Michigan completed a survey (N = 164), and 22 teachers attended focus groups. In the survey, teachers reported oral reading fluency to be highly acceptable for universal screening and progress monitoring. Teachers reported that oral reading fluency was slightly more acceptable for universal screening than progress monitoring. Knowledge of oral reading fluency and perceived time for assessment were significantly related to teachers' attitudes toward oral reading fluency. In the focus groups, teachers reported many of the reasons for their attitudes towards oral reading fluency. Implications for future research and educational practice are presented. Copyright by SARAH STEBBE ROWE 2013

DEDICATION

To my husband, Jeremy, who supports and loves me always. To my parents, Paul and Karen Stebbe, for encouraging me to pursue lifelong learning.

ACKNOWLEDGEMENTS

First, I want to praise God for giving me the opportunity, the energy, and the support I needed to complete this dissertation. I also want to thank my advisor, Sara Witmer, for spending many hours across the past several years discussing this research with me, reviewing research drafts, and providing critical feedback to encourage the development of my research skills. My dissertation committee members were also very supportive and helpful through this process. Thank you Dr. Evelyn Oka, Dr. Ed Roeber, Dr. Gary Troia, and Dr. Tanya Wright. Also, thank you to Elizabeth Cook and Katie daCruz who helped me transcribe and code the focus group data. I also appreciate the assistance from several graduate students (Paul Curran and Monthien Satimanon) at the Center for Statistical Training and Consultation who guided my learning about statistical methods.

This research was supported in part by grants from the Michigan State University College of Education, the Michigan State University Graduate School, and the Society for the Study of School Psychology.

TABLE OF CONTENTS

LIST OF TABLES	X
LIST OF FIGURES	xi
LIST OF ABBREVIATIONS	xii
CHAPTER I	
INTRODUCTION	1
Theoretical Framework	6
Purpose of this Study	9
CHAPTER II	
LITERATURE REVIEW	10
Defining Oral Reading Fluency	10
Oral Reading Fluency for Universal Screening and Progress Monitoring	13
Oral Reading Fluency Assessment Considerations	16
Reliability	16
Validity	19
Content validity	19
Construct validity	
Criterion validity	20
Other assessment considerations	
Face validity	
Consequential validity	24
External validity	
Conclusions	
Social Validity and Acceptability	27
School psychologists' acceptability	
Variables related to school psychologists' acceptability	30
Teacher acceptability	30
School characteristics influencing acceptability	
Teacher characteristics influencing acceptability	35
Conclusions	
The Present Study	
Research Questions and Hypothesis	
CHAPTER III	
METHOD	44
Research Design	44
Sampling Procedures	44
Sample Size and Power	47
Participant Recruitment and Characteristics	

Measures	
Dependent Variables	51
Assessment acceptability	51
Independent Variables.	
Time and resources	
Years teaching	54
Teaching certification	54
Teacher knowledge	54
Frequency of use	56
Training	
Procedures	
Survey pilot study	
Survey administration	
Focus groups	
Data Analysis	60

CHAPTER IV

RESULTS	64
Descriptive Statistics	64
Missing Data	69
Statistical Assumptions	73
Research Question One	75
Factors influencing accuracy of oral reading fluency	79
Resources needed for oral reading fluency	80
District or school issues	. 81
Influence on students	. 81
Use of data	. 81
Limitations of oral reading fluency	. 82
Research Question Two	83
Research Question Three	85
Research Question Four	88
Research Question Five	89
Standard measurement	90
Decision making with oral reading fluency	90
Consult with other teachers	90
Demonstrate growth or set goals	90
Parent communication	. 91
CHAPTER V	
DISCUSSION	93
Overview of the Study	93
Research question one: Do teachers view oral reading fluency as an acceptable	e
reading measure?	93
Research question two: Is there a significant difference between teachers'	-
acceptability of oral reading fluency for the purposes of universal screening ar	nd
progress monitoring?	. 95

Research question three: To what extent are teacher characteristics and teacher perceptions of school characteristics related to their acceptability of oral real	her ding
fluency?	97
Research question four: What elements of reading do teachers find valuable good reading?	for 99
Research question five: How do teachers use oral reading fluency data?	100
Future Research.	100
Limitations	102
Educational Implications	104
Conclusions	106
APPENDICES	107
Appendix A: Sample Administrative Support Emails	108
Appendix B: Teacher Survey Email Invitation	110
Appendix C: Teacher Focus Group Email Invitation	111
Appendix D: Teacher Demographic Form and Survey	112
Appendix E: Teacher Focus Group Question Protocol	127
Appendix F: Teacher Pilot Study Survey Consent Form	129
Appendix G: Teacher Survey Consent Form	131
Appendix H: Teacher Focus Group Consent Form	133
Appendix I: Variables	135
Appendix J: Focus Group Code Frequency	136
REFERENCES	138

LIST OF TABLES

Table 1: Percentage of Schools or Individuals from Cohorts 4-7	46
Table 2: Demographic Information of the Survey Participants (N = 164)	49
Table 3: Demographic Information of the Focus Group Participants (N=22)	50
Table 4: Knowledge Scale Items	56
Table 5: Data Analysis Methods.	62
Table 6: Descriptive Statistics for Survey Data	64
Table 7: Ranges for Survey Data	65
Table 8: Knowledge Questions Correct and Incorrect	67
Table 9: Missing Data Summary	70
Table 10: Mean and Standard Deviations Before and After EM Procedures	73
Table 11: Universal Screening with Oral Reading Fluency Acceptability Ratings by Item	77
Table 12: Progress Monitoring with Oral Reading Fluency Acceptability Ratings by Item	78
Table 13: Positive and Negative Attitudes and Representative Quotes	82
Table 14: Reading Value Descriptive Statistics	89
Table 15: Use of Data and Representative Quotes.	91
Table 16: Focus Group Code Frequency	.136

LIST OF FIGURES

Figure 1: Teaching experience	66
Figure 2: Knowledge	68
<i>Figure 3</i> : Venn diagram of themes	80
<i>Figure 4</i> : Major codes within the use of oral reading fluency theme	92

LIST OF ABBREVIATIONS

Assessment Profile Rating- Revised = APR-R Curriculum-based assessment = CBA Dynamic Indicators of Basic Early Literacy Skills = DIBELS General outcome measure = GOM Individuals with Disabilities Education Act = IDEA Intermediate School District = ISD Michigan Integrated Behavior and Learning Initiative = MiBLSi No Child Left Behind = NCLB Oral Reading Fluency = ORF Reading curriculum-based measurement = R-CBM Regional Education Service District = RESD Response to Intervention = RTI Standards Error of Measurement = SEM Words Read Correct per Minute = WCPM

CHAPTER I

INTRODUCTION

Not all children in the United States have proficient reading skills by the time they reach eighth grade (National Center for Educational Statistics, 2009). According to the 2009 Nation's Report Card for reading, only one third (33%) of fourth grade students and one third (32%) of eighth grade students demonstrated proficient reading levels on national standardized tests (National Center for Educational Statistics, 2009). This means that around two thirds of American students did not demonstrate complete mastery of grade level reading tasks in fourth and eighth grade. In an effort to improve student achievement in this country, legislation, government programs, and educational frameworks have been created to increase accountability, testing, and scientifically based practices in education.

To improve student achievement, the No Child Left Behind Act (NCLB) of 2001 emphasizes frequent assessment of student achievement using scientifically based tools. This legislation requires high-quality academic assessments that measure progress (No Child Left Behind Act of 2001, Sec. 1001). Additionally, NCLB supports universal screening, which is a data collection process that involves collecting brief measures of reading achievement from all students for the purpose of identifying students in need of additional intervention. According to NCLB, reading assessment tools must be scientifically based and must use brief procedures (No Child Left Behind Act of 2001, Sec. 1208). As a result of this legislation, many schools are now collecting brief reading assessment measures from all students (Dorn, 2010). One such brief reading test is called oral reading fluency. The purpose of this study is to explore the acceptability of oral reading fluency for universal screening and progress monitoring. In addition to NCLB, a government funded reading program, called Reading First, required that student progress and achievement be documented through frequent data collection. Reading First was an initiative to improve reading instruction in schools, which ended in 2009. While the U.S. Department of Education did not require specific assessment tools for Reading First, the assessments used in this program were required to be based on scientific reading research (U.S. Department of Education, 2009). Examples of formal assessments used in this program include Clay Observational Survey, Dynamic Indicators of Basic Early Literacy Skills (DIBELS), Developmental Reading Assessment (DRA), and STAR Reading (U.S. Department of Education, 2009). Reading First teachers in over 40 states adopted Dynamic Indicators of Basic Early Literacy Skills (DIBELS) as the required scientifically based assessment tool to screen students and monitor progress (Manzo, 2005). The frequent use of DIBELS in the Reading First initiative raised public awareness about oral reading fluency as a measure of reading achievement.

DIBELS is a commercially produced set of tools that is often used for universal screening and progress monitoring purposes. These tools are adapted from curriculum-based measurement (CBM), which was developed by Stanley Deno in 1985. DIBELS are similar to traditional CBM procedures because they are intended to be brief and easy to administer. CBM tools are administered in a standardized way, and they were developed to be sensitive to academic growth. Curriculum-based measurements typically consist of standard directions, a timer, materials (i.e., passages, sheets, lists), scoring guidelines and standards, and a record sheet (Hosp, Hosp, & Howell, 2007). DIBELS and CBM are intended to provide teachers and schools with information about their students' reading achievement.

DIBELS (or DIBELS Next, which is the name of the newest edition) is slightly different from traditional curriculum-based measurement in several ways. In DIBELS Next the reading passages are the same regardless of school curricular material (i.e., passages do not come from the local reading curriculum). Also, DIBELS Next uses multiple measures of reading to determine composite scores that determine students' risk levels. In first through sixth grade, oral reading fluency scores are collected along with other measures (e.g., retell and maze comprehension). Oral reading fluency is one measure within the DIBELS Next composite that is particularly important for determining the students' overall score and level of risk for failure.

In addition to legislation and programming that have changed assessment practices, an educational framework called Response to Intervention (RTI) has also emphasized the importance of universal screening and progress monitoring (Fuchs & Fuchs, 2006). Response to Intervention is an educational framework that promotes high quality instruction, assessment, and evidence-based intervention (2006). In nearly all response to intervention initiatives, universal screening and progress monitoring are essential components (Shinn, 2007).

According to George Batsche and colleagues, one core RTI principal is to screen all children to identify those who are not meeting academic expectations (2005). Universal screening is the systematic testing of all students in a classroom, school, or district on an academic skill (e.g., oral reading fluency) that the school or community deems important (Ikeda, Neessen, & Witt, 2008). In an RTI model, universal screening data are used to identify the students who need additional academic support. Universal screening helps identify what level of support a child needs within a multi-tier intervention system. This means that children with more intensive needs receive multiple levels of instructional support at increasingly higher levels of intensity at each tier (Fuchs & Fuchs, 2006). Once these students are identified, schools provide

the students with supplemental, targeted instruction in the area of their deficit. Within a response to intervention model, the progress of students is monitored for academic growth (Fuchs & Fuchs, 2004). Progress monitoring involves a test of an academic skill that occurs more frequently (such as monthly or weekly) to inform instructional changes for at-risk students (Shinn, 2008). Progress monitoring goals may be developed to close the gap between underachieving students and their peers. Through frequent data collection, progress towards these goals can be monitored.

One government funded RTI initiative in Michigan is called Michigan Integrated Behavior and Learning Support Initiative (MiBLSi). In this program, schools are required to collect universal screening data in reading and are highly encouraged to monitor progress for students receiving additional intervention support. DIBELS Next and another similar commercialized set of tools called AIMSweb are listed as recommended tools for these purposes (Michigan Integrated Behavior and Learning Initiative, n.d.). MiBLSi supports schools in the collection and use of these data to evaluate achievement at the school, grade, class, and individual student levels.

The increasing emphasis on collecting reading data from all students for universal screening or progress monitoring is an example of an educational innovation. Innovations are defined as "a departure from current practice... which is novel. Innovations include novel practices, tools or technologies, and knowledge and ideas" (Cohen & Ball, 2007, p. 2). The use of oral reading fluency for universal screening and progress monitoring is an educational innovation because many teachers are changing the way that they evaluate student reading achievement. More and more teachers are required to collect data through the use of measures such as DIBELS Oral Reading Fluency or AIMSweb Reading Curriculum-based Measurement

(R-CBM). In this case, R-CBM refers specifically to an oral reading fluency measure. Teachers who were previously unfamiliar with these reading assessments are now required or highly encouraged to learn the new methods and integrate them into their assessment practices. This first stage of innovation implementation is called "exploration" of the new practice by Fixsen & Blasé (2009), and they consider this stage to be imperative for successful implementation.

With these changes in education, a concern with the social validity of oral reading fluency has surfaced. In the 1940s, validity was considered to be an unchanging property of a tool, including content, criterion-related, and construct validity (Goodwin & Leech, 2003). Over the past several decades, the conceptualization of validity has changed to be considered a property of the *uses* of a test, and this has been broadened to include a consideration of the appropriateness, meaningfulness, and usefulness of a measurement tool for specific purposes (Goodwin & Leech, 2003). This shift in the conceptualization of validity." Although social validity is not considered to be a part of validity as described by the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999), it is still a construct worth considering when evaluating an assessment tool.

Social validity was originally defined by Wolf (1978) as including three parts: the social significance of the goals, the social appropriateness of the procedures, and the social importance of the effects. The second portion of this definition has been coined "treatment acceptability" (Kazdin, 1980). Kazdin (1980) said that a treatment is acceptable when it is appropriate to the problem, fair, reasonable, and non-intrusive. Similarly, the NASP position statement on *Prevention and Intervention Research in the Schools* defines acceptability as "the degree to which consumers find the procedures and outcomes acceptable in their daily lives" (2008, p.

cxx). Teachers and literacy experts have questioned the acceptability of oral reading fluency as an appropriate measure of reading achievement (Deeney, 2010; Samuels, 2007). Therefore, this study explores teacher acceptability of oral reading fluency.

Theoretical Framework

In this study, assessment acceptability is considered to be an attitude. Attitudes are thought to be a substructure formed from many deeply held beliefs (Pajares, 1992). This study rests on the assumption that attitudes and beliefs in general guide behavior (Dewey, 1910; Pajares, 1992). Psychologists have long been interested in the extent to which attitudes and behavior are related. Even dating back to the early twentieth century, this topic was hotly debated. Gordon Allport (1934 as cited in Forgas, Cooper, & Crano, 2010) suggested that attitudes directly influence behaviors. During this same time period, a sociologist named Richard LaPiere argued that there was a weak relationship between attitudes and behaviors (1934 as cited in Forgas et al., 2010).

This debate has extended to educational research regarding teachers' attitudes and behaviors in the classrooms (Fang, 1996). The two competing theories on teacher beliefs and behavior include a theory that these two are generally consistent (Brophy & Good, 1974, Rupley & Logan, 1984, Richardson, Anders, Tidwell, Lloyd, 1991) and a theory that these two are inconsistent (Duffy & Anderson, 1984). Richardson and colleagues (1991) suggested that teachers' beliefs about teaching reading related to their classroom instructional practices. On the other hand, Duffy (1982) reviewed literature on teacher thinking and instructional practices and concluded that teacher practices do not always align with beliefs due to situational constraints. This study operates under the theory that teachers' beliefs and behavior are somewhat congruent, but inconsistencies are a result of the complexities of classroom life (Duffy, 1982). Recognizing

the importance of both attitudes and contexts, this study will focus on teachers' attitudes towards assessment and the factors that influence those attitudes.

A majority of the research on the theories relating beliefs and behaviors considers teachers' attitudes toward instruction or interventions and their related classroom behaviors (e.g., Lentz, Allen & Ehrardt, 1996). The specific theoretical framework for this study that supports the connection between attitudes and behavior is the Treatment Acceptability Model proposed by Witt and Elliot (1985). This model describes the cyclical nature of treatment acceptability, use, integrity, and effectiveness. Acceptability is considered to be the attitude that relates to behaviors in the use of interventions. Within this model, treatments that are more likely to be accepted by consumers are also more likely to be implemented with integrity and to lead to better student outcomes. This theory provides recognition that the relationship between attitudes (i.e., acceptability) and behaviors (i.e., use and integrity) is reciprocal, indicating that attitudes and behavior may influence one another.

There is some research to support the theory that acceptability and intervention integrity are connected. Most of the research suggesting a connection between integrity and acceptability comes from research on behavioral interventions, and a smaller amount of academic focused research has supported this theory. One study reported a significant correlation of .35 (p = .02) between teacher acceptability ratings for reading interventions and treatment integrity checks (i.e., the mean of four integrity checks) (Mautone et al., 2009). The author of another study reported similar results, which indicated a positive relationship between reading treatment acceptability and treatment integrity (r = .12, p < 0.05) (Henninger, 2010). In this intervention study, ten percent of the variance in student outcomes was accounted for by the intervention's complexity, acceptability, and treatment integrity (Henninger, 2010). On the other hand, several

researchers have found no connection between intervention acceptability and intervention integrity, perhaps indicating that other variables were more influential in regards to teachers' behaviors (Peterson & McConnell, 1996; Sterling-Turner & Watson, 2002). These moderate or absent correlations suggest that acceptability is perhaps one factor that influences teacher behavior, but certainly not the only factor. This research might be explained by the theory that at times beliefs and behaviors are inconsistent due to contextual variables (Duffy, 1982).

This study rests on an extension of the Treatment Acceptability Model, with the more specific assumption that teacher attitudes towards oral reading fluency are related to their use of this assessment measure. This study assumes that if teachers accept an assessment measure, the measure is more likely to be used effectively and thus to improve student achievement. Allinder (1997) found that teachers who reported a higher level of acceptability for mathematics curriculum-based measurement implemented the assessment procedures with a higher level of fidelity than teachers who reported a low acceptability level. Therefore, assessment acceptability is considered to be one of several important factors that might influence the assessment behaviors of teachers.

Lastly, this study draws from the theoretical framework on educational innovations and their adoption, use, and implementation (Rogers, 2003). The use of oral reading fluency data for universal screening is considered to be an educational innovation that is influenced by 1) the teachers who adopt the innovation, and 2) the environment in which the innovation is adopted (Cohen & Ball, 2007, p. 4-10). Therefore, the variables of interest in this study include teacher characteristics and school characteristics that may relate to teachers' attitudes toward oral reading fluency as a universal screening tool. The theoretical frameworks in this study

emphasize that multiple factors including teacher attitudes and contextual variables influence teacher behavior.

Purpose of this Study

The purpose of this study is to understand teacher attitudes towards oral reading fluency as a measure of reading. Specifically, this study will measure teachers' acceptability of oral reading fluency (ORF) when it is used for universal screening and progress monitoring, will seek to understand why teachers find this measure acceptable or unacceptable, and will investigate how teachers use oral reading fluency. This research will also identify and quantify factors related to teachers' acceptability of oral reading fluency in schools. This study will investigate what teachers believe about the importance of various reading elements.

CHAPTER II

LITERATURE REVIEW

This literature review draws on research from several relevant areas. First, oral reading fluency is defined and its relationship to other assessment tools is described. Then the use of oral reading fluency data for universal screening and progress monitoring is explored. Next, current research evaluating oral reading fluency measures is presented, including an overview of the validity construct and several other important assessment considerations. Lastly, research about teacher acceptability is reviewed. Currently, very few studies focus specifically on oral reading fluency acceptability; therefore, literature on the acceptability of similar assessment tools, treatment methods, and educational innovations are reviewed as well.

Defining Oral Reading Fluency

Oral reading fluency is considered to be a general outcome measure (GOM). This means that oral reading fluency requires students to use several contributing skills, such as phonological awareness, knowledge of letter-sound correspondences, vocabulary, syntax, and content knowledge, in order to perform well (Hosp et al., 2007). The theoretical basis for using oral reading fluency to measure overall reading achievement is based on the assumption that accurate and efficient word reading skills free up higher order thinking skills for processing the text (Fuchs, Fuchs, Hosp, & Jenkins, 2001). Given that oral reading fluency is a general outcome measure, it is expected to demonstrate progress in a variety of reading skills and to accurately identify at risk readers. Identifying the specific skill deficit when a student is not reading fluently requires further assessment.

There are several advantages to using GOMs. First of all, they reduce the number of individual measures that teachers need to administer, score, and record (Hosp et al., 2007).

GOMs also avoid isolating and testing skills out of context. For instance, reading nonsense words or the sounds of letters in isolation are not typical classroom tasks for students, and therefore the results may be less valid; these are not considered to be general outcome measures. Instead of focusing on one skill, GOMs such as oral reading fluency require the coordination and use of multiple reading skills in order to read fluently. Thus, GOMs may be able to measure higher level skills than more specific measures such as nonsense word fluency. Lastly, visual displays of progress on a GOM allow for long-term progress monitoring of the same skill and for data-based decision-making regarding instructional changes (Hosp et al., 2007). Given these advantages, general outcome measures, such as oral reading fluency, are often used as universal screening tools.

As a global outcome measure, oral reading fluency measures rate and accuracy although this method of measurement is debated among reading experts. The National Reading Panel included reading fluency as one of the key components necessary for successful reading comprehension (National Reading Panel, 2000). This panel defined fluency as consisting of speed, accuracy, and proper expression (National Reading Panel, 2000). Similar to the National Reading Panel definition, many others have included prosody in the definition of reading fluency; prosody is the expressiveness of oral reading, including intonation, stress patterns, and phrasing (Daane, Campbell, Grigg, Goodman, & Oranje, 2005; Hudson, Lane, & Pullen, 2005; Moskal & Blachowicz, 2006). Unfortunately, as researchers have noted, prosody is more difficult to measure than rate and accuracy (Daane et al., 2005). Currently research remains inconclusive regarding whether the measurement of prosody is an important contributor above and beyond decoding skills for predicting reading comprehension (Schwanenflugel, Hamilton, Kuhn, Wisenbaker, & Stahl, 2004). Authors of one study suggested that fluent word decoding

skills free up cognitive resources for prosodic reading; however, the authors did not find that prosody was an additional support for reading comprehension (Schwanenflugel et al., 2004). Other scholars have defined reading fluency as only the rate and accuracy of reading aloud (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Shinn, Good, Knutson, Tilly, & Collins, 1992). In this study, oral reading fluency will be defined as accuracy and rate or correct words per minute read aloud.

Oral reading fluency scores measure students' reading skills with grade level reading passages. When the reading passages are selected directly from the students' curriculum, oral reading fluency is an example of reading curriculum-based measurement (R-CBM). Reading CBM is defined specifically as a standardized set of procedures used to measure student achievement in the area of reading using curriculum-specific reading passages (Hosp et al., 2007). Reading curriculum-based measurement (e.g., oral reading fluency) is often used within a larger process called curriculum-based assessment (CBA). Curriculum-based assessment, broadly defined, is the process for determining students' progress towards mastering skills taught in the curriculum (Howell, Hosp, & Kurns, 2008).

In contrast to traditional R-CBM, oral reading fluency data may also be collected through the use of pre-selected grade level reading passages that do not come specifically from the curriculum. Several commercialized assessments such as DIBELS and AIMSweb include a measure of oral reading fluency using standard passages. Fuchs and Deno reviewed the literature on oral reading fluency and concluded that generic passages are as effective as curriculum-specific passages for monitoring student progress in reading (1994). In this review, they described the major advantage of using curriculum-specific passages to be the face validity of teacher-developed assessments and the usefulness for instructional decision-making (Fuchs &

Deno, 1994). On the other hand, they described the disadvantages of curriculum-specific passages such as students' familiarity with the test material and the wide variety of reading difficulty levels within one text that reduces the accuracy and utility of the data collected (Fuchs & Deno, 1994). Currently, oral reading fluency data are collected using either local curriculum materials or passages from commercialized programs such as DIBELS or AIMSweb. This study will focus on oral reading fluency data collection using pre-selected passages from outside of the school curriculum through packages such as DIBELS or AIMSweb.

Oral Reading Fluency for Universal Screening and Progress Monitoring

Within a response to intervention model, universal screening is intended to identify students who need additional academic support (Batsche et al., 2005). Once data are collected from all students, those who do not meet expectations are provided with additional academic support until they demonstrate adequate grade level skills. Academic growth is monitored through frequent progress monitoring using oral reading fluency. Universal screenings occur periodically throughout the year (e.g., three times a year) to select students in need of these supplemental services, and progress monitoring typically occurs once a week or every other week.

A perfect universal screening tool for reading would accurately distinguish between students who need additional instruction and those who do not (Jenkins & Johnson, 2011). Unfortunately, a perfect screener does not exist, so schools must weigh the strengths and weaknesses of various tools before selecting one measure. Oral reading fluency is one measure that some schools and teachers have found to be useful for universal screening in order to identify students in need of more intensive instruction (Deno et al., 2009).

In order to evaluate the accuracy of oral reading fluency for screening, a tool must have both sensitivity for identifying students who will struggle without additional interventions ("true positives") and specificity for identifying those students who are not at risk for later failure ("true negatives") (Jenkins & Johnson, 2011). Ideally, a universal screener has both high sensitivity and specificity and produces no "false negatives" or "false positives." "False negatives" refers to students who achieve the benchmark score but really need additional support, and "False positives" refers to students who do not meet the expected benchmark score but are not truly at risk for reading failure. Given that no screening tool is perfect, schools must select a screener that meets their decision-making needs. More specific research results regarding this topic are reviewed later in the section on criterion validity.

In the universal screening process, oral reading fluency scores are used to create benchmarks for expected student achievement (Hosp & Fuchs, 2005). However, oral reading fluency scores differ between various groups of students (e.g., special education, general education, honors) (Fewster & Macmillan, 2002; Shinn, Tindal, Spira, & Marston, 1987). This indicates that oral reading fluency data may be used to identify students who are meeting or exceeding expectations and to identify students who are in need of additional support. While oral reading fluency data are useful pieces of information, the research suggests that this measure does not perfectly identify students at risk for later reading failure, and it does not perfectly differentiate between different groups of students (Jenkins, Hudson, & Johnson, 2007).

If oral reading fluency scores generally can differentiate between different types of students, such as special education and general education students, then researchers argue this tool is useful for universal screening purposes. One study utilized discriminant function analysis to determine factors that accounted for group membership (general education, Title I, or special

education) (Shinn et al., 1987). The results suggest that 98.29% of the variance in group membership was accounted for by oral reading fluency scores (p=.0001) (Shinn et al., 1987). The authors of this study suggested that the adequate discriminant validity of oral reading fluency supports the use of these measures in educational planning and placement decisions. However, this study does not report whether oral reading fluency scores perfectly discriminated between different achievement levels within those student groups.

Oral reading fluency may also be used to monitor students' reading growth (Fuchs & Fuchs, 2004). Yell, Deno, and Marston (1992) collected survey data about special education teachers' views on using CBM for screening and progress monitoring. The authors found that special education teachers expressed generally favorable views towards using CBM for both purposes (Yell et al., 1992). On a scale of 1 (good idea) to 5 (bad idea), 85% of teachers were favorable towards using CBM for monitoring student progress (rating of 1 or 2) and 78% of teachers were favorable towards using CBM for screening (Yell et al., 1992). At this point, research has not compared general education teachers' views towards specific reading measures such as oral reading fluency for progress monitoring and screening.

