THE EFFECTS OF DECODABILITY AND HIGH-FREQUENCY WORDS ON CURRICULUM-BASED MEASUREMENT ORAL READING FLUENCY OUTCOMES FOR AT-RISK STUDENTS

By

STACI KANOELANI CUMMING

Bachelor of Arts
Oral Roberts University
Tulsa, Oklahoma
2002

Master of Science
Oklahoma State University
Stillwater, Oklahoma
2004

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the degree of
DOCTORATE of EDUCATION
July, 2008

THE EFFECTS OF DECODABILITY AND HIGH-FREQUENCY WORDS ON CURRICULUM-BASED MEASUREMENT ORAL READING FLUENCY OUTCOMES FOR AT-RISK STUDENTS

Dissertation Approval

Dr. Judy Oehler-Stinnett
Dissertation Advisor

Dr. Eric Mesmer

Dr. Gary Duhon

Dr. Heidi Anne Mesmer

Dr. A. Gordon Emslie
Dean of the Graduate College

Acknowledgements

Throughout the process of completing this dissertation, and my graduate studies in general, there have been several groups of people who have supported me and without whom I would not have been successful in this endeavor.

Ka waihona o ka na'auao. Thank you to my dissertation committee for your guidance, wisdom, and patience. Dr. Oehler-Stinnett, thank you for stepping in to help me get this finished. I would not have made it to the end without you. Dr. E. Mesmer, thank you for getting me started in all of this. I would probably still be in Oklahoma trying to figure out what to do for a dissertation if it were not for you. Dr. H.A. Mesmer, I have learned more from you than you will ever know. Your knowledge in this area is incredible and I can only hope that one day I will be as knowledgeable as you. Dr. Duhon, thank you for helping to keep me on track by asking all those questions that I did not want to hear and for giving me a reminder to get moving when I needed one.

Hapapa hewa ka malihini makamaka 'ole. To my dissertation buddies, I could not have made it without our "dissertation parties" at the library, the numerous all-nighters that we pulled together, and our frequent "debriefing" sessions. Where would I be if I did not have you to share in my feelings of despair, anxiety, and now, my feelings of joy?

'A'ohe pauka 'ike I ka halau ho'okahi. Ka 'ike a ka makua he hei na ke keiki. Finally, I would not have made it without my family. I thank my parents for helping to develop the persistence and determination that it took to get through this process. I would

not have survived this without your love, support, prayers, and the frequent reminders that this is much better than working in the pineapple fields. Many thanks need also go to my sister, who put up with me through this process. This work is dedicated to my wonderful family.

Ola no ka mea ke Akua, make no ka mea ke Akua ʻole

TABLE OF CONTENTS

Chapter III (continued) Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION TO THE RESEARCH

Within the last forty years the field of learning disabilities has emerged as one of the most prominent areas within the larger discipline of special education. Extensive work has been completed within the field yet many areas remain under debate. One area within the field that has been disputed for years is the process by which students are identified as having a learning disability. The predominant model of determining a learning disability is the discrepancy model (Hallahan & Mercer, 2002). In this method, school personnel rely on norm-referenced intellectual ability and achievement measures to judge whether students are sufficiently discrepant to require special education services under the category of learning disabled. The discrepancy model posits that a learning disability is present when a student demonstrates a severe incongruity between his or her observed intellectual ability and academic ability within a specific skill area.

Opposition to this model has existed since its inception with concerns regarding faulty foundational data, possible confounding of variables, inappropriateness for some age groups, and poor discriminative value (Hallahan & Mercer, 2002). More importantly, however, assessments conducted in this manner give little information relevant for instructional planning or determining progress towards attaining proficiency within an area. Moreover, the little information that can be gleaned from these standardized

assessments may lead to erroneous decision-making (Jenkins & Pany, 1978). Due to growing discontent with this model a new means of identification has recently risen to the forefront of both discussion and research.

Recently the focus of much attention has been the responsiveness to intervention (RTI) model. This model has been accepted under the *Individuals with Disabilities Education Improvement Act,* as a new paradigm for identifying learning disabilities (U.S. Department of Education, 2005). Although RTI does not replace the traditional IQ-Achievement discrepancy model within the law, school districts have the option of implementing RTI instead of, or in conjunction with, the discrepancy model. Responsiveness to intervention is broadly defined as the extent to which a change in behavior or performance ensues as the result of an intervention (Gresham, 1991). Central to the notion of RTI are the concepts of instructional utility and continual progress monitoring in "direct indicators of child learning" to ascertain growth in response to validated interventions (Van Der Heyden, 2006). The aim is that students will be identified as learning disabled when they do not respond to reliably implemented, research-based interventions which have been delivered with enough strength that results are expected.

The most common means for progress monitoring within the RTI framework is through curriculum-based measurement (CBM), and in fact, RTI came out of the research and work with CBM. CBM is a standardized assessment procedure used for screening, monitoring progress, and evaluating instruction (Deno, 1985). When applied within the RTI framework, CBM is utilized to make decisions regarding a given student's level of response to educational intervention.

CBM procedures were developed to monitor the progress of students over time (Deno, 1985). CBM has been established as a formative assessment procedure that is predictive of general education outcomes and predictive of performance on criterion referenced assessments (Good & Jefferson, 1998). Curriculum based measures have also been recognized as reliable, valid, and sensitive measures for monitoring student growth (Fuchs, Fuchs, Hamlett, & Stecker, 1991).

The most commonly employed CBM measure for measuring reading ability is oral reading fluency (ORF; Good, Simmons, & Kame'enui, 2001). A student's ORF rate is determined by having the student read aloud from a passage for a predetermined time while the examiner records the student's accuracy on a separate probe. The total number of words read correctly per minute is that student's ORF rate (Shinn et al., 1989). ORF can be charted to track progress over time to determine slope of growth (Shinn, Good, and Stein, 1989) or can be compared to benchmarks to determine his or her relative standing in relation to national or local norms (Good, et al., 2001).

Research examining the technical adequacy of oral reading fluency CBM has established that it is closely related to performance on standardized reading tests (Deno, Mirkin, & Chiang, 1982), it is a reasonable estimate of reading comprehension (Fuchs, Fuchs, & Maxwell, 1988), it has strong discriminative validity for differentiating between student reading level (Shinn, Tindal, Spira, & Marston, 1987) and is sensitive to student growth (Good & Jefferson, 1998). Furthermore, Slocom, Street and Gilbert (1995) have determined that CBM ORF coordinates well with the theoretical underpinnings of the behavior analytic and cognitive theories of reading theory. By and large, CBM ORF

procedures have an extensive history of technical adequacy for evaluating and monitoring student progress.

Recently; however, researchers have questioned the continued applicability of CBM procedures, pointing out that most, if not all, of the research related to the validity and reliability of CBM was conducted in materials which are significantly different than the mainstream curricula of today (Hintze & Shapiro, 1997). As discussed by Hiebert and Fisher (2002) mainstream curricula has undergone significant change in the last century and even within the last thirty years. The primary change during this time has been a movement away from the controlled vocabulary of traditional basals and towards the uncontrolled literature based curricula (Hiebert & Fisher, 2002). While most of the research on CBM and fluency was completed with traditional basals, current curriculum series mix both literature passages and decodable passages (Foorman et al., 2004; Hiebert, 2005). This change in theoretical orientation of texts could have significant implications for *curriculum* based measurements.

Hiebert and Fisher (2002) examined the differences presented in each of the two major text types, traditional basal and literature-based. They found significant differences in the number of unique words presented in each series, with literature-based materials containing considerably more unique words than the traditional basal. Literature based texts were also observed to contain drastically less decodable words and high frequency words than the controlled traditional basal series. Finally, literature based series included significantly more difficult words and "singletons" (words only presented once throughout a passage). Hiebert and Fisher (2002) concluded that the higher repetition and more controlled texts of the traditional basal allowed for more practice with frequently

occurring words and letter patterns, whereas the uncontrolled texts of the literature based series lacked this repetition and control. These differences could significantly impact fluency development with struggling readers as more exposure to the literature based texts would be required to develop proficiency with the many uncontrolled words within the curriculum (Hiebert & Fisher, 2002).

With text differences in mind, researchers began examining the effects that these differences would have on CBM measures. Several researchers examined student performance as measured with CBM ORF probes developed from traditional basal passages and passages developed from literature based curricula (Bradley-Klug, Shapiro, Lutz, & DuPaul, 1998; Fuchs & Deno, 1991; Hintze & Shapiro, 1997; Hintze, Shapiro, & Lutz, 1994; Powell-Smith & Bradley- Klug, 2001). With the exception of Hintze et al. (1994) results indicated that probes developed from either curriculum would produce reliable estimates of growth as students demonstrated similar growth slopes on either probe type (Bradley-Klug, et al., 1998; Fuchs & Deno, 1991; Hintze & Shapiro, 1997; Powell-Smith & Bradley-Klug, 2001). Contrary to these results Hintze, et al. (1994) observed significant differences in student performance on CBM probes developed from literature-based and traditional basal series. Specifically, student performance on literature based probes tended to decrease over time while performance on traditional basal probes demonstrated growth over time. However, all studies indicated differential absolute rates of performance with the preponderance of data indicating that students performed at a higher level (more words correct per minute) with probes selected from the traditional basal than from the literature based series. Therefore, while type of text may not matter for measuring the rate of growth, it may matter when overall levels of

student performance are compared to benchmarks such as those proposed by Hasbrouck and Tindal (1992). One likely explanation for these findings is that the traditional basals contain a more stringent vocabulary control which results in increases performance across time as students are repeatedly exposed to these controlled words.

There are several issues that are of concern with these studies. First, of all the "traditional basals" utilized in these studies are ill defined. "Traditional basals" have taken many forms over the years including the "traditional" decodable series of the 1960's and the "traditional" vocabulary-controlled series of the 1940's (Hiebert, 1999). These texts approach reading from different perspectives and do not all share similar text characteristics. In these studies, curriculum distinctions were made on the basis of publisher's reports of theoretical orientation. In other words, the studies took curriculum which were identified by the publishers as "literature based" and compared them with any variety of "traditional basals" which had been abandoned in favor of the new literature-based series. Furthermore, while these studies purported to evaluate the differential effects of traditional basals and literature based series on reading fluency, none of the studies directly measure how, and if, important text characteristics varied between the two types of curricula. Major limitations result from these poor definitional issues and subsequently reduce the applicability of these findings.

A further examination of the effects of various texts characteristics on CBM outcomes can be completed by examining passage difficulty. Passage difficulty is generally characterized as the relation between the grade-level from which materials are drawn and the students' grade level. When passages are drawn from the students' grade level then it is considered less difficult, whereas materials drawn from higher grade level

materials are considered more difficult. Initially, Deno (1985) presented the notion that one passage level (i.e. third grade) should be utilized for all students across grades in order to track growth across grades and to facilitate comparisons between grades; however, this was never enacted. Over the years several researchers have set out to determine the optimal difficulty level for progress monitoring. Three studies examined the use of goal-, or challenge-level materials to monitor students' growth (Shinn, Gleason, & Tindal, 1989; Hintze, Daly & Shapiro, 1998; Dunn & Eckert, 2002). The findings of these articles present somewhat conflicting information. Two studies (Shinn et al., 1989; Dunn & Eckert, 2002) indicated passage difficulty did not significantly impact the slope of growth, however, Hintze et al. (1998) found that at the lower grades slope may be impacted by passage difficulty, as passages above students' level were not as sensitive to growth as passages on the students' grade level.

The impact of two other text characteristics, decodability and percentage of high-frequency words, have yet to be examined in relation to CBM outcomes. Decodability is defined by two factors, (a) phonetic regularity of the words, and (1) the match between letter/sound relationships occurring in the text and that which has been taught in the classroom (Allington & Woodside-Jiron, 1998; Mesmer, 2005). Limited research exists relating the positive impact of practice within decodable texts to increased student reading outcomes. Research has also established that limited exposure and practice with appropriate texts, such as decodable texts, impedes automaticity and word recognition speed which are essential for fluent reading (Stanovich, 1992). Extensive debate continues as to the necessity of decodable texts within the classroom (Allington, 1997; Beck, 1997), however, several states have mandated that decodable texts be included in

early reading curriculum (Hiebert, 2000). Seeing that research has reported the instructional utility of practice within decodable texts it is likely that student performance within such texts will be significantly different than that which is displayed in uncontrolled texts, however, this notion has not been tested.

Practice with and instruction in high-frequency words has been advocated as one means of improving reading fluency (Cunningham & Allington, 2003). Several high-frequency word lists are currently employed for both instructional purposes (i.e. Dolch Basic Sight Vocabulary; Dolch, 1948) and research purposes (i.e. Word Zones; Hiebert, 2005). The notion behind high-frequency word instruction and the few existing high-frequency texts is that repeated exposure and practice will help students to develop proficiency with these commonly occurring words, subsequently increasing automatic word recognition, and ultimately enhancing both reading fluency and comprehension (Hiebert, 1998). Current research has established that practice within high frequency texts produces beneficial results for struggling readers (Hiebert, 2005); however, no direct study has measured students' performance within high frequency texts.

To date, two studies (Hiebert & Fisher, 2002; Compton, Appleton, & Hosp, 2004) have provided preliminary indications that passage decodability and percentage of high-frequency words may significantly impact student oral reading fluency. Hiebert and Fisher (2002) examined the combined effect of passage decodability and percentage of high-frequency words on reading fluency, accuracy, and comprehension. Their results indicated that students display higher levels of fluency, accuracy, and comprehension within passages that had higher levels of both decodability and high-frequency words. Similarly, Compton, Appleton, and Hosp (2004) correlated student accuracy and fluency

outcomes with the decodability and percentage of high frequency words within the passages. Their findings revealed that both passage decodability and percent of high frequency words were significantly correlated with both student accuracy and fluency. These two studies indicate that text characteristics may significantly impact student outcomes as measured by CBM.

Summary

To date there remains considerable debate over appropriate identification of reading disabilities. One promising approach to identification is the responsiveness to intervention approach which utilizes curriculum-based measurement for monitoring student growth in response to empirically validated interventions. CBM has a long-established history of validity and reliability; however, several recent changes in the curricula used in the nations' schools have given rise to questions of the impact of text characteristics on reading outcomes. A handful of studies have examined the impact of various text characteristics on reading fluency. Initial studies proposed that the theoretical orientation of curriculum (i.e. literature-based or traditional basal) does not impact slope but does influence absolute reading level. Research on the difficult level of probes is less conclusive with studies varying in their findings. However, studies of these two aspects of text failed to evaluate important text features. Resent research has indicated that text characteristics such as decodability and high-frequency words may significantly impact student performance. These differences may be even more salient with struggling readers. However, these two theories have not been fully evaluated. If curriculum-based

assessments are to be used in conjunction with RTI to make educational decisions, it is imperative that the effects of characteristics of the texts are understood given the many changes in curriculum that have taken place in recent years.

Statement of Purpose

The purpose of this study is to investigate the effects of text characteristics, specifically text decodability and high frequency words, on oral reading fluency curriculum-based measurement across time in a group of at-risk readers. Specifically, participants' oral reading fluency will be monitored with two different types of CBM: highly decodable texts and a passage with a high percentage of high-frequency words. The proposed study will attempt to determine whether a difference in oral reading fluency rate and growth rate exists as a function of text characteristics.

Study Rationale

Under the new Individuals with Disabilities Education Improvement Act (IDEIA) states now have the option of implementing responsiveness to intervention procedures to identify students as learning disabled (U.S. Department of Education, 2005). The most commonly employed method for measurement in the RTI procedures is curriculum based measurement (CBM), and specifically for reading, oral reading fluency CBM (Gresham, 2002). While CBM has a long-standing history of reliability and validity (Fuchs, Fuchs, & Maxwell, 1988), recent studies have brought into question various curriculum

variables which might impact fluency outcomes (Compton, Appleton, & Hosp, 2002; Hiebert & Fisher, 2002). Specifically, these initial studies indicate that decodability and percentage of high-frequency words may impact student performance on CBM ORF measures. Given that educational decisions can now be made based on CBM ORF outcomes, it is important that educators and examiners understand those curriculum variables which may significantly impact student outcomes to circumvent faulty decision making.

CHAPTER II

LITERATURE REVIEW

Introduction

This chapter focuses on a review of related literature following the major areas that support the research questions guiding this study. This review will include a brief background of the field of learning disabilities, followed by a discussion of the discrepancy model and response to intervention model for identification of learning disabilities. The review will continue with a discussion of measurement systems, primarily curriculum-based measurement and the variables that affect the outcomes. Finally the chapter will conclude with a discussion of certain variables that have been acknowledged as having a significant impact on reading fluency outcomes as demonstrated by curriculum based measurement, and those which may possibly have an impact on oral reading fluency outcomes.

History of Learning Disabilities and the Discrepancy Model

In 1963 the term *learning disabilities* was formally recognized as a descriptive title for children who demonstrated difficulties in skill acquisition in a particular area

while demonstrating proficiency in other areas. Kirk (1962), who is generally recognized as the originator of the term (Hallahan & Mercer, 2002), defined learning disabilities as follows:

> A learning disability refers to a retardation, disorder, or delay in development in one or more of the processes of speech, language, reading, writing, arithmetic, or other school subject resulting from a psychological handicap caused by a possible cerebral dysfunction and/or emotional or behavioral disturbances. It is not the result of mental retardation, sensory deprivation, or cultural and instructional factors (Kirk, 1962, p. 263).

Following Kirk's work, advocacy groups succeeded in pushing for the formal recognition of learning disabilities. The Learning Disabilities Act and subsequently the passage of Public Law 94-142 in 1975 required all states to provide appropriate services for children with learning disabilities. Shortly thereafter the Office of Education developed criteria for the identification of learning disabilities utilizing the Discrepancy Model (Weintraub, 2005).

Today the most widely used and accepted definition of a learning disability in reading is that which is associated with PL 94-142 (Frankenberger & Fronzaglio, 1991):

> The term 'specific learning disability' means a disorder in one or more of the psychological processes involved in understanding or in using language spoken or written, which may manifest itself in an imperfect ability to listen, speak, read, write, spell, or to do mathematical calculations. The term includes such conditions as perceptual handicaps, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. The term does not include children who have learning disabilities which are primarily the result of visual, hearing, or motor handicaps, or mental retardation, or emotional disturbance, or of environmental, cultural, or economic disadvantage (U.S. Office of Education, 1977, p. 65083).