An argument against the use of oral reading fluency for either purpose states that teacher judgments are just as accurate as using oral reading fluency to identify students in need of additional support. Teacher judgments of student achievement, regardless of accuracy, are important because they can affect the selection of reading materials, instructional groupings, and expectations for achievement. In one study, the correlation between teachers' holistic ratings and reading fluency scores for students in grades one to six was strong (average r=.86) (Fuchs & Deno, 1981). In another study, while teachers accurately nominated strong readers, they had more difficulty estimating student skills in the low to average skill level range (Begeny, Eckert,

Montarello, & Storie, 2008). Other studies suggest that teachers sometimes over or under estimate student reading achievement (Feinberg & Shapiro, 2003; Hamilton & Shinn, 2003). These studies suggest that teacher judgment does not perfectly identify students at risk for reading failure. Some teachers may question the reliability and validity of oral reading fluency because these measures are also not 100% accurate.

Oral Reading Fluency Assessment Considerations

Reading curriculum-based measurement has been studied for decades, beginning with the work of Stanley Deno and colleagues in the 1980s (Deno, 1985). Research on the reliability and validity of reading curriculum-based measurement, such as oral reading fluency, is important to establish before these data are used in schools. Additionally, it is important that this research focus specifically on the purposes of the tool for which it is designed (e.g., universal screening or progress monitoring) (American Educational Research Association, 1999).

In addition to traditional reliability and validity measures, Wolf (1978) called for more subjective measures of social validity to support the use of behavioral treatments and procedures. While there is little research or theory on assessment acceptability, this call may be extended to reading assessment research. Research on the reliability, validity, and social validity of oral reading fluency may improve or reduce research support for this measure.

Reliability. This criterion seeks to establish that oral reading fluency provides relatively accurate scores across different contexts and time. Error in measurement should be minimal to support the use of oral reading fluency for universal screening and progress monitoring. Also, evidence for the reliability of a test should be viewed in light of the intended purpose because the standard for the reliability of tests for screening decisions is different than reliability standards for other purposes such as progress monitoring (Salvia, Ysseldyke, & Bolt, 2010). Salvia,

Ysseldyke, and Bolt suggest that the reliability of a test for screening should be at least .80 and reliability of a test for progress monitoring should be at least .70 (to allow for slight fluctuation that might occur when measuring a skill frequently) (Salvia, Ysseldyke, & Bolt, 2010).

Research suggests that oral reading fluency has adequate test-retest reliability according to those standards. One study selected 30 student participants to retake the same oral reading fluency test one to three weeks after the initial data collection point. The authors of this study found that the test-retest reliability for oral reading fluency using standard reading passages was strong (between .92 and .97) (Hosp & Fuchs, 2005). Another study also reported a high level of test-retest reliability (.93) with delayed retesting four months later (Christ & Silberglitt, 2007). This indicates that the oral reading fluency measure meets reliability recommendations for universal screening and progress monitoring set by Salvia and colleagues (2010).

Another key element of technical support for oral reading fluency includes the interrater reliability of the measures. In one study, scorer agreement for words read correctly per minute was between 96-100%, with a mean reliability of 99.9% (Dunn & Eckert, 2002). The test administrators were doctoral level school psychology students with direct training in administering and scoring oral reading fluency. They also had additional practice administering and scoring oral reading fluency with feedback on their performance (Dunn & Eckert, 2002). This provides support that when training and practice are provided, results of oral reading fluency tests are highly consistent. However, it is still unclear how much support teachers might need to reach similar levels of consistency.

Other studies suggest that variation in the specific reading passages used may account for some variance in student performance. One study collected oral reading fluency scores from 37 students and manipulated the probes used with each student. Alternate reading passages

accounted for 10% of variance in the final scores (Poncy, Skinner, & Axtell, 2005). When comparing differences in performance on hard and easy passages for second and third grade students, another study reported the average difference to be 46 words read correctly per minute (Christ & Ardoin, 2009). This research suggests that it would be problematic if different passages were used for different students in universal screening or progress monitoring due to the possible variation in scores based on the passage provided.

Another measure of the accuracy of testing is the standard error of measurement (SEM). Poncy and colleagues (2005) reported that the standard error of measurement for oral reading fluency ranged from 4-18 word read correctly per minute (WCPM), depending on the number of passages provided. When the median score from performance of reading several alternate passages was collected, the standard error of measurement decreased to 4-12 WCPM (Poncy et al., 2005). The standard error of measurement is largest when only one score is collected, but the variation is reduced when the median score from three passages is used. Thus the SEM of oral reading fluency is reduced for universal screening, which requires higher levels of reliability due to its importance for instructional planning or grouping.

In another study, oral reading fluency data had an average standard error of measurement (SEM) of five to nine WCPM in controlled settings with consistent passages for each student. The authors did find that in more controlled settings (e.g., quiet rooms and consistent passages) the SEM was lower than in less controlled setting (e.g., noisy classrooms and variable passages) (Christ & Silberglitt, 2007). Current research provides evidence for the adequate reliability of oral reading fluency in controlled research settings. However, if teachers are required to collect these measures without adequate training, without controlled settings, or without equivalent passages, the reliability of oral reading fluency may be put into question.

Validity. According to the *Standards for Educational and Psychological Testing*, test validity refers to "the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests" (AERA et al., 1999, p. 9). The Standards indicate that validity is an essential consideration in developing and evaluating tests. This definition is also significant because it emphasizes the importance of considering the validity for the specific purpose of a measurement tool instead of validating the tool itself (AERA et al., 1999). Originally, test validity was defined in terms of content, criterion-related, and construct validity instead of a unitary construct (Goodwin & Leech, 2003); therefore a majority of research continues to provide support for the use of assessment tools through these types of validity.

Content validity. Content validity refers to the extent to which oral reading fluency actually measures the skills it purports to measure. The *Standards for Educational and Psychological Testing* (AERA et al., 1999) indicate that test content includes the "themes, wording, and format of items, as well as the guidelines for procedures regarding administration and scoring" (p. 11). Perhaps the largest concern from literacy experts regarding the use of oral reading fluency data as a measurement of reading fluency is that this tool does not include scoring procedures to formally measure prosody (Peverly & Kitzen, 1998). As mentioned previously, prosody is an essential element of reading fluency for many scholars and teachers (Daane et al., 2005; Hudson et al., 2005; Moskal & Blachowicz, 2006). By removing the measurement of prosody, some argue that oral reading fluency curriculum-based measurement is not a complete measure of reading fluency.

Also, some argue that oral reading fluency might actually measure cognitive abilities such as processing speed, not reading fluency (Kranzler, Brownell, & Miller, 1998). However, research has found that the significant relationship between oral reading fluency and reading

comprehension cannot be explained by general cognitive ability or by processing speed and efficiency (Kranzler et al., 1998). Despite this research, some literacy experts and teachers still question the content validity of oral reading fluency as a measure of reading achievement.

Construct validity. There are concerns that the oral reading fluency measure is not a good measure of the broader theoretical constructs of reading fluency or overall reading ability. Construct validity indicates that a test score represents a person's ability level on some psychological construct (AERA et al., 1999, p. 174) and to some this term is synonymous with validity. It is represented by evidence from theory, content, and interrelations with other test scores of similar constructs (AERA et al., 1999).

Some are concerned that for a global outcome measure that is intended to represent a student's reading achievement level, the oral reading fluency construct is limited. While research suggests that oral reading fluency may predict later reading comprehension, this measure does not directly assess comprehension (Shinn et al., 1992). Students could possibly read very quickly to meet the benchmark without understanding what was read. The concern is that, with this tool, some students with high fluency (but weak prosody or comprehension) may not be identified for additional support. On the other hand, some students with low fluency who are capable of comprehending the material may be unnecessarily identified. Without a direct measure of prosody or comprehension, some argue that the content of oral reading fluency is limited as a reading assessment construct and tool (Schilling, Carlisle, Scott, & Ji, 2007). Teachers may not value the construct of oral reading fluency compared to other reading measures, but research is needed to study what constructs teachers do value.

Criterion validity. Establishing the criterion validity for oral reading fluency can provide support for the technical adequacy of this measure for use in the schools. Criterion validity

measures the correlation between scores on a screening measure and another reading performance test, either concurrently or at a later time. Educators are interested in this tool because it is thought to measure and predict achievement in a more efficient manner than longer tests.

Universal screening is based on the premise that by collecting data from all students, those students at risk for failure may be identified early and then may receive supplemental instruction. In one study, the author identified benchmarks for fall student achievement based on initial reading fluency scores and found that these were reasonably accurate for identifying students below the 25th percentile in reading achievement on the Iowa Tests of Basic Skills (ITBS) at the end of the year (Schilling, Carlisle, Scott, & Ji, 2007). In third grade, the fall oral reading fluency DIBELS scores alone accounted for 45% of winter ITBS total reading score variance. However, Schilling et al. (2007) also found that 32%-37% of students who were identified as low-risk at the beginning of first grade did not meet benchmarks at the end of the school year in second or third grade. This is an example of universal screening results that produce "false negatives."

Additionally, a meta-analysis of oral reading fluency CBM found a moderately strong correlation between oral reading fluency and other standardized tests of reading (weighted average r= .67 across 289 coefficients) (Reschly, Busch, Betts, Deno, & Long, 2009). When oral reading fluency was used to predict reading achievement within the current grade, the average correlation was .71. The correlation was slightly lower (r = .63) when oral reading fluency was used to predict reading achievement scores at least one year after the oral reading fluency score was collected (Reschly et al., 2009).

Authors of another study reported that oral reading fluency data in first grade accounted for 67% of the variance in second grade oral reading fluency scores (p < .001, N = 342) (Good, Simmons, & Kame enui, 2001). The authors also reported that 96% of students who met the third grade oral reading fluency screening benchmark also met or exceeded expectations on the Oregon Statewide Assessment that year (Good et al., 2001). This research suggests that oral reading fluency can be used to predict current or later achievement but that some variance remains unexplained.

Additionally, authors of a longitudinal study with students in sixth and seventh grade found that oral reading fluency scores were significantly correlated to later high school achievement (Fewster & Macmillan, 2002). For students in eighth, ninth, or tenth grade, the correlations between oral reading fluency in sixth and seventh grade and English and Social Studies grades ranged from .30 to .46 (*p*<.005). The authors of this study suggested that these results support the use of oral reading fluency to screen for students at risk (Fewster & Macmillan, 2002). Another study confirmed that DIBELS oral reading fluency in first grade significantly predicted performance on the TerraNova California Achievement Test (CAT) and the Pennsylvania System of School Assessment (PSSA) in second and third grade (Goffreda, Diperna, & Pedersen, 2009). Several other studies have also supported that oral reading fluency may predict later achievement on statewide achievement tests (Crawford, Tindal, & Stieber, 2001; Jones, 2009). Research on the criterion validity of oral reading fluency indicates that this tool can be used to measure current or predict later reading achievement but that some variation in student achievement still remains unexplained without additional data.

There are a variety of explanations for the moderate correlations between oral reading fluency and other tests and the imperfect predictions of later achievement from the results of

universal screening. The results of these studies indicate that while oral reading fluency may identify a large portion of students who are at risk for later reading failure, it may also miss some students who are in need of more intensive instruction. Another explanation is that students are not provided with appropriate, differentiated instruction throughout the school year to continue their academic development at the expected level even when they are clearly identified as at risk. Another possibility is that students may be able to decode fluently but they are struggling with language comprehension; they may struggle with vocabulary, background knowledge, or some other aspect of comprehension.

With any explanation, it is clear that using oral reading fluency for identifying at risk students has the chance to produce "false negatives" or students who meet the benchmark but fail to meet the desired achievement level at a later date (Jenkins & Johnson, 2011). Other research has supported that when using oral reading fluency benchmark scores, both false negatives and false positives (students who do not meet the benchmark but meet the desired achievement level later) are found (Valencia et al., 2010). This lack of sensitivity and specificity also presents concerns for the use of oral reading fluency for progress monitoring if its accuracy at predicting later achievement is questionable.

Other assessment considerations. In addition to the reliability and validity research described above, several other elements may be considered when evaluating an assessment measure including face validity, consequential validity, and external validity. While these are not technically considered to be validity measures, they provide important information related to the potential strengths or weaknesses of a measure such as oral reading fluency. These considerations may also influence teacher attitudes towards oral reading fluency.

Face validity. There is evidence that puts the face validity of CBM measures into question. Research suggests that teachers are concerned that a one-minute reading measure at one point in time may not be an accurate measure of overall reading ability, given that it doesn't measure comprehension (Foegen et al., 2001; Yell et al., 1992). Furthermore, there is a concern that oral reading fluency is a measure of speed, not fluency, and will be misused or misunderstood by students and teachers (Samuels, 2007). Regardless of other evidence for the validity of oral reading fluency, without higher face validity there may continue to be a lack of teacher buy-in for oral reading fluency measures for universal screening.

Consequential validity. Consequential validity is another test characteristic that some use to evaluate an assessment measure. Some teachers and literacy experts have concerns with the consequential validity of oral reading fluency for universal screening and progress monitoring. First, authors of several studies have found that using oral reading fluency as a universal screener produces many false negatives (Riedel, 2007; Schilling et al., 2007; Valencia et al., 2006). This raises the concern that this tool may misidentify students and that these students will not receive the academic support they need. Another major concern is that teachers will teach to the test by incorporating activities into the curriculum that include reading short pieces of text very quickly (Deeney, 2010). Instead of maintaining a balanced reading curriculum that focuses on a variety of important reading skills (e.g., vocabulary and comprehension), some are concerned that teachers will focus only on teaching and practicing speed in order to see improvement on universal screening or progress monitoring measures (Deeney, 2010).

External validity. Another important question regarding the use of oral reading fluency in schools is the generalizability of the scores (Roth, 2004). External validity is the extent to

which research results may be generalized to other populations, settings, treatment variables, and measurement variables (Campbell, Stanley, & Gage, 1963). In a way, this characteristic looks at the predictive validity of a measure for diverse people and settings and therefore is not necessarily considered a separate type of validity. Test users are expected to use assessments that are appropriate for use with the population being tested (Salvia et al., 2010).

Some research has reported that oral reading fluency test results differ significantly by setting and tester, thus limiting generalizability (Derr-Minneci & Shapiro, 1992). However, the authors of another study found strong reliability across the administration of different probes and across different trained examiners (Kranzler et al., 1998). Furthermore, whether reliability and validity of oral reading fluency extend to older grade levels is inconclusive given that the majority of research focuses on elementary aged students. Authors of another research study found low to moderate correlations between tenth grade reading fluency and achievement measured by course grades (Espin & Deno, 1993). However, stronger correlations were found with the Test of Academic Proficiency reading subtest (Espin & Deno, 1993). The results of this study indicate that while oral reading fluency may not strongly predict grades, which are influenced by a variety of situational and motivational factors, it may predict reading performance on standardized tests. Research on generalizability across testers, settings, and student ages is inconclusive, and this too may influence the acceptability of the measure.

Additionally, assessment tools used to evaluate academic achievement and growth should be fair and free from linguistic and cultural biases (Roth, 2004). To demonstrate this external validity, oral reading fluency measures should be free of test bias, and correlations between oral reading fluency and other tests of reading achievement should be equivalent across gender, SES, racial/ethnicity, and general and special education students.
To test the fairness for diverse learners, research has been conducted on oral reading fluency for students with special needs, such as those who are deaf or hard of hearing (Allinder & Eccarius, 1999). In this study, oral reading fluency procedures were altered to allow students to sign instead of read aloud. Interrater reliability was acceptable (average 78% agreement); however, evidence for validity was not as strong, with small, non-significant correlations between reading fluency and a standardized reading assessment (Allinder & Eccarius, 1999). These results indicate that oral reading fluency through signing may not be an appropriate way to monitor the reading progress of deaf or hard of hearing students.

Additionally, some research has raised concerns regarding ethnic and gender bias in oral reading fluency (Kranzler, Miller, & Jordan, 1999). In this study there were significant group differences for gender and race, and the authors cautioned educators that the use of oral reading fluency with diverse populations might be inappropriate. On the other hand, research has provided evidence for the reliability and validity of oral reading fluency measures with bilingual Hispanic students and English-only speakers (Baker & Good, 1994). Still, research regarding the technical adequacy of oral reading fluency with diverse populations is mixed, and this may relate to teacher concerns for the use of oral reading fluency.

Conclusions. In summary, research provides evidence for the reliability and validity of oral reading fluency. However, when used alone as a measure of reading achievement, oral reading fluency is not able to identify students at risk for later failure with perfect accuracy. Additionally, teachers and scholars alike have questioned the face, consequential, and external validity of oral reading fluency measures. Still, the research base for using oral reading fluency is strong enough that many schools now require the collection of these scores. The acceptability

of oral reading fluency has yet to be confirmed for specific purposes, especially in light of the concerns surrounding the reliability and validity of these scores.

Social Validity and Acceptability

In addition to demonstrating that an assessment tool is technically adequate for use, it is important to demonstrate the social validity and acceptability of the goals, procedures, and outcomes of the assessment tool (Foster & Mash, 1999). "Social validity" may be considered a softer or more subjective measure of validity but nonetheless presents important information to include in the evaluation of a reading test (Wolf, 1978). A majority of social validity and acceptability research has focused on establishing that clinical or educational interventions are acceptable to the parents, students, and teachers involved (Finn & Sladeczek, 2001; Gresham & Lopez, 1996).

Much less is known about the acceptability of assessment tools such as oral reading fluency for various purposes. Based on the theoretical assumption that attitudes are related to behavior, it is reasonable to assume that attitudes about the acceptability of oral reading fluency data may influence the administration and use of this assessment tool. As Wolf said, "If the participants don't like the treatment then they may avoid it, or run away, or complain loudly. And thus, society will be less likely to use our technology, no matter how potentially effective and efficient it might be" (1978, p. 206). Therefore, research is needed to establish when teachers find oral reading fluency acceptable (for universal screening or progress monitoring) and to identify what factors relate to this acceptability.

School psychologists' acceptability. Much of the research on acceptability has sampled school psychologists, rather than teachers, who are the professionals often expected to administer and use oral reading fluency for universal screening and progress monitoring. In the research on

assessment acceptability, the method most widely studied is called curriculum-based assessment (CBA). Curriculum-based assessment, broadly defined, is the process for determining students' progress towards mastering skills taught in the curriculum (Howell et al., 2008). Curriculum-based measurement is one type of CBA (Hosp et al., 2007). Research on the acceptability of CBA is relevant to the acceptability of oral reading fluency because much of this research asks professionals about a CBA process that utilizes oral reading fluency.

Several studies have suggested that school psychologists view curriculum-based assessment (CBA) as more acceptable than other comprehensive, individually administered standardized tests. For instance, one study showed that curriculum-based assessment was rated as significantly more acceptable than standardized tests (Shapiro & Eckert, 1994). In this study, 249 school psychologists were presented with an assessment scenario for an individual child. Then they completed the Assessment Rating Profile (ARP), a 18-item 6-point Likert scale measure of assessment acceptability (Shapiro & Eckert, 1994). Using factor analysis, the authors determined that 14 items of the ARP measured one construct, and thus these 14 items were used in the comparison between conditions. This measure asked questions about the acceptability of the measure for the student's problem, and total scores could range from 14-84. The CBA condition utilized oral reading fluency to determine the students' instructional reading level, and the standardized testing condition used the Peabody Individual Achievement Test to evaluate the student's difficulties. Scores on the ARP indicated that school psychologists rated CBA (M =54.84, SD = 14.08) as significantly more acceptable than standardized testing (M = 41.08, SD =13.41), *t*(214) = 7.34, *p* < .001 (Shapiro & Eckert, 1994).

Another more recent study confirmed the findings from Shapiro and Eckert (1994). In this study, school psychologists rated curriculum-based assessment methods significantly higher

than brief experimental analysis (BEA) or norm-referenced assessment methods (Chafouleas, Riley-Tillman, & Eckert, 2003). In this study, 189 school psychologists participated and were randomly assigned to one of three conditions (brief experimental analysis, norm-referenced assessment methods, or curriculum-based assessment). Participants were provided with a scenario describing the use of each assessment method with a particular student, and then they completed the Assessment Rating Profile-Revised (Chafouleas et al., 2003). A univariate analysis (ANOVA) was conducted, and the results indicated a significant difference in the acceptability of the three different assessment conditions, F(2, 174) = 9.38, p<.001 (Chafouleas et al., 2003). A Scheffe' post hoc analysis was conducted, and the results indicated that CBA was rated as significantly more acceptable (M = 57.93, SD = 9.12) than BEA (M = 52.22, SD = 12.78; p = .045) and norm-referenced assessment (M = 47.97, SD = 13.86; p < .001). BEA and normreferenced assessment (NRA) did not significantly differ from one another (p = .156). The effect size of the difference between CBA and the other conditions, BEA and NRA, ranged from .2 to 1.08 (Chafouleas et al., 2003). This information may be helpful for understanding how these assessment methods compare to one another, but it does not reveal information about the acceptability of oral reading fluency for different purposes.

This is valuable information demonstrating that oral reading fluency is acceptable to school psychologists for identifying a student's specific reading skill deficit. This study does not provide information on why school psychologists find CBA to be more acceptable than standardized testing or nor does this research provide information on what factors influence the acceptability of those two assessments. Also, as indicated earlier, it is important to validate and demonstrate support for the use of assessment tools for specific purposes. This research does not investigate the acceptability of oral reading fluency for other purposes.

Variables related to school psychologists' acceptability. Several models of acceptability include training and use as factors related to acceptability (Witt & Elliott, 1985). One study attempted to determine whether the frequency of use and training of school psychologists were related to the acceptability of certain assessment methods. This study asked school psychologists to rate their level of use and training on specific assessment methods on a four-point Likert scale (none, minimal, moderate, frequent) (Chafouleas et al., 2003). The authors used Pearson correlations to study the relationships between variables and reported that the level of school psychologists' training in CBA significantly related to their level of acceptability for CBA, r(53) = .34, p < .05 (Chafouleas et al., 2003). School psychologists' use of CBA was not related to the acceptability of CBA, however frequency of use was related to the acceptability of norm-referenced assessment (NRA), r(62) = .45, p < .05 (Chafouleas et al., 2003). The power to detect significant differences was not provided, so the non-significant relationship between acceptability and use of CBA may be due to a lack of power and a smaller sample size. This research suggests that the level of training on and use of an assessment method may influence assessment acceptability. However, this study does not provide information about whether or not training and frequency of use will influence the acceptability of oral reading fluency for other purposes.

Teacher acceptability. Some research has supported that educators find curriculumbased measurement (CBM) procedures to be generally acceptable. Research showed that four participants, an administrator, general education teacher, reading specialist, and student resource assistant, found CBM procedures to be useful for identifying students who were at risk for academic failure (Druckman, 1997). In this study, a school psychologist, an intern, or the reading specialist administered CBM measures to all students in teachers' classrooms. Through

open-ended questionnaires, these teachers also described a preference for administering and interpreting CBM measures compared to criterion-referenced tests. They indicated that the time spent preparing for administration and administering the measure was not a concern. Three of the respondents thought that reading CBM was an accurate measure, but one teacher was unsure of the correlation between reading fluency and reading comprehension (Druckman, 1997). All four respondents indicated that CBM was a useful tool for progress monitoring or making instructional groups. These participants were all selected based on their experience preparing, administering, scoring, and interpreting CBM. Therefore, this research supporting the acceptability of CBM may not extend to teachers with less experience with CBM, and also this research does not consider the use of CBM for universal screening.

Another study using qualitative methods to study the acceptability of CBM involved training a teacher to use oral reading fluency CBM to inform instruction (Rimstidt, 2001). After the intervention, the author interviewed the teacher to learn about her views of the treatment validity, treatment utility, and instructional utility. After the intervention, the teacher reported that she valued reading speed as an important element of reading fluency. Also, the teacher indicated that CBM data graphs were motivating and useful, but she also expressed the concern that charting progress-monitoring data could be discouraging to students who do not meet the teacher's goals (Rimstidt, 2001). Lastly, the teacher reported that she did not believe that the instructional changes made as a result of R-CBM were helpful for improving student achievement (Rimstidt, 2001). This interview data provides rich data from one teacher who was provided with specific training, but the acceptability of teachers with less training and exposure to using oral reading fluency is still unknown. Also, this study does not consider teacher acceptability for an

specific purpose, it would be valuable to have research on teachers' attitudes towards oral reading fluency for various types of purposes.

A quantitative study sought to compare the teacher acceptability of CBA to normreferenced tests (Eckert, Shapiro, & Lutz, 1995). Two hundred twenty-four teachers participated in the study and were assigned to read a scenario describing the use of CBA or a scenario describing the use of a norm-referenced test in an individual evaluation. Teachers were stratified according to grade level taught and teacher type (general education or special education) to ensure that the results were not influenced by these variables. Teachers then completed the Assessment Rating Profile (APR), with 15 items measuring assessment acceptability. Results of a two-tailed t-test indicated that teachers consistently reported significantly higher levels of acceptability for curriculum-based assessment (M = 80.24, SD = 15.21) than for published, norm-referenced tests (M = 56.72, SD = 15.96), t(223) = 11.26, p = .001 (Eckert et al., 1995). Possible ratings of acceptability on the APR ranged from 18-90, and the full scale effect size for the difference between scenarios was 1.51 (Eckert et al., 1995).

This research again confirmed that oral reading fluency in the context of curriculumbased assessment is more acceptable than other assessment methods. Still, this research does not provide information about the acceptability of oral reading fluency for universal screening and progress monitoring. Also, this study sampled from districts in two states, but no information about the number of schools or the way these schools used student data was reported. There is a possibility that teachers at these schools could be unique in their level of training or experience with oral reading fluency.

Overall, research indicates that teachers view reading curriculum-based measurement as more favorable than other assessment methods for individual evaluations (Eckert et al., 1995).

However, teachers also report concerns with the use of oral reading fluency measurement (Druckman, 1997; Rimstidt, 2001). The quantitative research on the acceptability of oral reading fluency and curriculum-based measurement did not report information on the school contexts or the teachers' level of training or exposure to R-CBM (Eckert et al., 1995). Also, the qualitative studies did not provide measurable information on the level of teacher acceptability for CBM (Druckman, 1997; Rimstidt, 2001). Research is still needed to identify if teachers view oral reading fluency as an acceptable tool for the purpose of universal screening and progress monitoring.

School characteristics influencing acceptability. Using oral reading fluency for progress monitoring and universal screening requires a large amount of time and resources from school professionals. In order to find an assessment acceptable, teachers should approve of the procedures used in the assessment administration and have the necessary materials to carry out these procedures. To collect oral reading fluency data from all students, teachers need resources such as grade level passages, scoring booklets, a timer, space to test individual students, and personnel to assist with teaching students. Also, several times a year more teacher time is required in order to conduct a universal screening using oral reading fluency. Additionally, to monitor students' progress teachers would need to allocate time and materials to the process on a regular basis. Teachers are expected to balance time between collecting assessment data and providing instruction to students. Therefore, school factors including the time and resources available to teachers at each school may relate to the acceptability of oral reading fluency. Previous research and theory supports the premise that time and resources relate to teacher's attitudes and use of curriculum-based measurement (Rimstidt, 2001; Wesson, King, & Deno, 1984; Yell et al., 1992). Additionally, the time and resources needed differ for using oral reading

fluency for progress monitoring or universal screening, and as a result the acceptability may differ for each purpose.