Furthermore, PL 94-142 states that a student has a specific learning disability if:

> (1) The child does not achieve commensurate with his or her age and ability levels in one or more of in one or more of the areas listed in paragraph (a) (2) of this section, when provided with learning experiences appropriate for the child's age and ability level, and

(2) The team finds that the child has a severe discrepancy between achievement and intellectual ability in one or more of the following areas:
    i) Oral expression
    ii) Listening comprehension
    iii) Written expression
    iv) Basic reading skills
    v) Reading comprehension
    vi) Mathematical calculation
    vii) Mathematics reasoning. (U.S. Office of Education, 1977; p. 65083)

Ultimately, the major components involved in the definition presented by PL 94-142 involve the concepts of disorder in psychological processes, difficulty in learning, the elimination of other disabilities as a causal factor, and most influentially, the existence of a severe discrepancy between potential and achievement (Lerner, 1988).

Although other definitions have been proposed the discrepancy model remains the most widely utilized despite its frequent criticism in the literature (Hallahan & Mercer, 2002). Researchers have devised numerous methods for operationalizing underachievement and determining a "severe" discrepancy between aptitude and achievement (Mellard, Deshler, & Barth, 2004). Generally, each of the various formulas represented an attempt to address the shortcomings of this model

Opposition to the discrepancy model has existed since its inception (Hallahan & Mercer, 2002). Dialogues detailing the faults of the discrepancy model have been discussed thoroughly over the last twenty-five years with many authors citing the poor definition of learning disabilities as one of the main causes of these difficulties (Keogh, 2005).

Hallahan and Mercer (2002) identified four major concerns with the discrepancy model. The first concern identified is that there are flaws inherent in the original studies

on which the discrepancy model was developed. Numerous authors (e.g. Van der Wissel & Zegers, 1985; Vellutino, Scanlon, & Lyon, 2000; Share, McGee, McKenzie, Williams, & Silva, 1987) argue that data from the original studies on which the discrepancy model was fashioned, chiefly that of Rutter & Yule (1975) and Yule, Rutter, Berger & Thompson (1974), were inherently flawed leading to inflated results, and consequently, flawed conclusions.

The second concern with the discrepancy model surrounds issues of the likelihood of the Matthew effect (Siegel, 1989; Stanovich, 1986). The Matthew effect is the phenomena whereby over time skilled individuals become more competent while the unskilled fall further and further behind (Stanovich, 1986). The concern is that good readers will have more exposure to the world, as a result of their increased reading, thereby possibly inflating and overestimating their IQ scores while the IQs of poor readers will be underestimated due to their underexposure to the world (Siegel, 1989; Vellutino, Scalon, & Lyon, 2000). Siegel (1989) noted that many intelligence tests evaluate acquired knowledge or cognitive abilities that can either be adversely affected by reading ability or adversely affect this ability. This compounds the issues already at hand and makes identification with a discrepancy model all the more fallible.

The third concern presented by Hallahan and Mercer (2002) is that the discrepancy model is not appropriate for identifying learning problems among children in early elementary grades (Fletcher et al., 1998; Shaywitz et al. 1992). The concern here is that children must demonstrate a sizeable discrepancy before qualifying as learning disabled, and as such, students in the lower grades frequently have not had the opportunity to be exposed to an adequate amount of information to show a discrepancy.

As noted by Fletcher et al., (1998) students "must stay at the floor of the achievement test to achieve at sufficiently low levels to be defined as having learning disabilities" (p. 201). Furthermore, there is evidence that discrepancy scores at early ages are not reliable over time and have weak predictive validity (Shaywitz et al., 1992). Under the discrepancy model the average age for identification is 10 years of age. Unfortunately, research has shown that students identified after 8 years of age are less likely to benefit from remediation (Fletcher et al., 1998). In a related manner, research has also found that early diagnosis of learning disabilities was linked with superior reading outcomes after two years regardless of the amount of intervention received (Muehl & Forell, 1973). Overall, this difficulty with making decisions for children at lower grades is a significant problem with long-term ramifications.

The final concern surrounds the issue that research has been unable to successfully distinguish between students with a discrepancy and those who are low achievers but not discrepant on measures of core reading behaviors, such as phonological awareness (Fletcher et al., 1994; Fletcher et al., 1998; Foorman, Francis, Fletcher, & Lynn, 1996; Francis et al., 1996; Stanovich & Siegel, 1994). Fletcher et al. (1998) found that performance on eight cognitive ability composites hypothetically related to reading disabilities (e.g. phonological awareness, speech production, verbal short-term memory, etc) did not differentiate among groups of struggling readers, (those who displayed low-achievement but no discrepancy, those who were discrepant using a regression method, and those who were discrepant using regression and standard score methods). In general, the severity of reading difficulties and the pattern of differences among groups were undifferentiated among the groups of students with reading difficulties (Fletcher et al.,

1998). Similarly, Stanovich and Siegel (1994) found that children who demonstrated a discrepancy between reading achievement and IQ and those who did not display a discrepancy did not differ on tasks of phonological or orthographic processing. In a replication study of Stanovich and Siegel's (1994) work Foorman, Francis, Fletcher, and Lynn (1996) further substantiated these claims. Fletcher et al. (1994) found, as Share and Silva (1986) had previously proposed, that it was possible to reliably define a group of children with reading difficulties, however once that distinction is made, it is not possible to reliably separate students on the basis of performance on cognitive processes related to reading.

Furthermore, Vellutino et al. (1996) made the important observation that the discrepancy approach does not screen out children whose underachievement might be the result of limited or ineffective reading instruction. Finally, when an IQ-achievement discrepancy approach is utilized the resulting representation of functioning is decontextualized, time-limited, and ultimately only presented a snapshot of discrete functioning (Vellutino, 1996). However, the most critical shortcoming, as argued by several authors (Reschly, 1988; Fuchs & Fuchs, 1997; Share, McGee, & Silva, 1989) is that it fails to provide adequate information regarding appropriate interventions that should be developed for students.

Overall, these arguments present a strong case against the use of aptitude and achievement discrepancy approaches for identifying learning disabilities and even challenges the theoretical foundation which purports that there is a distinction between poor readers and learning disabled readers who exhibit a discrepancy between IQ and achievement. This resistance to the discrepancy model, which continues to persists even

today (Adhearn, 2003), has resulted in a broadening of identification criteria to now include the notion of responsiveness to intervention.

Response to Intervention

Amendments to the Individuals with Disabilities Education Act (IDEA; now called the Individuals with Disabilities Education Improvement Act of 2004 [IDEIA]) states that "a local educational agency shall not be required to take into consideration whether a child has a severe discrepancy between achievement and intellectual ability…In determining whether a child has a specific learning disability, a local educational agency may use a process that determines if the child responds to scientific, research-based intervention as part of the evaluation procedures" (U.S. Department of Education, 2005). In other words, the new regulations allow for local education agencies (LEA's) to replace the traditional ability-achievement discrepancy assessments with a more formative, longitudinal process that evaluates student's response to evidence-based practices as a means of determining a specific learning disability (Kovaleski, 2005). Although the specific term *response to intervention* is not found in IDEIA the concept is unmistakably present.

While the term response to intervention is relatively new, Heller, Holtzman, and Messick (1982) presented the idea over twenty years ago. Gresham (1991) and Fuchs and Fuchs (1998) furthered the ideas presented by these initial authors and since then research in the area has grown exponentially. With the reauthorization of IDEIA, and its inclusion of RTI language, the field is bound to see a proliferation of the topic in the literature.

Generally, responsiveness to intervention (RTI) is simply defined as the extent to which a change in behavior or performance ensues as the result of an intervention (Gresham, 1991). As applied to identification procedures, Deshler, Mellard, Tollefson, & Byrd (2005) define RTI as "individual, comprehensive student-centered assessment models that apply a problem-solving framework to identify and address a student's learning difficulties" (Deshler, Mellard, Tollefson, & Byrd, 2005).

Essentially, RTI utilizes a discrepancy-based approach, however, unlike the discrepancy formulas typically employed with IQ and achievement, the discrepancy is between levels of performance pre- and post-intervention (Gresham, 2001). Students are identified as having a learning disability when their response to validated instruction is dramatically less than that of their peers (Fuchs, Fuchs, & Compton, 2004; Gresham, 2001). This picture of unresponsiveness is sometimes referred to as *resistance to intervention* (Gresham, 2001). It is important to note that in attempting to determine responsiveness, or resistance, the interventions utilized must have been implemented with integrity and also have been found to be generally effective within the larger peer group.

One of the basic assumptions behind RTI is that it can appropriately differentiate between underachievement as the result of poor instruction or as the result of a disability (Fuchs, Fuchs, & Compton, 2004). If a child does not respond to intensive instruction, which benefits the majority of students, then poor instruction is eliminated as the explanation for low achievement and a diagnosis of a learning disability may be appropriate. However, the RTI model does not serve primarily as a means of identifying internal cognitive deficits, but rather focuses on the remediation of difficulties through increasingly intense instruction (Fletcher, Coulter, Reschly, & Vaughn, 2004). Utilizing

the RTI approach transfers the focus from deciding eligibility to providing effective instruction. Conceptualizing the identification and severity of disorders in terms of a student's responsiveness to intervention takes the focus away from the within-child view of a problems (MacMillian & Speece, 1999) and focuses on the contribution of the environment and instruction to which a student is exposed (Gresham, 1991). In utilizing the RTI method, educators would approach identification of students from a "risk" perspective, whereby large numbers of students who were at risk for significant academic problems would receive interventions and those whose academic difficulties persisted despite increased intervention support would be identified as having a learning disability.

In a report prepared by the National Joint Committee on Learning Disabilities (2005) three core concepts of RTI are identified. The first of these is the application of scientific, research-based interventions that are implemented in general education. The second core concept is the measurement of a student's response to the interventions. Finally, imperative in the RTI process is the use of data gathered from the interventions to inform subsequent instruction. To determine response to intervention educators must first provide early intervention, determine and match instruction to student's academic needs, and use ongoing data-based decision making with progress monitoring (Vaughn, Linan-Thompson, & Hickman, 2003).

In keeping with the regulations set forth by IDEIA, the notion of "scientific, research-based intervention" is of particular importance in the discussion of RTI (Danielson, Doolittle, & Bradley, 2005). There are several models of validated intervention which might be utilized in the response to intervention approach to

identifying learning disabilities. The most frequently employed approaches are: predictor-criterion models, applied behavior analytic models, and dual-discrepancy models.

In brief, predictor-criterion models (e.g. Vellutino, 1996) use and teach those skills that best predict reading performance, such as phonemic awareness and word recognition. Problems with this approach arise due to the fact that this model is intended to "normalize" specific reading skills rather than identify those individuals who inadequately respond to validated intervention (Gresham, 2002).

The applied behavior analytic model (e.g. Daly & Martens, 1994) focuses on manipulating the antecedents and consequences of environmental events to improve reading competence (e.g. opportunities to respond, reinforcement of accurate responses). This approach has been validated as effective intervention for remediating reading difficulties; however, the use of the ABA approach for eligibility determination creates some measurement challenges because of its reliance on single-case experimental design data (Gresham, 2002).

Dual-discrepancy models (e.g. Fuchs & Fuchs, 1998) measure a student's responsiveness to intervention. Generally, the general education teacher delivers interventions in the general education classroom. The most frequently employed measurement system utilized with interventions developed for the dual-discrepancy approach is curriculum-based measurement (Gresham, 2002). This approach fits well with the RTI model and has been the most frequently discussed intervention approach utilized with RTI.

While there are many approaches to RTI implementation they all stress improved results for students and follow the philosophy that services should be needs driven (Van

Der Heyden, 2006). Furthermore, the effects of these needs-driven interventions should be evidenced in improvements in "direct indicators of child learning" (Van Der Heyden, 2006, p. 18) rather than solely in decontextualized measures such as standardized assessments. Of importance to this philosophy is the definition of problems and use of assessments which will identify solutions effective at remediating the identified problem. In this manner, the way in which problems are identified should also point to the way in which they should be resolved (Van Der Heyden, 2006).

Fuchs and Fuchs (1997) presented an influential conceptualization of the response to intervention model for eligibility determination. In their discourses the authors presented a dual discrepancy model that emphasized the importance of treatment validity and curriculum based measurement as a means of progress monitoring. Treatment validity is "the degree to which assessment is shown to contribute to beneficial treatment outcome" (Hayes, Nelson, & Jarrett, 1987; p. 963) and is based on its capacity to inform, cultivate, and detail treatment effectiveness (Fuchs & Fuchs, 1997). Hayes, Nelson, & Jarrett (1987) argue that measures utilized for eligibility decision making should possess treatment validity. Without treatment validity eligibility decisions have no utility beyond acknowledging the presence of a problem. Treatment validity is a central tenet in RTI (Gresham, 2002) and when curriculum based measurement is used to quantify progress, RTI has been found to demonstrate effectiveness in modeling academic growth (Fuchs & Fuchs, 1998), discriminating between insufficient instruction and unsatisfactory rates of learning (Fuchs, Fuchs, & Compton, 2004), contributing to improved instructional decisions (Fuchs, Fuchs, & Compton, 2004), and sensitivity to growth for monitoring treatment effects (Fuchs & Fuchs, 1999).

Central to the notion of RTI are the concepts of continual progress monitoring to ascertain growth, or lack there of, in response to validated interventions and the use of that data to inform future practice. The most frequently employed, and most well researched, program for progress monitoring and data-based decision making in instruction is curriculum based measurement (Fuchs & Fuchs, 1999).

Curriculum-Based Measurement

Measurement of student's academic progress and general achievement has been at the forefront of educational interest for numerous years and is even more relevant now in the discussions of responsiveness to intervention. This interest in evaluating growth has resulted in many different approaches for assessment.

At one point, informal measures, such as teacher observation and scores on daily assignments, were the common form of measurement (Salmon-Cox, 1981). Gradually these methods fell out of favor as the primary means for assessing growth due to the fact that teachers' judgments were frequently found to be flawed (Coladarci, 1992; Fuchs & Fuchs, 1984; Madelaine & Wheldall, 2005).

In the 1960's assessment moved to more formal techniques, such as mastery measurement procedures, where specific skill hierarchies and related objectives were identified and used as the means for determining growth (Fuchs & Deno, 1991). This approach was deemed inadequate for progress monitoring on account of problems with measurement shifts, limits in determining broad growth, and a restricted instructional focus, among other things (Fuchs & Deno, 1991).

Emerging from this period came the dependence on broader measures of achievement, specifically, commercially developed, standardized, and norm-referenced tests. In time these too were determined unsound for monitoring progress. The appropriateness of such measures has been highly criticized over the years for numerous reasons including futility for making instructional decisions (Salmon-Cox, 1981), infrequent measurement and limits on increasing the frequency of administration (Freeman, Kuhs, Knappen, & Porter, 1982), and incongruence between test and curriculum content (Jenkins & Pany, 1978).

A movement towards assessments based in the curriculum with focused concentration on the direct assessment of academic skills began in the 1980's (Fuchs & Deno, 1991). As defined by Tucker (1985) this new technique of curriculum-based assessment was a system where assessments were drawn directly from curriculum and administered in equivalent forms across a school year for the purpose of informing educational decisions. Several versions of curriculum based measurement have been developed over the years (Blankenship, 1985; Deno, 1985; Gickling, & Thompson, 1985; Howell, Fox, & Moorehead, 1993), nevertheless they all have in common the assessment of that which is directly taught in the classroom (Shapiro, 1996).

Fuchs and Deno (1991) demarcated two different forms of curriculum-based assessments: (a) the specific subskill mastery measurement and (b) the general outcomes measurements. Specific subskill mastery measurements CBMs are equivalent to the mastery measurements of the 1960s, and consequently, possess the same drawbacks as the aforementioned system. On the other hand, general outcome measurement CBMs are broader in perspective as it targets the assessment of growth over time rather than solely

focusing on the mastery of specific skills. General outcome measurement CBMs assess overall outcome goals and rely on "standardized, prescriptive measurement methodology that produces critical indicators of performance" (Fuchs and Deno, 1991, p.493). It is these types of CBM which have persisted and are found repeatedly throughout research literature.

In 1985, Deno published a landmark article delineating a system for continuous evaluation of students' progress toward educational goals. Although the concepts behind CBM were not new (Tucker, 1985), Deno (1985) began the process of determining its technical adequacy and provided empirical evidence of the advantages of CBM. The goal in development of the CBM procedures was to obtain a system which was: (a) reliable and valid, (b) uncomplicated and efficient, (c) easily understood, and (d) economical to administer (Deno, 1985; Deno, Fuchs, Marston, & Shinn, 2001). As described by Deno (1985) CBM meets all of these criteria and can be conceptualized as a curriculum-referenced, individually-referenced, and peer-referenced system which allows for continuous measurement of students achievement. Deno's model has received considerable attention and has the most substantial research base of all CBM methods (Shapiro, 1996), and as such, it has been frequently utilized in research studies.

Modern Curriculum-Based Measurement

CBM as conceptualized by Deno (1985) and furthered by others (Fuchs & Fuchs, 1984, 1986; Shinn et al., 1989) is a standardized measurement system that allows for frequent measurement of students' performance and progress on various academic skills

within a school's curriculum. The CBM procedure requires that one first identify measurement tasks that are representative of cumulative end of the year performance within an academic domain. Two means for achieving this goal have been identified (Fuchs, 2004) and will be discussed in subsequent sections of this paper.

Regardless of the measurement task that is chosen, CBM probes are administered to the students at various times throughout the school year. A student's performance on a CBM measure at any one point is reflective of current achievement in relation to overall curricular outcome goals in that academic domain (Fuchs and Deno, 1991). When these scores are compared over time an educator is able to assess both the academic growth of the individual and the student's progression toward the cumulative goals of the curriculum (Fuchs, Fuchs, Hamlett, and Stecker, 1991).

There are several techniques for determining the CBM task. The first method of task identification is to take the various academic domains (e.g. reading, math, spelling, and written expression) and identify the different skills which constitute that year's curriculum (e.g. single digit multiplication and simple digit division). These skills are then complied on individual subject probes and parallel probes are developed for subsequent administrations (e.g., Fuchs, Fuchs, Hamlett, Phillips, Karns, & Dutka, 1997; Fuchs, Fuchs, Hamlett, & Stecker, 1991). This is different than the mastery measurements of former times as the mastery measurements assessed only one skill at a time, whereas with CBM, students are presented with all tasks simultaneously. However, this form of CBM is similar to mastery measurements in that it can be used to guide instructional planning for educators by identifying skills which need subsequent reinforcement and those that have been mastered by a class (Marston, 1989).