In several studies, time was identified as a major barrier to implementing curriculumbased measurement. Collecting and using assessment data is time consuming for teachers, and this may limit the acceptability of the measure. A survey of 136 teachers of students with learning disabilities showed that only 43.8 percent of the sampled teachers used direct and frequent measurement (e.g., CBM measurement) (Wesson et al., 1984). These teachers reported that they spent ten percent of their time collecting, scoring, and analyzing CBM data. Forty-six percent of teachers reported not using CBM due to perceived lack of time (Wesson et al., 1984). Other reasons for not utilizing direct and frequent measurement included a lack of knowledge (24.2%) and lack of materials (4.2%).

Another study using survey methodology identified the major barriers to CBM implementation from the perspective of special education teachers (Yell et al., 1992). The special education teachers studied reported logistical concerns, such as not having the time and resources to collect the data (Yell et al., 1992). Also, a teacher using oral reading fluency for progress monitoring cited time as a barrier for using formative assessment data to inform instruction (Rimstidt, 2001).

Additionally, theory suggests that educational innovations are influenced by the resources that teachers and schools possess (Cohen & Ball, 2007). More elaborate innovations are more taxing on the time and resources of teachers and innovators. The use of oral reading fluency for universal screening may require more elaborate implementation plans and more time, resources, and people to complete successfully. For progress monitoring purposes, teachers need to not only collect the data but also analyze the data to inform instructional decisions. The use of oral

reading fluency data most likely differs for teachers who have ample time and resources to accomplish collecting and analyzing data without detracting from instructional time.

Therefore, time and resources are school variables that relate to how teachers perceive and use assessment. Research supports the premise that teachers of students with learning disabilities and teachers using oral reading fluency for progress monitoring are concerned with the use of CBM because of a lack of time and materials. This research does not test whether time and materials would relate to the perspectives of general education teachers who use oral reading fluency for universal screening. Whether there is a different level of acceptability for oral reading fluency in progress monitoring or universal screening is also unknown.

Teacher characteristics influencing acceptability. Teacher characteristics, including knowledge, training, years of experience, and frequency of use, have been reported to relate to teachers' perspectives of assessment methods.

Knowledge relates to teachers' attitudes towards academic assessments. Some teachers use their assessment knowledge to guide what they believe about how best to measure reading skills (Hall, 2005). These beliefs then also influence attitudes. The relationship may be bidirectional, with knowledge influencing attitudes or vice versa, with attitudes influencing knowledge (Hall, 2005). For instance, attitudes towards particular assessment methods may have a filter effect for gaining new knowledge (Hall, 2005). Teachers with negative attitudes toward oral reading fluency measurement may look for examples to support their beliefs and ignore instances when their beliefs are disconfirmed. On the other hand, new knowledge about a particular assessment tool might influence teachers' attitudes towards the assessment tool. From a theoretical standpoint, considering and measuring teacher knowledge about reading fluency.

In the survey research mentioned previously, teachers of students with learning disabilities also reported not using direct and frequent measurement due to lack of knowledge about how to use this type of measurement (Wesson et al., 1984). This study did not specify a definition for direct and frequent measurement, but from the literature review it is assumed that they were referring to assessment techniques using curriculum-based measurement. Teachers indicated that they did not have enough training or practice with the use of CBM-like measures to use them effectively (Wesson et al., 1984). Also, a qualitative study of teachers' use of formative literacy data showed that teachers reported that teacher knowledge about how to use data influenced successful data usage (Roehrig et al., 2008). The author of a study on intervention acceptability found that when teachers were presented with research efficacy information regarding a treatment, they were slightly more likely to indicate that the intervention was acceptable than those who did not have that research information (Fischl, 2008). This literature suggests that knowledge about curriculum-based measurement may be lacking, and this could relate to the acceptability of this measure. Additionally, more knowledge about how to use data is needed when collecting oral reading fluency for progress monitoring. If knowledge influences acceptability, it is possible that differences exist in the acceptability of oral reading fluency for progress monitoring or universal screening.

In addition to direct knowledge about oral reading fluency, training may relate to teacher acceptability. A literature review reported that undergraduate coursework can help teachers develop positive attitudes towards topics such as reading instruction (Hall, 2005). In particular, Stieglitz (1983) found a statistically significant difference between the attitudes towards reading instruction of teachers who completed a reading methods course and those who had not, f(78) = 3.69, p < .001. Therefore, an extension of this research finding would suggest that undergraduate

instruction in reading assessment may also influence attitudes towards reading assessment. However, one study reported that there was no difference in teacher acceptability of CBA between special education and general education teachers (Eckert et al., 1995). This is interesting given that special and general education teachers often complete a different set of undergraduate courses. However published studies on teachers' specific coursework were not identified in this review.

Formal training is not the only method of gaining knowledge about a subject matter; professional development may also contribute to the implementation of educational innovations such as using oral reading fluency for universal screening. Educational innovations, such as the mandate for collecting and using oral reading fluency data, require teachers to change their assessment practices. In some schools numerous learning and training opportunities may scaffold this change; whereas in other schools teachers may have few opportunities for formal training in this innovative assessment strategy (Cohen & Ball, 2007). Training is often necessary for change in attitudes or behavior to occur; mere exposure to curriculum-based assessment is not related to acceptability of this measurement process (Eckert et al., 1995). Special education teachers reported through surveys that more training in the use of CBM would help improve the administration and use of these measures (Yell et al., 1992). Training, both through a teaching degree program and through professional development, may influence teachers' attitudes towards oral reading fluency data.

Additionally, years of teaching experience may relate to teachers' acceptability of oral reading fluency. The longer a teacher has been teaching and using other assessment strategies, the more difficult it may be to change the attitudes and practices of that individual (Pajares, 1992). Authors of a study using survey data found that years of experience influenced beliefs

and practices regarding standardized testing (Urdan & Paris, 1994). Teachers who had been teaching longer than ten years held more positive beliefs about students' participation and effort in assessment than teachers who reported having less than five years of teaching experience (Urdan & Paris, 1994). Thus an extension of this finding would suggest that years of experience might also influence other teacher attitudes, such as their attitude towards the use of oral reading fluency data for various purposes.

Furthermore, the frequency of use may relate to teachers' acceptability of oral reading fluency. One teacher reported higher levels of acceptability for oral reading fluency when she was using the measure frequently (Rimstidt, 2001). More research is needed to confirm if the frequency of use relates to teachers' attitudes about using oral reading fluency.

Conclusions. The literature reviewed suggests that the acceptability of oral reading fluency is mixed, and research specifically on the use of oral reading fluency for different purposes (i.e., universal screening and progress monitoring) is lacking. The research suggests that several variables (i.e., time, resources, knowledge, training, years of experience, and frequency of use) may relate to teachers' acceptability of assessment procedures. Further research focusing on understanding why teachers find oral reading fluency acceptable or not is warranted.

The Present Study

While several studies have shown that teachers are more accepting of CBM or CBA than of other assessment methods (e.g., Druckman, 1997; Eckert et al., 1995), there are still many concerns that surround the fairness, appropriateness, and usefulness of oral reading fluency (Rimstidt, 2001; Wesson et al., 1984; Yell et al., 1992). Authors that have reported that school psychologists and teachers find CBM or CBA to be acceptable did not report surveying teachers

from various school contexts. Also, these studies did not evaluate the acceptability of CBM when used for the purpose of universal screening or progress monitoring. Lastly, these studies generally used either quantitative or qualitative methodology, instead of a mixed methods approach.

This study fills this gap by surveying teachers through cluster and stratified sampling in order to gather survey data from a variety of school contexts. Also, this study measures teachers' level of acceptability for oral reading fluency for the purposes of universal screening and progress monitoring and compares the two to determine if teachers find oral reading fluency more acceptable for one purpose or another. This mixed methodology provides a quantitative measurement of acceptability as well as rich qualitative data that provides for a deeper understanding of the teacher perspective.

Previous research has established that time, available resources, knowledge, level of training, years of experience, and frequency of use relate to teachers' acceptability of assessment methods. No research has measured how these variables relate to teachers' perceptions of oral reading fluency for universal screening or progress monitoring. This study will use focus group data to understand the attitudes of teachers and survey data to quantify the factors related to the acceptability of oral reading fluency data. Survey research will also explore teacher's opinions about what elements of reading are valuable and focus group data will provide information on how teachers use oral reading fluency data.

Research Questions and Hypotheses

This study will focus on five major research questions:

(1) Do teachers view oral reading fluency as an acceptable reading measure?

It was hypothesized that teachers have a moderate level of acceptability for oral reading fluency data for the purpose of universal screening and progress monitoring. Previous research is mixed regarding teachers' acceptability of oral reading fluency and similar measures. Some research supports that teachers find reading curriculum based measurement to be acceptable (Druckman, 1997; Eckert et al., 1995), while other studies show that teachers are concerned with the use of these types of assessment tools (Rimstidt, 2001; Wesson et al., 1984; Yell et al., 1992). Due to the mixed research support for the acceptability of measurements similar to oral reading fluency, it was hypothesized that there would be some teachers with high acceptability and some teachers with low acceptability ratings for this type of assessment. It was thought that these variable responses might lead to an average overall acceptability level. Therefore, the spread of the data would provide more valuable information than a measure of central tendency such as mean or median. Therefore, it was hypothesized that there would be a high level of spread or variability in teacher acceptability for this measure.

(2) Is there a significant difference between teachers' acceptability of oral reading fluency for the purposes of universal screening and progress monitoring?

It was hypothesized that teachers would find oral reading fluency for the purpose of universal screening more acceptable than the use of this measure for progress monitoring. Previous research has suggested that special education teachers find both purposes generally acceptable (Yell et al., 1992). Currently, administrators often require oral reading fluency data to be collected for universal screening, and sometimes additional support is provided for this data collection process. Progress monitoring with this measure would require an additional level of time and commitment from the teacher and therefore might not be as acceptable.

(3) To what extent are teacher characteristics (i.e., years of experience teaching, teaching certification status, training, knowledge, and frequency of use) and teacher perceptions of school characteristics (i.e., time and facilities/resources) related to their acceptability of oral reading fluency?

It was hypothesized that these seven variables would significantly relate to teacher acceptability for oral reading fluency for universal screening and progress monitoring. Previous quantitative and qualitative research has supported that these variables may relate to the acceptability of assessment methods (Chafouleas et al., 2003; Fischl, 2008; Rimstidt, 2001; Roehrig et al., 2008; Spear-Swerling, Brucker, & Alfano, 2005; Urdan & Paris, 1994; Wesson et al., 1984; Yell et al., 1992).

Theory and research indicates that years of teaching experience influences attitudes and behavior (Pajares, 1992; Urdan & Paris, 1994). Given that teachers with more experience may have been using other reading assessments for a longer amount of time, it was expected that if a teacher reported more teaching experience then their level of acceptability for oral reading fluency data would be lower. Also, how often the teacher uses oral reading fluency data was expected to relate to the level of acceptability. Previous research suggests that the frequency of use relates to the level of acceptability that educational professionals have for assessment methods (Chafouleas et al., 2003; Rimstidt, 2001). Therefore, it was hypothesized that teachers who use oral reading fluency more frequently would also have higher levels of acceptability for this assessment tool.

Training has been linked to influencing attitudes towards a particular topic (Hall, 2005; Stieglitz, 1983), and teachers have requested more training in the use of curriculum-based measurement (Yell et al., 1992). Therefore, it was hypothesized that training through a specific

educational major or endorsement (i.e., special education major/minor or endorsements in English, Language Arts, Reading, Reading Specialist), educational coursework, and/or professional development work would relate to teachers' level of acceptability for oral reading fluency. It was expected that students trained in a special education program and those with more exposure to instruction on educational assessment would have a more positive view of the oral reading fluency measure. Also, teachers who have more exposure to professional development that includes information on oral reading fluency were expected to have a higher level of acceptability for this data.

In addition to training, several authors have suggested that knowledge about an assessment or intervention influences perceptions of that assessment or intervention method (Fischl, 2008; Roehrig et al., 2008). Teachers also indicate that knowledge about an assessment method influences whether or not they choose to use that tool (Wesson et al., 1984). Therefore, it was hypothesized that teachers with more knowledge about oral reading fluency, curriculum-based measurement, universal screening, and progress monitoring would have higher levels of acceptability for the use of oral reading fluency.

Additionally, many teachers have indicated that time and resources influence how they use assessment tools (Cohen & Ball, 2007; Rimstidt, 2001; Wesson et al., 1984; Yell et al., 1992). Therefore, it was hypothesized that the more time and resources teachers have, the higher the level of acceptability they will have for using oral reading fluency in universal screening and progress monitoring.

Qualitative data was collected and analyzed to compliment and expand on the factors related to teachers' acceptability. It was hypothesized that teachers would report a variety of contextual variables and experiences that influence their attitude towards oral reading fluency

including the factors studied through quantitative survey data. This data was used to illuminate themes related to the factors studied in the quantitative data.

(4) What elements of reading do teachers find valuable for good reading?

Concerns with the construct and face validity of this measure make it important to understand what reading measures teachers find valuable. Social validity includes the consumers (e.g., teachers) opinions regarding the social importance of intervention or assessment goals and outcomes (Kazdin, 1980; Wolf, 1978). Therefore, it is important to understand if teachers value oral reading fluency as a part of good reading or if teachers value other reading constructs more. It was hypothesized that teachers would report comprehension as the most valuable measure of reading achievement, given the concerns often raised about the focus of oral reading fluency on speed rather than understanding (Samuels, 2007).

(5) How do teachers use oral reading fluency data?

It was expected that there are a variety of experiences with using oral reading fluency for universal screening and progress monitoring. This research describes teachers' experiences using oral reading fluency for various purposes.

CHAPTER III

METHOD

Research Design

To expand and improve the research on assessment acceptability, this study used mixed methods. Mixed methods is a class of research that uses both quantitative and qualitative research approaches in a single study (Johnson & Onwuegbuzie, 2004). By mixing qualitative and quantitative methods, it is expected that a study will draw on the strengths and address the weaknesses of each method (Johnson & Onwuegbuzie, 2004). The rationale for mixed methods in this study was to provide elaboration and illustration of quantitative survey data with qualitative data from focus group interviews (Greene, Caracelli, & Graham, 1989).

The quantitative and qualitative data were collected concurrently using a convergent design. The purpose of a convergent design is "to obtain different but complementary data on the same topic" (Morse, 1991, p. 122). In the convergent design, quantitative and qualitative data are collected simultaneously, data are analyzed separately, and finally the two sources of data are combined for interpretation. This mixed method design is outlined in more detail in the work of Johnson and Onwuegbuzie (2004) and Creswell and Clark (2011).

Sampling Procedures

School sites were identified for inclusion in this study based on the following characteristics: geographic region, grade levels present, Michigan Initiative for Behavior and Learning Initiative (MiBLSi) participation, and administrative support for participation. The sampling frame (Fowler, 2009) for this study was any elementary or intermediate school serving children in grades one to six within the Ingham County Intermediate School District (ISD), Clinton County Regional Education Service Area (RESA), and Shiawassee Regional Education

Service District (RESD). One large district in Ingham County was excluded from the sampling frame due to the difficulty of obtaining the district's approval for research through its own Institutional Review Board. Oral reading fluency scores were collected from students in grade one through grade six in many area schools, therefore the sampling frame included teachers of grades one to six. This sampling frame included schools from urban, suburban, and rural areas.

As mentioned earlier, schools were also selected based on their MiBLSi participation. MiBLSi is a state-wide project that assists schools in the development of schoolwide support systems in reading and behavior. The project began in 2003 and each year a new cohort of schools has been added to the project. "Cohort" refers to the group of schools that start MiBLSi in each subsequent year. Participating schools have received a \$3,000 stipend to assist with system implementation and have also received a series of didactic training sessions for school leadership. The collection of reading curriculum-based measurement (R-CBM), including oral reading fluency, from all students is a requirement for participation in MiBLSi. This made participating schools good candidates for inclusion in this study. Schools that did not participate in MiBLSi were also selected to participate so that teachers with a variety of training and experience with oral reading fluency would be included in this study.

Schools were chosen using multistage sampling and a combination of cluster and stratified sampling (Fowler, 2009; Newman & McNeil, 1998). First, clusters identified for possible selection were elementary and intermediate (grades 5-6) schools in the sampling frame; this included 36 schools in Ingham ISD, 13 schools in Clinton County RESA, and 15 schools in Shiawassee RESD.

Stratified sampling was used to select clusters based on their participation in MiBLSi. Schools were identified from two lists: a list of MiBLSi schools that is available to the public

online (Michigan Integrated Behavior and Learning Support Initiative, n.d.) and a list of non-MiBLSi schools developed from school district websites within each county. No schools from cohorts 1-3 were located in the mid-Michigan geographic region, so schools from Cohort 4-7 and schools not participating in MiBLSi were selected to participate.

Table 1 provides information about the stratification of the sampling frame and the stratification of the schools and teachers that participated in the study. In order to achieve a representative sample for the survey, the researcher's goal was to keep the percentage of schools from each cohort similar to the percentage of schools from each cohort in the sampling frame. Selection of schools from each strata was as follows: Cohort 4- two schools, Cohort 5- four schools, Cohort 6- eight schools, Cohort 7- twelve schools, No MiBLSi- two schools. The percentage of schools and teachers from each MiBLSi cohort that participated was similar to the overall stratification of the sampling frame.

Table 1

Percentage of Schools or Individuals from Cohorts 4-7

	Cohort 4	Cohort 5	Cohort 6	Cohort 7	No MiBLSi
Schools in the sampling frame	6%	11%	30%	47%	6%
Schools that participated	7%	11%	30%	44%	7%
Teachers that participated	4%	18%	30%	44%	4%

For the eligible schools, the researcher used emails with follow-up phone calls to solicit administrative support for participation (see Appendix A). Once an administrator agreed to allow the school faculty to participate, every general and special education teacher who taught reading in grades one to six in that school was invited by an email to participate in the survey portion of this study (see Appendix B).

After the survey invitation was e-mailed, teachers (N=72) from five different schools were

selected and invited through email to participate in a focus group interview (see Appendix C). Teachers were not required to complete the survey in order to participate in a focus group. One school from each stratified group (i.e., MiBLSi cohort or no MiBLSi participation) was selected in order to include teachers with a variety of experiences with universal screening and progress monitoring using oral reading fluency. The selected schools all collected oral reading fluency data for universal screening. All of the teachers from the five selected schools who taught grades 1-6 and used oral reading fluency were invited to participate.

Sample Size and Power

To determine the minimum sample size necessary to find significant results, power analyses for each type of statistical analysis were considered. To detect a small effect size, with a .05 significance level, and a power of .80, the necessary total sample size was 148 participants. The expected response rate for the full survey was about 38% based on pilot study data so at least 390 teachers needed to be invited to take the survey. To achieve this desired sample size, 27 schools were selected from the sampling frame, and the total number of teachers invited to take the survey was 402.

Participant Recruitment and Characteristics

Participants (N = 402) for the survey were recruited in the spring of 2012 through an email invitation sent to their inbox. Reminder emails were sent after one and two weeks. After three weeks, the survey was closed. One hundred sixty-six teachers started the survey; the response rate was 41.3%. One hundred forty-six teachers completed the survey; the completion rate was 36.3%. This is slightly above the typical response rate to online surveys, which Nulty (2008) reported to be approximately 33%. The descriptive analyses of demographic data (e.g., gender, ethnicity, highest level of education, years teaching, grade level, teaching position, and school

location) of those who completed and those who partially completed the survey were compared and found to be similar.

One week after the survey invitation was sent, seventy-two teachers from the survey sample were invited to attend a focus group. Twenty-two teachers attended one of the four focus groups, which represents 30.6% of those invited. Due to scheduling difficulties, the group size was uneven; two focus groups consisted of four teachers and two focus groups consisted of seven teachers.

Specific survey participant demographics are provided in Table 2. A majority of the survey participants were female (87.0%) and Caucasian/white (97.6%). There were teachers in the sample who reported either little or considerable years of teaching experience (ranging from less than 5 years to more than 20 years). Also, teachers from all grade levels participated, with approximately equal proportions of teachers in first through fifth grade and fewer sixth grade teachers participating. A majority of the teachers who participated taught in either suburban or rural schools (35.4% and 57.1%, respectively) while a smaller percentage of teachers taught in urban settings (7.5%). Also, a large percentage of teachers who participated were general education teachers (75.8%).

Specific focus group participant demographics are provided in Table 3. The ratio of males to females and the proportion of teachers in general education or other teaching positions was similar in the focus groups and survey. However, there were no fifth or sixth grade teachers participating in the focus groups, there was a larger number of highly experienced teachers in the focus groups than in the survey (40.9% of focus group participants had 20 or more years of teaching experience), and more teachers from suburban settings participated in the focus groups than in the survey (63.3% of focus group participants taught in suburban settings).

Table 2

Demographic Information of the Survey Participants (N = 164)

Descriptive Information	n (%)
Gender	
Male	21 (13.0%)
Female	141 (87.0%)
Ethnicity	× ,
Caucasian/White	160 (97.6%)
Black or African American	1 (0.6%)
American Indian or Alaskan Native	1 (0.6%)
Native Hawaiian or other Pacific Islander	1 (0.6%)
Highest Education Level	
B.A. or B.S.	39 (23.8%)
M.A. or M.S.	123 (75.0%)
Specialist/Ed.S.	2 (1.2%)
Total Number of Years Teaching	. ,
1-5	21 (13.0%)
6-10	40 (24.8%)
11-19	56 (34.8%)
20 +	44 (27.3%)
Current Grade Level	~ /
1	22 (13.6%)
2	25 (15.4%)
3	19 (11.7%)
4	27 (16.7%)
5	28 (17.3%)
6	6 (3.7%)
Mixed	35 (21.6%)
Current Teaching Position	. ,
General Education	122 (75.8%)
Special Education	24 (14.9%)
Reading interventionist/specialist	5 (3.1%)
Title I	8 (5.0%)
Other	2 (1.2%)
School Location	
Urban	12 (7.5%)
Suburban	57 (35.4%)
Rural	92 (57.1%)

Table 3

Demographic Information of the Focus Group Participants (N=22)

Descriptive Information	n (%)		
Gender	× ×		
Male	3	(13.6%)	
Female	19	(86.4%)	
Highest Education Level		()	
B.A. or B.S.	6	(27.3%)	
M.A. or M.S.	16	(72.7%)	
Total Number of Years Teaching			
1-5	5	(22.7%)	
6-10	3	(13.6%)	
11-19	5	(22.7%)	
20 +	9	(40.9%)	
Current Grade Level			
1	3	(13.6%)	
2	6	(27.3%)	
3	3	(13.6%)	
4	4	(18.1%)	
5	0	(0%)	
6	0	(0%)	
Mixed	6	(27.3%)	
Current Teaching Position			
General Education	16	(72.7%)	
Special Education	2	(9.1%)	
Reading interventionist/specialist	3	(13.6%)	
Title 1	1	(4.5%)	
School Location			
Urban	5	(22.7%)	
Suburban	14	(63.6%)	
Rural	3	(13.6%)	
MiBLSi Cohort			
4	2	(9.1%)	
5	9	(40.9%)	
6	5	(22.7%)	
7	5	(22.7%)	
No MiBLSi	1	(4.5%)	

Measures

Participants completed an online survey, which included a demographics section and multiple questions to measure dependent and independent variables. A pilot study was conducted

as a part of this research study to evaluate the clarity of survey items and to calculate reliability statistics for the revised scales. Revisions made to the survey based on pilot study data are described in the following section. The final complete survey included 57 questions and took approximately 30-45 minutes to complete (see Appendix D).

Dependent Variables

Assessment acceptability. The Acceptability Rating Profile- Revised (ARP-R) (Eckert, Hintze, & Shapiro, 1999) is a survey instrument that measures a consumer's acceptability of an assessment tool. First, the participant is provided with a written description of a specific assessment scenario. Then the participant completes a twelve-item, 6-point Likert scale measure and each item is rated from Strongly Disagree to Strongly Agree. An example of one APR-R item is "This assessment would be appropriate for a variety of students." Overall scores for this measure are calculated by adding together item responses, which are scored from one (strongly disagree) to six (strongly agree). Scores are on an interval scale and range from twelve (indicating low acceptability) to 72 (indicating high acceptability).

This measure has been used in research, and previous researchers refined the original version to create the APR-R. Eckert, et al. (1999) reported that the internal consistency ranged from .94 to .99, and the test-retest reliability ranged from .82 to .85 across 1, 3, 6, and 12 months. In previous research, content analysis was used to support the theory that nine items on this scale evaluate whether the assessment process is generally acceptable, two items evaluate whether the tool is appropriate for a variety of problems, and one item evaluates the effectiveness of the assessment method (Eckert et al., 1999). To provide evidence to support the construct validity of this survey, confirmatory factor analysis was used to estimate how well the twelve items measured assessment acceptability. Goodness of fit statistics suggested a very good fit for this

model and that the APR-R was an appropriate measure of general assessment acceptability (Eckert et al., 1999).

For this study, the APR-R was revised to make the wording reflect using oral reading fluency for the purposes of universal screening and progress monitoring. In Appendix D, the instrument is provided with the original wording in parenthesis and revised wording bolded. The first assessment scenario and several items were changed to describe universal screening using oral reading fluency. The second assessment scenario and several items were altered to represent the purpose of progress monitoring. In the pilot study, Cronbach's alpha for the universal screening scale was .98, and Cronbach's alpha for the progress monitoring scale was .96. In the pilot study, universal screening acceptability scale scores ranged from 41 to 72 (M = 62.55, SD = 8.14) and the progress monitoring acceptability scale scores ranged from 37 to 71 (M = 58.90, SD = 8.62), suggesting adequate variability. In the full study, the internal consistency reliability estimates for the universal screening scale was .98 and for the progress monitoring scale was .99. These results suggest adequate reliability for the revised versions of the APR-R.

Independent Variables

Time and resources. Teachers' perceptions of school working conditions, including time, facilities, and resources, was measured using the North Carolina Teaching and Working Conditions Survey (TWC) (New Teaching Center, 2011). This survey was originally developed by the North Carolina Professional Teaching Standards Commission in 2002 and has since been refined and used in eleven different states as a tool to measure teacher perceptions of working conditions. This study used two scales from the 2010 North Carolina Teaching and Working Conditions Survey, the scales that measure time and facilities/resources.

Each question is rated by a teacher on a scale of one (strongly disagree) to five (strongly agree). Item scores were added together for each scale; seven questions measured time, and nine questions measured facilities and resources. These scores were on an interval scale with time scores ranging from seven to 35 and facilities and resources scores ranging from nine to 45. Higher scores indicated that teachers perceived that teachers at their school have adequate time, resources, and facilities, while lower scores indicate a lack of appropriate time, resources, and facilities.