The second approach for identifying CBM measurement tasks is to isolate a task which can be shown to "correlate robustly" with criterion measures or with the component skills of the academic domain, (Fuchs, 2004). This second approach has been applied extensively in the area of reading.

Deno, Mirkin, and Chiang (1982) identified three different procedures for measuring reading which were found to be closely related to performance of standardized reading tests: reading aloud from basal readers, reading aloud from lists of words, and cloze procedures (a procedure in which students supply words that had been deleted from passages). Validity coefficients were found to be high and reliable across all three measures with correlation coefficients between .73 and .91 with most coefficients in the .80s. The three measures identified were all found to be more highly correlated with performance on standardized word recognition tasks than comprehension measures; however, correlation between the three tasks and comprehension measures were still statistically significant (Deno, Mirkin, and Chiang, 1982). The authors concluded that "[s]ince the validity coefficients for measures based on these behaviors are all high and reliable, data on any one, or combination, of these behaviors can be used to estimate proficiency in both decoding and comprehension" (Deno, Mirkin, & Chiang, 1982; p. 43). Other researchers have validated these findings as well (e.g. Fuchs, Fuchs, & Maxwell, 1988; Shinn & Good, 1992).

Currently, the most common CBM method employed for measuring reading ability, and the one which has been most frequently used in research, is oral reading fluency. Broadly defined, reading fluency is the ability to read texts with speed, accuracy, and expression, and is a critical component in comprehension (National Institute of Child

Health & Human Development, 2000). In terms of CBM, oral reading fluency examines only the speed and accuracy of oral reading and does not include a direct measure of expression. However, research has established ORF as an appropriate estimate of overall student reading ability (Good & Jefferson, 1998).

Oral reading fluency (ORF) CBM are completed by having students read aloud from grade-level text passages for a predetermined time while the examiner records total words read correctly (Good, Simmons, & Kame'enui, 2001). A student's oral reading fluency then is defined as the number of words read aloud and correctly per minute (i.e. WC/M). Studies examining the minimal time necessary for reliable results have examined reading times between 30 seconds and three minutes and have found reliable results at all three levels (Ardoin, et al., 2004; Deno, Mirkin, & Chiang, 1982). Most researchers utilize a one-minute interval for measuring oral reading fluency and if other interval lengths are employed, scores are generally transformed to indicate total words read correct per minute (e.g. Deno, Fuchs, Marston, & Shin, 2001; Deno, Marston, Shinn, & Tindal, 1983; Hintze, Callahan, & Matthews, 2002; Hosp & Fuchs, 2005).

Research examining the technical adequacy of oral reading fluency measures increased in the 1980s when CBM procedures began receiving increased attention. Fuchs, Fuchs, and Maxwell (1988) completed an extensive study of the criterion, construct, and concurrent validity of various informal reading comprehension measures, one of which was oral reading fluency. The results indicated that, despite some discussion of limited face validity, oral reading fluency is a valid correlate of comprehension (Fuchs, Fuchs, & Maxwell, 1988). Various other studies have also established reading fluency measures as predictive of performance on reading comprehension tasks (Fuchs, Fuchs, Hosp, &

Jenkins, 2001; Jenkins & Jewell, 1993; Markell, & Deno, 1997). CBM oral reading fluency measures have also been utilized to predict performance on state wide achievement tests with moderate to moderately strong correlations, and subsequently also to identify those students at risk of failing the state tests (Crawford, Tindal, & Stueber, 2001; McGlinchey & Hixson, 2004; Stage & Jacobsen, 2001; Wiley & Deno, 2005).

Oral reading fluency tasks have also been established as having strong discriminative validity for differentiating between students with reading disabilities, low achieving students who receive remedial services, and average achieving students who do not receive remedial services (Deno, Mirkin, & Chiang, 1982; Shinn, Tindal, Spira, & Marston, 1987; Marston & Magnusson, 1988; Deno, Marston, Shinn, & Tindal, 1983; Shinn & Marston, 1985; Shinn, Tindal, & Stein, 1988; Shinn, Ysseldyke, Deno, & Tindal, 1986). Additionally, some research has found that when oral reading fluency tasks are utilized to make eligibility decisions the resulting incidence figures closely parallel national averages (Germann & Tindal, 1985).

By and large, curriculum based measurement procedures have an extensive history of technical adequacy for evaluating and monitoring student achievement, making educational decisions, and aiding in instructional decision making. In terms of reading, oral fluency measures have been the traditional measurement for reading achievement and have demonstrated ample validity for continued use.

Variables Affecting CBM Outcomes

Although CBM has a longstanding history of validity and reliability, most of the founding studies evaluating CBMs applicability in reading were conducted in traditional basal series (Hiebert & Fisher, 2005). Widespread changes in reading curricula have resulted in questions regarding the effect of various text characteristics on fluency measures typically utilized in curriculum based measures. These questions are of particular concern if curriculum based measurements are going to continue to be developed from actual texts and used to make decisions regarding student performance and growth. In this section a discussion of various studies on the topic of the influence of curriculum and text variables on CBM outcomes will be examined and compared and contrasted.

Curriculum Type

Over the last century, and even within the last thirty years, extensive changes have occurred in reading curriculum. Curriculum in the early to mid-1900's, frequently referred to as traditional basals, generally controlled for vocabulary, emphasized isolated skills, and utilized generic stories which were designed for practice of particular skills (Koskinen, McCarthey, & Hoffman, 1995). In the early 1970's and 1980's basals moved away from the strict vocabulary controls and explicit skills instruction towards a heavy reliance on sight word instruction. The adoption of literature-based curriculum by the early-1990's placed emphasis on predictable text structures and engaging authentic

literature without decodability or vocabulary-control (Hoffman, Sailors, & Patterson, 2002). Most recently, several states have mandated that beginner texts include a high level of decodable words (Hiebert, 2002). These changes are of significance for the discussion of CBM outcomes and RTI.

Hiebert and Fisher (2005) examined the nature of these changes in curriculum and found several significant differences in each curriculum's approach to reading instruction. First, literature based curriculum tend to have more unique words per 100 words of running text with a ratio 35:100, while controlled vocabulary texts feature a ratio of about 25:100. Secondly, controlled vocabulary texts contained significantly more decodable words and high frequency words than the literature based curricula. The authors noted that this fact allowed for "increase[d] practice with a large number of words that students are expected to read in content area texts in the middle grades and high school" (Hiebert & Fisher, 2002). Other noted differences between the curricula were a higher percentage of unique words per 100, difficult words, and singletons (words only repeated once throughout a passage) in the literature based than in the traditional basal. Hiebert and Fisher (2002) concluded that these differences in text design may be critical in the development of struggling readers' fluency.

Various other studies have examined the effect of different types of curricula on the outcomes of curriculum based measurement. If educational decisions are to be made based on students' responsiveness to intervention as measured by performance on CBM tasks, it is imperative that educators and those involved in the decision making process be informed of the various aspects of CBM development and its usage as it pertains to outcomes measures.

It has been established that CBM reading fluency measures are strongly related with socially important, widely used criterion measures of reading (Fuchs, Fuchs, Maxwell, 1988). Thus far, two dimensions of reading curriculum have been inspected in relation to its impact on CBM. One of the salient dimensions of curriculum is the particular basal reading series used, especially in regards to the question of curriculum bias. Curriculum bias is the extent to which a particular curriculum series affects the validity or usefulness of measurement given that the curriculum used might affect the criterion validity of CBM. The second salient dimension of curriculum is the difficulty of the materials presented to the student.

The most widely recognized study focusing on the impact of curriculum in CBM is that of Fuchs and Deno (1991a). In this study the authors examined two dimensions of curriculum, difficulty and basal series, on the technical features of CBM in reading. Two types of measures were employed with study participants. Students who were instructed in a traditional skills-based curriculum were first administered a commercial, standardized reading achievement test, then each child read orally for one minute from 19 passages, one from each grade level of two reading series. The two curricula utilized in this study were the Ginn 720 which utilizes an eclectic approach to teaching reading, and the Scotts-Foresman which is a literature based series that employs strategies which stress comprehension. Results demonstrated that correlations between the oral reading samples and the test of reading comprehension were similar across difficulty levels and across the two different series. Growth rates detailing reading development also remained strong regardless of difficulty level and series. The authors concluded that "technical features of measurement may not be influenced in major ways by curriculum dimensions" (Fuchs &

Deno, 1991a, p. 241). The authors state, however, that the applicability of these findings for other curricula might be limited, especially for those with highly controlled vocabulary. Furthermore, this study utilized only students instructed in a traditional-basal series and generalizations to students instructed in a literature-based series cannot be made.

Hintze, Shapiro, and Lutz (1994) evaluated the effects of curriculum on the sensitivity of oral reading CBM. In this study, participants were drawn from two third grade classes, one which utilized a literature-based curriculum (Scotts Foresman) and one which utilized a more tradition basal series which highlights decoding/phonics, and vocabulary skill development (Houghton Mifflin). In this study students from the two different instructional classrooms read CBM probes developed from the third-grade passages of each basal series twice weekly over a 9-week period. Overall, statistical analyses indicated two significant main effects in the results, instructional type and probe type. Of importance for this review is the significance of probe type. Results from their study demonstrated that, probes selected from the literature-based basal series were less sensitive to indexing growth over time than those from the traditional basal series. Of concern is the fact that regardless of the series in which the students were instructed, data from the literature-based passages showed an overall decrease in reading fluency over time while data from the traditional-based passages showed a general increase. More specifically, students instructed in the literature-based series showed a decrease in words correct per minute (WCPM) of approximately one word per week while students instructed in the traditional series increased by roughly .35 words per week. The authors general conclusions indicated that “the type of measurement probe employed for progress

monitoring may indeed affect the data generated" (Hintze, Shapiro, & Lutz, 1994, p.196). This information is contrary to the findings of Fuchs and Deno (1991a) which suggested that type of curricula utilized for probe development is inconsequential in CBM.

Hintze and Shapiro (1997) presented a further attempt to clarify the findings of the previous studies. Their study focused on the "extent to which curriculum-based measurement (CBM) could be implemented in nonbasal reading curricula" (Hintze & Shapiro, 1997, p.351) and evaluated progress with challenging materials and the relation of passage readability to outcomes. All students were monitored with CBM passage probes developed from both curricula. Of importance is the fact that rather than utilizing passages at the students' instructional level, passages were demonstrative of "long-term *challenging*-level material" (Hintze & Shapiro, 1997, p. 358), this was defined as material one grade-level above students current grade level. The authors cited three advantages of using probes at the "challenging-level": (a) it is a better representation of ultimate performance, (b) correlation with global achievement tests is stronger than short term measures, and (c) it reduces the risk of a ceiling effect.

Results of Hintze and Shapiro (1997) indicated that regardless of the instructional reading series, students demonstrated positive growth on both types of reading probes, with the exception of second graders. Second-grade students from both categories of reading instruction indexed positive slopes as measured by the literature-based probes but displayed negative slopes when monitored with the traditional-basal. The results also demonstrated that measures utilizing the literature-based series evidenced significantly greater growth in comparison to probes developed from the traditional-basal series. However, both curricula were sensitive to changes in oral reading rate as a function of

time and instruction. Furthermore, results indicated that progress varied significantly as a function of grade level. However, unlike previous studies which found negatively decelerating growth slopes as grades progressed (Fuchs, Fuchs, Hamlett, Walz, & German, 1993; Marston & Magnusson, 1985), Hintze and Shapiro (1997) found a positively accelerating curve with a slight leveling off at the fifth grade. Finally, these results suggest that using materials which are at the "challenge-level" may produce differential effects for students at the lower grade levels (first and second grade) and students at the upper elementary grades (third through fifth grade). Specifically, probes developed at the challenging-level are less sensitive to student growth at the lower grade levels than materials derived from similar-grade level materials. Conversely, at the upper grades challenging-material probes continue to demonstrate sensitivity to growth. This finding may not be of significance if one is utilizing local norms for comparison of growth outcomes; however, if results are to be compared to empirically derived estimates of growth, as suggested by Fuchs et al (1993), students at the younger grades may be misidentified as unresponsive to instruction.

Bradley-Klug, Shapiro, Lutz, and DuPaul (1998) presented another study evaluating the impact of curriculum on CBM. This study investigated the utility of oral reading rate as a metric in monitoring students' progress over time when instruction was occurring in a literature-based curriculum. In addition, the effects of passage readability on oral reading performance were examined. CBM probes were selected from end-of-the-year goal material from the instructional curriculum and a traditional-basal series (Harcourt Brace Jovanich). Passages were evaluated and controlled for readability using the Fry readability index. Probes from both series were administered to students twice a

week for ten weeks. The results indicated that the oral reading fluency slopes of the students who read from traditional-basal and the literature-based series were not significantly different at the second or fifth grade. No significant correlations were found between passage readability and student performance. Overall, the authors found that curriculum probes developed from either literature-based or traditional-basal series were effective indexes of student growth over time. However, of importance is the fact that overall student performance (e.g. absolute reading rate) on literature-based measures tended to be significantly less than the mean level of performance on traditional-basal probes. Therefore, again, some caution may be needed when CBM data are used for educational decision-making as in this case literature-based probes may have actually underestimated the performance of students.

Powell-Smith and Bradley-Klug (2001) completed a follow up study on the analysis discussed above. Poor readers in the second-grade were drawn from four school districts which utilized traditional basal series (Scribner Reading Series and MacMillian Reading Program). Two forms of assessment passages were utilized in the study, probes from the curriculum of the school and passages from the *Test of Reading Fluency* (TORF). Each participant in the study was first screened with curriculum-probes and the TORF to determine each student's current reading instructional level. This information was then utilized to develop each student's long-term goal materials, the materials with which each student would be monitored. Twice a week for five weeks each student was monitored with probes developed from the basal series and the TORF. Results indicated that slopes for both probes were similar. This information indicates that either type of reading probe functioned equally well for progress monitoring. However, students'

reading performance was significantly higher on the TORF passages (about twelve more words correct on the TORF probes) indicating that probe type may influence ultimate outcome results. The relation between readability and oral reading fluency was found to be non-significant. Therefore, probes controlled for difficulty using readability formulas do not appear to be correlated significantly with student fluency. Overall, these findings support the suppositions of Hintze and Shapiro (1997) and Bradley-Klug, Shapiro, Lutz, and DuPaul (1998) and suggest that it is not essential for CBM reading materials to be pulled directly from the curriculum of instruction.

Overall, research examining the effects of literature-based and traditional basal curricula on oral reading fluency outcomes indicates that curriculum type does not differentially affect growth slopes. However, these articles demonstrate that there are differences in overall fluency rate associated with different curriculum types, as students generally performing higher on probes generated from traditional basals than on probes developed from literature-based series. These articles indicate that a number of questions remain unanswered relative to reading rates, most important for the current study, the effect of word-level text characteristics on reading outcomes. Furthermore, it is imperative that research goes beyond the simple comparison of growth as measured with curriculum based measurements developed from non-specific "traditional basal" and "literature series" to examining growth as a function of specific, measurable text variables.

Passage Difficulty

Recently the effects of passage difficulty on reading fluency rates have been addressed by a handful of research projects. Generally, two questions are addressed in these studies. The first concerns issues of the effect of passage readability on outcomes while the second surrounds issues of which level of material is most appropriate for assessment of growth through progress monitoring.

The effects of passage difficulty were first examined by Shinn, Gleason, and Tindal (1989). The progress of 30 students in middle-elementary and junior high was monitored in one of two different measurement conditions. The first condition included probes pulled from one level below and one level above the students' instructional placement. The second condition utilized probes developed from curriculum two and four levels above instructional placement. CBM reading data were collected for 4 days per week for 4 weeks. No significant differences were found in the slope of improvement as a function of difficulty level or the curriculum from which progress-monitoring reading probes were selected. Based on these findings, Shinn, Gleason, and Tindal (1989) suggest that goal level material (materials one year above current placement) should be utilized for progress monitoring. A significant limitation of the findings of these authors is that neither condition included materials from students' current grade level and information was based on students with formally diagnosed reading difficulties.

In 1998, Hintze, Daly, and Shapiro presented a discussion of the effect of difficulty level on oral reading fluency outcomes in progress monitoring as they evaluated the use of grade-level versus challenging- or goal-level materials. Participants

for this study were drawn from grades 1 through 4 in one elementary school. Grade- and goal-level CBM probes were developed from the instructional curriculum (e.g. *Silver, Burdett, & Ginn*). Goal-level material was defined as materials one year above current grade placement. For example, students in second grade would be evaluated with materials from the beginning of third grade. All probes were evaluated for readability. For a probe to be included in the grade-level series, the readability score had to be within the grade level for which the probe was designed. Results indicated that regardless of the level from which progress-monitoring material was drawn, students showed positive growth in both grade and goal level material on average. The study also revealed that the amount of progress that could be expected (i.e. slope of improvement) varied as a function of grade and difficulty level of the reading passages. Statistical analyses demonstrate that at grades 1 and 2 the level from which the CBM progress-monitoring material is selected has a significant role in observed student outcomes. However, at grades 3 and 4 differences become significantly reduced. Based on the conclusions of this study, Hintze, Daly, and Shapiro (1998) propose that students in the lower grades (grades 1 and 2) should be monitored with materials at their current grade placement in order obtain results which are sensitive to growth *and* with goal-level materials for greater outcomes measurement. In contrast, the authors propose that in grades 3 and 4 students should be assessed using materials at a goal-level as doing such would allow for estimates of oral reading fluency growth during longer periods of time and across reading levels.

Dunn and Eckert (2002) continued this line of research and directly compared similar and challenging-materials with participants controlled for reading ability. Twenty

students from grades 2 and 3 in one school were selected for participation in this study. All participants met entry criteria of demonstrating (a) a frustrational level with material from grades 3 and 4, and (b) an instructional level with second grade material. Selecting participants in this manner allowed for an examination of students who shared similar reading abilities but who were performing at grade level and below grade level. CBM materials were selected from the instructional curriculum (*Silver, Burdett, and Ginn*) at both the similar- and challenge-level. Similar-level materials were required to have a readability index between 2.5 and 3.0, while challenge-level materials fell between 3.5 and 4.0 according to the Spache Readability Formulas (Spache, 1953). The results of this study indicate that no differences were found in WCPM rates between average-achieving students in second grade and low-achieving students in third grade. However, significant differences in reading fluency levels were observed as a result of material type, with students performing better on similar level material than challenging material. Analyses of slope indicated no significant differences for type of measurement material, indicating that measures were equally sensitive to growth. The authors note that although each measure utilized in this study was sensitive, both types appear to have significant error associated with the data, suggesting that "progress monitoring over time using CBA accounts for very little systematic variance in student data" (Dunn & Eckert, 2002, p. 41).