In a study that used the 2006 version of this scale, factor analysis was used to support that these two scales measured unique constructs at the elementary and middle school level, with each question loading onto the expected factor (Ladd, 2011). Slight changes to wording and order were made on the 2006 version to improve the 2010 version. The Cronbach's alpha for the time and facilities/resources scales from the 2010 version was reported to be .86 and .88 respectively (New Teaching Center, 2011).

For the current study, several items on the facilities/resources scale were modified to focus on the resources teachers have for assessment practices. The modified version of these measures is presented in Appendix D as a part of the complete teacher survey, with the original wording in parenthesis. One example item from the time scale is "Teachers have time available to collaborate with their colleagues" and an example item from the resources scale is "Teachers have sufficient access to appropriate assessment materials." In the pilot study for the current research, Cronbach's alpha for the time scale was .88 and Cronbach's alpha for the resources scale was also .88. Scores on the time scale in the pilot study ranged from 7 to 25 (M= 16.40, SD = 5.67). The mean of the resources scale was 29.35 (SD = 7.55) with a range of scores from 14 to 42. These results in the pilot study suggested adequate variability and reliability. In the full

study, Cronbach's alpha for the time scale was .84, and Cronbach's alpha for the resources scale was .80.

Years teaching. Information was collected about the teachers' number of years teaching. Years teaching was measured using an ordinal scale with 5 possible categories (1-2 years, 3-5 years, 6-10 years, 11-19 years, 20 or more years). In the pilot study, there were teachers from each category of teaching experience, indicating adequate variability.

Teaching certification. Teaching certification status was measured using a nominal scale. Teachers reported if they had earned a major, minor, or endorsement in one the following categories: Learning Disabilities, English, Language Arts, Reading, or Reading Specialist. The teaching certification variable was computed as a dummy variable with 0=no special education certification or reading endorsements and 1=special education certification or reading endorsements.

Teacher knowledge. Additionally, teachers answered eighteen survey items about their knowledge of universal screening, progress monitoring, oral reading fluency, and educational assessment (see Table 4). Questions on this scale were drawn from content in the books *The ABCs of CBM* (Hosp et al., 2007) and *Best Practices in School Psychology* (Thomas & Grimes, 2008), questions on the *Assessment Literacy Inventory* used in previous research (Mertler, 2009), and questions on the *Teacher Knowledge Survey* used in previous research (Spear-Swerling & Cheesman, 2011).

To establish the content validity of the knowledge scale, several assessment experts reviewed the items in this scale and provided feedback. Teachers in the pilot study also provided feedback on any confusing items. From this feedback, items were refined to be clear to the reader. Incorrect items were scored as a zero and correct items were scored as a one and then the

item scores were added together. The knowledge scale was measured on an interval measurement scale with possible scores ranging from zero to 18. For the overall scale in the pilot study, mean teacher knowledge score was 13.26 (SD = 1.48). Scores ranged from 11 to 17, and the most frequent score was 13. Knowledge items 5, 11, 12, 16 had no variation. Items 5, 11, and 16 teachers all answered correctly and item 12 all teachers answered incorrectly. The descriptive statistics suggested that other items had adequate variability. This type of variation is desirable, indicating that some of the items were easy and some were difficult.

Once the items were developed, each items' content were compared with the *Standards for Teaching Competence in Educational Assessment of Students* developed in 1990 by the American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), & National Education Association (NEA). The theoretical categories for the knowledge scale reflect three standards from this document: (1) "Teachers should be skilled in choosing assessment methods appropriate for instructional decisions," (2) "The teachers should be skilled in administering, scoring, and interpreting the results of both externally-produced and teacher-produced assessment methods," and (3) "Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement" (AFT, NCME & NEA, 1990). Items chosen for the knowledge scale targeted these standards of competence with a focus on oral reading fluency. Table 4 describes how each knowledge item aligns with one of these standards. The categories were shortened and called: (1) Choosing assessment methods for various purposes (2) Administering, scoring, and interpreting tests (3) Decision-making with assessment results.

Table 4

Knowledge Scale Items

Knowledge Category	Survey Numbers	Number of items
Choosing assessment methods for various purposes	23, 25, 30-33	6
Administering, scoring, and interpreting tests	17-20, 21, 26, 34	7
Decision-making with assessment methods	22, 24, 27-29	5

Frequency of use. Several items were used to collect information on how often teachers collected or used oral reading fluency data for universal screening and progress monitoring. In the pilot study, participants reported a wide-range of use for oral reading fluency, with scores ranging from 7 to 24, suggesting adequate variability. In the pilot study the mean score was 15.77 (SD = 5.07). The original five items did not account for teachers who did not collect data themselves but who worked at schools where data were collected. Therefore, this scale was altered to identify teachers at schools where data were not collected, teachers who did not participate in data collection but used it once it was collected by others, and teachers who were actively involved in data collection and analysis. The final scale included seven items.

Frequency of use was measured by asking how involved teachers were in data collection, how many students the teachers progress monitor using oral reading fluency, the frequency with which they use oral reading fluency for both progress monitoring and universal screening, the frequency with which they use progress monitoring data to inform instruction, and the frequency with which they consult with other teachers about the results of a universal screening (7 items total). First, teachers were asked if oral reading fluency was collected for universal screening at their school and if not, they skipped the related items and moved to a question asking if oral reading fluency was collected for progress monitoring at their school. If not, they skipped the related items and went to the next portion of the survey. If this type of data was collected for universal screening and/or progress monitoring in their school, teachers were directed to a page where they answered more specific questions about their level of involvement in the collection and use of the data. Five of the items on this scale were scored from one (indicating no use of oral reading fluency) to five (indicating a high level of use). Two of the items were scored 0 to 3. The items were added together and resulted in one overall score for frequency of use that ranged from five points (indicating data were not collected) to thirty-one points (indicating frequent use). If teachers reported this data were not collected, then they received a score of 0 on two items and a score of 1 on the other five items. The lowest possible score on this scale was a five, indicating that this type of data was not collected at their school. A low score indicated that the data may have been collected, but the teachers were mostly uninvolved in collecting or using that data. A high score indicated that the teacher was frequently engaged in collecting and/or using the oral reading fluency data.

Training. Lastly, training in administering and using oral reading fluency data was measured through survey items regarding instruction in undergraduate coursework, graduate coursework, and professional development opportunities (see Appendix D). Additionally, teachers were asked to rate their level of preparation for administering and using oral reading fluency data from each of those instructional experiences (6 items total). Instructional experiences were rated from one to four (2 items) and from one to five (1 item), and the level of preparation was indicated on a scale from one to four (from very unprepared to very prepared) (3 items). In total, the lowest possible score on this scale was a score of six, indicating a low level of training. The highest possible score on this scale was a score of 25, indicating a high level of 21, suggesting adequate variability.

Procedures

Survey pilot study. A pilot study was conducted to determine the time necessary to complete the survey, to test the survey items, to seek feedback on confusing items, and to check the reliability of revised scales (i.e., APR-R, Time, Resources). First, several assessment experts reviewed the content of the knowledge scale and provided feedback. Once the measures were solidified, IRB approval was granted under the category of exempt. An invitation to the pilot study survey online was sent to 53 elementary school teachers from two mid-Michigan schools. Teachers who participated were compensated with twenty-dollar Amazon gift cards sent via email. Of those invited, 23 teachers agreed to participate, and 20 completed the survey. According to Rea & Parker (2005), 20-40 respondents is an acceptable number for pilot study were used to confirm the reliability of revised scales and to alter any confusing items before full survey administration.

Survey administration. The revised survey was administered using a common online survey tool called Survey Monkey. Teachers were emailed a unique link to the survey and provided information regarding the purpose of the study in the email (see Appendix B). In this email, teachers were informed that upon completion of the survey they would be given a twentydollar gift card to Amazon as a token of appreciation for their time. After reading the consent form, they continued to complete the survey. The consent form for the pilot study is presented in Appendix F, and the consent form for the full study is presented in Appendix G.

Focus groups. The researcher led several focus groups to gain a deeper understanding of teacher acceptability of oral reading fluency. Focus groups are appropriate to use when the research purpose is to shed light on quantitative data collection, to understand the range of ideas

and feelings people have about a topic, and/or to uncover factors that influence attitudes and behavior (Krueger & Casey, 2009). Focus group interviews in this study were used to explore teachers' attitudes about the collection and use of oral reading fluency data and used to complement quantitative data for research questions one, two, and three. Additionally, focus group data were used to answer research question five regarding how teachers use these data. At the beginning of each focus group, important terms were defined to clarify the use of language in the focus group questions (e.g., universal screening, progress monitoring, oral reading fluency). The questioning protocol, including definitions provided to teachers, is presented in Appendix E.

Procedures for planning and moderating focus groups were based on suggestions from the book *Focus Groups: A Practical Guide for Applied Research* (Krueger & Casey, 2009). The purpose of including teachers from various schools in each focus group was to increase teachers' comfort level to share diverse opinions and to increase a feeling of safety that the information shared would not be reported back to their school administrators (Krueger & Casey, 2009). Each group included teachers from two or more different schools. Two groups had four teachers and two groups had seven teachers, for a total of 22 participants. Groups of four to eight are considered ideal by Krueger and Casey (2009).

Informed consent was required through a signed consent form before the start of each focus group. The informed consent form used for teachers is presented in Appendix H. All interviews were audiotaped. The meetings took place in the evening hours at Erickson Hall on the campus of Michigan State University. Each focus group lasted ninety minutes. The researcher was the moderator for all teacher focus groups. For participating in a focus group, teachers were offered a 50-dollar Amazon gift card plus fifteen dollars in cash to pay for transportation costs.

Two trained school psychology graduate research assistants transcribed the focus group

interviews. Each transcript was reviewed and compared to audio by the other trained research assistant and the principal investigator. Any edits were made using track changes and then accepted by the principal investigator in the final review. Focus group documents were blinded by replacing names with "Speaker #" and specific information such as school name with codes such as "School #."

Data Analysis

An overview of data analysis is presented in Table 5. First, all assumptions for the statistical analyses were reviewed. For the first research question, descriptive statistics for the acceptability of oral reading fluency for universal screening and progress monitoring were calculated.

For the second, a repeated measures ANOVA was used. The repeated measure was the assessment acceptability of oral reading fluency for universal screening and for progress monitoring because each teacher completed both scales. The seven independent variables (i.e., time, resources, knowledge, training, years of teaching, teaching certification, and frequency of use) from research question three were used as covariates to determine if there was an interaction effect. A *p* value below .05 indicated statistical significance. For the third research question, two multiple linear regressions with the acceptability of oral reading fluency for universal screening and progress monitoring as dependent variables were used.

For the fourth research question, descriptive statistics were used to indicate how teachers value each element of reading. Teachers rated each reading element from one (very unimportant) to seven (very important).

Focus group data were used to complement the quantitative data from the first three research questions and to address the fifth research question. When analyzing focus group data,

an analytic framework for identifying key concepts was used (Krueger & Casey, 2009, p. 125). The objective in this framework was to understand how participants viewed a topic (e.g., the acceptability of oral reading fluency), and to discover core ideas (e.g., factors that influence this level of acceptability). The key task in this analysis procedure was to identify a limited number of important ideas that shed light on the topic of teacher acceptability (Krueger & Casey, 2009).

After all focus groups were completed and transcribed, the researcher and two trained graduate research assistants coded the data into categories using qualitative data analysis software (NVIVO). The researcher attended a full day workshop at the University of Michigan regarding how to conduct qualitative data analysis using NVIVO, and this information was shared with the graduate research assistants through a training session. Each transcript was coded independently, once by the principal investigator and once by one of the graduate research assistants, using descriptive coding (Saldaña, 2009). Descriptive coding summarizes the basic topic of a passage using a word or short phrase. Then the researchers worked together to draw connections between codes and to create a categorical organization to the codes. Following this, each transcript was coded again using the finalized list of codes and focused coding (Saldaña, 2009). Focused coding is a method that identifies the most relevant categories in the data, and the goal is to develop major themes from the data. In the next stage of analysis, the primary researcher and two assistants held four meetings to go through each transcript and to reach agreement regarding coding differences. No attempt to develop a numerical reliability rating was made because the goal was agreement. Each coding difference was discussed and debated until the group agreed on appropriate codes. This method has been used in previous educational research (Harry, Sturges, & Klingner, 2005). Summary statements and visual illustrations were created to describe the qualitative focus group data and supplement the first, second, and third
research questions. Additionally, the data analysis focused on identifying how teachers used oral

reading fluency in their daily practice.

Table 5

Data Analysis Methods

Research Question	Measures	Variables	Data Analysis
1. Do teachers view oral reading fluency as an acceptable reading	APR-R for universal screening	Acceptability of oral reading fluency for universal screening	Mean, Median, Mode Visual analysis
measure?	APR-R for progress monitoring	and progress monitoring	of data Inter-quartile range
	Focus group interviews		Descriptive and focused coding
2. Is there a significant difference between teachers' acceptability	APR-R for universal screening	Acceptability of oral reading fluency for universal screening	Repeated measures ANOVA
of oral reading fluency for the purposes of universal screening and	APR-R for progress monitoring	and progress monitoring	
progress monitoring?	Focus group interviews		Descriptive and focused coding
3. To what extent are teacher characteristics (i.e., years of	Demographic form	Teaching certification status, years teaching	Multiple linear regressions
experience teaching, teaching certification status, training,	NC Teaching and Working Conditions	Time and Resources	
knowledge, and frequency of use) and teacher perceptions of	Survey	Knowledge, frequency of use, training	
school characteristics (i.e., time and facilities/resources)	APR-R	Acceptability of oral reading fluency	
related to the acceptability of oral reading fluency?	Focus group interviews		Descriptive and focused coding

Table 5 (cont'd)			
4. What elements of reading do teachers find valuable for good reading?	Likert scale ratings on importance of reading elements	Value of reading elements	Descriptive statistics (mean and median of each reading element)
5. How do teachers use oral reading fluency data?	Focus group interviews	Use of oral reading fluency	Descriptive and focused coding

Chapter IV

RESULTS

This chapter presents the results of data analysis for the survey and focus group data. First, the descriptive statistics for the independent and dependent variables are presented. Additionally, the method for handling missing data, the analysis of statistical assumptions, and the results for each research question are presented.

Descriptive Statistics

Data analysis for this study was conducted using SPSS Version 20. The survey data were imported from Survey Monkey directly into SPSS. Values for each item were set, and missing items were marked. Then, the variables described in chapter three were computed. Next, descriptive statistics for the independent and dependent variables were reviewed. When one or more items for a scale were missing, the total variable was marked missing and not calculated in the descriptive statistics. Table 6 contains the means, skewness, kurtosis, and standard deviations of each of the variables. Table 7 contains the range, minimum, and maximum score for each scale.

Table 6

Variable	Ν	Mean	Standard	Skewness	Kurtosis
			Deviation	(Standard	(Standard
				Error)	Error)
Dependent Variables					
Universal Screening	142	60.73	11.00	-1.80 (0.20)	5.21 (0.40)
Acceptability					
Progress Monitoring	142	58.15	13.51	-1.50 (0.20)	2.55 (0.40)
Acceptability					
Independent Variables					
Years teaching	161	3.73	1.08	-0.62 (0.19)	-0.21 (0.38)
Certification	150	0.73	0.44	-1.07 (0.20)	-0.88 (0.39)

Descriptive Statistics for Survey Data

Table 6 (cont'd)					
Knowledge	139	13.52	2.04	-0.41 (0.21)	-0.31 (0.41)
Time	146	18.40	5.44	0.01 (0.20)	-0.27 (0.40)
Resources	146	33.08	5.55	-0.10 (0.20)	-0.24 (0.40)
Training	144	14.45	3.84	0.249 (0.20)	-0.25 (0.40)
Frequency of Use	146	21.62	6.60	-0.68 (0.20)	-0.78 (0.40)

Table 7

Ranges for Survey Data

Variable	Ν	Range	Minimum	Maximum
Dependent Variables				
Universal Screening Acceptability	142	60	12	72
Progress Monitoring Acceptability	142	60	12	72
Independent Variables				
Years teaching	161	4	1	5
Certification	150	1	0	1
Knowledge	139	9	8	17
Time	146	27	7	34
Resources	146	28	17	45
Training	144	19	6	25
Frequency of Use	146	25	5	30

The mean scores for the acceptability of oral reading fluency (M = 60.73 and M = 58.15) indicate that on average teachers held positive attitudes toward oral reading fluency for both universal screening and progress monitoring. Possible scores on these scales were from 12 to 72, and a score of 42 indicates a neutral attitude toward oral reading fluency. A score of 24 or lower means that on average the teacher disagreed moderately or strongly with statements about the acceptability of oral reading fluency. A total score of 60 or above on the APR-R means that, on average, the teacher agreed moderately or strongly with statements about oral reading fluency acceptability. This is discussed in more detail in the results for research question one.

More than half of the teachers who participated reported either a special education background or a teaching endorsement in reading (73.3%). Teachers were asked to endorse all

teaching major, minors, and endorsements on the survey that applied. The responses were as follows: Learning disabilities major or endorsement (N = 26), English elementary level minor (N = 30), Language Arts group major, minor, or endorsement (N = 52), Reading endorsement (N = 21), Reading Specialist endorsement (N = 13), and Other (N = 77). Also, more than half of the teachers reported having 11 or more years of teaching experience (62.1%). Figure 1 shows the years of teaching experience that was reported by participants in the survey.



Figure 1. Teaching experience. Number of teachers reporting different years of teaching experience.

For the knowledge scale, most teachers answered more than half of the items correctly. There was a possible range of 0 to 18 items correct on this scale. The mean number of items correct was 13.52 items. Table 8 shows the percentage of teachers who answered each item correctly or incorrectly, and the number of teachers who did not answer each item. The full text of each item is included in Appendix D. The table is organized by the conceptual topics described in the methods. Figure 2 provides a visual display of teachers' total knowledge scores.

Table 8

Knowledge Questions Correct or Incorrect

Item Number and Brief Description	N	% Missing	% Incorrect	% Correct
Choosing assessment me	ethods t	for various pur	poses	
23. Reason for conducting universal screening	160	2.4	8.1	91.9
25. Characteristics of universal screening	159	3.0	16.4	83.6
30. Characteristics of progress monitoring	155	5.5	46.5	53.5
31. Definition of reliability for assessment	154	6.1	29.9	70.1
32. Considerations for choosing a universal	154	6.1	15.6	84.4
33. Definition of construct validity for assessment tools	149	9.1	60.4	39.6
Administering, scorir	ng, and	interpreting te	sts	
17. Reading skills measured with oral reading fluency	162	1.2	16.0	84.0
18. Relationship between oral reading fluency and reading comprehension	160	2.4	28.1	71.9
19. Relationship between oral reading fluency and reading comprehension	154	6.1	31.8	68.2
20 Characteristics of oral reading fluency	160	24	18 1	81.9
21. Universal screening with oral reading fluency administration guidelines	160	2.4	5.6	94.4
26. Interpreting cut-score points	159	3.0	15.1	84.9
34. Interpreting scores from standardized, norm-referenced tests for comparison to peers	149	9.1	24.2	75.8
Decision-making wi	th asse	ssment method	ls	
22. Identifying at-risk students with	160	2.4	0.6	99.4
24. Universal screening using oral reading fluency data for decision making	159	3.0	36.5	63.5
27. Using progress monitoring oral reading fluency data	159	3.0	3.1	96.9
28. Detecting academic skill growth when progress monitoring with oral	156	4.9	87.2	12.8
29. Decision-making with progress monitoring data	158	3.7	5.1	94.9



Figure 2. Knowledge. Histogram depicting the total number of items teachers answered correctly on the knowledge scale (N = 139).

Overall, teachers reported a lack of time to assess and teach students, but the perception of time varied among teachers (M = 18.40, SD = 5.4). The range of scores was 7 to 35, and a score of 21 suggests that teachers felt neutral about the amount of time they were given to use at school. Scores of 28 or above indicate that teachers mostly agreed (with item scores of 4 or 5 on average) with statements about the positive use of time at their school. Scores of less than 14 indicated that teachers on average disagreed with statements about the positive use of time at their school. The mean of 18.40 indicates teachers were slightly negative about the use of time at their school. Visual analysis and the Shapiro-Wilk test indicated a normal distribution, and no extreme outliers were detected.

On the other hand, teachers overall reported that their school had adequate resources to conduct assessments (M = 33.08, SD = 5.56). The possible range of scores was 9 to 45, and the reported range was 17 to 45. A score of 27 means that the teacher reported a neutral attitude about resources (with item scores around three), and a score of 36 indicates that the teacher reported a positive attitude about resources (with item scores (with item scores around four). The mean of 33.08 suggests that teachers in general were slightly positive about the resources available at their schools. Visual analysis and the Shapiro-Wilk test indicated a normal distribution, and no extreme outliers were detected.

Additionally, teachers reported a moderate amount of training in and preparation for the use of oral reading fluency (M = 14.45, SD = 3.84). The possible range of scores was from 6 (low level of training) to 25 (high level of training). Visual analysis and the Shapiro-Wilk test indicated a normal distribution, and no extreme outliers were detected. Teachers who participated in this survey reported a moderately frequent use of oral reading fluency (M = 21.62, SD = 6.60). However, there were also ten teachers who reported a lower level of use for oral reading fluency (i.e., a score of 14 or below). The possible range of scores was 5 to 31, and the actual range of scores was 5 to 30. The Shapiro-Wilk test was significant for a non-normal distribution and the distribution appears to be bimodal through a visual analysis.

Missing Data

A review of the descriptive statistics revealed that some participants did not respond to every question. One hundred sixty-six people agreed to complete the survey online. Two respondents were removed from analysis because they were outside of the sampling frame (e.g., teachers of kindergarten and early childhood). Eighteen participants did not complete the survey. There were additional cases of item nonresponse because some teachers skipped

particular items. Due to the additive nature of several measures, when a teacher skipped one item on a longer scale, this led to a missing value for the overall scale. This pattern of missing data is called a unit nonresponse pattern (Enders, 2010). For the seven independent variables and two dependent variables used in this analysis, 112 participants had a complete data set. Table 9 provides more information on how many participants failed to respond to all items on a scale. Table 9

Variable	N	Participants	% Missing
		Missing	
Years teaching	161	3	1.83
Certification	150	14	8.54
Knowledge	139	25	15.24
Universal Screening Acceptability	142	22	13.41
Progress Monitoring Acceptability	142	22	13.41
Time	146	18	10.98
Resources	146	18	10.98
Training	144	20	12.20
Frequency of Use	146	18	10.98

Missing Data Summary

Missing data are a concern because they may influence the outcomes of a study (Knight, Knight, Sidani, Figueredo, 2007). This is particularly a concern for the inferential research questions that require particular sample sizes to detect significance. Multiple options are available to deal with item nonresponse, such as listwise deletion, pairwise deletion, simple imputation, or model-based imputation procedures (Raykov & Marcoulides, 2008). Each method for dealing with missing data has challenges and benefits.

The decision regarding how to handle missing data should be influenced by the reason for the missing data and the statistical tests to be used (McKnight, McKnight, Sidani, & Figueredo, 2007). Therefore, it is first important to establish if the data are likely missing completely at random, missing at random, or missing not at random. Little's MCAR test was used to test the null hypothesis that the data were missing at random. To test this hypothesis, all items used to compute the independent and dependent variables were included. The results of Little's MCAR test, χ^2 (1777, N = 164) = 1816.82, *p* = 0.25, suggest that the data were indeed missing at random (i.e., no known pattern exists to the missing data). Given that the data are most likely missing completely at random, it is appropriate to consider listwise deletion, pairwise deletion, or replacing data with predicted values (e.g., imputation).

If the data are considered to be missing completely at random, then listwise deletion is a method that may be appropriate because the missing data should not be influenced by the phenomenon of interest (Enders, 2010). Listwise deletion is the simplest method for dealing with missing data, but in the current case it would significantly reduce the power to detect significance by reducing the sample size to 112 participants (a reduction of 32%). Additionally, a report by the American Psychological Association Task Force on Statistical Inference (Wilkinson & American Psychological Association Task Force on Statistical Inference, 1999, p. 600) called listwise deletion "among the worst methods available for practical applications." Pairwise deletion (e.g., available case method) on the other hand, maintains more data for research question one and four, which used only descriptive statistics. Pairwise deletion was used to answer question one and four because this method uses all available data and does not alter the data used to report descriptive statistics.

Another highly acceptable method for dealing with missing data is called expectation maximization (EM), which is based on maximum likelihood estimation (Enders, 2010). According to Little and Rubin (2002), this method is a model-based procedure. Methods using maximum likelihood estimation, such as EM, are considered superior to single imputation and

listwise or pairwise deletion (Baraldi & Enders, 2010; McKnight, McKnight, Sidani, & Figueredo, 2007). Mean imputation or regression substitution result in biased estimates and also underestimate standard errors; EM procedures minimize this problem when data are missing completely at random or missing at random. Therefore, expectation maximization was the chosen method to deal with the missing data for research questions two and three because it uses available data to predict missing data, and it maintains more power than other methods.

For this method, the SPSS Missing Values Analysis EM estimation method was used. The steps included Analyze-> Missing Values Analysis and entering the items for each scale one at a time (leading to eight different transformations). This is because the maximum likelihood procedure uses two steps to build a regression equation, and then replace missing data. This method uses information about the mean and central tendency of each variable and also uses information about the covariance between variables. The estimated regression equation relates each variable to the other related variables, and then it estimates missing values based on existing data. Therefore, each set of related items from one scale were estimated together. This created a new data set with no missing data for each scale. Subsequently, the datasets for each variable were copied into a new complete dataset for use in the analysis to address the research questions. (Categorical variables may not be replaced using EM procedures, and therefore the sample size for research questions two and three was reduced to 150 because 14 participants did not respond to the question asking about teacher certification). Table 10 compares the mean and standard deviations for each variable before and after EM procedures. The expectation maximization procedures resulted in a larger sample size with complete data, altered the means by no more than two tenths, and altered the standard deviations by less than nine tenths. Also, analysis of the influence of estimation on the individual items for the dependent variables demonstrated that

estimation altered the means by less than 0.01 and altered the standard deviations by less than

0.1. A comparison of descriptive statistics for the variables of interest in this study before and after EM procedures is presented below.

Table 10

Mean and Standard Deviations Before and After EM Procedures

		Before Es	timation	After Estimation		
	Ν	Mean	Standard	Ν	Mean	Standard
			Deviation			Deviation
Universal Screening Acceptability	142	60.73	11.00	164	60.56	10.36
Progress Monitoring Acceptability	142	58.15	13.51	164	57.99	12.64
Years teaching	161	3.73	1.08	164	3.73	1.07
Certification	150	0.73	0.44	150	0.73	0.44
Knowledge	139	13.52	2.04	164	13.50	1.99
Time	146	18.40	5.44	164	18.39	5.14
Resources	146	33.08	5.55	164	33.00	5.28
Training	144	14.45	3.84	164	14.41	3.64
Frequency of Use	146	21.62	6.60	164	21.50	6.3

Statistical Assumptions

The statistical assumptions for analysis of variance and regression analysis are relevant for this study. The data with estimated and replaced values were analyzed to identify the extent to which assumptions necessary for the chosen analyses (e.g., repeated measures ANOVA and multiple linear regression) appear to have been met.