Of these three studies which directly examined the effect of passage difficulty on CBM results, the preponderance of data suggests that while there may be differences in overall student reading outcomes, slope of improvement does not differ significantly CBM probes of various difficulty levels. In other words, students demonstrate positive growth on the various difficulty levels utilized in these studies. The one exception to this

finding was the provided by Hintze, Daly, and Shapiro (1998). They demonstrated that while this conclusion was true for students in third and fourth grade, the same could not be said for first and second graders. Results indicated that for students in grades one and two the difficulty level of the CBM probe *did* significantly impact the slope of growth, with greater growth (in terms of both the rate of acquisition and overall acquisition) evidenced in materials developed from grade-level materials.

Other Curriculum Variables

One tentative conclusion from the literature reviewed is that aspects of text, such as passage difficulty, may have greater impact on curriculum based measures of oral reading fluency with younger students. It is possible that as these readers develop early literacy skills, features of curriculum and curriculum measures may be more evident in influencing the reading performance of these students. Of concern then, are other curriculum differences such as passage decodability and percentage of high-frequency words, which could possibly impact early readers. Data related to these factors are severely lacking within current literature. To date only two studies have evaluated the effects of these text characteristics on reading fluency and accuracy, and only one study has examined these factors in relation to curriculum based measurement.

These textual characteristics are of particular importance given that reading instruction for struggling readers, who are more likely to enter into the RTI process, are typically completed within curriculum materials that stress such word controls (e.g. Engelmann & Bruner, 1978; Wilson, 1996; Herman, 1995). If educators and evaluators

are to select CBM fluency probes from the curriculum in which students are currently being instructed, there is the chance that selection of probes from controlled texts will occur. As such, it is important to determine the effects that such curriculum variables have on fluency outcomes, especially given that there has been some initial indication that such text characteristics do indeed impact outcomes. Furthermore, initial studies suggest that phonics knowledge and word recognition are precursors of fluency. These factors will be discussed in the following section

Decodability

Traditionally, the notion of determining text difficulty has been completed on the sentence level with measures of readability. However, although readability formulae provide useful information about text difficulty at the higher levels of difficulty, results for the earlier stages of reading are somewhat unstable. (Hiebert & Fisher, 2002). This is expected as word level differences are not as influential when students have passed the acquisition phases as generally occurs in the mid- to late-second grade. However, when students are still in the acquisition phase these issues may be important. Furthermore, struggling readers are typically in the acquisition phase longer than average readers. The notion of controlling for text by decodability, or word difficulty, is another means of determining the difficulty of a given text. Decodability is generally determined by two factors: phonetic regularity of the words and the match between those letter/sound relationships occurring in the text and that which has been taught in the classroom (Allington & Woodside-Jiron, 1998; Groff, 1999; Hiebert, 1999; Mesmer, 2005; Stein,

Johnson, & Gotlohn, 1999). Information regarding the degree of congruence between text and previous instruction, also known as lesson-to-text-match (Mesmer, 2005), is typically derived from estimates of instruction based on the teachers manual.

To determine the phonetic regularity of words Menon and Hiebert (1999) developed a system comprised of eight different levels of difficulty. Levels one through three are considered easily decodable and covers words up to the consonant-consonant-vowel-consonant-consonant (CCVCC) pattern. Levels four through seven are deemed within the moderate difficulty level, and include silent –e endings, double vowels, and diphthongs. These levels are often termed the Complex Vowel levels. At the highest level of difficulty, level eight, the MultiSyllabic level, consists of multisyllable words. Using Menon and Hiebert's (1999) method, words are assigned to the appropriate phonetic regularity level and text estimates of phonetic regularity are established by calculating the percentage of words at each of the eight decodability levels. Passages are contrasted based on its percentages at each level, with passages having higher percentages at higher decodability levels being considered less decodable than passages with higher percentages at lower decodability levels (Menon & Hiebert, 1999).

Extensive debate continues as to the necessity of decodable text within the classroom (Allington, 1997; Beck, 1997, Fletcher, Francis, & Foorman). Opponents to its use suggest that texts controlled for decodability are unnecessary for reading development, and that in actuality, the use of such texts will hamper students' growth (Allington, 1997; Moustafa, 1997). Conversely, proponents of decodable texts highlight the necessity of practice within decodable texts for strengthening the fundamental letter-sound connections which form the basis of our written language (Beck & Juel, 1995;

Fletcher, Francis, & Foorman, 1997). The theory supporting the use of decodable texts is that such texts exposes readers to words that follow relationships taught in phonics lessons, reinforces the phonics skills presented during phonics lessons, and provides them with reading practice in texts.

To date there are significant gaps in research on decodable texts, and definitive statements cannot be made regarding the optimal level of decodability that should be employed in reading texts (Beck, 1997). Most of the existing literature on decodable texts discusses text characteristics rather than empirical evaluations of its impact on reading (Hiebert, 2002).Despite the limited amount of research, state mandates in two influential states are now requiring beginning level texts to have high percentages of decodable words (Hiebert, 2000; Hoffman et al., 2002; Texas Education Agency, 2000). As decodable texts increase in prevalence in the schools so to does the likelihood that such texts will be employed for CBM development. With this possibility in mind it is essential that the impact of decodable texts on oral reading fluency are understood.

Two studies have directly examined the effects of decodable texts on reading development in isolation from other factors. Juel and Roper-Schneider's (1985) article is likely the most frequently referenced article addressing such issues. In this research endeavor the authors compared reading outcomes for two groups of students who were both instructed in the same phonics program but one of whom practiced reading within a phonics-based, decodable text, and the other that practiced within a text which had equal levels of decodable and non-decodable words. Results from this study revealed that the group of students who practiced within the decodable text developed significantly stronger decoding skills as applied to both nonsense words and decodable words for

which they received no instruction. Similarly, Jenkins et al. (2004) examined the effects of reading more or less decodable texts with students who received the same intensive phonics tutoring. Texts decodability varied from 85% in the highly decodable text to 11% in the less decodable texts. The findings of this study were contrary to the findings of Juel and Roper-Schneider, and indicated that "decodable texts do not add value to supplemental tutoring programs" (Jenkins et al., 2004; p. 81). Furthermore, the authors concluded that text differences do not always "power through" other instructional factors, and that the study "cannot, by itself, settle an instructional issues as complex as text decodability" (Jenkins et al., 2004; p. 81).

Other studies have examined decodable text in conjunction with other instructional factors. Torgenson et al. (1999) found that students who were tutored with explicit phonics instruction and who practiced decodable texts performed significantly better than students who received an implicit phonics program and practiced with less decodable texts. It is evident that in this situation the effects of decodable texts cannot be separated from the instructional intervention, yet one goal of this examination, as well as several others that have been conducted in this manner (e.g. Foorman, et al., 1998; D. Fuchs, et al. 2001; Vadasy, Jenkins, & Poor, 2000) is to closely coordinate instruction with practice texts to subsequently increase reading outcomes. Ultimately, this can be conceptualized as increasing the decodability of text in that there is a focus on lesson-to-text-match which is a major component of determining decodability. Together with the studies discussed above, these researchers present a discussion of whether practice within decodable texts produces differential growth. By extension, it can also be questioned whether oral reading fluency measurement materials that are highly decodable will

indicate fluency growth rates that differ from that which is observed when less decodable measures are utilized.

High Frequency Words

Imbedded in fluent reading is rapid recognition of words. It is generally regarded that in individual reading students should encounter no more that one unknown word for every ten known words in a passage to ensure that students continue to understand a passage (Rasinski, 1999). After this limit students begin to spend their energy attempting to decipher words rather than on processing the text to develop an understanding of the passage. Given that a mere 300 words accounts for 65% of all words in texts (Zeno et al. 1995), explicit training and practice with these words is likely to increase fluency (Singh & Singh, 1988). These words, typically called high frequency words, are those which occur repeatedly throughout texts, but which often have irregular letter-sound patterns and frequently lack meaning (Cunningham & Allington, 2003). The theoretical basis behind high frequency word instruction is that students will become automatic with recognition of whole words through repeated exposures to that word. In turn this would decrease the amount of time required to decipher the word thus allowing for more fluent reading, and subsequently increased comprehension (Hiebert, 1998).

Training and practice with high frequency words is one advocated means of increasing student fluency (Cunningham & Allington, 2003; Hiebert et al., 1998; Rashotte & Torgensen, 1985). The prevalence of this method is evidenced in the popularity of the Dolch High-Frequency word lists and the fact that some reading

materials systematically introduce high frequency words into their texts (Hiebert & Fisher, 2002). However, while students frequently practice with high frequency words, there is a dearth of information related to how students perform when they are *measured* with texts that include percentages of high frequency words which exceeds that of the typical 65 percent. In view of the fact that high frequency words are instructed in classrooms, whether through word walls, flash-card practice, or other activities, it is likely that students will perform differentially on texts which include high levels of these words. Menon and Hiebert (1999) state that "the acquisition of this core set of words (often taught as "sight words") is linked to children's ability to begin reading independently" (p. 5). Although high frequency texts are not currently prominent in schools, with the push for more controlled texts, it is likely that the percentages of high-frequency will increase within beginning reading curriculum.

Several lists of high frequency word lists have been presented over the years. Popular lists include Edward Fry's List of Instant Words (Fry, 1980) the Dolch Basic Sight Vocabulary of 220 Service Words (Dolch, 1949), and Zeno et al.'s (1995) Word Frequency Guide. Recently, Hiebert (2005) devised a system for rating texts in terms of high frequency levels based on earlier work by Zeno et al. (1995). There are six levels, or *zones,* in this system that is comprised of 5586 most frequent words. Those words that fall outside of these 5586 words are considered to be rare words. The high frequency rating is based on the first 100 words of a passage and is determined by the number of words within the passage that fall into each of the six zones. Zones 1 through 4 correspond to the curriculum of grades one through four.

Research has established that practice with texts that have high percentages of high frequency words is more beneficial for struggling readers than those texts which have a higher percentage of rare words (Hiebert, 2005). This hypothesis has been supported by studies which utilized different levels of word frequency control in conjunction with repeated readings (Rashotte & Torgenson, 1985; Faulkner & Levy, 1994; Dowhower, 1989). A meta-analysis complete by Hiebert and Fisher (2005) determined that initial studies which demonstrated the benefits of repeated and guided oral reading were based on texts which had a higher degree of vocabulary control. Subsequently, studies which utilized literature without vocabulary control did not demonstrate significant increases in fluency despite repeated practice (Hiebert & Fisher, 2005). At this point, it is evident that high frequency words are important to fluency development; however, the impact of this factor has not been studied in curriculum based measurement studies.

Effects of Decodability and High Frequency Words on Outcome Measures

Recently, researchers have begun to examine the effects of word-level text characteristics on student reading outcomes. Two innovative studies by Hiebert and Fisher (2002) and Compton, Appleton, and Hosp (2004) have examined the effects of decodability and high frequency words on fluency outcomes. However, in both of these studies the effects of these two word characteristics were combined, making analysis of the individual influence of these variables impossible. In spite of this dilemma, these

studies provide valuable information related to the effect of word-level text variables on fluency measures.

Hiebert and Fisher (2002) evaluated the effects of what they termed critical word factor (CWF) on students reading fluency, accuracy, and comprehension. The critical word factor analyzes the word recognition demands of texts as indicated by two text characteristic: (a) the match between high frequency and phonetically regular words with students' current stage of reading development and (b) the number of words that do not match with the students' current reading development. (See Menon & Hiebert (2005) for a quasi-experiment evaluating the validity and efficacy of the concept.) Using these criteria high- and low-CWF passages were selected from first grade texts with the same readabilities. High- and low-CWF was determined with the level of middle of first grade as the students' current reading development level. At this level, students are expected to have been exposed to and have gained proficiency with the 25 most frequent words and vowel patterns in consonant-vowel, vowel-consonant, and consonant-vowel-consonant words. Students were asked to read each of the four texts to the researcher and each student's speed and accuracy was recorded. Results indicated that texts with different CWFs significantly influenced students' performance in terms of both reading speed and reading accuracy. Specifically, higher reading speeds and accuracy were associated with the high CWF passages. These finding indicate that the match between students' current reading development and factors such as decodability and percentage of high-frequency words significantly influences students' reading. However, generalizations about the individual effects of decodability and high-frequency words cannot be made due to the fact that they were combined in this study.

Compton, Appleton, and Hosp (2004) presented another evaluation which assessed factors which influence second-grade students' reading accuracy and fluency within curriculum based measurement. Both low and average achieving students' performance on oral reading CBM probes were assessed and correlated with passage attributes including readability, decodability, percentage of high frequency words, percentage of multisyllabic words, and average sentence length. Readability of passages was determined with the use of both Flesch-Kincaid and Spache readability formulas. The Flesch-Kincaid readability formula uses the average number of syllables per word and number of words per sentence in its calculation of readability, while the Spache formula employs both of these factors as well as a leveled vocabulary list to calculate readability. Decodability of passages was assessed using the decodability system developed by Menon and Hiebert (1999). As previously discussed, this system evaluates the linguistic difficulty of the decoding pattern of a word and assigns a level between one and eight to the word based on its pattern. To determine the decodability of a passage, the percentage of decodable words per passage was calculated for each level of the decodability levels. To calculate the percentage of high frequency words, the 500 most frequently printed words (Zeno, Ivens, Millard, and Duvvuri, 1995) were identified in each passage and divided by the total number of words in that passage. Results from this evaluation indicated that students' fluency and accuracy was not significantly correlated with passage readability, percent of multisyllabic words, or sentence length. However, passage decodability was found to be highly associated with reading fluency, and percent of high frequency words was highly correlated with both reading fluency and accuracy

(Compton, Appleton, and Hosp, 2004). The findings of this study provide further evidence that student outcomes are differentially affected by text characteristics.

Text Variables and Reading Ability

Studies which have examined the impact of text-level variables on fluency are very limited in number, and yet there are even fewer studies which examined the affect of these factors on students of different reading abilities. Of six studies which have examined the impact of decodability and/or word frequency on the *development* of reading skills (Hiebert & Fisher, 2002; Hoffman, et al., 2002; Mesmer, 2005; Juel & Roper-Schneider, 1985; Compton, Appleton, & Hosp, 2004; Jenkins, et al., 2004) only three studies included students who were considered to be reading below expectancy levels (Juel & Roper-Schneider, 1985; Compton, Appleton, & Hosp, 2004; Jenkins, et al., 2004). These three studies developed different conclusions regarding the impact of text variables on development.

Juel & Roper-Schneider (1985) determined that practice with different types of texts early on in one's reading development differentially influence students word identification strategies. Compton et al. (2004) identified that student accuracy within texts was predicted by the percentage of high-frequency words in text for both average- and low-achieving students. However, they found that while decodability was correlated with accuracy and reading fluency for the average-achieving group, it was not significant for the low-achieving group. It is important to note that this study, conducted with second grade students, found a differential outcome for students with different ability levels.

Finally, Jenkins et al. (2004) determined that while students who had been exposed to extensive practice within highly decodable texts performed better on controlled texts the increase in proficiency did not translate into better performance on uncontrolled texts. Limited conclusions can be drawn from these studies partially due to the fact that these studies did not all examine these factors in the same way. Nevertheless, these studies demonstrate the veritable dearth of information related to these factors.

In view of the fact that there has been a renewed push to include decodable texts in curriculum (Foorman et al., 2004; Hoffman et al., 2002; Hiebert, Martin, & Menon, 2005) and that practice with high frequency words is advocated for struggling readers, the effects of these variables must be examined in relation to the fluency outcomes that are typically associated with CBM. This need for information is intensified by the inclusion of RTI criteria for determining learning disabilities.

Expected Growth Rates

Over the years researchers have proposed various estimates of expected reading growth. Generally, research has demonstrated a negatively accelerating curvilinear trend in reading acquisition rates across grades, where the slopes of students in the lower grades (first, second, and third grade) are steeper than those of students in higher grades (fourth, fifth, and sixth grade) (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, et al, 1993). Fuchs et al., (1993) suggest that this observation is congruent with developmental reading theory, as during the early elementary grades students make great gains in basic decoding and fluency that are not as evident in the upper grades. Table 1 presents a summary of the

generic growth estimates proposed by Fuchs et al., (1993) and Deno et al., (2001). A negatively accelerating pattern is clearly evident in these figures.

Slopes are presented as week gains in words correct per minute and are based on annual student growth (i.e. 36 weeks). Given that the above figures were established on yearly growth outcomes, some authors have questioned the appropriateness of applying these standards to short term assessments (i.e. 10 weeks; Hintze & Christ, 2004).

Summary

A review of the literature reveals that research related to the influence of particular text characteristics, such as curriculum orientation and difficulty level, on reading fluency outcomes has been established. Findings indicate that text characteristics do not necessarily impact slope of growth as measured by CBM ORF, however, differences in absolute levels of performance are frequently observed.

Table 1

Expected Growth Rates

Grade	Slope (Fuchs et al., 1993)	Slope (Deno et al., 2001)	
	General Education	General Education	Special Education
1	2.10	1.80	.83
2	1.46	1.66	.57
3	1.08	1.18	.58
4	.84	1.01	.58
5	.49	.58	.58
6	.32	.66	.62

Research examining the effect of text characteristics such as decodability and percentage of high-frequency words on fluency outcomes has not been completed. Two studies (e.g. Compton, Appleton, and Hosp, 2004; Hiebert and Fisher, 2002) have indicated that these factors may indeed impact profiles of growth. Furthermore, research indicates that both of these characteristics significantly impact development of reading proficiency and lend support to the argument that these characteristics may impact outcomes.

In light of the new IDEIA procedures and its inclusion of RTI it is important that the effects of text characteristics on CBM measures be understood in order to optimize potential for accurate decision making.

Research Questions

The following questions will guide this research.

Research Question #1: Did students demonstrate growth in oral reading fluency performance on the two probe types over the course of the eight trials which took place over four weeks?

It was hypothesized that students would demonstrate growth in oral reading fluency, as measured by words read correctly per minute, on both probe types over the course of the four weeks. Given that the passages take into consideration the reading skills of the subjects, the passages should be sensitive to oral reading fluency growth.

Research Question #2: Are there differences between the absolute oral reading fluency rates of students who read the high-frequency probes and those who read the highly decodable probes?

It was hypothesized that there would be differences in absolute oral reading fluency rate as a function of probe type as evidenced by significant differences in group means across decodable and high-frequency conditions.

Research Question #2: Is oral reading fluency slope of growth significantly impacted by text characteristics?

It was hypothesized that estimates of growth (i.e. slope) would vary as a function of probe type within the at-risk population as evidenced by significant differences in growth profiles across decodable and high-frequency conditions.

CHAPTER III

METHODOLOGY

Subjects and Sampling Procedures

Previous research using repeated measures analysis of variance has indicated the need for approximately 15 participants per group in order to detect small to moderate effect sizes with an alpha level of .05 (Hintze, Christ, & Keller, 2002). An a priori power analysis (Cohen, 1988) was conducted suggesting that a sample size of 46 students would provide adequate power (.80) for main and interaction effects assuming a large effect size (.35) and an alpha level of .05.

The schools from which participants were drawn were two rural mid-west elementary schools, one of which was participating in a state-funded education improvement project. These schools were selected for the proposed study on the basis of convenience. Consent from the district and school administrators were obtained before the commencement of research activities.

The study population consisted of a convenience sample in which participants were solicited from six first, second, and third grade classrooms within two rural south Midwest schools. Students were eligible for participation based on previous DIBELS assessments indicating an At-Risk/Some-Risk status. A solicitation and parent consent

form was sent home to the parents/guardians of the students in these grades. Included in the solicitation letter and consent form was a description of the study and its purpose, an outline of what the students would be asked to do, a description of the potential risks and benefits to the students, a statement that participation would be voluntary and results kept confidential. The contact information of the researcher, information for filing complaints, and a place for the consenting adult to grant or deny permission for inclusion in the study were also included in the letter.

The schools in which data collection occurred were located within two small towns of less than a thousand people and were located within equally small school districts. One of the schools was participating in a school wide state-funded education improvement grant. Student performance data were collected within the elementary school buildings in an area where traffic and noise were minimal, and where disruption to other classes was nominal.

Design

This study examined the natural variation of student oral reading performance on CBM passages developed from curriculum with different word controls. One within group variable, trial, and one between group variable, text type, was utilized in this study. CBM text type served as the independent variable. Two levels of CBM type were included in this study: (1) DIBELS high-frequency passage and (2) highly decodable text. Students were randomly assigned to read from either the high-frequency or decodable passages. Passages were administered to each student twice a week for four weeks to

assess growth over time. A single dependent variable, oral reading fluency, was collected in this analysis.

Independent Variables

One independent measure was utilized in this research endeavor, CBM text type. Two levels of text type were employed (a) a highly decodable text and (b) a text controlled for high-frequency words.

Text Type

Particular variables of texts have been established as influencing reading development; however, limited research has evaluated the effects of these text differences on curriculum based measurement outcomes. Research has suggested that particular characteristics of passages used for curriculum based measurement may influence oral reading fluency outcomes (Compton, Appleton, & Hosp, 2004; Hiebert & Fisher, 2002).

As previously mentioned, this research endeavor employed two levels of CBM passage types controlled for various text characteristics. The first of these CBM text types was controlled for decodability. Decodability is defined as phonetic regularity of the words and the match between those letter/sound relationships occurring in the text and that which has been taught in the classroom (Allington & Woodside-Jiron, 1998; Groff, 1999; Hiebert, 1999; Mesmer, 2005; Stein, Johnson, & Gotlohn, 1999).

Passage decodability was established in two stages. First, a review of the intervention materials used with students and the Open Court teacher's manuals for the

first and second grade levels was reviewed to determine the letter-sound patterns that should have been taught up through the beginning of second grade. These were the letter-sound correspondences which were considered "decodable" for the students given that all students should have been exposed to these patterns. Secondly, all words within a passage were fitted to Menon and Hiebert's (1999) eight levels of phonetic regularity. Words in levels one through four, which covers words up to the consonant-consonant-vowel-consonant-e (CCVCe) pattern, were considered decodable in keeping with the letter-sound patterns covered in the curriculum.

The second type of text utilized served as a high-frequency word contrast passage and was drawn from the Dynamic Indicators of Basic Early Literacy (DIBELS; Good, Simmons, and Kame'enui, 2001). The DIBELS are a set of standardized individually administered measures designed to monitor students reading skills. The measures have been demonstrated to be reliable and valid indicators of literacy development and predictive of reading proficiency (Elliott, Lee, & Tollefson, 2001; Hintze, Ryan, & Stoner, 2003). Included in this set of measures are oral reading fluency passages. DIBELS passages have been controlled for readability level, using the Spache readability formula (Good & Kaminski, 2002). First grade progress monitoring passages all have a Spache readability between 2.0 and 2.3 (Good & Kaminski, 2002). These passages served as a high-frequency comparison passage, as passages are comprised of higher levels of high-frequency words then the decodable passages.

Each DIBELS passage was evaluated for percentage of high-frequency words and decodable words. High frequency words are those words which occur repeatedly throughout texts, but which often have irregular letter-sound patterns and frequently lack

meaning (Cunningham & Allington, 2003). The percentage of high frequency words was determined by the number of words which are found in the first Word Zone (Hiebert, 2005).

A Chi-square analysis was completed to ensure that statistically significant differences existed between the high-frequency and decodable passages in terms of high-frequency and decodable words. Results indicated that a statistical difference existed between the passage types in terms of both percentage of high-frequency ($X^2 = 32.213$, $p < .001$) and decodable words ($X^2 = 187.962$, $p < .001$).

Dependent Variable

Oral reading fluency (ORF) is the single dependent variable and was operationalized by measuring words correct per minute on designated reading passages. The reliability and validity of ORF tasks have been well established. Oral reading fluency has been established as a valid measure of reading achievement, which is highly correlated with standardized reading tests (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Shinn & Good, 1992), as well as standardized comprehension measures (Fuchs, Fuchs, & Maxwell, 1988; Fuchs, Fuchs, Hosp, & Jenkins, 2001; Jenkins & Jewell, 1993; Markell, & Deno, 1997). ORF tasks have also been established as having strong discriminative validity for differentiating between students with reading disabilities, low achieving students who receive remedial services, and average achieving students who do not receive remedial services (Deno, Mirkin, & Chiang, 1982; Shinn, Tindal, Spira, & Marston, 1987; Marston & Magnusson, 1988; Deno, Marston, Shinn &

Tindal, 1983; Shinn & Marston, 1985; Shinn, Tindal, & Stein, 1988; Shinn, Ysseldyke, Deno, & Tindal, 1986).

In keeping with the procedures delineated by Shinn et al. (1989) oral reading fluency was calculated as the number of words read correctly in one minute. Words were counted as correct if they were pronounced correctly or were self-corrected within three seconds of the initial incorrect pronunciation (Shinn et al., 1989). Substitutions, omissions, and mispronunciations were considered errors, while repetitions and addition of words were not scored as errors (Shinn et al., 1989).

Materials

Eight different passages were developed from each of the two text types. Passages were controlled according to the procedures described above. The high-frequency control passages were taken from the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) fluency progress monitoring materials for the first grade. The decodable probes were developed from the Basic Reading Series (Science Research Associates, 1985) which is a series developed to maximize decodable words. All decodable passages were controlled in accordance with the methods described in previous sections.

Procedures

Step 1: Screening Risk Status

The DIBELS was utilized to determine risk status. This measure identifies students as either At-Risk, Some-Risk, or Low-Risk (Good et al., 2002). For the purposes of this research endeavor, the At-Risk and Some-Risk categories were collapsed. All participants were administered the DIBELS, Benchmark Two, reading fluency passages to determine their risk status. Students identified under the At-Risk/Some-Risk categories range from having some word attack skills to having no reading skills (Good et al., 2002). Students were randomly assigned to read from either the highly-decodable passages or the high-frequency control passages.

Step 2: Training

Data collection was completed by the author. The experimenter had received prior instruction in CBM administration procedures. Prior to data collection the experimenter received a refresher course in administering and scoring the reading fluency probes. A set of written instructions detailing how the reading probes were to be administered and scored was reviewed prior to data collection.

Step 3: Data Collection

Data was collected on student oral reading fluency in the two text types using the procedures for curriculum-based measurement delineated by Deno (1985). The reading passages were administered to participants twice a week for four weeks at the selected school during regular school hours. As such, students participated in a maximum of eight sessions. At each session, depending on their group membership, participants were presented with one of eight decodable or high-frequency passages to read. The order of presentation of reading passages was counterbalanced across sessions for all students. No student read the same passage more than once during the study.

Before the beginning of each session a standard script was read giving a brief explanation of the procedures, followed by specific directions for completion. The experimenter then instructed the students to begin reading aloud. Students were allowed one minute to read as many words possible. The reading probes were scored immediately by the experimenter. As previously mentioned words were counted correct if they were pronounced correctly, or if the students self-corrected within three seconds after making an error. Errors constituted substitutions, omissions, and mispronunciations. Repetitions and added words were not scored as errors. The experimenter provided the word when the student hesitated or struggles to pronounce a word for three seconds. These supplied words were counted as errors. After each student read to the experimenter, he or she was thanked for their participation and returned to class.

Step 4: Data Analysis

Data collected for each group was analyzed with a 2 (text) x 8 (trials) mixed model repeated measures analysis of variance. Such a design was selected to compare within-student performance across the different text types, as well as evaluate possible interaction effects. Text type (high-frequency and decodable passages) served as a between-subjects variable, while trial (8 sessions) served as the within-subjects variable.

Limitations

Several limitations associated with the methodology of the current study should be noted. First, the population for the study consists of a convenience sample which may not represent true population statistics and may limit the generalizibility of the study results. Secondly, the repeated measures design of the current study may ultimately function as an intervention in which exponential growth will result simply because students are repeatedly exposed to the material. However, this is the nature of progress monitoring with curriculum based measurement and this effect would be expected with any set of passages. A third limitation that should be noted is that there is the possibility of a floor effect with very low or non-readers. To limit the likelihood of this occurring, the DIBELS first grade progress monitoring passages was employed.

CHAPTER IV

RESULTS

This study investigated the effects of curriculum-based measurement text features on students' oral reading performance. Specifically, two types of texts were utilized in this study, one which loaded on high-frequency words and one with a high percentage of decodable words. The texts were administered in eight sessions to 43 students who had previously been identified by their schools as struggling readers. Students' oral reading performance, in terms of words read correctly per minute, served as the datum for analysis.

The results for the three research questions which guided this study are presented in this chapter. All data were analyzed using the Statistical Program for Social Sciences version 12.0 (SPSS 12.0). The significance level was set a priori at $p<.05$ for all statistical analyses.

Sample Characteristics

To obtain the sample for the study students were screened by their teachers with the DIBELS screening measure. Those students identified as performing below level were solicited for participation in the study and permission letters were mailed to the

parents of the identified students. Of the 56 letters sent out, 50 letters (89%) were returned granting consent for their child to participate in the study. Because of absences, seven students (14%) were dropped from the study before the end of data collection. In total, 43 subjects (86%) completed the data collection process and were included in the study.

The final subject population was comprised of 26 males (61%) and 17 females (40%). Nine of the participants (21%) were first graders, ten (23%) were second graders, and twenty-four (56%) were third graders. Furthermore, 34 (80%) of the participants attended one of the schools, and an additional 9 (21%) students participated from a second elementary school. All students were participating in daily supplemental remedial reading instruction aimed at increasing reading fluency outcomes.

Descriptive Statistics

Descriptive statistics for each of the independent variables (text type and trial) are presented in Tables 2 and 3.

Table 2

Descriptive Statistics for Text Type

	Decodable	High Frequency
N	22	21
Mean	67.24	76.09
Median	67	76
Mode	65	94
Minimum	11	14
Maximum	153	146
Range	142	132
SD	27.84	28.60

Table 3

Descriptive Statistics for Trial

	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8
N	43	43	43	43	43	43	43	43
Mean	70.30	72.07	70.93	72.26	72.74	71.95	71.19	71.05
Median	65	68	73	69	75	65	67	69
Mode(s)	61	52	39, 75	69, 82	77, 81	54	67	46
Minimum	14	15	13	13	15	18	11	18
Maximum	136	153	140	144	142	136	149	136
Range	122	138	127	131	127	118	138	118
SD	26.34	30.20	29.89	32.06	27.86	25.90	30.09	28.29

Research Question 1

1. Did students demonstrate growth in oral reading fluency performance on the two probe types over the course of the eight trials which took place over four weeks?

To evaluate whether growth was observed over the course of the data collection period the main effects of Trial was examined. A 2 (type) x 8 (trial) mixed-model repeated-measures analysis of variance (ANOVA) was utilized to analyze the data. This analysis allowed for the evaluation of growth in oral reading fluency across the two measures. No significant main effects for Trial were observed ($F [7, 41] = .28$, $p > .05$; see Table 4), indicating that students in both groups did not demonstrate significant growth over the eight trial periods. Post hoc analyses were not run due to the insignificant results of the initial repeated-measures ANOVA.

Table 4

Analysis of Variance Results for Text Type and Testing Trial

Source	df	SS	MS	F
		Between subjects		
Text	1	6733.08	6733.08	1.13
Error 1				
		Within subjects		
Trial	7	196.53	28.08	.28
Text x Trial	7	114.88	157.84	1.56
Error 2				

Research Question 2

2. Are there significant differences between the absolute oral reading fluency rates of students who read the high-frequency probes and those who read the highly decodable probes?

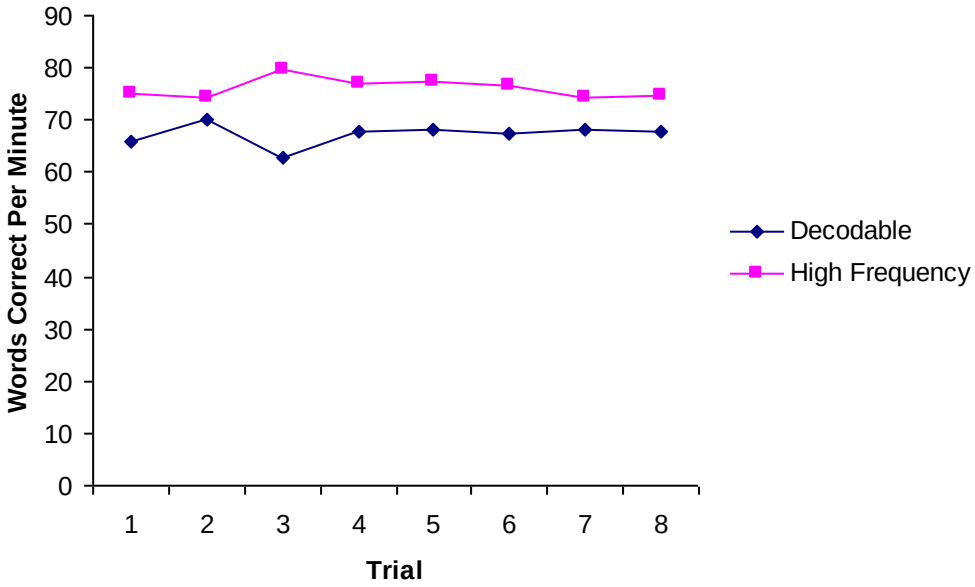
Data were analyzed to determine the presence of significant main effects for probe type using a 2 x 8 mixed-model repeated-measures ANOVA. Data from this analysis indicated no significant main effects for probe type ($F [1, 41] = 1.13$, $p > .05$; see Table 4). Because of the failure to meet significance with the initial repeated-measures ANOVA no post hoc analyses were run on these data.

Research Question 3

3. Is oral reading fluency slope of growth significantly impacted by text characteristics?

Using a 2 x 8 mixed-model repeated-measures ANOVA, data were analyzed to determine the presence of an interaction effect between probe type and number of trials. Data from this analysis indicated no significant interaction effects ($F [7, 41] = 1.56$, $p > .05$; see Table 4). Graphical representation of this data is also presented in Figure 1. No post hoc analyses were run on these data due to the insignificant findings of the initial repeated-measures ANOVA.

Figure 1. Average words correct per minute by group across trials.



CHAPTER V

DISCUSSION

Overview

The purpose of this research project was to assess the impact of word-level factors on struggling readers' oral reading fluency within the progress-monitoring setting. Specifically, this study examined the effect of differential levels of high-frequency and decodable words within progress-monitoring passages on poor readers' oral reading fluency performance across time. The effects of these variables were assessed by examining the reading rate (words correct per minute) and growth trends of the students who read from passages with increased levels of high-frequency words as compared to those who read from passages with higher percentages of decodable words. The absolute reading rate and growth trends of the two groups were assessed to determine (a) if students demonstrated growth, (b) if, in terms of absolute reading rate, students performed differentially on the two probe types, and (c) if the growth trends of the students reading the two different passage types differed significantly.

Statistical analyses indicated no significant performance differences between the two groups of students, in terms of either absolute reading rate or growth. These findings suggesting that over the course of the four weeks in which students were monitored, they

did not demonstrate statistically significant growth as measured by either set of progress monitoring probes. Furthermore, although the passages had included statistically different levels of decodable and high-frequency words, the findings of this study demonstrate that these factors did not significantly alter the speed at which students read the passages, nor was the absolute reading rate significantly different between the two groups. A more thorough discussion of these finds will be presented in this chapter along with a discussion of the possible weaknesses that impacted this study. Finally, directions for future research and implications for practice will be offered.