First, the data were screened for potential univariate and multivariate outliers. One

participant reported an unusually high level of time, and one participant reported an unusually low level of resources, as indicated by z-scores above 3 or below -3. Additionally, Cook's Distance was reviewed to evaluate the potential influence of outliers on the regression analysis. Results suggest that only one multivariate outlier may be problematic for regression analysis, due to unusually low acceptability for oral reading fluency for both purposes. Also, the Mahalanobis distance measures for all the variables of interest indicated three multivariate outliers. Review of these multivariate outliers revealed that two of the participants reported unusually low acceptability for oral reading fluency for progress monitoring only (e.g., scores of 12 on the APR-R).

Given the large sample size, it is assumed that these extreme attitudes are representative of the larger population, and therefore will be retained in the analysis. Raykov and Marcoulides (2008) noted that there are often extreme observations found in large samples (greater than 100). Several researchers have noted that it is possible that these observations are truly representative of the population and should not be removed from the analyses (Orr, Sckett, Dubois, 1991; Osborne & Overbay, 2004; Raykov & Marcoulides, 2008). The inferential statistics were tested with and without the multivariate outliers, and the results were not changed notably.

A visual analysis of the residuals suggested that the error distributions violate the assumption of a normal distribution. Also, the Shapiro-Wilk test of normality on this data suggested that the residual distributions were significantly different from normal, p < .05. The Koenker test for homoscedasticity, which is robust for violations of normality in the dependent variable, was not significant, X^2 (7, N = 164) = 2.40, p = 0.93. Thus the data used in the regression analysis meets the assumption of homoscedasticity. The results of the Durbin-Watson

(DW) test suggest no violations of the assumption of independence of errors. Additionally, examination of tolerance levels and multicollinearity among independent variables indicated no problems. Also, residual plots with the standardized predicted value and the standardized residual showed a rectangular pattern, indicating a linear relationship.

Additionally, repeated measures analysis of variance (ANOVA) assumes that each level of the dependent variable is normally distributed in the population. Visual analysis (P-P plots, histograms, and box plots) and statistical analyses (Shapiro-Wilk's W, skewness, kurtosis) of the sample were used to evaluate whether the data were normally distributed. Visual analysis and the skew statistics confirmed that both dependent variables had a left (e.g., negative) skew.

The statistical techniques selected are robust for violations of normality, so the statistical methods chosen for this study were not altered. The repeated measures ANOVA is robust against violations of normality with moderate or large sample sizes (larger than 30) as long as there are equal numbers of participants on the repeated measures (Glass, Peckham, Sanders, 1972; Green & Salkin, 2003). Some researchers choose to use transformations to bring data closer to a normal distribution; however, other statisticians argue that these transformations are not advisable due to the error introduced by altering the data (Glass, Peckham, Sanders, 1972). For the closely related t-test, Boneau concluded that with a sample size over thirty and when the distributions of the two populations are the same shape (normal or non-normal), then the t-test is robust for non-normality (Boneau, 1960). For these reasons, the chosen analyses were not altered despite the violation of normality.

Research Question One

To address the first research question (Do teachers view oral reading fluency as an acceptable reading measure?), both survey and focus group data were analyzed. Acceptability is

the degree to which teachers find oral reading fluency appropriate to the problem, fair, reasonable, and non-intrusive (Kazdin, 1980), or in other words, the degree to which teachers find the procedures and outcomes acceptable in practice (NASP, 2008). Both survey data and focus group data were used to evaluate teacher acceptability of oral reading fluency.

Descriptive statistics from the two acceptability scales for oral reading fluency were used to address the first research question. Given that no inferential statistics were needed to address this research question, the original dataset without estimation maximization was used. Given that data was missing completely at random, it was appropriate to use listwise or pairwise deletion for the analysis of descriptive statistics. Pairwise deletion was the method used for dealing with missing data at the item and scale level for this research question. For the overall scale, cases were dropped prior to data analysis if there was missing data at the item level, and 142 participants had complete data for the universal screening and progress monitoring scale.

With 12 items and a 6-point Likert scale, a higher rating on each item (e.g., 6) indicates that the assessment is more acceptable (Eckert, Hintze, Shapiro, 1999). The highest possible acceptability rating was 72, and the lowest possible rating was 12 for each scale. A score of 42 indicates a neutral attitude towards this assessment tool. Less than 42 suggests lower acceptability and more than 42 indicates higher acceptability. Due to the non-normal distribution of the dependent variables, all three measures of central tendency (mean, median, and mode) are reported for these variables.

Overall teachers reported that using oral reading fluency for universal screening or progress monitoring was moderately to highly acceptable. The three measures of central tendency for universal screening with oral reading fluency acceptability ratings suggested moderate to high acceptability (M = 60.72, SD = 11.00, Mdn = 61, Mode = 72, N = 142). The

mean, median, and mode acceptability scores for using oral reading fluency in progress monitoring were also fairly high (M = 58.15, SD = 13.51, Mdn = 60, Mode = 60, N = 142). Ninety-six percent (137 out of 142) of teachers found oral reading fluency acceptable for universal screening, and 91% (129 out of 142 participants) of teachers found oral reading fluency acceptable for progress monitoring. There was a wide range of attitudes on the acceptability of oral reading fluency, with at least one teacher endorsing a one and at least one teacher endorsing a six on each item. The ratings for both scales ranged from 12 (low acceptability) to 72 (high acceptability). Table 11 and 12 demonstrate the mean, standard deviation, minimum, and maximum rating for each specific item on the scales.

Table 11

Item	Ν	Mean	Standard	Minimum	Maximum	
			Deviation			
1	148	5.22	1.05	1	6	
2	148	5.10	1.01	1	6	
3	148	5.10	0.97	1	6	
4	148	5.05	1.00	1	6	
5	148	5.26	0.92	1	6	
6	147	5.07	0.98	1	6	
7	147	4.86	1.08	1	6	
8	146	5.16	0.98	1	6	
9	148	4.90	1.14	1	6	
10	147	4.90	1.08	1	6	
11	148	4.76	1.18	1	6	
12	147	5.18	0.95	1	6	

Universal Screening with Oral Reading Fluency Acceptability Ratings by Item

Table 12

Item	Ν	Mean	Standard	Minimum	Maximum
			Deviation		
1	146	4.92	1.17	1	6
2	146	4.79	1.09	1	6
3	146	4.74	1.20	1	6
4	145	4.80	1.22	1	6
5	145	5.09	1.12	1	6
6	145	4.86	1.17	1	6
7	146	4.74	1.26	1	6
8	145	4.81	1.22	1	6
9	146	4.70	1.28	1	6
10	146	4.83	1.19	1	6
11	146	4.79	1.27	1	6
12	146	4.92	1.20	1	6

Progress Monitoring with Oral Reading Fluency Acceptability Ratings by Item

Stem and leaf plots and box plots demonstrated that there were several outliers in each of these scales. Analysis of z-scores indicated that there were three outliers on the universal screening scale, indicating that three teachers reported extremely low levels of acceptability for oral reading fluency for this purpose, compared to the majority of other teachers (z-scores less than -3). There were five outliers on the progress monitoring scale, indicating that five teachers reported unusually low levels of acceptability compared to the majority of other teachers (z-scores less than -3). However, the majority of teachers reported that oral reading fluency for universal screening and progress monitoring was generally acceptable. Analyzing spread through the use of quartiles is appropriate if the data are skewed or have several outliers, which is the case with these two variables (Gonick & Smith, 1993). The interquartile range was 13.25 for the universal screening scenario; the score at the 25th percentile was 70. The interquartile range was 17.50 for the progress monitoring scale; the score at the 25th percentile was 51.75 to the score at the 75th percentile was 69.25.

Qualitative data analysis identified both positive and negative attitudes towards oral reading fluency. The focus group data provide more detailed information regarding specifically why teachers found oral reading fluency acceptable or unacceptable. Through the coding and analysis process, eleven broad thematic categories were identified, including a code for positive attitudes and code for negative attitudes. Table 16 in Appendix J presents the full coding scheme, with information on how many focus groups discussed a code and how many total references to that code were made. Six of these themes related to teachers' acceptability of oral reading fluency, district or school issues, and use of data). The themes suggested that teachers found some characteristics of the measure acceptable but had concerns about other features. Figure 3 demonstrates the positive, negative, and mixed themes by category. Additionally, Table 13 presents some representative positive and negative quotes from teachers. In the following section, specific codes are put in parenthesis.

Factors influencing accuracy of oral reading fluency. Teachers expressed concerns that some teachers may administer oral reading fluency differently and thus decrease the reliability of the scores (nature of the assessor). One teacher commented that "It is hard to consistently train new people and to make sure they're using, doing it with fidelity" (Speaker 5). Multiple teachers across all focus groups also mentioned a concern that oral reading fluency scores are not accurate reading measurements for some students (nature of the student). For instance, teachers were concerned for English language learners, non-verbal students, or students with mental health difficulties. A teacher described her mixed feelings about the appropriateness of oral reading fluency and said, "There are circumstances that it isn't a good measure….for lots of our students it's a wonderful way to get that dipstick piece of information" (Speaker 8). Additionally,

teachers mentioned that sometimes a busy environment or a "bad passage" led to inaccurate scores (nature of environment and nature of passage). Another teacher commented that assessments "need to be run like machines. If you're going to truly assess you put them in a single room, no distractions, with one on one monitoring, not 40 kids running around going up and down hallways" (Speaker 15).



Figure 3. Venn diagram of themes. Venn diagram relating the qualitative themes to positive and negative attitudes toward oral reading fluency.

Resources needed for oral reading fluency. Teachers expressed concerns regarding the numerous resources, including time, people, space, knowledge, training, and funding, needed to use oral reading fluency. Some teachers felt they did not have adequate resources to make meaningful use of oral reading fluency data. One teacher commented on the time and knowledge needed to make the data useful when she said, "DIBELS is just really nothing if you don't take

time to analyze it and pull something out of it and get something to make your teaching better, and I think that's what's really frustrating" (Speaker 7). However, teachers also commented on the time saved by using oral reading fluency for universal screening instead of other longer assessments.

District or school issues. Each group of teachers also brought up the concern that oral reading fluency scores may be used as a part of their teacher evaluation. The majority of teachers disliked this prospect, but one teacher commented that he liked this idea because his students always make progress on oral reading fluency. Another teacher was concerned that she would be told "I'm not effective," due to the large number of students in her class who have intensive needs, even if a large portion of them made progress (Speaker 9).

Influence on students. Another major theme in the data indicated that teachers were concerned about the influence of using oral reading fluency on some students. Teachers were concerned that using oral reading fluency made some struggling students anxious, sad, or upset. Speaker 6 said, "By second grade there's some of them that have real anxiety about it [the oral reading fluency test]." They were also concerned that oral reading fluency scores were being used by students to compare themselves to others. Teachers mentioned that using oral reading fluency with students had a positive effect on some students' motivation and a negative effect on other students. Lastly, this theme revealed that teachers were concerned that using oral reading fluency frequently would teach students that reading with speed is important. The majority of discussion on this topic surrounded concerns for students, while several teachers commented that using oral reading fluency may benefit the students.

Use of data. One positive theme about the acceptability of oral reading fluency that emerged from the focus group data was how the data can be used for a variety of purposes (e.g.,

parent communication, standard measurement, to make decisions, to demonstrate growth and set goals, and to communication with other teachers). Generally, teachers were positive about the various ways that they can use oral reading fluency in their job. This theme will be discussed more in detail under the results section for research question five.

Limitations of oral reading fluency. Lastly, teachers expressed concerns about the limitations of oral reading fluency. Several people commented on how oral reading fluency is not the only beneficial assessment tool, and sometimes it is limited in how it can be used. Other teachers expressed a desire to use other assessments, such as running records or to rely on their judgment as a skilled teacher. Still another concern was that the assessment tool did not match classroom instruction. A teacher said, "Does the measure correlate with the way we are instructing? Should it? We're saying it doesn't necessarily match" (Speaker 3). Teachers discussed how comprehension and vocabulary were often the focus of instruction, and oral reading fluency does not directly measure those skills.

Table 13

Positive and Negative Attitudes and Representative Quotes

Theme	Sample Quotes
Positive attitude	"I mean it's so cool to see it on paper and have those numbers you know the numbers are a blessing and a curse but you know when you, when you use it the right way and you know you use it as a snapshot it can really be helpful to have for parents and you know to guide your instruction a little further." (use of data) (Speaker 12)
	"One thing that, in comparison if I do a DIBELS assessment versus trying out eight different reading levels to see which one they are not accurate on, it is quicker it is more efficient to do a real quick read three passage and I can kind of still gauge like based on my knowledge of the reading levels gauge where they would fall so I think at the beginning of the year it is a bit quicker process." (use of data) (Speaker 21)

Table 13 (cont'd)	
Negative attitude	"I have an issue with the data collection and administration to me I feel that if you're going to administer DIBELS you should not administer your own children I think its very important cause you need to take out the bias." (factors that influence ORF scores) (Speaker 15)
	"I think the concerns go back to the pieces that we've talked about in terms of the English language learner [or] you have a student who is a stutterer it simply isn't a valid measure so there are circumstances that it isn't a good measure." (factors that influence ORF scores) (Speaker 8)
	"There's so many aspects of literacy that you have to look at that if you're only looking at oral reading fluency you're not getting the whole picture. You're getting one piece of the puzzle." (limitations of ORF) (Speaker 2)
	"The more I work with it [oral reading fluency] the more negative my feelings are because I don't think it really tests all the aspects of reading." (limitations of ORF) (Speaker 4)

Research Question Two

To address the second research question (Is there a significant difference between teachers' acceptability of oral reading fluency for the purposes of universal screening and progress monitoring?), both survey research and focus group data were analyzed.

A repeated measures ANOVA was used to address this research question and to determine if any of the independent variables studied accounted for the difference between the two dependent variables. Repeated measures ANOVA requires that the design be balanced; therefore only complete cases from the estimated dataset (N=150) were used to calculate the results. The repeated within-subjects factor for this test was considered to be the purpose (universal screening or progress monitoring); the categorical variable of teacher certification was entered as the between subject factor; the continuous independent variables were entered as covariates. The achieved power to detect significance in the omnibus test with the repeated measures ANOVA was 0.60.

Results indicated that there was a significant difference in acceptability for oral reading

fluency based on the purpose, F(1, 142) = 4.98, p < .05, $\eta^2 = 0.03$. Analysis of the means for these variables and the pairwise comparison table demonstrates that teachers found oral reading fluency for universal screening (M = 60.71) to be slightly more acceptable than oral reading fluency for progress monitoring (M = 57.49). The small partial eta squared ($\eta^2 = 0.03$) indicates that the practical significance of this difference is small. Only three percent of the variance in attitudes towards oral reading fluency was accounted for by the purpose of the test.

Additionally, the repeated measures ANOVA tested whether or not the difference in attitudes toward oral reading fluency for universal screening or progress monitoring depended on any of the independent variable. The knowledge scale significantly interacted with the difference in attitudes toward oral reading fluency, F(1, 142) = 7.25, p < 0.05, $\eta^2 = 0.05$. Higher knowledge corresponded with greater attitude differences between the two purposes. Higher knowledge related to higher acceptability of oral reading fluency for universal screening; higher knowledge did not relate to higher acceptability of oral reading fluency for progress monitoring. Training, resources, time, teaching certification, frequency of use, and years of experience were not found to interact significantly with the difference in acceptability of oral reading fluency for a reading fluency for various purposes (p > .05).

In the focus groups, teachers discussed the relative benefits and drawbacks of universal screening and progress monitoring with oral reading fluency. Several teachers commented that they found using oral reading fluency for universal screening more acceptable than progress monitoring (Speaker 18, 21, 22). One teacher said:

I like to use it [oral reading fluency] as a screener to kind of see, but I just think there are better ways to progress monitor them. Like, I mean, I would rather monitor their comprehension and their reading strategies they are utilizing for unknown words or

whatever rather than monitor their fluency. (Speaker 21)

Another teacher also mentioned preferring another method for progress monitoring. She said, "I could see the benefits of doing it [oral reading fluency] three times a year to get a quick snapshot of where your kids are, but then rather than progress monitoring with that I would like to progress monitor with running records." (Speaker 18). Another teacher also expressed a desire to use other assessments in addition to oral reading fluency when she said, "I think you have to throw some other stuff in there, so when we meet I don't want to see just the [oral reading fluency] data. Bring me the running records" (Speaker 19). Teachers in each focus group commented on desiring to use other assessment methods to get more information on struggling students or to monitor their progress; running records was the most frequently mentioned other assessment method.

Speaker 22 also preferred using oral reading fluency for universal screening rather than for progress monitoring. She commented on the concern that using the test every week with struggling readers would overemphasize the importance of speed. She also said, "I'd rather... give them a little one to one time for goodness sakes from that 5 minutes [of progress monitoring] and then every third week progress monitor them." Another teacher described a desire to use oral reading fluency only once a month to monitor progress instead of once a week (Speaker 19). These comments correspond with the survey data results, which suggest that progress monitoring acceptability on average was slightly lower than universal screening acceptability.

Research Question Three

To address the third research question (To what extent are teacher characteristics (i.e., years of experience teaching, teaching certification status, training, knowledge, and frequency of

use) and teacher perceptions of school characteristics (i.e., time and facilities/resources) related to their acceptability of oral reading fluency data?), both survey research and focus group data were analyzed.

The tests of between-subjects effects in the repeated measures ANOVA partially addressed this research question. Both time F(1, 142) = 7.51, p < .05 and knowledge F(1, 142) = 3.84, p < .05, significantly influenced the acceptability of oral reading fluency. Two follow-up linear multiple regressions, with all seven independent variables and the each dependent variable entered in a separate test, were run. These tests were used to determine how time and knowledge were related to the acceptability of oral reading fluency for universal screening and progress monitoring.

The regression results suggest that the model with all seven independent variables significantly predicted the oral reading fluency for universal screening F(7, 142) = 3.75, p < .05, $R^2 = 0.16$ and progress monitoring F(7, 142) = 3.12, p < .05, $R^2 = 0.13$. Analysis of significance levels for specific variables within the models indicated that knowledge significantly influenced teachers' attitudes toward oral reading fluency for universal screening, t (150) = 3.25, p < .05, but not toward progress monitoring, t (150) = 0.65, p = .52. A follow-up correlational analysis revealed a positive relationship (r = .28, p < .05) between knowledge and universal screening acceptability. This suggests that the higher a teacher scored on the knowledge scale, the higher the level of acceptability for oral reading fluency for universal screening. The relationship between knowledge and acceptability of oral reading fluency for universal screening is small in magnitude, and the effect size of knowledge on universal screening acceptability was 0.07.

Teachers' perceived time was also significantly related to their attitudes toward oral reading fluency for both universal screening t(150) = 2.03, p < .05 and progress monitoring

t(150) = 2.93, p < .05. A follow up correlation analysis indicated a positive correlation between time and universal screening (r = .24, p < .05) and progress monitoring (r = .25, p < .05). Therefore, the more time a teacher reported, the higher the reported acceptability for oral reading fluency. However, the relationship between time and acceptability of oral reading fluency is small in magnitude. Training, resources, teaching certification, frequency of use, and years of experience were not found to relate significantly to teachers' acceptability of oral reading fluency for either purpose (p > .05).

Teachers commented on these factors and their relationship to the acceptability of oral reading fluency in the focus groups. Multiple teachers in each focus group commented on how resources (e.g., funding, time, people, space, training, knowledge) were important for the data collection process and for supporting the use of the data. One teacher requested more time to use the data. She said, "How can we make this meaningful for what we do? I believe we need more time to work as grade levels or work as a building to understand how to then use this data for instruction" (Speaker 3). Speaker 8 described a more positive attitude toward oral reading fluency as she described how her schools use two hour late start days for teachers to hold grade level meetings about oral reading fluency data. Another concern was raised regarding how assessment time might take away from instructional time. Other teachers expressed concerns that they did not have the necessary space or support from other adults to conduct universal screenings or to monitor progress with oral reading fluency.

Teachers mentioned that with changes in staff every year, training needs to occur frequently. One teacher said, "It's hard to consistently train new people and to make sure they're using, doing it with fidelity" (Speaker 5). One participant's role at school was as a part-time coach and part-time title teacher, and she played a large role in reading assessment at her

schools. She described how she was involved in training other teachers on how to use oral reading fluency. She said, "I spent a lot of time in the fall helping, and you know, 'This is how you progress monitor, this is how you set up their goals' and, and the whole procedural piece" (Speaker 8). Another teacher commented that the administrator should play a role in training and supporting teachers in the use of oral reading fluency. She said, "DIBELS is an assessment, but it really does need to be paired with an administrator who knows the purpose and who's gonna lead their staff in that direction... follow through with telling them how to progress monitor, why are you progress monitoring" (Speaker 7).

Also, many teachers expressed a change in attitude toward oral reading fluency as time progressed and they became more familiar with the tool. One teacher reported that the more he used the assessment, the more negative he became regarding its usefulness (Speaker 4). Multiple other teachers reported more positive attitudes now that they have used oral reading fluency more frequently. For instance, one teacher said:

I think I've changed a lot over the years and having experience using it... now I really am in favor of using it and just really understanding what the purpose of it is and what it can do and can't do. (Speaker 1)

Another teacher expressed a similar sentiment, "I think now that I've done it awhile... I like it because it's standard" (Speaker 7). One teacher described her observation that more experienced teachers were resistant to using oral reading fluency at first. She said:

I was a new teacher so it was all very interesting but I remember... a lot of the teachers had been there awhile were so against it and were so negative and I just remember having like arguments after arguments as a staff. (Speaker 6)

Although not all of the variables of interest from this research question were found to be

statistically significant for influencing attitudes, focus group data revealed that teachers found time, resources, knowledge, training, and experience to be relevant in a discussion of the use of oral reading fluency.

Research Question Four

The fourth research question asked: What elements of reading do teachers find valuable for good reading? The mean of value items suggests that teachers find all seven elements included on the survey to be valuable. On a scale of 1 to 7, all items had a mean of higher than 6. The lowest mean value was for concepts about print (6.35), and the highest mean value was for comprehension (6.73). Fluency was the second lowest valued reading element on this survey, with a mean rating of 6.37. Table 14 provides information on the descriptive statistics for teachers' responses.

Table 14

Item	N	Mean	Standard Deviation	Minimum	Maximum
Phonological Awareness	146	6.51	1.19	1	7
Concepts about Print	146	6.35	1.24	1	7
Alphabetic Principle	146	6.44	1.21	1	7
Phonemic Awareness	146	6.60	1.17	1	7
Fluency	146	6.37	1.14	1	7
Vocabulary	145	6.57	1.13	1	7
Comprehension	146	6.73	1.09	1	7

Reading Value Descriptive Statistics

Research Question Five

The fifth research question asked: How do teachers use oral reading fluency data?

Five major themes surfaced from the focus group data regarding how teachers used oral

reading fluency data. These five themes included the following:

Standard measurement. First, teachers in two of the focus groups mentioned how oral reading fluency is a measurement tool that they use to communicate with other teachers in their grade level, building, or district. They both commented on the benefit of using a "common language" to describe students' reading progress. One teacher said:

I like it because it's standard. We can measure my class to someone else's... if we didn't have that it would be like, 'Well, I do a different assessment than you do', so we wouldn't be able to measure apples to apples. But it gives us that data, school-wide data. (Speaker 7)

Decision making with oral reading fluency. Furthermore, teachers discussed how they use oral reading fluency data to make decisions. Teachers described using the data to make instructional groups, plan interventions, inform instructional changes, and use as a part of special education evaluations. This teacher used the data to make decisions about interventions, "We use it mainly at the beginning of the year to decide who needs really to have additional services" (Speaker 21).

Consult with other teachers. Teachers also discussed how oral reading fluency was helpful when teachers needed to communicate about the reading level of a particular student. They also used the data in grade level team meetings to make decisions about instructional plans or instructional groups.

Demonstrate growth or set goals. Teachers in all four focus groups described how they use oral reading fluency to set reading goals and to demonstrate growth in their students. Some teachers commented that this was a great benefit of using oral reading fluency because it was possible to document student learning very clearly. One teacher said:

Our teachers set goals like by the end of the year I want, like they'll look at their mid-year

data and... say 'alright, what's reasonable for me? Where can I get these kids by the end of the year?' and they set goals. Then we also have class or school-wide goals that we set as grade levels on how many kids we want to move up. (Speaker 6)

Parent communication. In one of the districts, oral reading fluency screening and progress monitoring data were used to communicate with parents about their child's progress. Teachers mentioned that it required writing a letter or explaining the scores at a teacher conference. There were some concerns that parents might misinterpret the data they were given, so teachers described how they made sure to communicate clearly, what the numbers meant, in writing or in person. This method for using oral reading fluency was not used in all of the schools that were represented in the focus groups.

Table 15 shows representative quotes for each theme (i.e., standardized measurement, demonstrate growth or set goals, to make decisions, communication with teachers, parent communication). Figure 4 presents a visual display of the five themes related to teachers' use of oral reading fluency data.

Table 15

Use of Data Themes	Sample Quote
Standard Measurement	"I think its just helpful one to see where they're at compared to
	where an average third grader should be."
Demonstrate growth or set goals	"I'm writing goals so that it actually shows whether or not the student is making progress at a level that they will eventually catch up to their same age peers it's more precise. And I like that I would say it's very helpful to see whether or not they're closing the gap." (Speaker 1)
To make decisions	"I think its just helpful one to see where they're at compared to where an average third grader should be To kind of give me a guide as to you know who do I need to follow up onSo that's kind of helpful to have a quick glance tool to help me inform the other pieces that I need to put into place." (Speaker 2)

Use of Data and Representative Quotes

Table 15 (cont'd)	
	"I do know the progress monitoring data and phase lines on it has been a very useful tool you know for when we want to push forward with kids for special education and has been very useful in helping to push through kids as evidence and such." (Speaker 5)
Consult with other teachers	"As a district it's nice to know to have a common language, to have everybody using the same measures so when we talk about things we all understand." (Speaker 3)
Parent communication	"I love the progress monitoring because it takes the guesswork out of it. So that I'm not sitting there telling a parent "well, fingers crossed we make some progress, right?" so I like the progress monitoring for me to make the dot, for me to either, you know, tell myself to keep going, or you need to change this up a bit." (Speaker 7)



Figure 4: Major codes within the use of oral reading fluency theme.

Chapter V

DISCUSSION

This chapter reviews the results of the current study and compares them to previous research. The results are interpreted, the limitations of this study are reviewed, and recommendations are made for future research and educational practice.

Overview of the Study

This study provides information on the teacher acceptability of oral reading fluency for progress monitoring and universal screening. Results suggest that teachers find using oral reading fluency slightly more acceptable for universal screening than for progress monitoring; however both purposes for oral reading fluency are acceptable to the majority of teachers. Also, the results suggest that knowledge and time are factors that influence teachers' attitudes toward oral reading fluency. A more detailed discussion of the findings for each research question is provided below:

Research question one: Do teachers view oral reading fluency as an acceptable reading measure? The current study examined whether the teachers studied found oral reading fluency to be an acceptable measure for universal screening and progress monitoring. It was hypothesized that on average there would be a moderate level of acceptability but a large amount of variability. In the survey, teachers reported moderately high levels of acceptability for oral reading fluency as a universal screening and progress monitoring tool. There was also a wide range of attitudes about oral reading fluency; several teachers reported extremely low acceptability, but the majority of teachers reported high acceptability for oral reading fluency.