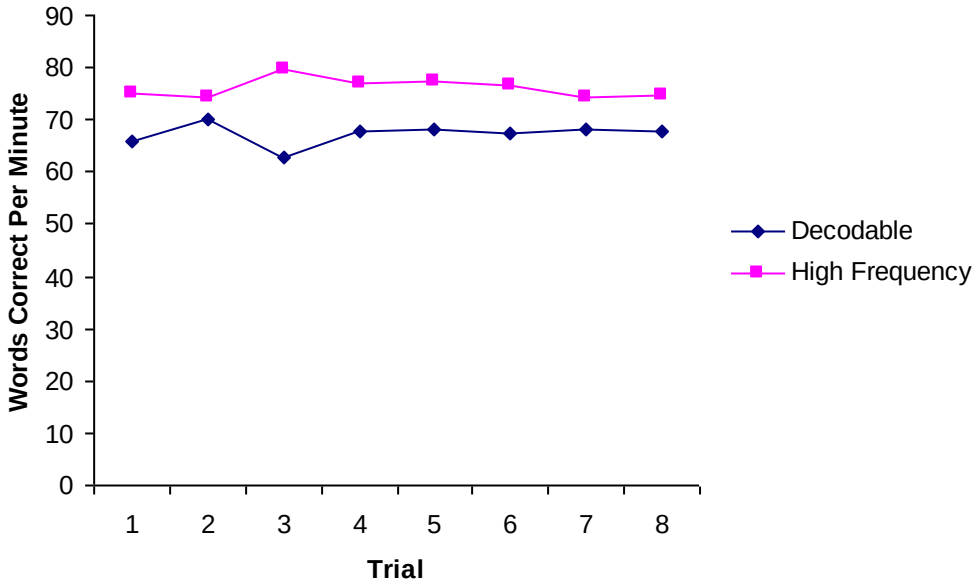
Assessing Growth

One of the more surprising findings of this study was that over the course of the four weeks in which students were monitored, students failed to demonstrate statistically significant growth on either of the two measures ($F [7, 41] = .277$, $p > .05$). Group means across the eight trials ranged from 70.30 at the first trial, to 72.26 at the fourth trail, and 71.05 at the eighth trail. Because neither group demonstrated significant growth on either sets of probes comparisons could not be made between the differences in growth for those who read from the decodable passages and those who read from the high frequency passages.

Research has established that in the primary grades average readers are expected to increase their reading fluency at a rate of 1-2 words per week (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993). This same slope of growth is the goal for struggling readers who are also receiving specific reading

interventions, as was the case with the students who participated in this study. Thus, over the course of the four weeks of data collection one would have expected that students would have increased their words read correctly per minute by at least four to eight words per minute; however, across the four weeks of this study there was an average increase of 2.5 words correct per minute. Graphical representation of the weekly averages, and the trendline associated with those averages, indicates an almost completely flat growth profile (see Figure 2).

Figure 2. Average words correct per minute by group across trials.



In Hintze, Daly, and Shapiro's (1998) discussion of selecting appropriate progress monitoring material, the authors demonstrated that for average readers in the primary grades (1st and 2nd grade) on-level materials were more sensitive to students' oral reading fluency growth than goal-level materials. Thus, given that in the current study materials were selected to be at the reading level of the participants, one would have additional

reason to expect that student's would have demonstrated significant growth over the course of the four weeks of progress monitoring. A similar study (e.g. Shinn et al., 1989) demonstrated that grade level and above grade level materials were equally effective in terms of sensitivity to growth.

On-level material was selected for this study based on the assumption that students would be able to directly apply the decoding skills with which they were currently gaining proficiency. Subjects utilized in this study were participating in intensive phonics interventions programs, and given that probes were selected to include a high percentage of words which were in line with their decoding skills, students should have been able to demonstrate growth on these materials. A discussion of factors which could have impacted student progress will be presented in latter sections of this chapter.

To further evaluate growth between the groups, data was visually inspected with relation to grade level differences in performance within each group. As such, growth in reading performance for each probe type was broken down by the first, second, and third grades. Figure 3 presents all grade-level data. Data has been separated by each grade level in Figure 4, 5, and 6, and trend lines have been added to facilitate the inspection of the data. Due to the limited number of subjects in each group (i.e. first grade n = 10, second grade n = 10, third grade n = 23) data could not be evaluated statistically. Visual inspection of the trendlines associated with the data reveals that third grade students who read from the high-frequency passages and all second graders demonstrated no growth in reading rate or exhibited a decrease in overall reading rate. However, all first grade students demonstrated approximately a four-word growth over the course of the eight trials, regardless of the passage type from which they read. Similarly, based on trendlines,

third graders reading from the decodable passages demonstrated a growth of approximately three words correct per minute over the eight trials.

When evaluated by grade level, it is apparent that students in the first grade demonstrated some growth, indicating that the probes were sensitive to growth at this level. Growth was likely observed with the first-grade readers as the difficulty of the passages was more appropriate than it was for the second and third graders. That is, the difficulty level was the closest to the first-grades independent reading levels, and thus, more sensitive to the changes in their reading ability. Probes typically fell within the mastery range for second and third graders, and as such, were less sensitive to changes in reading ability as these students read at a high level due to the relative ease of the passages. For these second and third grade students there was likely a ceiling effect which limited the demonstration of growth.

Figure 3. Average words correct per minute by grade level and group across trials.

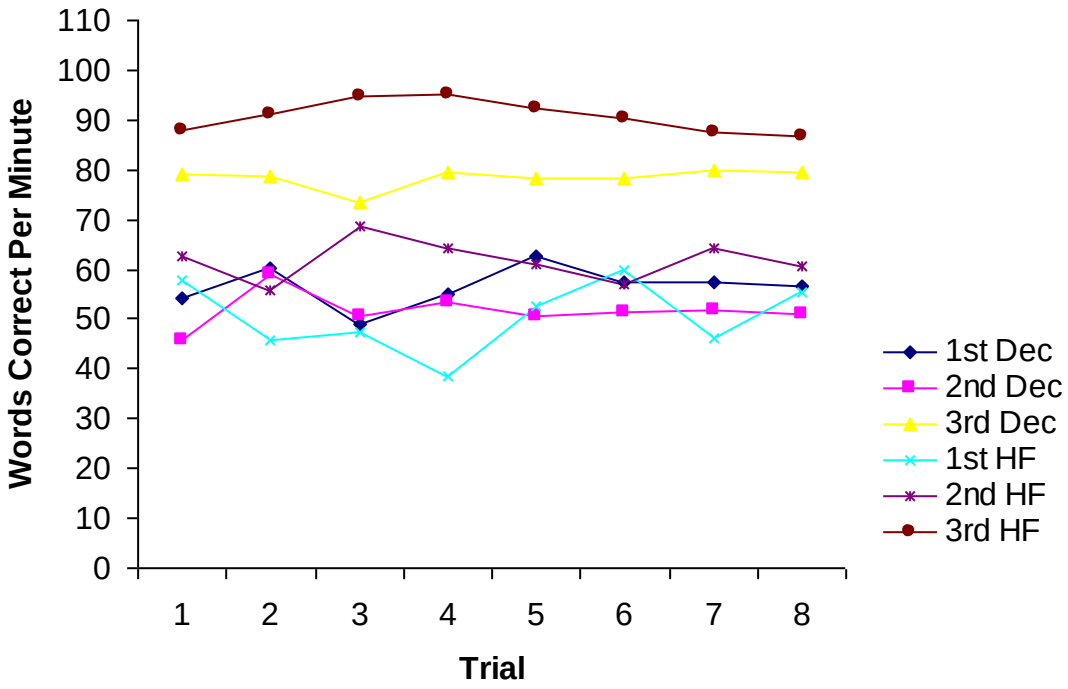


Figure 4. Average first grade reading rate by group across trials.

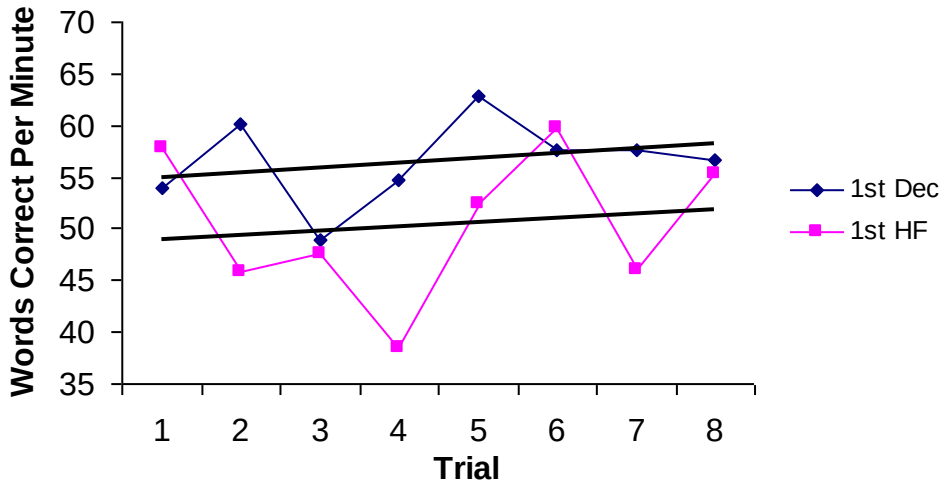


Figure 5. Average second grade reading rate by group across trials.

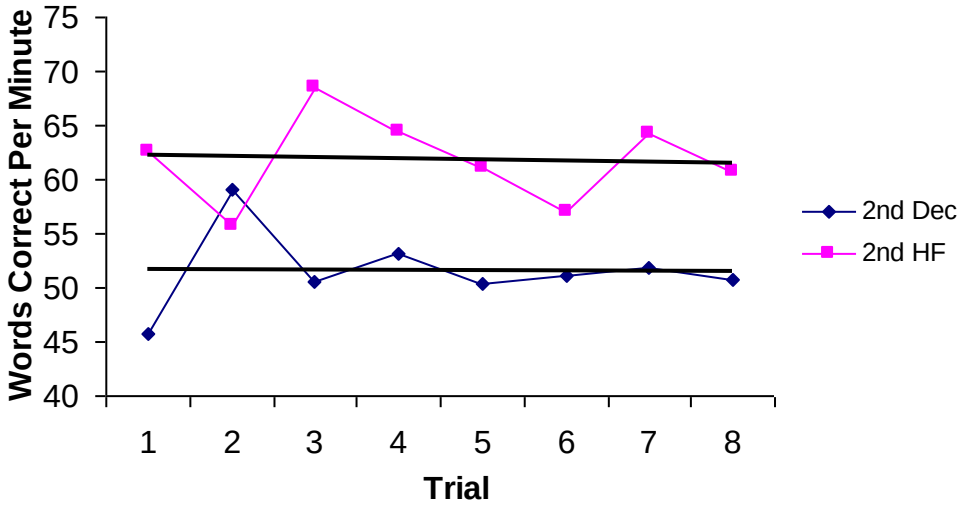
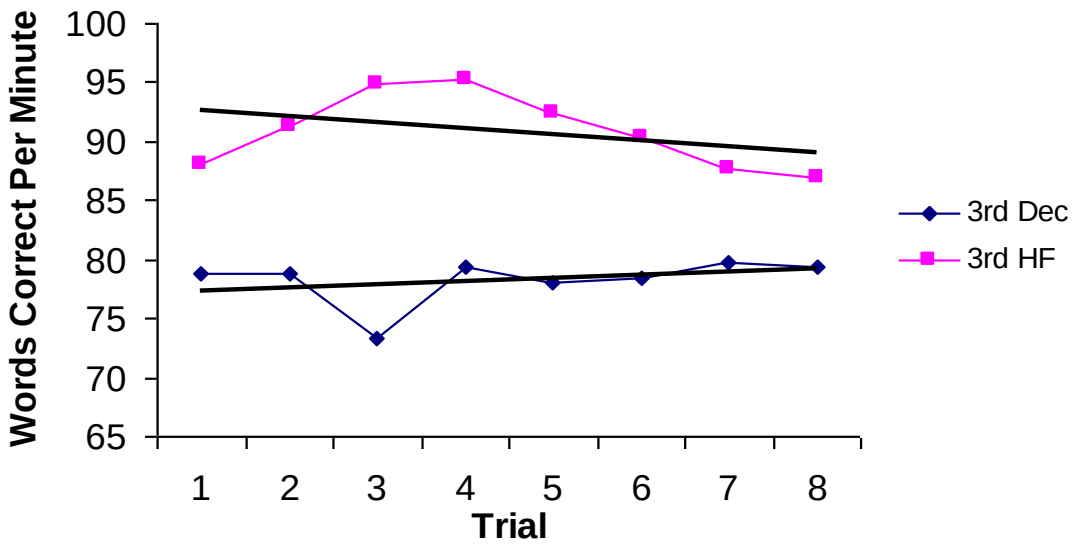


Figure 6. Average third grade reading rate by group across trials.



Absolute Reading Rate

The second objective of this research study was to determine if absolute reading rates varied significantly between those who read from the high-frequency passages versus those who read from the decodable passages. One would have expected that given the significant differences in the composition of the passages, it would be likely that students would have performed differentially. Statistical analysis indicated that the effects of differential levels of high-frequency words and decodable words did not significantly impact student's absolute reading rate. The average absolute reading rate for the high-frequency group was 76 words read correctly per minute, while for the decodable group it was 67 words per minute. It should be noted, however, that although the differences did not reach statistical significance, visual inspection of the reading rates of the two groups at each of the different trial times demonstrates that the high-frequency

group consistently out-performed their decodable counterparts. This can be observed in the graphical representation of the reading rates of both groups in Figure 5.

Compton, Appleton, and Hosp (2004) identified that reading accuracy and reading fluency were significantly correlated with two predictors, the percentage of high-frequency and decodable words found in a passage. In their study statistical analysis indicated that the percentage of high frequency words accounted for 20 percent of variance in reading fluency, while percentage of decodable words accounted for 23 percent of variance in reading fluency. They also compared the reading accuracy and reading fluency performance of below-average and average readers. Here, Compton, et al. (2004) found that for average readers, reading fluency and reading accuracy were significantly correlated with percentage of high-frequency words and decodable words. In their study, below-average readers' performance was only significantly correlated with percentage of high-frequency words. Based on this last finding, one would hypothesize that while increasing the percentage of high-frequency words would likely result in an increase in reading fluency for struggling readers, increasing the percentage of decodable words would not equal an increase in reading fluency as decodable words were not significantly correlated with reading fluency for this group. Although it was not a statistically significant result, visual inspection of the data demonstrates that the student's who read the high-frequency passages consistently performed at a higher level that those who read from the decodable passages.

To further evaluate potential differences in reading performance, data was visually inspected by grade levels. These visual inspections indicated that there were differences in the reading performances at the various grade levels. Importantly, when

inspecting the trendlines associated with first-graders' performance it is evident that the students who read from the decodable probes typically read at a higher rate (i.e. they read more words correct per minute) than those who read from the high-frequency probes. Conversely, second and third graders typically read faster on the high-frequency probes than on the decodable probes. This information is congruent with reading development theories (Ehri & McCormick, 1998) which indicate that in the early stages of reading development, students' primary identify words through the application of their knowledge of the grapheme-phoneme connection, that is, through the application of phonics and decoding skills. Additionally, at this point, first-grade students have not developed a broad repertoire of high-frequency words, whereas, second and third graders have become fluent with most high-frequency words.

The information obtained in the current study would seem to support this developmental theory as first grade students performed better on passages which were easier to decode and for which they possessed the decoding skills, than on those passages which relied more heavily on high-frequency words with which they may not yet have become proficient. Second and third graders, on the other hand, generally performed better on the high-frequency passages. At this level, students have transitioned to reading by recognizing whole words and decoding words in larger chunks rather than by individual letters (Ehri & McCormick, 1998), and this was facilitated in the high-frequency passages. However, the decodable passages included many words which, although easily decodable, would be considered uncommon (e.g. huff, bits) and would require more time to decode than common words.

Additional Data Evaluations

In addition to the above described statistical evaluations, supplementary visual inspection of data was completed to determine if any patterns in data could be identified. Although, these data could not be evaluated statistically due to the limited number of subjects and statistical power, it is useful in determining possible areas which might have reached significance had power not been limited in the current study.

The first of such evaluations involved evaluating students' oral reading error rates associated with the high-frequency and decodable probes. The average associated with each probe types was practically equivalent, at 2.88 and 2.77 for decodable and high-frequency probes respectively. In addition, growth trends for the two passage types were similar as well (see Figure 7). Furthermore, errors were further analyzed by grade level (see Figures 8, 9, and 10). Visual analysis of this information indicated that errors were fairly consistent for the high-frequency and decodable groups at each grade level. However, it should be noted that as was observed in with the oral reading fluency analyses, first graders tended to perform better (e.g. have fewer errors) on the decodable probes, while second and third graders typically performed better on the high-frequency probes. Again, these differences could not be evaluated statistically due to insufficient power.

Figure 7. Average errors per minute by group across trials.

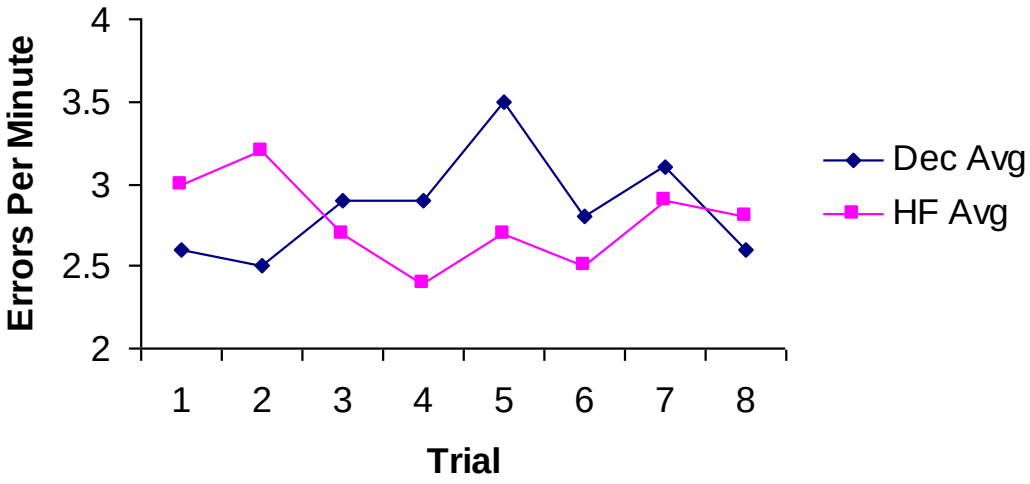


Figure 8. Average first grade errors by group across trials.

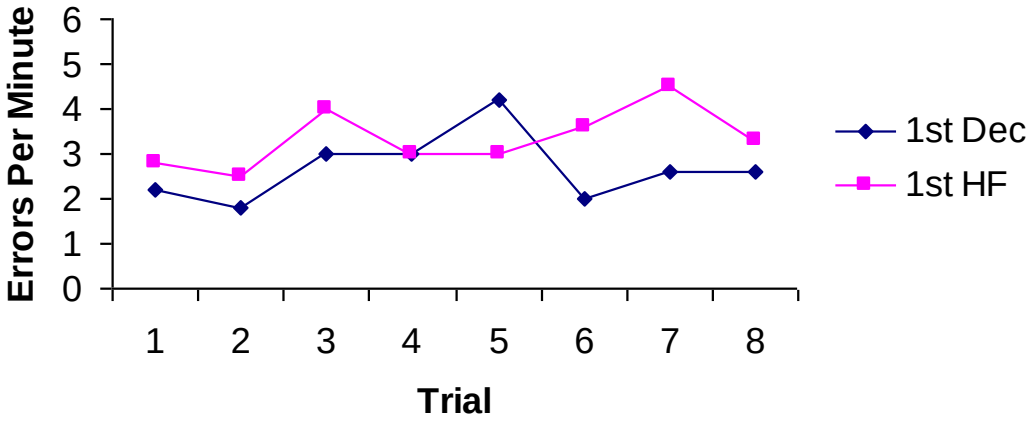


Figure 9. Average second grade errors by group across trials.