Previous researchers also found oral reading fluency to be acceptable to school psychologists and teachers (Chafouleas, Riley-Tillman, & Eckert, 2003; Druckman, 1997;

Eckert, Shapiro, & Lutz, 1995; Rimstidt, 2001; Shapiro & Eckert, 1994). In previous survey research using the APR-R, the mean acceptability ratings for oral reading fluency in curriculumbased assessment (M = 57.93, SD = 9.12) (Chafouleas, Riley-Tillman, & Eckert, 2003) was comparable to the mean acceptability ratings for oral reading fluency in the current study (universal screening: M = 60.72, SD = 11; progress monitoring: M = 58.15, SD = 13.51). In previous survey research, over seventy-five percent of the teachers surveyed reported that using curriculum-based measurement (including oral reading fluency) was a good idea for progress monitoring and screening (Yell et al., 1992). In this study, 96% of teachers found oral reading fluency acceptable for universal screening, and 91% of teachers found oral reading fluency acceptable for progress monitoring. The current study supported previous research because oral reading fluency for universal screening and progress monitoring purposes was acceptable to most teachers.

This is an important finding because universal screening and progress monitoring are two key components of Response to Intervention (Fuchs & Fuchs, 2006). Ensuring that an assessment tool is acceptable to teachers is also important because previous researchers have reported significant relationships between teacher acceptability for interventions and treatment integrity (Henninger, 2010; Mautone, et al., 2009). Therefore, to ensure that oral reading fluency is administered and used with fidelity within an RTI service delivery model, it is important to know how acceptable teachers find the measure. Not all teachers found oral reading fluency to be acceptable, so it is important to know more about their concerns.

The current study extended previous research by including qualitative data focusing on why teachers found oral reading fluency acceptable or unacceptable. Teachers reported appreciating how oral reading fluency documents student growth, saves time compared to other

tests, and can be used to communicate with other teachers about students' progress. However, despite the overall positive attitudes, there were several teachers who reported very low levels of acceptability for oral reading fluency on the survey. The focus groups revealed information about why some teachers hold negative attitudes toward this assessment. Negative themes relating to oral reading fluency included concerns with the accuracy of the measure, its negative influence on students, the resources needed for administration and use, the limitations of the test, and the potential use of oral reading fluency in teacher evaluations.

In the focus groups, teachers mentioned that other assessments are still necessary to measure reading ability completely. These results support previous research indicating that teachers are concerned that oral reading fluency may not accurately measure overall reading ability or reading comprehension (Foegen et al., 2001; Yell et al., 1992). In both the current study and previous research, teachers expressed a concern that oral reading fluency measures speed, not good reading (Samuels, 2007).

The concerns expressed by the teachers in focus groups are important to consider because using oral reading fluency for universal screening and progress monitoring is a new practice that requires a change in teacher behavior. Teachers expressing negative or mixed attitudes toward oral reading fluency may still be in the exploration stage of implementation (Fixsen & Blasé, 2009) and not fully ready to implement the new practice with fidelity. In this study, some teachers explained how they were initially skeptical about this new practice, but later adopted and supported the use of oral reading fluency. The attitudes that teachers hold toward oral reading fluency has implications for how schools may choose to administer and use this assessment tool.

Research question two: Is there a significant difference between teachers'

acceptability of oral reading fluency for the purposes of universal screening and progress monitoring? This study also examined whether there was a significant difference between acceptability for oral reading fluency for two purposes: universal screening and progress monitoring. It was hypothesized that teachers would find universal screening slightly more acceptable than progress monitoring; this hypothesis was confirmed in the current study. The focus groups also revealed that teachers held more concerns for progress monitoring with oral reading fluency than for universal screening. In the focus groups, teachers were concerned about the frequent use of oral reading fluency for progress monitoring because they reported that this took away from instructional time. Also, several teachers reported that using oral reading fluency frequently for progress monitoring with struggling students was inappropriate. They believed that it might be harmful to the student's affect or motivation for reading; these concerns were not previously reported in research.

Teachers also mentioned several other assessments that they would prefer to use for progress monitoring, such as running records, which was also preferred by a teacher who was the focus of another study (Rimstidt, 2001). The preference teachers' expressed to use other assessments for progress monitoring may be appropriate. As described earlier, oral reading fluency is considered a global outcome measure (GOM), and therefore in some instances it may be appropriate to monitor progress of a specific skillset with skills-based measures or mastery measures instead of oral reading fluency (Hosp, Hosp, & Howell, 2007).

This research finding is important because this means that teachers may be more ready and willing to use oral reading fluency for universal screening than for progress monitoring. Salvia, Ysseldyke, & Bolt (2010) recommend that research on the reliability and validity of tests be conducted with a focus on the purpose of the test. Likewise, decisions regarding assessment

practices with oral reading fluency should be made in light of the specific purpose (e.g., universal screening or progress monitoring), given the differing teacher attitudes for the two purposes.

Research question three: To what extent are teacher characteristics and teacher perceptions of school characteristics related to their acceptability of oral reading fluency? This study also examined whether several factors (time, resources, knowledge, years of teaching experience, teaching certification, training, and frequency of use) were significantly related to the acceptability of oral reading fluency. Teacher knowledge and time were found to significantly predict attitudes toward oral reading fluency. Resources, training, years of teaching experience, teaching certification status, and frequency of use did not significantly relate to attitudes toward oral reading fluency.

Authors of previous research also suggested that the barriers of time and knowledge were related to attitudes about assessment, sometimes inhibiting the use of formative assessments (Wesson et al., 1984; Yell et al., 1992). In the current study, teachers commented that a lack of understanding for how to use oral reading fluency data was frustrating. One previous study reported a teacher's concern that she needed more time to analyze progress monitoring data to make it valuable (Rimstidt, 2001). In this study, teachers also mentioned that time was needed to gather and use oral reading fluency data, but sufficient time was not always provided. This study supports the findings of previous researchers who suggested that knowledge and time are related to teachers' attitudes about oral reading fluency.

Training and teacher certification type did not significantly relate to attitudes toward oral reading fluency in this study. Chafouleas et al. (2003) found that school psychologists' training was positively related to acceptability for CBA. The differing results between previous research
and this study could be a result of different methods to measure training. Chafouleas et al. (2003) asked school psychologists to indicate their level of training on a four point Likert-type scale (no, minimal, moderate, frequent) whereas this study included six items to measure the level of training from various training methods (i.e., undergraduate education, graduate courses, and other professional development). Alternatively, it is possible that training is related to school psychologists' acceptability for CBA but not related to teachers' acceptability for oral reading fluency.

In the current study, teachers mentioned that resources were necessary for using oral reading fluency effectively. However, the quantitative results demonstrate that the amount of resources a teacher has does not significantly influence attitudes. In the past, researchers have reported that variables such as resources were related to teachers' attitudes about and use of oral reading fluency (Wesson, King, & Deno, 1984). Wesson, King, & Deno (1984) asked teachers to report factors (such as lack of materials) that inhibited their use of the measurement, but there was no direct measure of resources, attitudes, or use to suggest a relationship. The difference in results between previous research and the current study could be a result of different methodology and measurement. Also, the qualitative category of resources was broader than what was measured in the survey and may account for the difference between qualitative and quantitative data.

In previous research, authors have suggested that how often teachers use a test may relate to their attitudes toward that test (Rimstidt, 2001; Roehrig et al., 2008). Roehrig et al. (2008) used qualitative analysis to study how teachers used assessment data for progress monitoring. Rimstidt (2001) suggested that as teachers become more familiar with an assessment, their attitudes toward the assessment become more positive. Neither of these studies used inferential

statistics to study the relationship between frequency of use and acceptability, which may account for the different results between this study and previous research.

In the current study, the years of teaching experience was not often discussed in the focus groups as a factor that influenced teachers attitudes about oral reading fluency and this variable also was not found to significantly influence attitudes. Pajares suggested that beliefs, which influence attitudes, are formed early and may be difficult to alter (1992). However, several teachers did comment that their attitudes toward oral reading fluency changed across time. Yet there was inconsistency in the direction of this change, with some teachers becoming more negative and others becoming more positive with time. Another reason for non-significant findings related to years of experience may be that some teachers' early formed beliefs aligned with the use of oral reading fluency for universal screening and progress monitoring and some early formed beliefs did not. Given that time and knowledge were the most influential variables, educational professionals may consider how to increase teacher time for and knowledge about oral reading fluency in order to improve the use of this measure within a RTI service delivery model.

Research question four: What elements of reading do teachers find valuable for good reading? This study also examined the value that teachers' placed on various elements of reading. Teachers reported placing the highest value on reading comprehension and placing the lowest value on concepts about print and oral reading fluency. It is particularly interesting that oral reading fluency is one of the lowest rated reading elements because some of the other skills (e.g., alphabetic principle, phonemic awareness) are necessary skills for successful oral reading fluency. Samuels described how some teachers are concerned with assessment methods that do not include comprehension, and he suggested that comprehension is highly valued by teachers

(2007). This study found that teachers place a high value on reading comprehension. These beliefs about the most valuable reading elements may relate to beliefs about the most appropriate types of assessment. Therefore, it is important to note that teachers indicated that comprehension was very important and that oral reading fluency was rated lower than many other reading skills.

Research question five: How do teachers use oral reading fluency data? In the focus groups, teachers reported using oral reading fluency for a variety of purposes. Teachers reported using oral reading fluency to identify struggling readers, communicate with parents, make instructional decisions, compare the reading achievement of students or classrooms, and monitor the progress of students. Teachers reported using the data to support their classroom decisions and be transparent with parents and other teachers about student achievement in their classrooms. Previous researchers had not described how teachers communicate oral reading fluency data with parents. The activities described by teachers in the focus groups align well with how data might be used in a response to intervention educational framework (Fuchs & Fuchs, 2006). These activities are similar to the uses for progress monitoring that Shapiro (2008) described as best practice, including using data to create instructional groups, identify at risk students, and assist in eligibility decision-making.

Future Research

Future research might build off of the results of this study to improve our understanding of teacher acceptability for oral reading fluency. This study supports a connection between knowledge and acceptability. Future researchers may study the connection between knowledge and assessment acceptability to identify specifically what types of knowledge are helpful for improving acceptability of an assessment tool. Research could include an intervention element where teachers are provided with additional instruction regarding oral reading fluency to

determine if the instruction improves knowledge and attitudes toward oral reading fluency. This future research may also provide deeper understanding of the current study's results, which suggest knowledge influences acceptability for universal screening but not for progress monitoring, by focusing on specific purposes for oral reading fluency.

Also, researchers might study the connection between time and assessment acceptability. For example, a researcher might study the level of support for oral reading fluency administration that general education teachers receive and investigate if extra time and support influences the attitudes that teachers hold toward oral reading fluency. In the current study, perceived time was related to attitudes about oral reading fluency; future research could identify alterable variables that might add time for teachers to conduct and use oral reading fluency assessment data.

Additionally, this study identified that some teachers hold very negative attitudes about oral reading fluency for progress monitoring and universal screening. It would be beneficial to study teachers with particularly low levels of acceptability in order to identify if their negative attitudes are unique from the concerns that teachers in this study expressed (given that this study interviewed teachers with both positive and negative attitudes toward oral reading fluency). If the reasons are unique, that research might highlight additional ways for school psychologists to work with teachers who do not find oral reading fluency acceptable.

In this study, teachers reported that oral reading fluency was valued slightly less than other skills such as phonemic awareness and comprehension. Researchers might study how the value of oral reading fluency or other reading elements relates to the types of assessments that teachers find acceptable and how those assessments are used in the classroom. Furthermore, future researchers might develop more detailed measures for the value of various reading

elements to build upon this study.

Another important area for future research would be to examine the connection between assessment acceptability, the fidelity of administration and use, and student outcomes. In this study, teachers reported varying levels of acceptability for oral reading fluency. It would be valuable to know if measures of teacher acceptability could accurately predict whether a teacher would administer oral reading fluency accurately, use the data effectively, and improve student outcomes. If research confirms that teacher acceptability for oral reading fluency is strongly linked to student literacy outcomes, then this would have important implications for the collection and use of acceptability data.

Also, given that teachers indicated in this study that oral reading fluency data is shared with parents, it would also be worthwhile to study parents' attitudes or knowledge about the data. The researchers might study whether parents' attitudes or knowledge about the data influences how much time parents spend on reading activities at home with their child and whether those attitudes and behaviors influence student outcomes.

With the advent of new teacher evaluation practices in many states, it will be important to study how this influences the use and perceptions about oral reading fluency. Oral reading fluency has been suggested as one tool for measuring teacher effectiveness. With this new purpose, the teacher attitudes toward oral reading fluency may change. Also, teachers in this study noted that there is a potential threat to the validity of the data collected if teachers' evaluations are connected to their students' progress. New research may consider how schools may use oral reading fluency for a variety of purposes (teacher evaluations, universal screening, progress monitoring) while maintaining the integrity of the scores.

Limitations

The interpretation of these results must be viewed in light of the current study's limitations. First, oral reading fluency was the only academic assessment studied, and therefore whether these same results would apply to other types of academic tests for universal screening or progress monitoring is unclear. Future research may examine the acceptability of math curriculum-based measurement or other reading assessments such as the comprehension measure called MAZE. Also, in practice oral reading fluency is rarely used in isolation as the survey scenarios presented; often it is used in combination with a comprehension measure (e.g., retell quality or MAZE). Therefore teachers' attitudes toward oral reading fluency used in conjunction with another assessment may be slightly different than the attitudes reported in this study.

Additional research that gathers information about teachers' assessment practices through observations and interviews may shed light on teachers' attitudes toward the current assessment practices in their school context. The survey data may be biased because the information is self-report and subject to desirability factors. Given that using oral reading fluency for universal screening and progress monitoring is becoming a common practice, teachers may have been hesitant to express negative attitudes toward this tool because they feared that their responses may be shared with school administrators. Also, in the focus groups all teachers may not have felt comfortable sharing their opinions in the presence of other teachers for this same reason. Research incorporating an additional data collection method such as individual interviews would be beneficial to get more detailed and contextual information from each individual without the influence of other teachers.

Additionally, the data were collected from a mid-Michigan geographic area that is highly involved with an RTI initiative called MiBLSi, which mandates universal screening. Also, most of the teachers who responded reported teaching in rural or suburban school settings. Therefore,

it is unclear if the results would generalize to an urban school setting or to a setting that does not mandate the use of oral reading fluency for universal screening. The number of participants who did not respond to the survey (more than 50%) may also limit the generalizability of the survey results. It should be noted that a high percentage of teachers reported an education background in special education or language arts/reading. It is possible that the survey respondents have more education than the general teacher population regarding reading and assessment. Missing survey data limited the power of the statistical tests, and therefore an imputation method was used to replace missing values. Ideally, there would have been a high completion rate and low rate of missing values in order to ensure the accuracy of the results. Future researchers might consider alternative strategies for improving response and completion rates, such as having teachers complete the survey at their schools during a staff meeting.

It is also important to note that although many of the assumptions for regression and analysis of variance were met (i.e., homoscedasticity, no multicollinearity, independence of errors), the dependent variables were not normally distributed. Also, several teachers reported extremely low levels of acceptability for oral reading fluency. However, given that these outliers are likely true members of the larger population, it was considered important to retain their responses in the data analysis. The inferential statistics used in this research are considered to be robust for slight violations of statistical assumptions (Glass, Peckham, & Sanders, 1972). Future research could aim to include more teachers in the survey in order to have a better chance of satisfying statistical assumptions.

Educational Implications

In this study, teachers reported both positive and negatives attitudes toward oral reading fluency. The majority of teachers reported quite positive attitudes toward oral reading fluency,

which may be a reflection of how long they have worked within a school that uses a response to intervention service delivery model. With these teachers, a school psychologist might support the analysis and the use of their data in an effort to make the data collection valuable. Given that these already teachers find the measure to be acceptable, the school psychologist also may complete fidelity checks to make sure the data is collected with fidelity.

On the other hand, several reported extremely negative attitudes, and many teachers expressed concerns about oral reading fluency in the focus groups. School psychologists often are involved in consultation regarding universal screening and progress monitoring, so it important to be aware of the varying attitudes that teachers hold toward oral reading fluency, particularly if they are negative because they may adversely influence how the data is collected and used (Dewey, 1910; Pajares, 1992; Richardson, Anders, Tidwell, Lloyd, 1991). The results of this study suggest that survey methodology may highlight particular teachers with very negative attitudes, and interviews may provide more detailed information regarding why those negative attitudes persist. Therefore, educational professionals may decide to use a variety of methods for measuring acceptability.

Through consultation with teachers, school psychologists may play a role in shaping what teachers know or think about oral reading fluency and thereby play a role in how effectively the data are used. This research suggested that some teachers have slightly lower levels of acceptability for progress monitoring with oral reading fluency. Therefore, the school psychologist may pay attention to the needs of teachers who collect oral reading fluency progress monitoring data. School psychologists may consult with teachers about progress monitoring to ensure that teachers have an adequate amount of knowledge and time for this assessment task. If needed, the school psychologist may help the teachers analyze and interpret progress monitoring

data to make it useful for their instruction. Alternatively, when teachers express difficulty with managing the administration and collection of the data, the school psychologist may brainstorm solutions with the teacher or offer to assist with some element of data collection.

School psychologists may also work with school administrators to determine which assessments are imperative, in order to reduce the redundant requirements that use up teachers' limited time. If possible, the psychologist may consider consulting with the school administration to determine if teachers have the necessary knowledge, training, time, and support needed to collect and use the data effectively. Particularly in settings where oral reading fluency has more recently been introduced as a universal screening or progress monitoring tool, the school psychologist may offer assistance in the form of professional development on this topic.

Conclusions

Oral reading fluency is one measure of reading that is commonly used for universal screening and progress monitoring. In this study, a majority of teachers found oral reading fluency to be an acceptable measure for universal screening and progress monitoring. Acceptability for oral reading fluency as a progress monitoring tool was slightly lower and therefore this may call for greater attention to how teachers are trained and asked to use oral reading fluency for progress monitoring. Progress monitoring informs instructional decision making in schools using multi-level systems of support, and therefore understanding teachers' concerns with oral reading fluency for progress monitoring is important. Also, time and knowledge were significantly related to teachers' attitudes toward oral reading fluency. With the information provided in this study, administrators and school psychologists may consider ways to support teachers as they use oral reading fluency for universal screening and progress monitoring.

APPENDICES

Appendix A

Sample Administrative Support Emails

District support letter Dear XXXX,

My name is Sarah Stebbe Rowe, and I am a doctoral candidate in the School Psychology Program at Michigan State University. Through my graduate education, I have worked with local school psychologists in Clinton and Ingham County. I am currently developing my dissertation proposal, and I am writing to ask for your support for my data collection process.

My research focuses on teacher perspectives of oral reading fluency data. At the end of the study I would be willing to present a report of the findings (with no identifying information) to you and/or the XXX staff.

The study will utilize teacher surveys and teacher focus groups. Participation in my study would be completely voluntary. The surveys and focus groups will most likely occur between April and May 2012 and will occur outside of regular school hours.

I would like to request your approval for other principals in XXX to allow their staff to participate if they are willing. Would you consider allowing elementary schools in the XXX district to participate? If you are interested in participation, you may contact me by email (stebbesa@msu.edu) or phone (248-214-9697). Also, I'd be happy to give you more details if needed. Thanks for your time.

Sincerely, Sarah Stebbe Rowe Email: stebbesa@msu.edu Phone number: 248-214-9697

School support letter

Dear XXXX,

My name is Sarah Stebbe Rowe, and I am a doctoral candidate in the School Psychology Program at Michigan State University. Through my graduate education, I have worked with local school psychologists in Clinton and Ingham County. I am currently developing my dissertation proposal, and I am writing to ask for your support for my data collection process.

My research focuses on teacher perspectives of oral reading fluency data. At the end of the study I would be willing to present a report of the findings (with no identifying information) to you and/or the XXX staff.

The study will utilize teacher surveys and teacher focus groups. Participation in my study would be completely voluntary. The surveys and focus groups will most likely occur between April - May 2012 and will occur outside of regular school hours.

I would like to request your approval for staff at your school to participate if they are willing. Would you consider allowing staff at your school to participate? If you are interested in participation, you may contact me by email (stebbesa@msu.edu) or phone (248-214-9697). Also, I'd be happy to give you more details if needed. Thanks for your time.

Sincerely,

Sarah Stebbe Rowe Email: stebbesa@msu.edu Phone number: 248-214-9697

Appendix B

Teacher Survey Email Invitation

Dear teacher,

My name is Sarah Stebbe Rowe, and I am conducting an online survey as a part of my dissertation research. The survey will investigate attitudes toward reading assessment, and your response would be appreciated.

Your response will be kept confidential and participation is completely voluntary. For participation, you will receive a 20-dollar Amazon gift card through email.

If you are interested in participating in this survey, please follow this link and you will be provided with more information and the option to participate in the survey.

Here is a link to the survey: https://www.surveymonkey.com/s.aspx

Thanks in advance,

Sarah Rowe MSU School Psychology sarah.rowe.msu@gmail.com

This link is uniquely tied to this survey and your email address. Please do not forward this message.

Please note: If you do not wish to receive further emails from us, please click the link below, and you will be automatically removed from our mailing list. https://www.surveymonkey.com/optout.aspx

Appendix C

Teacher Focus Group Email Invitation

Dear teacher,

My name is Sarah Stebbe Rowe, and I am a graduate student at Michigan State University in the School Psychology Program. I am completing my doctoral requirements by examining the acceptability of oral reading fluency data. You were recently invited to participate in a research survey at your school. Your school was selected because many teachers at this school collect oral reading fluency data.

You have also been selected to participate in a ninety-minute small group discussion about oral reading fluency in May. In these discussions, we hope to gain a better understanding of teacher perspectives on oral reading fluency and the factors that influence teachers' opinions. You are one of many teachers from Ingham, Clinton, and Shiawassee County who have been asked to participate in the small group discussions, so by participating you will have the opportunity to meet and interact with teachers from other schools. Participation is completely voluntary.

For participating in this focus group, you will receive a fifty-dollar Amazon gift card plus fifteen dollars to cover transportation costs. The groups will take place in Erickson Hall on the campus of MSU and refreshments will be provided. The times and dates of the groups are listed below. I'd love to have you participate- if you're interested please let me know.

Discussion group times- choose one: Saturday, May 12: 12-1:30 pm Monday, May 14: 5-6:30 pm Tuesday, May 15: 5-6:30 pm Thursday, May 17: 5-6:30 pm

I hope to meet you in May!

Thanks in advance, Sarah Stebbe Rowe sarah.rowe.msu@gmail.com

Appendix D

Teacher Demographic Form and Survey

Initials:

Grade Level Taught:

School:

1. Consent form

Demographics & Background Information

- 2. Gender:
 - a. Male
 - b. Female

3. Race/Ethnicity:

- a. Caucasian or White
- b. Black or African American
- c. American Indian or Alaska Native
- d. Native Hawaiian or Other Pacific Islander
- e. Asian
- f. Latino or Hispanic
- g. Other:

4. Which best describes the education level you have attained?

- a. High school diploma/GED
- b. B.A. or B.S.
- c. M.A. or M.S.
- d. Specialist/ Ed.S.
- e. Ed.D./Ph.D

5. What year did you receive a bachelor's degree in teaching?

- a. Before 1970
- b. 1970 1989
- c. 1990 1999
- d. 2000 2005
- e. 2006 2012
- f. I did not receive a bachelor's degree in teaching

6. What year did you receive your highest degree of education in teaching?

- a. Before 1970
- b. 1970 1990
- c. 1991 2000
- d. 2001 2005
- e. 2006 2012

- 7. What school did you teach at this school year?
- 8. Which best describes the location of your school?
 - a. urban
 - b. suburban
 - c. rural
- 9. How many students are in your classroom? If you are not a general education teacher please select the number that best represents how many students you teach at one time.
 - a. Less than 10
 - b. 11-20
 - c. 21-30
 - d. 31-35
 - e. 36 or more
- 10. Which of the following is the most appropriate description of the level at which you teach currently?
 - a. Kindergarten
 - b. First grade
 - c. Second Grade
 - d. Third Grade
 - e. Fourth Grade
 - f. Fifth Grade
 - g. Sixth Grade
 - h. Mixed (please specify):
- 11. Including the current year, how many years of experience do you have as a classroom teacher at the current grade level you teach?
 - a. 1-2 years
 - b. 3-5 years
 - c. 6 10 years
 - d. 11 19 years
 - e. 20 or more years
- 12. Including the current year, how many years of experience do you have as a classroom teacher at your current school:
 - a. 1-2 years
 - b. 3-5 years
 - c. 6-10 years
 - d. 11-19 years
 - e. 20 or more years
- 13. Including the current year, how many years of experience do you have as a classroom teacher?

- a. 1-2 years
- b. 3-5 years
- c. 6-10 years
- d. 11-19 years
- e. 20 or more years

14. Which best describes the role you hold at your school:

- a. General education
- b. Special education
- c. Title I teacher
- d. Reading specialist/support
- e. Other:

15. Which <u>best</u> describes your teacher certification status: (check all that apply)

- a. Provisional Certificate, Provisional Renewal, or Two-Year Extended Provisional
- b. Professional Education Certificate
- c. 18-Hour and 30-Hour Continuing Certificate or Permanent Certificate
- d. Emergency Certificate or Temporary Teacher Employment Authorization
- e. Substitute Permit
- f. Other (please specify):

16. Teaching Majors, Minors, and Endorsements (check all that apply)

- a. Learning disabilities- major or endorsement
- b. English elementary level- minor
- c. Language Arts- group major, minor or endorsement
- d. Reading- endorsement
- e. Reading Specialist- endorsement
- f. Other (please specify):

Teacher Knowledge¹

- 17) Oral reading fluency is a term that refers to a score based on the number of words read correctly and the number of errors in one minute from a grade level passage. Words omitted, words substituted, and hesitations of more than three seconds are scored as errors. In this test, oral reading fluency evaluates a student's:
 - a. Rate only.
 - b. Accuracy only.
 - c. Rate and accuracy.
 - d. Rate, accuracy, and prosody. (Hosp, Hosp, Howell, 2007, p. 31)
- 18) A fourth-grade teacher notices that the students in his class who have the best reading comprehension for informational texts tend to be students who also have excellent oral reading fluency.² Which of the following is the most likely reason for this relationship?
 - a. Fluent readers have more interest in reading informational texts than do other children.
 - **b.** Fluent readers can focus more of their attention and mental resources on comprehension.
 - c. Fluent readers have extensive background knowledge that enables them to understand the difficult vocabulary in informational texts.
 - d. Fluent readers use context cues to help decode most words, which improves their comprehension of all types of texts.