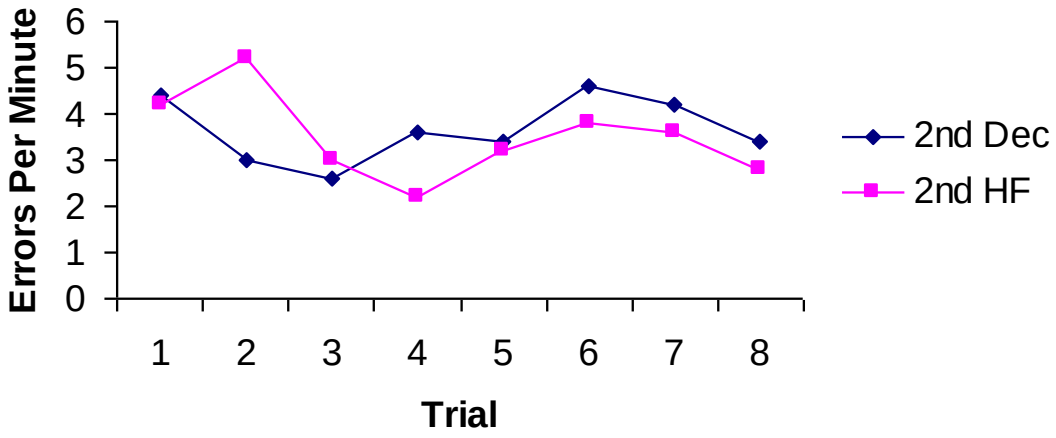
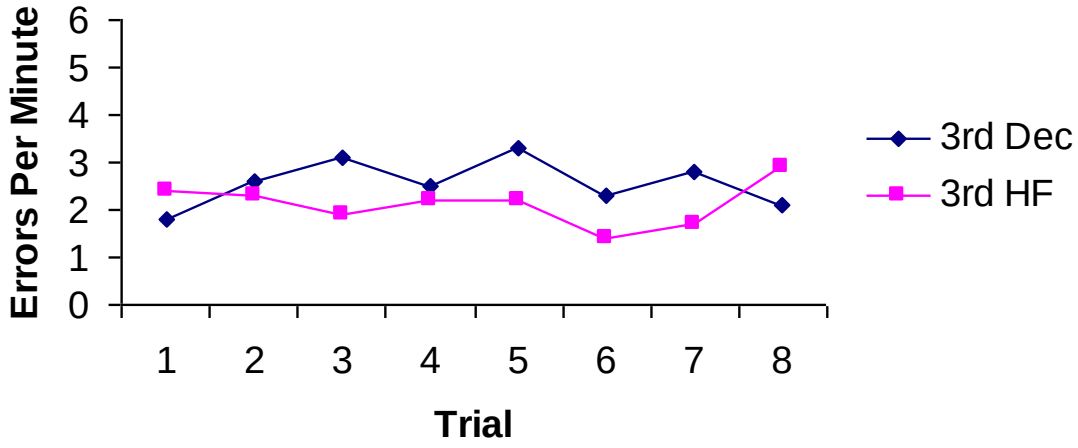


Figure 10. Average third grade errors by group across trials.



Passages were also evaluated for differences in Critical Word Factor to determine if differences might account for insignificant results. Hiebert and Fisher (2002) discussed the importance of what they termed "Critical Word Factor" (CWF) in beginning-readers' texts. Critical word factor utilizes the percentage of both high-frequency and decodable words in a passage to determine the difficulty of a passage. The CWF of a passage is

determined by examining (a) the number of words which fall outside of a pre-specified list of high-frequency words and (b) the percentage of words at each of the eight decodability levels (for the purposes of the current study words were considered decodable if they fell in the first four decodability levels.) Thus, the higher the CWF the more difficult the passage would be.

While a chi-square analysis was completed prior to the beginning of the study, this analysis only guaranteed that the percentages of high-frequency were statistically different between the two groups of passages and that the percentages of decodable words were statistically different between the two types of passages. There was no attempt to evaluate the collective percentages of high-frequency and decodable words in the passages before they were administered. A priori analysis of the two groups of passages utilized in the current study revealed that for the decodable passages the CWF was generally around 1%, indicating that 99% of the words would have been considered high-frequency words or decodable words. For the high-frequency passages, the CWF was generally between 20% and 30%, suggesting that only 70% to 80% of words were decodable or high frequency. Therefore, the high-frequency passages included a significantly higher percentage of words which were neither decodable nor high-frequency words. Given this differential one would have further expected that students who read from the decodable passages would have outperformed their high-frequency counterparts.

Limitations and Implications for Future Research

When discussing the limitations of a study there are two factors to address, the threats internal and external validity. In the discussion of threats to internal validity, factors other than the independent variable which might have influenced the results must be addressed. Similarly, limitations in the generalizibility of the findings must be addressed in the discussion of external validity. A discussion of both the threats to the internal and external validity of this study follows in the subsequent section. A discussion of the implications for future research will also be interspersed throughout this section as it relates to the various threats to internal and external validity.

Threats to Internal Validity

In sum, the findings from this study indicated that there were no statistically significant differences in the oral reading fluency performance of struggling readers who read from passages with differential percentages of high-frequency and decodable words. This finding is acceptable and might suggest that both passage types measured reading fluency in a highly reliable fashion. However, the fact that neither group demonstrated any growth over the course of the four weeks suggests that there were other factors at work which could have possibly impacted students' performance. This section will discuss possible factors which could have impacted the outcomes of this study.

One hypothesis for the lack of significant differences in the findings of the current study, is that it is not solely high-frequency words *or* decodable words that is more

important for reading fluency, but perhaps the importance lies in percentage of high-frequency *and* decodable words, as suggested by Hiebert and Fisher (2002). Although the current study would suggest that the differences in CWF between the two types of passages did not significantly influence performance (e.g. although the CWF was higher in one set of passages than the other, students performed equally on the two passages), future research should attempt to hold the CWF steady across the two passages while systematically manipulating the percentage of high-frequency and decodable words which account for the remainder of the words in the passages. This would further guarantee the equivalence of the passage, as it would parse out additional confounding factors which could have influenced the current study, and would also allow for greater power to evaluate the effects of the decodable and high-frequency words.

Another likely influencing factor is related to the time of the year when data was collected. Data was collected within the last month of the school year for the majority of participants and during the special summer reading program for an additional eight participants. As one teacher described it, by this point in the year students had "already checked out for the summer." This same teacher indicated that, for all intents and purposes, classroom instruction had stopped, and that the teachers had transitioned away from teaching and into classroom management. Although not the case for the eight summer school participants, if intensive interventions had stopped it is unlikely that one could expect continued growth from the struggling readers. Future research should be conducted while the school year and reading interventions are in full swing. If the students are reliably receiving the reading interventions, then growth would be expected over the course of four weeks.

As previously mentioned, another factor which could possibly account for the insignificant results is a possible ceiling effect. The group averages for the students reading the high-frequency and decodable passages were 76 and 67 words correct per minute respectively. On screening measures, students' average performance was 38 and 43 words per minutes respectively for the decodable and high-frequency groups. Given that students were reading in materials that were *at* their instructional level, rather than the challenging level that they were screened with, it is expected that they would perform at a slightly higher level. Yet, one would not expect that there would be such a large difference between the screening measure and performance on these probes. It is likely that, although the probes were designed to address the skills being taught in the reading interventions, the passages were actually too easy. If this is actually the case then it is possible that a ceiling effect might have affected the results. Future research should attempt to align more closely the passages with students' reading abilities.

Two of the primary factors which likely impacted the outcomes of the current study the most were the fact that participants in the current study were grouped across three different grade levels and there was insufficient power to evaluate grade level differences. Had first, second, and third grade students been separated, and more subjects obtained in each group, it is plausible that the results of the study would be different. As demonstrated by visual inspection of the grade-level data, it is possible that actual differences in performance were masked by the combined nature of the groups. Future evaluations should separate participants by grade level and increase the number of students to increase power for statistical evaluations.

Threats to External Validity

For the current study the population from which subjects were drawn consisted of a convenience sample, and as such, it might not have represented true population statistics. This factor is further compounded by the fact that the two schools from which subjects were drawn were rural mid-west schools, and it is highly unlikely that population statistics matched that of these two schools.

To address the issue of external validity, it is suggested, that for future research the current study could be replicated with a larger sample size from a population which closer approximates that of the general population.

Implications

Although statistical significance was not reached in the current study some implications for practice can be made based on information obtained from visual inspection of grade-level data. First, data suggests that at different levels readers may perform differentially on reading tasks which have varied levels of high-frequency and decodable words. This being so, it is important that when curriculum-based measurement probes are being selected for use in progress monitoring that attempts be made to keep the percentage of high-frequency and decodable words consistent across probes. Such endeavors would ensure that differences in reading performance could be attributed to true changes in reading ability rather than differences in reading probes.

Additionally, the data would seem to indicate that if educational decisions are to be made based on CBM data within the RTI framework, then it is important to realize that reading rates may be influenced by the word-level make up of the passages. Because this study did not include a control passage (i.e. a set of passages without word-level controls) comparisons cannot be made between performances on the high-frequency and decodable passages with typical reading performance within uncontrolled material. However, preliminary data obtained in this study indicate that at the early reading acquisition phases students would have a higher reading rate when reading from probes developed from decodable books, while students at later acquisition phases would tend to read "slower" from such passages. Thus, theoretically, if a criteria level is used to determine whether or not a student should qualify as learning disabled, then second and third grade students may tend to be over identified and first graders may be under identified if decodable probes are used for decision making. A similar effect may be observed if probes which have a high percentage of high-frequency words are selected for use in decision making. Further research needs to be completed to test out this notion; however, in the mean time, it is important that these possibilities be considered when using curriculum-based measurements to determine education needs.

Ultimately, the current study confirms that there are many factors which could impact the outcomes of curriculum-based measurement. Although this study does not provide a solution for eliminating such factors, it does open discussion related to identifying factors which may impact outcomes. Hopefully, such discussions will in turn lead to identification of means for reducing error variance associated with curriculum-

based measurement, and thus improving educational assessment strategies and educational outcomes.

References

Adhearn, E.M. (2003). Specific learning disability: Current approaches to identification and proposals for change. Report prepared for the National Association of State Directors of Special Education (NASDE) Project FORUM. Alexandria, VA: NASDE. Allington, R. (1997, October/November). Overselling phonics: Five unscientific assertions about reading instruction. *Reading Today,* p.15.

Allington, R., & Woodside-Jiron, H. (1998). Decodable text in beginning reading: Are mandates and policy based on research? *ERS Spectrum, 16,* 3-13.

Ardoin, S.P., Witt, J.C., Suldo, S.M., Connell, J.E., Koenig, J.L., Resetar, J.L., Slider, N.J., & Williams, K.L. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screenings. *School Psychology Review, 33*(2), 218-233.

Beck, I. L. (1997, October/November). Response to "Overselling phonics." *Reading Today,* p. 17.

Beck, I.L. & Juel, C. (1995). The role of decoding in learning to read. *American Educator, 8,* 39-42.

Blakenship, C.S. (1985) Using curriculum-based assessment data to make instructional decisions. *Exceptional Children, 52*(3), 233-238.

Bradley-Klug, K.L., Shapiro, E.S., Lutz, J.G., & DuPaul, G.J. (1998). Evaluation of oral reading rate as a curriculum-based measure within literature-based curriculum. *Journal of School Psychology, 36*(2), 183-197.

Coladarci, T. (1992). Teachers' knowledge of what students' know: The case of reading. *Reading Improvement, 29*, 34-39.

Compton, D.L., Appleton, A.C., & Hosp, M.K. (2004). Exploring the relationship between test-leveling systems and reading accuracy and fluency with second-grade students who are average and poor decoders. *Learning Disabilities Research and Practice, 19*(3), 176-184.

Crawford, L., Tindal, G., & Stueber, S. (2001). Using oral reading rates to predict student performance on statewide achievement tests. *Educational Assessment, 7,* 303-323.

Cunningham, P.M., & Allington, R.L. (2003). *Classrooms that work: They can all read and write.* (3rd Edition). Boston, MA: Pearson Education.

Daly, E.J., & Martens, B.K. (1994). A comparison of three interventions for increasing oral reading performance: Application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27*(3), 459-469.

Danielson, L., Doolittle, J., & Bradley, R. (2005). Past accomplishments and future challenges. *Learning Disabilities Quarterly, 28*, 137-139.

Deno, S.L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*(3), 219-232.

Deno, S.L., Fuchs, L.S., Marston, D., & Shin, J. (2001). Using curriculum-based measures to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4), 507-525.

Deno, S.L., Marston, D., Shinn, M.R., & Tindal, G. (1983). Oral reading fluency: A simple datum for scaling reading disability. *Topics in Learning and Learning Disabilities, 2,* 53-59.

Deno, S.L., Mirkin, P.K., Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children, 49,* 36-45.

Deno, S.L., Fuchs, L.S., Marston, D., & Shin, J. (2001). Using curriculum-based measures to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4), 507-525.

Deshler, D.D., Mellard, D.F., Tollefson, J.M., & Byrd, S.E. (2005). Research topics in responsiveness to intervention: Introduction to the special series. *Journal of Learning Disabilities, 38*(6), 483-484.

Dolch, E.W. (1949) *Problems in reading.* Champaign, IL: The Garrand Press.

Dowhower, S.L. (1989). Repeated reading: Research into practice. *Reading Teacher, 42,* 502-507.

Dunn, E.K., & Eckert, T.L. (2002). Curriculum-based measurement in reading: A comparison of similar versus challenging material. *School Psychology Quarterly, 17*(1), 24-46.

Ehri, L.C., & McCormick, S. (1998). Phases of word learning: Implications for instruction with delayed and disabled readers. *Reading and Writing Quarterly, 14*(2), 135-163.

Elliott, J., Lee, S.W., & Tollefson, N. (2001). A reliability and validity study of the Dynamic Indicators of Basic Early Literacy Skills – Modified. *School Psychology Review, 30*(1), 33-50.

Engleman, S., & Bruner, E. (1978). *Reading Mastery.* Chicago: Science Research Associates.

Faulkner H.J., & Levy, B.A. (1994). How text difficulty and reader skill interact to produce differential reliance on word and content overlap in reading transfer. *Journal of Experimental Child Psychology, 58,* 1-24.

Fletcher, J.M., Coulter, W.A., Reschly, D.J., & Vaughn, S. (2004). Alternative approaches to the definition of learning disabilities: Some questions and answers. *Annals of Dyslexia, 54*(2), 304-331.

Fletcher, J.M., Stuebing, K.K., & Shaywitz, B.A. (1994). Validity of the concept of dyslexia: Alternative approaches to definition and classification. In K.P. van den Bos, L.S. Siegel, D.J. Bakker, and D.L. Share (Eds.). *Current directions in dyslexia research* (pp 31-43). Liesse, Netherlands: Swets & Zeitlinger.

Fletcher, J.M., Francis, D.J., & Foorman, B., (1997, October/November). Only 15-20 percent. *Reading Today,* p.18.

Foorman, B.R., Francis, D.J., Fletcher, J.M., & Lynn, A. (1996). Relation of phonological and orthographic processing to early reading: Comparing two approaches to regression-based, reading-level-match designs. *Journal of Educational Psychology, 88,* 639-652.

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

Foorman, B.R., & Moats, L.C. (2004) Conditions for sustaining research-based practices in early reading instructions. *Remedial & Special Education, 25*(1), 51-60.

Francis, D.J., Shaywitz, S.E., Stuebing, K.K., Shaywitz, B.A., & Fletcher, J.M. (1996). Developmental lag versus deficit models of reading disability: A longitudinal, individual growth curve analysis. *Journal of Educational Psychology, 88*(1), 3-17.

Frankenberger, W., & Fronzaglio, K. (1991). A review of states' criteria for identifying children with learning disabilities. *Journal of Learning Disabilities, 24*, 495-500.

Freeman, D.J., Kuhs, T.M., Knappen, L.B., & Porter, A.C. (1982). A closer look at standardized tests. *The Arithmetic Teacher, 29*, 50-54.

Fry, E.B. (1980). The new instant word list. *Reading Teacher, 34*, 284-289.

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

Fuchs, Fuchs, L.S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N., et al. (2001). Is reading important in reading-readiness programs? A randomized field trial. *Journal of Educational Psychology, 93*, 251-267.

Fuchs, L.S. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review, 33*(2), 188-192.

Fuchs, L.S. & Deno, S.L. (1991a). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58*(3), 232-243.

Fuchs, L.S. & Deno, S.L. (1991b). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children, 57*(6), 488-500.

Fuchs, L.S. & Fuchs, D. (1984). Criterion referenced assessment without measurement: How accurate for special education? *Remedial and Special Education, 5*(4), 29-32.

Fuchs, L.S. & Fuchs, D. (1986). Effects of systematic formative evaluation on student achievement: A meta-analysis. *Exceptional Children, 53,* 199-208

Fuchs, L.S. & Fuchs, D. (1997). Use of curriculum-based measurement in identifying students with disabilities. *Focus on Exceptional Children, 30*(3), 1-16.

Fuchs, L.S. & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research & Practice, 13*, 204-219.

Fuchs, L.S. & Fuchs, D.F. (1999). Monitoring students progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28*(4), 659-671.

Fuchs, L.S., Fuchs, D.F., Hamlett, C.L., Phillips, N.B., Karns, K., & Dutka, S. (1997). Enhancing students' helping behavior during peer-mediated instruction with conceptual mathematical explanations. *Elementary School Journal, 97*, 223-250.

Fuchs, L.S., Fuchs, D.F., Hamlett, C.L., & Stecker, P.M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal, 28*(3), 617-641.

Fuchs, L.S., Fuchs, D.F., Hamlett, C.L., Walz, L. & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22*(1), 27-49.