(Spear-Swerling, Knowledge Survey)

- 19) Which of the following statements about measures of oral reading fluency² and reading comprehension is most accurate?
 - a. Measures of oral reading fluency are not particularly useful because they provide no information about comprehension.
 - b. Measures of oral reading fluency are more time-consuming to administer than are measures of reading comprehension.
 - c. Measures of reading comprehension are generally more reliable and valid than are measures of oral reading fluency.
 - d. Measures of oral reading fluency are useful predictors of reading comprehension in most elementary children.

(Spear-Swerling, Knowledge Survey)

20) Oral reading fluency² is not

- a. Individually administered.
- b. A global outcome measure.

¹ Knowledge items will each be presented on a different page

² At the top of the page the following text appeared: Note: Oral reading fluency measures the number of words read correctly and the number of errors in one minute from a grade level passage. Words omitted, words substituted, and hesitations of more than three seconds are scored as errors.

c. A comprehension check.

- d. Used for progress monitoring.
- 21) The recommended minimum number of times to screen all students using oral reading fluency³ is:
 - a. Once a year.
 - b. Two times a year
 - c. Three times a year.
 - d. Eight times a year. (Hosp, Hosp, Howell, 2007, p. 47)
- 22) The main purpose of universal screening (schoolwide process of collecting oral reading fluency³ data from all students) is to:
 - a. Identify students who may have a learning disability.
 - b. Identify students who may need more academic support.
 - c. Evaluate teachers' performance.
 - d. Identify students who may need to repeat a grade. (Hosp, Hosp, Howell, 2007, p. 48)
- 23) The best reason for conducting universal screening (schoolwide process of collecting oral reading fluency data from all students) using oral reading fluency³ is that it provides information about:
 - a. student motivation.
 - b. how to teach students.
 - c. which students possibly are at risk for reading failure.
 - d. which students possibly have a learning disability.
- 24) School-wide screening using oral reading fluency³ data can help answer all of the following questions <u>except</u>:
 - a. How many students need additional assessment or intervention?
 - b. How many students are at benchmark (i.e., the expected level)?
 - c. What type of instruction do students need?
 - d. How intensive are students' academic needs at this school?
- 25) A tool intended to be used as a universal screener (i.e., to be used to identify a set of students who may need extra support) is typically intended to be:
 - a. Brief.
 - b. Comprehensive.
 - c. Easy.
 - d. Difficult.

³ At the top of the page the following text appeared: Note: Oral reading fluency measures the number of words read correctly and the number of errors in one minute from a grade level passage. Words omitted, words substituted, and hesitations of more than three seconds are scored as errors.

26) Universal screening benchmarks are cut-point scores:

- a. Used to evaluate mastery of a grade level academic task.
- b. That indicate the presence of a learning disability.
- c. Used to identify students with low, moderate, and high risk for later reading problems.
- d. Used to split students into learning groups for instruction.
- 27) Progress monitoring (i.e., frequently collecting data from an individual student) using oral reading fluency⁴ can be used to:
 - a. Identify if a student has a processing disorder.
 - b. Determine if an intervention is working.
 - c. Determine intervention integrity.
 - d. Identify if a student is motivated.

(Hosp, Hosp, Howell, 2007, p. 28)

28) In order to detect academic skill growth when progress monitoring⁵ with oral reading fluency⁴, it is recommended to chose a reading passage that is

- a. always at the student's current grade level.
- b. always one grade level below the student's current grade level.
- c. at the highest level at which the student performs at least at the 75th percentile.
- d. at the highest level at which the student performs at least at the 25th percentile

(Best Practices in Progress Monitoring Academic Goals- Shapiro- page 149)

- 29) Emily, a second grader, has been receiving an initial intervention in reading, for a halfhour three times per week, for the past eight weeks. She has not previously received any reading interventions. A graph of her performance on assessments shows limited growth during the eight-week period, with Emily still well below benchmark in reading and not on a trajectory to catch up. In a typical three-tiered Response To Intervention model, Emily would:
 - a. be referred for comprehensive evaluation for special education.
 - b. be classified as having a learning disability.
 - c. receive pull-out remedial reading services.
 - d. receive a more intensified or individualized intervention.

(Spear-Swerling, Knowledge Survey)

⁴ At the top of the page the following text appeared: Note: Oral reading fluency measures the number of words read correctly and the number of errors in one minute from a grade level passage. Words omitted, words substituted, and hesitations of more than three seconds are scored as errors.

⁵ At the top of the page the following text appeared: Progress monitoring is frequently collecting data from an individual student.

- 30) Assessments used in progress monitoring in reading should have which of the following characteristics?
 - a. They should provide diagnostic information about a student's decoding skills, reading fluency, and reading comprehension.
 - b. They should yield accurate estimates of a student's current grade level functioning.
 - c. They should be relatively quick to administer and have multiple equivalent forms.

d. They should provide many different types of norm-referenced scores. (Spear-Swerling, Knowledge Survey)

- 31) When scores from a standardized test are said to be "reliable," what does this mean?
 - **a.** Student scores from the test can be used for a large number of educational decisions.
 - **b.** If a student retook the same test, he or she would get a similar score on each retake.
 - c. The test score is a more valid measure than teacher judgments.
 - d. The test score accurately reflects the content of what was taught.

(Assessment Literacy Inventory, Craig A. Mertler)

- 32) What is the most important_consideration in choosing an assessment tool for an academic universal screening?
 - a. The accuracy of identifying students at risk for reading difficulty.
 - b. The ease of administering and scoring an assessment tool.
 - c. The accuracy of identifying students who are not at risk for reading difficulty.
 - d. The length of time to conduct testing for all students.
- 33) An assessment tool is considered to have adequate construct validity if
 - a. It measures the theoretical construct it purports to measure.
 - b. It predicts later achievement on an assessment measuring a similar construct.
 - c. It predicts current achievement on an assessment measuring a similar construct.
 - d. Its results can be generalized to other constructs.
- 34) In interpreting scores from standardized, norm-referenced tests for comparison to sameage peers across the nation, most authorities recommend the use of:
 - a. grade equivalent or age equivalent scores.
 - b. standard scores or percentile ranks.
 - c. raw scores.
 - d. percentages.

(Spear-Swerling, Knowledge Survey)

Special Instructions: For the remainder of this survey, you will be asked questions about the use of oral reading fluency. Oral reading fluency is defined as a measure of the rate and accuracy of student reading. Oral reading fluency scores indicate the number of words read correctly and the number of errors in one minute from a grade level passage. Words omitted, words substituted, and hesitations of more than three seconds are scored as errors. Examples of systems that include oral reading fluency include DIBELS® and AIMSweb® Reading Curriculum-based Measurement (R-CBM). Running records are not considered oral reading fluency for the purpose of this study.

Universal screening is defined as a schoolwide process of collecting oral reading fluency data from all students. Three one-minute reading probes are collected from each student and the median (i.e., middle) score for both correct words and errors is recorded. Progress monitoring is defined as frequent data collection from students most at risk for reading failure. This study will gather information about your attitudes towards oral reading fluency used within the context of systems such as DIBELS® or AIMSweb® for universal screening and progress monitoring.

35. Acceptability For Universal Screening

Mrs. Lee is a fourth grade teacher at Woods Elementary. Not all students at this school are reading at the expected grade level. At the beginning, middle, and end of each school year teachers in her school assess the reading skills of all students using oral reading fluency.

In fourth grade, Mrs. Lee collects oral reading fluency data from all of her students. In this test, students read aloud for one minute from three different generic grade level passages. The teacher counts the number of words read correctly and the number of errors in each passage until one minute is complete. Words omitted, words substituted, and hesitations of more than three seconds are scored as errors. The teacher records all six scores and selects the median number of words read correctly (and the associate error score) for the final scores.

She then enters the scores into a school-wide assessment database. Soon after this universal screening process, teachers from each grade level meet to discuss the results of this screening and to identify students who are at risk and in need of further assessment and intervention.

The following information was obtained from this assessment:

- The number of words each student read correctly per minute from a grade level passage.
- The number errors per minute from a grade level passage.
- Categorization of each student into one of three risk categories for later reading failure: low risk, some risk, and high risk.
- Percentage of students in each risk level category for the school and each grade.

Please indicate the number that best describes your agreement or disagreement with each statement regarding this scenario.

1	2	3	4	5	6
Disagree	Disagree	Disagree	Agree slightly	Agree	Agree
strongly	moderately	Slightly		moderately	Strongly

- This would be an acceptable assessment strategy **for universal screening in reading** (for the child's problem)⁶.
- Most teachers would find this approach to assessment appropriate for **identifying students in need of further assessment or intervention** (problems in addition to the ones described).
- This assessment should prove effective in identifying children who need additional instruction (the child's problems).
- I would suggest the use of this assessment to other teachers (school psychologists).
- I would be willing to receive assessment results such as those described with a student transferring into my school district.
- This assessment would be appropriate for a variety of children.
- The assessment was a fair way to identify the **children at risk for reading failure** (child's problems).
- The assessment is reasonable to use **schoolwide** (for the problems described).
- I like the procedures used in this assessment.

⁶ Altered wording appears in bold print and original wording appears in parenthesis

- This assessment was a good way to **identify students at risk for reading failure** (to handle the child's problem).
- Overall, this assessment would be beneficial for all **children** (child).
- This assessment is likely to be helpful in **selecting students who may need additional intervention** (the development of intervention strategies).

36. Acceptability for Progress Monitoring

Chris is a child in the beginning of the fourth grade with difficulty learning. Chris is having trouble in class related to his academic subjects. He is below the grade level benchmark in oral reading fluency. The teacher has identified Chris's specific reading skill deficit and is implementing an intervention to improve his reading skills.

One way the teacher will evaluate the effectiveness of the instruction and monitor his reading progress is by collecting oral reading fluency data on a weekly basis. Chris is asked to read aloud for one minute from a progress-monitoring probe at the appropriate grade level. The following information is obtained from this process each week:

- The number of words Chris read correctly per minute from the reading passage.
- The number of errors per minute from a grade level passage. Words omitted, words substituted, and hesitations of more than three seconds are scored as errors.
- The improvement or decline in the number of words read correctly since the previous week.
- Progress towards the goal to close the gap between his reading performance and his peers.

D1	• • •	1	1 0		1 0 11	• ,		1.	.1 .	•
DIADGA	indianta	VOUR LOV	N ot oaro	nmont to t	tha talla	www.mag.ata	atomonta r	agarding 1	thia	anneria
FICASE	munuale		כו טו מצוכי		ньстони	JWINE SLA		ושמוחווצ	uns-	SUCHALIO

1	2	3	4	5	6
Disagree	Disagree	Disagree	Agree slightly	Agree	Agree
strongly	moderately	Slightly		moderately	Strongly

- This would be an acceptable assessment strategy for monitoring the child's reading growth (the child's problem).
- Most teachers would find this approach to assessment appropriate for problems in addition to the one described.
- This assessment should prove effective in **monitoring the child's reading growth** (identifying the child's problem).
- I would suggest the use of this assessment to other teachers.
- I would be willing to receive assessment results such as those described with a student transferring into my school district.
- This assessment would be appropriate for a variety of children.
- The assessment was a fair way to identify **if the child's reading skills are improving** (the child's problem).
- The assessment is a reasonable way to **monitor reading progress** (for the problem described).

- I like the procedures used in this assessment.
- This assessment was a good way to **monitor if the intervention is working** (handle the child's problem).
- Overall, this assessment would be beneficial for the child.
- This assessment is likely to be helpful in **deciding if changes need to be made to the intervention** (the development of an intervention).

37. Time

Please rate how strongly you agree or disagree with the following statements about the use of time in your school.

1=Strongly Disagree 2=Disagree 3=Neither Agree nor Disagree 4=Agree 5=Strongly Agree

- Class sizes are reasonable such that **the majority of** teachers **at your school** have the time available to meet the needs of all students.
- Teachers have time available to collaborate with their colleagues.
- Teachers are allowed to focus on educating students with minimal interruptions.
- The non-instructional time provided for teachers in my school is sufficient.
- Efforts are made to minimize the amount of routine paperwork teachers are required to do.
- Teachers have sufficient instructional time to meet the needs of all students.
- Teachers are protected from duties that interfere with their essential role of educating students.

38. Facilities and Resources

Please rate how strongly you agree or disagree with the following statements about your school facilities and resources.

1=Strongly Disagree 2=Disagree 3=Neither Agree nor Disagree 4=Agree 5=Strongly Agree

- Teachers have sufficient access to appropriate **assessment** (instructional)* materials (e.g., testing booklets, timers, scoring materials, and administration and scoring manuals)
- Teachers have sufficient access to **assessment** (instructional) technology, including computers, printers, software, and Internet access.
- Teachers have sufficient access to communications technology, including phones, faxes, and email.
- Teachers have sufficient access to office equipment and supplies such as copy machines, paper, and pens for use in classroom assessment.
- Teachers have sufficient access to a broad range of professional support personnel to help with assessment procedures.
- The school environment is clean and well maintained.
- Teachers have adequate space to **conduct assessments** (work productively).
- The physical environment of classrooms in this school supports assessment.
- The reliability and speed of Internet connections in this school are sufficient to support **assessment** (instructional) practices.

Frequency of Use⁷

Special note: The next several items will gather information about how often you use oral reading fluency within packages or systems such as DIBELS or AIMSweb.

Note: Universal screening means collecting data from every student in the classroom to identify those at risk for failure.

Note: Oral reading fluency scores indicate the number of words read correctly and the number of errors in one minute from a grade level passage. Words omitted, words substituted, and hesitations of more than three seconds are scored as errors.

39) Is oral reading fluency data collected from all students in your classroom as a universal screener to identify students at risk for reading failure?

- a) No
- b) Yes

(IF NO- GO TO QUESTION 43, IF YES- GO TO QUESTION 40)

40) Who collects oral reading fluency data as a part of universal screening?

- a. Other individuals collect all of the data.
- b. I collect some of the data and have assistance from others.
- c. I collect all of the data.

41) How often do you personally administer oral reading fluency for the purpose of universal screening?

- a. Never
- b. Once a year
- c. Twice a year
- d. Three times a year
- e. Four or more times a year

42) How often do you consult with other teachers about the outcomes of a <u>universal screening</u> using oral reading fluency data?

- a. Never
- b. Once a year
- c. Twice a year
- d. Three times a year
- e. Four or more times a year

43. Is oral reading fluency data collected to progress monitor student achievement in your classroom?

- a. No
- b. Yes

⁷ Researcher developed

(IF NO- GO TO QUESTION 48, IF YES- GO TO QUESTION 44)

44. Who collects the oral reading fluency progress monitoring data?

- a. Other individuals collect all of the data.
- b. I collect some of the data and have assistance from others.
- c. I collect all of the data.

45) How many students do you personally progress monitor at least once a month using oral reading fluency?

- a. Zero
- b. One
- c. Two
- d. Three
- e. Four or more

46) Which of the following best describes how often you administer oral reading fluency to progress monitor the reading progress of individual students in your classroom?

- a. Never
- b. Three times a year
- c. Once a month
- d. Twice a month
- e. Once a week or more

47) Which of the following best describes how often you use to results of individual students' oral reading fluency progress monitoring data to inform your instruction?

- a. Never
- b. Three times a year
- c. Once a month
- d. Twice a month
- e. Once a week or more

Training and Professional Development

48) To the best of your knowledge, how many courses in your undergraduate program included instruction on using oral reading fluency data?

None 1 2 3 or more

49) After receiving your teaching degree, how many college courses have you taken that included instruction on using oral reading fluency data?None 1 2 3 or more

50) After receiving your teaching degree, how many hours or days of non college-related professional development (e.g., school in-services, trainings, or conferences) have you attended that included instruction on using oral reading fluency data?

a) Noneb) 4 hours or lessc) 1 full dayd) 2 full dayse) 3 or more full days

51) Which best describes your level of preparation for administering and using oral reading fluency data from the following:

Very unprepared Somewhat unprepared Somewhat prepared Very prepared

- undergraduate teacher preparation program
- college coursework after receiving your undergraduate teaching degree
- professional development experiences (e.g. school in-services, trainings, or conferences)

Reading Value

52) Please rate how important teaching each of the following constructs is for promoting students' reading ability.

1	2	3	4	5	6	7
Very	Moderately	Somewhat	Neither	Somewhat	Moderately	Very
Unimportant	Unimportant	Unimportant	important	Important	Important	important
			or			
			unimportant			

- Phonological Awareness: detection and manipulation of sound structures (e.g., rhymes and syllables)
- Concepts about Print: what is known about written language (e.g., title location, author, directionality, punctuation has meaning, connections between letters, words, and sentences)
- Alphabetic Principle: the knowledge that letters are symbols of speech
- Phonemic Awareness: the ability to hear, identify, and manipulate phonemes (i.e., speech sounds) in spoken words
- Fluency: the ability to read with speed, accuracy, and proper expression
- Vocabulary: understanding of word meanings
- Comprehension: ability to understand text

School Use of Data⁸

53) What literacy assessments are required to be collected systematically from <u>all students</u> in the grade level you teach at your school:

- a. Developmental Reading Assessment (DRA)
- b. Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Next

⁸ Collected for the purpose of describing the sample, not for analysis in multiple regression

c. AIMSweb Reading Curriculum-based Measurement (R-CBM)

- d. Michigan Literacy Progress Profile (MLPP)
- e. EasyCBM
- f. Running records
- g. No assessment is required
- h. Other (please specify): _____

54) What literacy assessments are required to be collected systematically from <u>all students</u> in other grade levels at your school:

a. Developmental Reading Assessment (DRA)

- b. Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Next
- c. AIMSweb Reading Curriculum-based Measurement (R-CBM)
- d. Michigan Literacy Progress Profile (MLPP)
- e. EasyCBM
- f. Running records
- g. No assessment is required for other grade levels
- h. Other (please specify):

55) Is it required in your position to use oral reading fluency for universal screening students at your school?

- a. Yes
- b. No

(if no, go to end of survey, if yes, go to 56)

56) If it is required, how often?

- a. Never
- b. Once a year
- c. Twice a year
- d. Three times a year
- e. Four or more times a year

57) If your school requires you to collect oral reading fluency data to screen student reading progress in your classroom, how long has this been in place?

- a. Less than 1 year
- b. 1-2 years
- c. 3-4 years
- d. 5-6 years
- e. 7+ years

Thank you for participating in this study. As a token of appreciate for your time, you will be emailed a twenty-dollar Amazon gift card. If you would like to receive this gift, please provide an email address where the gift card may be sent. Thanks so much.

Email address:

Appendix E

Teacher Focus Group Question Protocol

Today we will be discussing the reading measure called oral reading fluency. What we mean by oral reading fluency is words read correctly per minute from passages in the DIBELS or AIMSweb packages. For the purpose of this discussion, we are just focusing on the measure of oral reading fluency, not other measures that are included in DIBELS and AIMSweb.

Universal screening is a data collection process where oral reading fluency scores are collected from all students in a classroom or school. This is typically done three times a year.

Progress monitoring is a more frequent data collection process used to monitor the progress of specific students who are at risk for reading failure. This might be done once a week, twice a month, or once a month.

- 1. Let's begin by finding out more about each other. Tell us your name, what grade level or type of classroom you teach, and what you enjoy doing most when you aren't teaching.
- 2. How did you learn about oral reading fluency? Think back to when you first started collecting the oral reading fluency measure from DIBELS or AIMSweb for universal screening or progress monitoring. What were your first impressions? What was the initial data collection process like for you?
- 3. How do you use oral reading fluency for universal screening? If you don't use or collect this data for universal screening, how does your school conduct universal screenings?
 - a. How is the data collected?
 - b. Who collects the data?
 - c. What is done with the data?
- 4. How do you use oral reading fluency for progress monitoring? If you don't use oral reading fluency for progress monitoring, how do other teachers use this for progress monitoring?
 - a. How is the data collected?
 - b. Who collects the data?
 - c. What is done with the data?
- 5. How acceptable is oral reading fluency as a measure used for the purpose of universal screening?
 - a. How appropriate is oral reading fluency?
 - b. How fair is oral reading fluency?
 - c. How reasonable is oral reading fluency?
- 6. How acceptable is oral reading fluency as a measure used for the purpose of progress monitoring?
 - a. How appropriate is oral reading fluency?
 - b. How fair is oral reading fluency?
 - c. How reasonable is oral reading fluency?
- 7. What is particularly helpful about using oral reading fluency as a measure of reading achievement?

- a. What supports your ability to collect oral reading fluency data from all students in your classroom?
- b. What supports your ability to collect oral reading fluency data from several students for progress monitoring?
- 8. What are your concerns about using oral reading fluency as a measure of reading achievement?
 - a. What hinders your ability to collect oral reading fluency data from all students in your classroom for universal screening?
 - b. What hinders your ability to collect oral reading fluency data from several students for progress monitoring?

Appendix F

Teacher Pilot Study Survey Consent Form

Dear Teacher:

My name is Sarah Stebbe Rowe, and I am a graduate student at Michigan State University in the School Psychology Program. I am completing my doctoral requirements by examining the acceptability of oral reading fluency data. This research may lead to improvements in reading assessment and intervention.

You are invited to participate in a research study at your school. Through this study, we hope to learn about teacher perspectives on the acceptability of a common reading assessment tool called oral reading fluency. Your school was selected because many teachers at this school collect oral reading fluency data. All of the first through sixth grade teachers at your school will be asked to participate. Participation is completely voluntary.

Study Title: Teacher Acceptability of Oral Reading Fluency *Researchers and Title:* Sarah Stebbe Rowe, M.A., Doctoral Candidate and Sara (Bolt) Witmer, Ph.D., Associate Professor *Department:* Department of Counseling, Educational Psychology, and Special Education, Michigan State University

Participation Procedures

If you decide to participate, you will be asked to complete a questionnaire online and provide feedback about the survey items. This process is estimated to take less than one hour. There are no risks involved by participating in this study and we cannot guarantee any direct benefits for participation. You may stop the survey at any point if you decide to stop participation by clicking the exit button in the top right corner of your screen.

Confidentiality and Privacy

The confidentiality of your survey responses will be protected to the maximum extent allowable by law. Survey data will be analyzed to understand teacher acceptability of oral reading fluency. All subjects will be assigned a confidential code by the principal researcher. Research data collected will use this code instead of teachers' names. Records of this study will be kept confidential, and you will not be identified in any written or verbal reports.

Data Protection

The researcher will provide the results of the study to each school reported in a summary with no individual data. Data will only be used for research purposes. All data will be stored on a password protected computer and external hard drive. Any hard copies of data will be stored in a locked room in a locked cabinet at Michigan State University only accessible to the researcher.

Compensation

For your time and effort, we are offering twenty-dollar Amazon gift cards. Once you have completed the survey, you will be asked to provide an email address where an electronic Amazon Gift Card may be emailed within two weeks of the completion of your survey.

Contacts for Questions

In research, participation is completely voluntary. Please understand that refusal to participate or if you consent and then later withdraw consent from the study, this will not result in any negative consequences for you. You may also refuse to answer any particular questions. The researcher is responsible for explaining risks and benefits of participation so that you may make an informed decision regarding participation. Please ask the researchers any questions you may have about this study.

If you have questions about this study, you may direct those to Dr. (Bolt) Witmer at sbolt@msu.edu or myself at 248-214-9697 or stebbesa@msu.edu. If you have questions about your rights as a participant, you may contact the MSU Human Research Protection Program at 517-355-2180, FAX 517-432-4503, or e-mail irb@msu.edu, or regular mail at 207 Olds Hall, MSU, East Lansing, MI 48824.

Sarah Stebbe Rowe, M.A.

Sara (Bolt) Witmer, Ph.D.

Appendix G

Teacher Survey Consent Form

Dear Teacher:

My name is Sarah Stebbe Rowe, and I am a graduate student at Michigan State University in the School Psychology Program. I am completing my doctoral requirements by examining the acceptability of oral reading fluency data. This research may lead to improvements in reading assessment and intervention.

You are invited to participate in a research study at your school. Through this study, we hope to learn about teacher perspectives on the acceptability of a common reading assessment tool called oral reading fluency. Your school was selected because many teachers at this school collect oral reading fluency data. All of the first through sixth grade teachers at your school will be asked to participate. Participation is completely voluntary.

Study Title: Teacher Acceptability of Oral Reading Fluency Researchers and Title: Sarah Stebbe Rowe, M.A., Doctoral Candidate and Sara (Bolt) Witmer, Ph.D., Associate Professor Department: Department of Counseling, Educational Psychology, and Special Education, Michigan State University

Participation Procedures

If you decide to participate, you will be asked to complete a questionnaire that will take 40-45 minutes to complete. There are no risks involved by participating in this study and we cannot guarantee any direct benefits for participation. You may stop the survey at any point if you decide to stop participation by clicking the exit button in the top right corner of your screen.

Confidentiality and Privacy

The confidentiality of your survey responses will be protected to the maximum extent allowable by law. Survey data will be analyzed to understand teacher acceptability of oral reading fluency. All subjects will be assigned a confidential code by the principal researcher. Research data collected will use this code instead of teachers' names. Records of this study will be kept confidential, and you will not be identified in any written or verbal reports.

Data Protection

The researcher will provide the results of the study to each school reported in a summary with no individual data. Data will only be used for research purposes. All data will be stored on a password protected computer and external hard drive. Any hard copies of data will be stored in a locked room in a locked cabinet at Michigan State University only accessible to the researcher.

Compensation

For your time and effort, we are offering twenty-dollar Amazon gift cards. Once you have completed the survey, you will be asked to provide an email address where an electronic Amazon Gift Card may be emailed within two weeks of the completion of your survey.

Contacts for Questions

In research, participation is completely voluntary. Please understand that refusal to participate or if you consent and then later withdraw consent from the study, this will not result in any negative consequences for you. You may also refuse to answer any particular questions. The researcher is responsible for explaining risks and benefits of participation so that you may make an informed decision regarding participation. Please ask the researchers any questions you may have about this study.

If you have questions about this study, you may direct those to Dr. (Bolt) Witmer at sbolt@msu.edu or myself at 248-214-9697 or stebbesa@msu.edu. If you have questions about your rights as a participant, you may contact the MSU Human Research Protection Program at 517-355-2180, FAX 517-432-4503, or e-mail irb@msu.edu, or regular mail at 207 Olds Hall, MSU, East Lansing, MI 48824.

Sarah Stebbe Rowe, M.A.

Sara (Bolt) Witmer, Ph.D.

I have read the above consent form. I agree to participate in this research study.

I have read the above consent form and would not like to participate in this research study.

Appendix H

Teacher Focus Group Consent Form

Dear Teacher:

My name is Sarah Stebbe Rowe, and I am a graduate student at Michigan State University in the School Psychology Program. I am completing my doctoral requirements by examining the acceptability of oral reading fluency data. This research may assist students and teachers to improve reading assessment and intervention.

You recently were invited to participate in a survey research study. In addition to the survey, you have been selected to participate in a focus group to share your perspectives on the acceptability of a common reading assessment tool called oral reading fluency. Your school was selected because many teachers at this school collect oral reading fluency data. Participation is completely voluntary.

Study Title: Teacher Acceptability of Oral Reading Fluency Researchers and Title: Sarah Stebbe Rowe, M.A., Doctoral Candidate and Sara (Bolt) Witmer, Ph.D., Associate Professor Department: Department of Counseling, Educational Psychology, and Special Education, Michigan State University

Participation Procedures

If you decide to participate, you will be asked to attend a ninety-minute focus group with 3-7 other teachers. Each participant is responsible for arranging transportation to the focus group location. The focus group will focus on soliciting your attitudes regarding oral reading fluency. A trained researcher from Michigan State University will moderate the group. This interview will be audio-taped. There are no risks involved by participating in this study and we cannot guarantee any direct benefits for participation. Refreshments will be served at this meeting.

Confidentiality and Privacy

Focus group data will be recorded, transcribed, and analyzed by university researchers to identify common themes among teacher perspectives on oral reading fluency. All subjects will be assigned a confidential code by the principal researcher. Research data collected will use this code instead of teachers' names. Records of this study will be kept confidential, and you will not be identified in any written or verbal reports. Specific quotes and paraphrased ideas may be used but no identifying information will be reported or published.

Data Protection

The researcher will provide the results of the study to each school reported in a summary with no individual data. Data will only be used for research purposes. All data will be stored on a password protected computer and external hard drive. Any hard copies of data will be stored in a locked room in a locked cabinet at Michigan State University only accessible to the researcher.
Compensation

For your time and effort, we are offering fifty-dollar Amazon gift cards. Also, each participant will be provided fifteen dollars in cash to cover transportation costs. Once you have completed the focus group, you will be provided with the gift card and cash and you will be asked to sign a statement indicating that you have accepted these items.

Contacts for Questions

In research, participation is completely voluntary. Please understand that refusal to participate or if you consent and then later withdraw consent from the study, this will not result in any negative consequences for you. You may also refuse to answer any particular questions. The researcher is responsible for explaining risks and benefits of participation so that you may make an informed decision regarding participation. Please ask the researchers any questions you may have about this study.

If you have questions about this study, you may direct those to Dr. Witmer at sbolt@msu.edu or myself at 248-214-9697 or stebbesa@msu.edu. If you have questions about your rights as a participant, you may contact the MSU Human Research Protection Program at 517-355-2180, FAX 517-432-4503, or e-mail irb@msu.edu, or regular mail at 207 Olds Hall, MSU, East Lansing, MI 48824.

Sarah Stebbe Rowe, M.A.

Sara (Bolt) Witmer, Ph.D

Your signature below indicates your voluntary agreement to participate in this study.

Name (Print please):

Signature:

Date:

Appendix I

Variables

Dependent Variables:

- Acceptability Rating Profile- Revised for Universal Screening
- Acceptability Rating Profile- Revised for Progress Monitoring

Independent Variables:

Teacher Characteristics:

- Teacher endorsement status
- Number of years teaching
- Frequency of use
- Training
- Knowledge

Teacher Perceptions of School Characteristics:

- Time
- Resources/Facilities

Appendix J

Table 16

Focus Group Code Frequency

Final Coding Scheme	Number of Groups Discussing Code	Total References
Change	4	27
Opinion	3	9
Behavior	3	8
Influence on Students	4	28
• Teaching speed	4	13
Motivation	3	6
• Comparison with peers	1	3
• Student mood	3	9
District or School Issues	4	36
• Teacher evaluations using ORF	4	16
• Role of administration in assessment	4	20
Data Collection Procedures	4	43
• Administration responsibility	4	26
• Scheduling	4	9
• Computer assisted testing	3	6
Factors influencing accuracy of the scores	4	52
Nature of environment	3	6
• Nature of passage	4	12
• Nature of students	4	25
• Nature of assessor	4	17
Purpose of ORF	4	54
Progress monitoring	4	25
• Universal screening	4	39
• US vs. PM	2	7
Resources needed	4	56
• Funding	4	7
• People	4	15
• Space	3	6
• Time	4	30
• Training	4	12
• Knowledge	4	13
Limitations of ORF	4	86
• Other assessments needed	4	40
• Teacher judgment needed	4	20
• Mismatch with instruction	4	17

Table 16 (cont'd)		
Use of Data	4	88
• Student growth and goal setting	4	31
• Using ORF to make decisions	4	46
Parent communication	3	15
Standard measurement	2	5
• Consult with other educators	4	13
Negative attitude	4	56
Positive Attitude	4	31

REFERENCES

REFERENCES

- Allinder, R. M., & Eccarius, M. A. (1999). Exploring the technical adequacy of curriculumbased measurement in reading for children who use manually coded English. *Exceptional Children*, 65(2), 271-283.
- Allinder, R. M., & Oats, R. G. (1997). Effects of acceptability on teachers' implementation of curriculum-based measurement and student achievement in mathematics computation. *Remedial and Special Education*, 18, 113-120. doi: 10.1177/074193259701800205
- AERA, APA, & NCME (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- American Federation of Teachers (AFT), National Council on Measurement in Education (NCME), & National Education Association (NEA) (1990). Standards for Teacher Competence in the Educational Assessment of Students. Retrieved from http://www.unl.edu/buros/article3.html
- Baker, S. K., & Good, R. (1994). *Curriculum-based measurement reading with bilingual hispanic students: A validation study with second-grade students.* Paper presented at the Annual Meeting of the Council for Exceptional Children, Denver, CO.
- Baraldi, A.N. & Enders, C.K. (2010). An introduction to modern missing data analyses. *Journal* of School Psychology, 48, 5-37. doi:10.1016/j.jsp.2009.10.001
- Batsche, G., Elliot, J., Graden, J., Grimes, J., Kovaleski, J., Prasse, D., et al. (2005). *Response to intervention: Policy considerations and implementation*. Alexandria, VA: National Association of State Directors of Special Education.
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23, 43-55. doi: 10.1037/1045-3830.23.1.43
- Boneau, C.A. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, *57* (1), 49-64. doi: 10.1037/h0041412
- Brophy, J. E., & Good, T. L. (1974). *Teacher-student relationships: Causes and consequences*. New York: Holt, Rinehart & Winston.
- Campbell, D. T., Stanley, J. C., & Gage, N. L. (1963). *Experimental and Quasi-experimental Designs for Research*. Boston, MA: Houghton Mifflin Boston.
- Chafouleas, S. M., Riley-Tillman, T. C., & Eckert, T. L. (2003). A comparison of school psychologists' acceptability, training, and use of norm-referenced, curriculum-based, and

brief experimental analysis methods to assess reading. *School Psychology Review*, 32(2), 272-281.

- Christ, T. J., & Ardoin, S. P. (2009). Curriculum-based measurement of oral reading: Passage equivalence and probe-set development. *Journal of School Psychology*, *47*, 55-75. doi: 10.1016/j.jsp.2008.09.004
- Christ, T. J., & Silberglitt, B. (2007). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review*, 36(1), 130-146.
- Cohen, D. K., & Ball, D. L. (2007). Educational innovation and the problem of scale. *Scale-up in Education: Ideas in Principle*. Washington, DC: Rowman & Littlefield Publishers.
- Crawford, L., Tindal, G., & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment*, *7*, 303-323. doi: 10.1207/S15326977EA0704_04
- Creswell, J. W., & Clark, V. L. P. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.
- Daane, M., Campbell, J., Grigg, W., Goodman, M., & Oranje, A. (2005). Fourth-grade students reading aloud: NAEP 2002 special study of oral reading (NCES 2006-469). *National Center for Education Statistics*. Washington, DC: Government Printing Office. Retreived from http://nces. ed. gov/nationsreportcard/pdf/studies/2006469. Pdf
- Deeney, T. A. (2010). One-minute fluency measures: Mixed messages in assessment and instruction. *The Reading Teacher*, 63(6), 440-450.
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, *52*(3), 219-232.
- Deno, S. L., Reschly, A. L., Lembke, E. S., Magnusson, D., Callender, S. A., Windram, H., et al. (2009). Developing a school-wide progress monitoring system. *Psychology in the Schools, 46*, 44-55. doi: 10.1002/pits.20353
- Derr-Minneci, T. F., & Shapiro, E. S. (1992). Validating curriculum-based measurement in reading from a behavioral perspective. *School Psychology Quarterly*, 7(1), 2-16.
- Dewey, J. (1910). How We Think. Lexington, MA: D.C. Heath & Co.
- Dorn, S. (2010). The political dilemmas of formative assessment. *Exceptional Children*, 76(3), 325-337.
- Druckman, L. E. (1997). An investigation of the utility of using local curriculum-based measurement norms versus group standardized norms to predict students at risk for

academic failure in reading and mathematics. (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses. (9715231)

- Duffy, G. G. (1982). Fighting off the alligators: What research in real classrooms has to say about reading instruction. *Journal of Literacy Research*, 14(4), 357-373.
- Duffy, G. G., & Anderson, L. (1984). Teachers' theoretical orientations and the real classroom. *Reading Psychology: An International Quarterly*, *5*, 97-104.
- Dunn, E. K., & Eckert, T. L. (2002). Curriculum-based measurement in reading: A comparison of similar versus challenging material. *School Psychology Quarterly*, 17, 24-46. doi: 10.1521/scpq.17.1.24.19904
- Eckert, T. L., Hintze, J. M., & Shapiro, E. S. (1999). Development and refinement of a measure for assessing the acceptability of assessment methods: The Assessment Rating Profile-Revised. *Canadian Journal of School Psychology*, 15, 21-42. doi: 10.1177/082957359901500103
- Eckert, T. L., Shapiro, E. S., & Lutz, J. G. (1995). Teachers' ratings of the acceptability of curriculum-based assessment methods. *School Psychology Review*, 24(3), 497-511.
- Enders, C.K. (2010). Applied Missing Data Analysis. New York, NY: Guilford Press.
- Espin, C., & Deno, S. (1993). Performance in reading from content area text as an indicator of achievement. *Remedial and Special Education*, 14, 47-59. doi: 10.1177/074193259301400610
- Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational research*, *38*(1), 47-65.
- Feinberg, A., & Shapiro, E. S. (2003). Teacher accuracy: An examination of teacher-based judgments of students reading with differing achievement levels. (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses. (3086945)
- Fewster, S., & Macmillan, P. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education*, 23, 149-156. doi: 10.1177/07419325020230030301
- Finn, C. A., & Sladeczek, I. E. (2001). Assessing the social validity of behavioral interventions: A review of treatment acceptability measures. *School Psychology Quarterly*, 16, 176-206. doi: 10.1521/scpq.16.2.176.18703
- Fischl, M. F. (2008). The influence of research-based efficacy information on teacher acceptability of a repeated reading/peer tutoring intervention for reading fluency. (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses. (3326326)

- Fixsen, D. L., & Blase, K. A. (2009, January). Implementation: The missing link between research and practice. NIRN Implementation Brief #1. Chapel Hill: The University of North Carolina, FPG, NIRN.
- Foegen, A., Espin, C., Allinder, R., & Markell, M. (2001). Translating research into practice: Preservice teachers' beliefs about curriculum-based measurement. *The Journal of Special Education*, 34, 226-236. doi: 10.1177/002246690103400405
- Forgas, J.P., Cooper, J., Crano, W.D. (2010). *The Psychology of Attitudes and Attitude Change*. London: Routledge.
- Foster, S. L., & Mash, E. J. (1999). Assessing social validity in clinical treatment research: Issues and procedures. *Journal of Consulting and Clinical Psychology*, 67, 308-319. doi: 10.1037/0022-006X.67.3.308
- Fowler, F. J. (2009). Survey research methods. Los Angeles, CA: Sage Publications, Inc.
- Fuchs, L. S., & Deno, S. L. (1981). A comparison of reading placements based on teacher judgment, standardized testing, and curriculum-based assessment. Report for the Minneapolis Institute for Research on Learning Disabilities. Minneapolis, MN: Minnesota University.
- Fuchs, L. S., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children*, *61*, 15-24. doi: 35400004113529.0020
- Fuchs, D., & Fuchs, L. S. (2006). Introduction to Response to Intervention: What, why, and how valid is it? *Reading Research Quarterly*, *41*, 93-99. doi: 10.1598/RRQ.41.1.4
- Fuchs, L., & Fuchs, D. (2004). What is scientifically-based research on progress monitoring. Washington, DC: National Center on Progress Monitoring, American Institute for Research, Office of Special Education Programs. Retrieved from http://www.studentprogress.org/library/What_is_Scientificall_%20Based_Research.pdf
- Fuchs, L. S., Fuchs, D., Hosp, M., & Jenkins, J. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, 5, 239-256. doi: 10.1207/S1532799XSSR0503_3
- Garson, G. D. (2011). Multiple regression. *Statnotes: Topics in Multivariate Analysis* Retrieved from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm
- Glass, G.V., Peckham, P.D., & Sanders, J.R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research*, 42, 237-288. doi: 10.3102/00346543042003237

- Goffreda, C. T., Diperna, J. C., & Pedersen, J. A. (2009). Preventive screening for early readers: Predictive validity of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS). *Psychology in the Schools, 46*, 539-552. doi: 10.1002/pits.20396
- Gonick, L., & Smith, W. (1993). *The Cartoon Guide to Statistics*. New York, NY: HarperCollins.
- Good, R., Simmons, D., & Kame enui, E. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Studies of Reading*, *5*, 257-288. doi: 10.1207/S1532799XSSR0503_4
- Goodwin, L. D., & Leech, N. L. (2003). The meaning of validity in the new standards for educational and psychological testing: Implications for measurement courses. *Measurement and Evaluation in Counseling and Development*, *36*(3), 181-191.
- Green, S. B., & Salkind, N.J. (2003). *Using SPSS for Windows and Macintosh*. Upper Saddle River, NJ: Prentice Hall.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. *Educational Evaluation and Policy Analysis*, 11, 255-274. doi: 10.3102/01623737011003255
- Gresham, F. M., & Lopez, M. F. (1996). Social validation: A unifying concept for school-based consultation research and practice. *School Psychology Quarterly*, 11, 204-227. doi: 10.1037/h0088930
- Harry, B., Sturges, K.M., & Kligner, J.K. (2005). Mapping the process: An exemplar of process and challenge in grounded theory analysis. *Educational Researcher*, 34, 3-13. doi: 10.3102/0013189X034002003
- Hall, L. A. (2005). Teachers and content area reading: Attitudes, beliefs and change. *Teaching and Teacher Education, 21*, 403-414. doi: 10.1016/j.tate.2005.01.009
- Hamilton, C., & Shinn, M. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review*, 32(2), 228-241.
- Henninger, K. L. (2010). Exploring the relationship between factors of implementation, treatment integrity and reading fluency. (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses. (3409588)
- Heeringa, S.G., West, B.T., & Berglund, P.A. (2010). *Applied Survey Data Analysis*. New York, NY: CRC Press.

- Hosp, M., & Fuchs, L. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review*, 34(1), 9-27.
- Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A Practical Guide to Curriculum-Based Measurement*. New York, NY: Guilford Press.
- Howell, K. W., Hosp, J. L., & Kurns, S. (2008). *Best practices in curriculum-based evaluation*. Bethesda, MA: The National Association of School Psychologists.
- Hudson, R. F., Lane, H. B., & Pullen, P. C. (2005). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher*, *58*, 702-714. doi: 10.1598/RT.58.8.1
- Ikeda, M. J., Neessen, E., & Witt, J. C. (2008). Best practices in universal screening. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V* (103-114). Bethesda, MA: The National Association of School Psychologists.
- 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.
- Jenkins, J. R., & Johnson, E. (2011). Universal screening for reading problems: Why and how should we do this. *Response to Intervention Network*. Retrieved November 25, 2011, from www.rtinetwork.org
- Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, 33, 14-26. doi: 10.3102/0013189X033007014
- Jones, D. L. (2009). Analyzing the effects of two response to intervention tools, oral reading fluency and Maze assessments, in the language arts classrooms of middle school students. (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses. (3378853)
- Kazdin, A. E. (1980). Acceptability of alternative treatments for deviant child behavior. *Journal* of Applied Behavior Analysis, 13, 259-273. doi: 10.1901/jaba.1980.13-259
- Knight, P.E., Knight, K.M., Sidani, S., Figueredo, A.J. (2007). *Missing Data: A Gentle Introduction*. New York, NY: Guilford Press.
- Kranzler, J., Brownell, M., & Miller, M. (1998). The construct validity of curriculum-based measurement of reading: An empirical test of a plausible rival hypothesis. *Journal of School Psychology*, 36, 399-415. doi: 10.1016/S0022-4405(98)00018-1
- Kranzler, J. H., Miller, M. D., & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. *School Psychology Quarterly*, 14, 327-342. doi: 10.1037/h0089012

- Krueger, R. A., & Casey, M. A. (2009). *Focus Groups: A Practical Guide for Applied Research*. Thousand Oaks, CA: Sage Publications.
- Ladd, H. F. (2011). Teachers' perceptions of their working conditions. *Educational Evaluation* and Policy Analysis, 33, 235-261. doi: 10.3102/0162373711398128
- Lentz, F. E., Allen, S. J., & Ehrhardt, K. E. (1996). The conceptual elements of strong interventions in school settings. *School Psychology Quarterly*, 11, 118-136. doi: 10.1037/h0088924
- Little, R.J.A., & Rubin, D.B. (2002). *Statistical Analysis with Missing Data*. Hoboken, NJ: John Wiley & Sons, Inc.
- Manzo, K. (2005). National clout of DIBELS test draws scrutiny. Education Week, 25(5).
- Mautone, J. A., DuPaul, G. J., Jitendra, A. K., Tresco, K. E., Vile Junod, R., & Volpe, R. J. (2009). The relationship between treatment integrity and acceptability of reading interventions for children with attention-deficit/hyperactivity disorder. *Psychology in the Schools, 46*, 919-931. doi: 10.1002/pits.20434
- McKnight, P.E., McKnight, K.M., Sidani, S., & Figueredo, A.J. (2007). *Missing Data: A Gentle Introduction*. New York, NY: The Guilford Press.
- Mertler, C. A. (2003). Patterns of response and nonresponse from teachers to traditional and web surveys. *Practical Assessment, Research & Evaluation, 8*(22).
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools*, *12*(2), 101-113.
- Michigan Department of Education (n.d.). MiBLSi: Michigan Integrated Behavior and Learning Supports Initiative. Retrieved from http://miblsi.cenmi.org/MiBLSiModel/Evaluation/Measures.aspx
- Michigan Integrated Behavior and Learning Support Initiative (n.d.). Participating MiBLSi schools. Retrieved from http://miblsi.cenmi.org/About/Contact/ParticipatingSchools.aspx
- Morse, J. M. (1991). Approaches to qualitative-quantitative methodological triangulation. *Nursing research*, 40(2), 120-123.
- Moskal, M. K., & Blachowicz, C. (2006). *Partnering for Fluency*. New York, NY: The Guilford Press.
- National Association of School Psychologists (2008). *NASP Position Statement: Prevention and Intervention Research in the Schools*. Bethesda, MD: NASP.

- National Center for Educational Statistics (2010). *Common Core of Data*. Retrieved from http://nces.ed.gov/ccd/bat/output.asp
- National Center for Educational Statistics. (2009). *The Nation's Report Card: Reading 2009*. Washington, D.C.: Institute of Educational Sciences.
- National Reading Panel. (2000). *Report of the National Reading Panel: Teaching Children to Read*. Bethesda, MD.
- Nesbary, D. K. (2000). Survey Research and the World Wide Web. Boston: Allyn and Bacon.
- Newman, I., & McNeil, K. (1998). *Conducting Survey Research in the Social Sciences*. New York: University Press of America, Inc.
- New Teacher Center. (2011). New Teacher Center. Retrieved from http://www.newteachercenter.org/research.php
- New Teacher Center. (2011). Reliability and Validity of the North Carolina Teacher Working Conditions Survey. Retrieved from http://www.ncteachingconditions.org/sites/default/files/attachments/validityandreliability. pdf
- No Child Left Behind Act of 2001, 20 U.S.C. § 6310 (2002).
- Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? Assessment & Evaluation in Higher Education, 33, 301-314. Doi: 10.1080/02602930701293231
- Orr, J.M., Sackett, P.R., Dubois, C. (1991). Outlier detection and treatment in I/O psychology: A survey of researcher beliefs and an empirical illustration. *Personnel Psychology*, *44*, 473-486. doi: 10.1111/j.1744-6570.1991.tb02401.x
- Osborne, J.W. & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation, 9*(6), 1-12.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*, 307-332. doi: 10.2307/1170741
- Peterson, C. A., & McConnell, S. R. (1996). Factors related to intervention integrity and child outcome in social skills interventions. *Journal of Early Intervention*, 20, 146-164. doi: 10.1177/105381519602000206
- Peverly, S. T., & Kitzen, K. R. (1998). Curriculum-based assessment of reading skills: Considerations and caveats for school psychologists. *Psychology in the Schools*, 35, 29-47. doi: 10.1002/(SICI)1520-6807(199801)35:1<29::AID-PITS3>3.0.CO;2-P

- Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculumbased measurement. *Journal of Psychoeducational Assessment*, 23, 326-338. doi: 1177/073428290502300403
- Raykov & Marcoulides (2008). *An Introduction to Applied Multivariate Analysis*. New York, New York: Routledge.
- Rea, L. M., & Parker, R. A. (2005). *Designing and Conducting Survey Research*. San Francisco, CA: Jossey-Bass.
- Reschly, A., Busch, T., Betts, J., Deno, S., & Long, J. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, *47*, 427-469. doi: 10.1016/j.jsp.2009.07.001
- Richardson, V., Anders, P., Tidwell, D., & Lloyd, C. (1991). The relationship between teachers beliefs and practices in reading comprehension instruction. *American Educational Research Journal*, 28(3), 559-586.
- Rimstidt, H. L. (2001). Investigation of alternative administration strategies for curriculum-based measurement: Maximizing cost-effectiveness. (Doctoral dissertation). Retrieved from Proquest Dissertations and Theses. (3020774)
- Roehrig, A. D., Duggar, S. W., Moats, L., Glover, M., & Mincey, B. (2008). When teachers work to use progress monitoring data to inform literacy instruction. *Remedial and Special Education, 29*, 364-382. doi: 10.1177/0741932507314021
- Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology*, 46, 343-366. doi: 10.1016/j.jsp.2007.06.006
- Rogers, E. M. (2003). Diffusion of Innovations. New York, NY: Free Press.
- Roth, F. P. (2004). Word Recognition Assessment Frameworks. In C. A. Stone, E. R. Silliman,
 B. J. Ehren & K. Apel (Eds.), *Handbook of Language and Literacy: Development and Disorders*. New York, NY: Guilford.
- Rupley, W. H., & Logan, J. W. (1984). Elementary Teachers' Beliefs about Reading and Knowledge of Reading Content: Relationships to Decisions about Reading Outcomes. Eric Document Reproduction Service No. ED 285162.
- Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. Thousand Oaks, CA: Sage Publications.

- Salvia, J., Ysseldyke, J. E., & Bolt, S. (2010). *Assessment in Special and Inclusive Education*. Boston, MA: Houghton Mifflin Company.
- Samuels, S. (2007). The DIBELS tests: Is speed of barking at print what we mean by reading fluency? *Reading Research Quarterly*, *42*, 563-566. doi: 10.1598/RRQ.42.4.5
- Schilling, S. G., Carlisle, J. F., Scott, S. E., & Ji, Z. (2007). Are fluency measures accurate predictors of reading achievement? *Elementary School Journal*, 107(5), 429-448.
- Schonlau, M., Fricker, R. D., & Elliott, M. N. (2002). *Conducting Research Surveys via E-mail and the Web*. Santa Monica, CA: Rand Corporation.
- Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stahl, S. A. (2004). Becoming a fluent reader: Reading skill and prosodic features in the oral reading of young readers. *Journal of Educational Psychology*, 96, 119-129. doi: 10.1037/0022-0663.96.1.119
- Shapiro, E. S., & Eckert, T. L. (1994). Acceptability of curriculum-based assessment by school psychologists. *Journal of School Psychology*, 32, 167-183. doi: 10.1016/0022-4405(94)90009-4
- Shinn, M. R. (2007). Identifying students at risk, monitoring performance and determining eligibility within response to intervention: Research on educational need and benefit from academic intervention. *School Psychology Review*, *36*(4), 601-617.
- Shinn, M. R. (Ed.). (2008). Best practices in using curriculum-based measurement in a problemsolving model. In A. Thomas & J. Grimes (Eds.), *Best Practices in School Psychology V*. Bethesda, MA: The National Association of School Psychologists.
- Shinn, M. R., Good, R. H., Knutson, N., Tilly, W., & Collins, V. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, 21(3), 469-479.
- Shinn, M.R, Tindal, G., Spira, D., & Marston, D. (1987). Practice of learning disabilities as social policy. *Learning Disability Quarterly*, *10*, 17-28. doi: 10.2307/1510751
- Solomon, David J. (2001). Conducting web-based surveys. *Practical Assessment, Research & Evaluation, 7*(19). Retrieved from http://PAREonline.net/getvn.asp?v=7&n=19
- Spear-Swerling, L., Brucker, P. O., & Alfano, M. P. (2005). Teachers literacy-related knowledge and self-perceptions in relation to preparation and experience. *Annals of Dyslexia*, 55, 266-296. doi: 10.1007/s11881-005-0014-7
- Spear-Swerling, L., & Cheesman, E. (2011). Teachers' knowledge base for implementing response-to-intervention models in reading. *Reading and Writing*, 1-33. doi: 10.1007/s11145-011-9338-3

- StatSoft. (2011). Electronic Statistics Textbook. Retrieved from http://www.statsoft.com/textbook/
- Sterling-Turner, H. E., & Watson, T. S. (2002). An analog investigation of the relationship between treatment acceptability and treatment integrity. *Journal of Behavioral Education*, 11, 39-50. doi: 1053-0819/02/0300-0039/0
- Stieglitz, E. L. (1983). Effects of a content area reading course on teacher attitudes and practices: A four year study. *Journal of Reading*, *26*(8), 690-696.
- Urdan, T. C., & Paris, S. G. (1994). Teachers' perceptions of standardized achievement tests. *Educational Policy*, 8(2), 137. doi: 10.1177/0895904894008002003
- U.S. Department of Education (2006). Reading First Implementation Evaluation: Interim Report. Retreived from www2.ed.gov/rschstat/eval/other/readingfirst-interim/readingfirst.pdf
- U.S. Department of Education (2009). Reading First. Retrieved from http://www2.ed.gov/programs/readingfirst/index.html
- Valencia, S. W., Smith, A. T., Reece, A. M., Li, M., Wixson, K. K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly*, 45, 270-291. doi: 10.1598/RRQ.45.3.1
- Wesson, C. L., King, R. P., & Deno, S. L. (1984). Direct and frequent measurement of student performance: If it's good for us, why don't we do it? *Learning Disability Quarterly*, 7, 45-48.
- Wilkinson, L. American Psychological Association Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604. doi: 10.1037/0003-066X.54.8.594
- Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. *Advances in School Psychology*, *4*, 251-288.
- Wolf, M. (1978). Social validity: The case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11(2), 203-214.
- Yell, M., Deno, S., & Marston, D. (1992). Barriers to implementing curriculum-based measurement. Assessment for Effective Intervention, 18, 99-112. doi: 10.1177/153450849201800109