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

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

Germann, G. & Tindal, G. (1985). An application of curriculum based assessment: The use of direct and repeated measurement. *Exceptional Children, 52*(3), 244-265.

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

Good, R.H. & Jefferson, G. (1998). Contemporary perspectives on curriculum-based measurement validity. In M. Shinn (Ed.). *Advanced applications of curriculum based measurement* (pp. 61-88). New York: Guilford Publications.

Good, R.H. & Kaminski, R.A. (2002). *DIBELS oral reading fluency passages for first through third grades* (Technical Report No. 10). Eugene, OR: University of Oregon.

Good, R.H., Simmons, D.S., Kame'enui, E.J., Kaminski, R.A., & Wallin, J. (2002). *Summary of decision rules for intensive, strategic, and benchmark instructional recommendations in kindergarten through third grade* (Technical Report No. 11). Eugene, OR: University of Oregon.

Good, R.H., Simmons, D.C., & Kame'enui, E.J. (2001). The importance of 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*(3), 257-288.

Gresham, F. (1991). Conceptualizing behavior disorders in terms of resistance to intervention. *School Psychology Review, 20,* 23-36.

Gresham, F. (2001). *Responsiveness to intervention: An alternative approach to the identification of learning disabilities.* Paper presented at the Learning Disabilities Summit: Building a foundation for the future, Washington, D.C. (ERIC Document Reproduction Service No. ED458755).

Gresham, F. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley, L. Danielson, & D. Hallahan (Eds.). *Identification of learning disabilities: Research to practice* (pp. 467-519). Mahwah, N.J.: Erlbaum.

Groff, p. (1999). Decodable words in reading textbooks: Why are they imperative? The National Right to Read Foundation. Available at http://www.nrrf.org/27_decode_textbooks.htm. Accessed April 7, 2006.

Hallahan, D.P., & Mercer, C.D. (2002). Learning disabilities: Historical perspectives. In R. Bradley, L. Danielson, & D.P. Hallahan's (Eds.). *Identification of learning disabilities: Research to practice.* Mahwah, NJ: Lawrence Erlbaum.

Hasbrouck, J.E. & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children, 24*(3), 41-44.

Hayes, S.C., Nelson, R.O., & Jarrett, R.B. (1987). The treatment utility of assessment: A functional approach to evaluating assessment quality. *American Psychologist, 42,* 963-973.

Heller, K.A., Holtzman, W.H., & Messick, S. (Eds.). (1982). *Placing children in special education: A strategy for equity*. Washington, D.C.: National Academy Press.

Herman, R. (1995). *The Herman Method Set A.* Lincoln, Massachusetts: Lexia Learning Systems.

Hiebert. E.H., Pearson, P.D., Taylor, B.M., & Paris, S.G. (1998). Every child a reader topic 4: High-frequency words and fluency. Ann Arbor, MI: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED429269).

Hiebert, E.H. (1999). Text matters in learning to read. *Reading Teacher, 52*(6), 552-566.

Hiebert, E.H. (2000, April). The task for the first-grade texts: Have state policies influenced the content? Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.

Hiebert, E.H. (2002). Textbooks and model programs: Reading reform in the United States (pp. 57-174). In R. Fisher, G. Brooks, & M. Lewis (Eds.). *Raising standards in literacy.* New York: Rutledge Farmer.

Hiebert, E.H. (2005). The effects of text difficulty on second graders' fluency development. *Reading Psychology, 26,* 183-209.

Hiebert, E.H., & Fisher, C.W. (2002). *The critical word factor in texts for beginning readers: Effects on reading speed, accuracy, and comprehension.* Paper presented at the annual meeting of the American Educational Research Association. New Orleans, L.A.

Hiebert, E.H., & Fisher, C.W. (2005). A review of the National Reading Panel's studies on fluency: The role of text. *Elementary School Journal, 105*(5), 443-499.

Hiebert, E., Martin, L., & Menon, S. (2005). Are there alternatives in reading textbooks? An examination of three beginning reading programs. *Reading & Writing Quarterly, 21*(1), 7-32.

Hiebert, E.H., Pearson, P.D., Taylor, B.M., & Paris, S.G. (1998). Every child a reader topic 4: High-frequency words and fluency.

Hintze, J.M., & Christ, T. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review, 33*(2), 204-217.

Hintze, J.M., Christ, T.J., & Keller, L.A. (2002) The generalizability of CBM survey-level mathematics assessment: Just how many samples do we need? *School Psychology Review, 31*(4), 514-529.

Hintze, J.M., Callahan, J.E., Matthews, W.J., Williams, S.A., & Tobin, K.G. (2002). Oral reading fluency and prediction of reading comprehension in African American and Caucasian elementary school children. *School Psychology Review, 31*(4), 540-554.

Hintze, J.M., Daly, E.J., & Shapiro, E.S. (1998). An investigation of the effects of passage difficulty level on outcomes of oral reading fluency progress monitoring. *School Psychology Review, 27*(3), 433-446.

Hintze, J.M., Ryan, A.L. & Stoner, G. (2003). Concurrent validity and diagnostic accuracy of the Dynamic Indicators of Basic Early Literacy Skills and the Comprehensive Test of Phonological Processing. *School Psychology Review, 32*(4), 541-556.

Hintze, J.M., & Shapiro, E.S. (1997). Curriculum-based measurement and literature-based reading: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology, 35*(4), 351-375.

Hintze, J.M., Shapiro, E.S., & Lutz, J.G. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *Journal of Special Education, 28*(2), 188-202.

Hoffman, J.V., Sailors, M., & Paterson E.U. (2002). Decodable texts for beginning reading instruction: The year 2000 basals. (CIERA Report #1-016) Retrieved April 11, 2006 from the Center for Early Reading Achievement website: http://www.ciera.org/library/reports/index.html

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

Howell, K.W., Fox, S.L., Moorehead, M.K. (1993). *Curriculum-based evaluation: Teaching and decision making* (2nd ed.). Pacific Grove, CA: Brooks/Cole.

Jenkins, J.R. & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59*(5). 421-431.

Jenkins, J.R., & Pany, D. (1978). Standardized achievement tests: How useful for special education? *Exceptional Children, 44*, 448-453.

Jenkins, J.R., Peyton, J.A., Sanders, E.A., & Vadasy, P.F. (2004). Effects of reading decodable texts in supplemental first-grade tutoring. *Scientific Studies of Reading, 8*(1), 53-85.

Juel, C. & Roper-Schneider, D. (1985). The influence of basal readers on first grade reading. *Reading Research Quarterly, 20*, 134-152.

Keogh, B.K. (2005). Revisiting classification and identification. *Learning Disabilities Quarterly, 28*, 100-102.

Kirk, S.A. (1962). *Educating exceptional children.* Boston: Houghton Mifflin.

Koskinen, P.S., McCarthey, S.J., & Hoffman, J.V. (1995). The new basals: How are they different? *The Reading Teacher, 49*(1), 72-75.

Kovaleski, J.F. (2005). IDEA reauthorization includes RTI: Now what? *Communiqué, 34*(3), 26.

Lerner, J. (1988). *Learning disabilities: Theories, diagnosis and teaching strategies.* Dallas, TX: Houghton-Mifflin.

MacMillian, D.L. & Speece, D. (1999) Utility of current diagnostic categories for research and practice. In R. Gallimore, L. Hernheimer, D., MacMillan, D. Speece, & S. Vaughn (Eds.). *Developmental perspectives on children with high-incidence disabilities* (pp. 111-133). Mahwah, NJ: Lawrence Erlbaum.

Madelaine, A. & Wheldall, K. (2005). Identifying low-progress readers: Comparing teacher judgment with curriculum-based measurement procedure. *International Journal of Disability, Development, and Education, 52*(1), 33-42.

Markell, M.A., & Deno, S.L. (1997). Effects of increasing oral reading: Generalization across reading tasks. *The Journal of Special Education, 31*(2), 233-250.

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

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children, 52*(3), 266-276.

Marston, D., & Magnusson, D. (1988). Curriculum-based assessment: District-level implementation. IN J. Graden, J. Zins, & M. Curtis (Eds.). *Alternative educational delivery systems: Enhancing instructional options for all children* (pp. 137-172). Washington, D.C.: National Association of School Psychologists.

McGlinchey, M.T. & Hixson, M.D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33*(3), 193-203.

Mellard, D.F., Deshler, D.D., & Barth, A. (2004). LD identification: It's not simply a matter of building a better mousetrap. *Learning Disability Quarterly, 27,* 229-242.

Menon, S. & Hiebert, E.H. (1999). *Literature anthologies: The task for first-grade readers* (Report No. CIERA-R-1-009). Ann Arbor, MI: Center for the Improvement of Early Reading Achievement. (ERIC Documentation Service No. ED436754).

Menon, S. & Hiebert, E.H. (2005). A comparison of first graders' reading with little books or literature-based basal anthologies. *Reading Research Quarterly, 40*(1), 12-38.

Mesmer, H.A.E. (2005a). Introduction: Text accessibility and the struggling reader. *Reading and Writing Quarterly, 21,* 1-5.

Mesmer, H.A.E. (2005b). Text decodability and the first-grade reader. *Reading and Writing Quarterly, 21,* 61-86.

Muehl, S., & Forell, E.R. (1973). A follow-up study of disabled readers: Variables related to high school reading performance. *Reading Research Quarterly, 9,* 110-123.

Moustafa, M. (1997, October/November). Bravo to Allington. *Reading Teacher,* p.18.

National Joint Committee on Learning Disability. (2005). Responsiveness to intervention and learning disability. *Learning Disability Quarterly, 28*, 249-260.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel: Teaching Children to Read.* Washington, D.C.: Author.

Powell-Smith, K.A., & Bradley-Klug, K.L. (2001). Another look at the "C" in CBA: Does it really matter if curriculum-based measurement reading probes are curriculum-based? *Psychology in the Schools, 38*(4), 299-312.

Rashotte, C.A., & Torgensen, J.L. (1985). Repeated reading and reading fluency in learning disabled children. *Reading Research Quarterly, 20*, 180-188.

Rasinski, T.V. (1999). Exploring a method for estimating independent, instructional, and frustrational reading rates. *Reading Psychology, 20*, 61-69.

Reschly, D.J. (1988). Special education reform: School psychology revolution. *School Psychology Review, 17*(3), 459-475.

Rutter, M., & Yule, W. (1975). The concept of specific reading retardation. *Journal of Child Psychology and Psychiatry, 16,* 181-197.

Salmon-Cox, L. (1981). Teachers and standardized achievement tests: What's really happening? *Phi Delta Kappa, 62*, 631-634.

Shapiro, E.S. (1996). *Academic skills problem: Direct assessment and intervention* (2nd ed.). New York: Guilford Press.

Share, D.L., McGee, R., McKenzie, D., Williams, S., & Silva, P.A. (1987). Further evidence relating to the distinction between specific reading retardation and general reading backwardness. *British Journal of Developmental Psychology, 5*, 35-44.

Share, D. L., McGee, R., & Silva, P.A. (1989). IQ and reading progress: A test of the capacity of notion. *Journal of the American Academy of Child and Adolescent Psychiatry, 28,* 97-100.

Share, D.L. & Silva, P.A. (1986). The stability and classification of specific reading retardation: A longitudinal study from age 7 to 11. *British Journal o Educational Psychology, 56,* 32-29.

Shaywitz, S.E., Escobar, M.D., Shaywitz, B.A., Fletcher, J.M., & Mukuch, R. (1992). Evidence that dyslexia may represent the lower tail of a normal distribution of reading ability. *New England Journal of Medicine, 326*(3), 145-150.

Shinn, M.R., Gleason, M.M., & Tindal, G. (1989). Varying the difficulty of testing materials: Implications for curriculum-based measurement. *Journal of Special Education, 23*(2), 223-233.

Shinn, M.R., & Good, R.H. III (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review, 21*(3), 459-480.

Shinn, M.R., Good, R.H. III, & Stein, S. (1989). Summarizing trend in student achievement: A comparison of methods. *School Psychology Review, 18,* 356-370.

Shinn, M.R. & Marston, D. (1985). Differentiating mildly handicapped, low-achieving, and regular education students: A curriculum-based approach. *Remedial and Special Education, 6,* 31-45.

Shinn, M.R., Tindal, G. & Stein, S. (1988) Curriculum-based assessment and the identification of mildly handicapped students: A research review. *Professional School Psychology,* 4,

Shinn, M.R., Tindal, G., Spira, D., & Marston, D. (1987). Practice of learning disabilities as social policy. *Learning Disability Quarterly, 10*, 17-28.

Shinn, M.R., Ysseldyke, J.E., Deno, S.L., & Tindal, G.A. (1986). A comparison of differences between students labeled learning disabled and low achieving on measures of classroom performance. *Journal of Learning Disabilities, 19*(9), 545-552.

Siegel, L.S. (1989). IQ is irrelevant to the definition of learning disabilities. *Journal of Learning Disabilities, 22*, 469-479.

Singh, N.N., & Singh, N.N. (1988). Increasing oral reading proficiency through over-correction and phonic analysis. *American Journal on Mental Retardation, 93*, 312-319.

Slocom, T.A., Street, E.M., & Gilbert, G. (1995). A review of research and theory on relation between oral reading rate and reading comprehension. *Journal of Behavioral Education, 5*(4), 377-398.

Spache, G. (1953). A new readability for primary grade materials. *Elementary English, 53*, 410-413.

Stage, S.A., & Jacobsen, M.D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review, 20*(3), 407-419.

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

Stanovich, K.E. (1992). Developmental reading disorder. In S.R. Hooper, G.W. Hynd, & R.E. Mattison's *Developmental disorders: Diagnostic criteria and clinical assessment* (173-208). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stanovich, K.E., & Siegel, L.S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of phonological-core variable-difference model. *Journal of Educational Psychology, 86*(1), 24-53.

Stein, M., Johnson, B., & Gotlohn, L. (1999). Analyzing beginning reading programs: The relationship between decoding instruction and text. *Remedial and Special Education, 20,* 257-287.

Torgensen, J.K., Wagner, R.K., Rashotte, C.A., Rose, E., Lindamood, P., Conway, T., et al. (1999). Preventing reading failure in young children with phonological processing disabilities: Group and individual responses to instruction. *Journal of Educational Psychology, 91,* 579-593.

Tucker, J.A. (1985). Curriculum-based assessment: An introduction. *Exceptional Children, 52*(3), 199-204.

U.S. Department of Education (2005). Assistance to states for the education of children with disabilities. *Federal Register, 70*(118), 35782-35892.

U.S. Office of Education (1977). Education of handicapped children: Assistance to states: Proposed rulemaking. *Federal Register, 41*, 52404-52407.

Vadasy, P.F., Jenkins, J.R., & Poor, K. (2000). Effects of a first-grade tutoring program in phonological and early reading skills. *Journal of Learning Disabilities, 33,* 126-139.

Van Der Heyden, A.M. (2006). RTI myths and misrepresentations: Looking at data and experience. *Communiqué: Newspaper of the National Association of School Psychologist, 34*(4), 18-21.

Van der Wissel, A., & Zegers, I.C. (1985). Reading retardation revisited. *British Journal of Developmental Psychology, 3*, 3-9.

Vaughn, S., Linan-Thompson, S., & Hickman, P. (2003). Response to instruction as a means of identifying students with reading/learning disabilities. *Council for Exceptional Children, 69*(4), 391-409.

Vellutino, F.R., Scalon, D.M., & Lyon, G.R. (2000). Differentiating between difficult-to-remediate and readily remediated poor readers: More evidence against the IQ-achievement discrepancy definition of reading disability. *Journal of Learning Disabilities, 33*, 223-238.

Vellutino, F.F., Scanlon, D., Sipay, E., Small, S., Pratt, A., Chen, R., & Denckla, M. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology, 88*, 601-638.

Weintraub, F. (2005).The evolution of LD policy and future challenges. *Learning Disabilities Quarterly, 28*, 97-100.

Wiley, H.I., & Deno, S.L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26*(4), 207-214.

Wilson, B.A. (1996). *The Wilson Reading System*. Millbury, MA: Wilson Language Training.

Yule, W., Rutter, M., Berger, M., & Thompson, J. (1974). Over and under achievement in reading: Distribution in the general population. *British Journal of Educational Psychology*, *44*, 1-12.

Zeno, S.M., Ivens, S.H., Millard, R.T., & Duvvuri, R. (1995). *The educator's word fluency guide.* New York: Touchstone Applied Science Associates, Inc.

Appendix A

Oklahoma State University Institutional Review Board

Date:	Monday, March 26, 2007
IRB Application No	ED0735
Proposal Title:	The Effects of Decodability and High-Frequency Words on Curriculum-Based Measurement Oral Reading Fluency Outcomes for At-Risk Students
Reviewed and Processed as:	Expedited (Spec Pop)

Status Recommended by Reviewer(s): Approved Protocol Expires: 3/25/2008

Principal Investigator(s)

Staci Cumming
1400 N. Perkins Rd. #L88
Stillwater, OK 74075

Eric Mesmer
420 Willard
Stillwater, OK 74078

The IRB application referenced above has been approved. It is the judgment of the reviewers that the rights and welfare of individuals who may be asked to participate in this study will be respected, and that the research will be conducted in a manner consistent with the IRB requirements as outlined in section 45 CFR 46.

☑ The final versions of any printed recruitment, consent and assent documents bearing the IRB approval stamp are attached to this letter. These are the versions that must be used during the study.

As Principal Investigator, it is your responsibility to do the following:

1. Conduct this study exactly as it has been approved. Any modifications to the research protocol must be submitted with the appropriate signatures for IRB approval.
2. Submit a request for continuation if the study extends beyond the approval period of one calendar year. This continuation must receive IRB review and approval before the research can continue.
3. Report any adverse events to the IRB Chair promptly. Adverse events are those which are unanticipated and impact the subjects during the course of this research; and
4. Notify the IRB office in writing when your research project is complete.

Please note that approved protocols are subject to monitoring by the IRB and that the IRB office has the authority to inspect research records associated with this protocol at any time. If you have questions about the IRB procedures or need any assistance from the Board, please contact Beth McTernan in 219 Cordell North (phone: 405-744-5700, beth.mcternan@okstate.edu).

Sincerely,

Sue C. Jacobs, Chair
Institutional Review Board

Name: Staci Kanoelani Cumming Date of Degree: July, 2008

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: THE EFFECTS OF DECODABILITY AND HIGH FREQUENCY WORLDS ON CURRICULUM BASED MEASUREMENT ORAL READING FLUENCY OUTCOMES FOR AT RISK STUDENTS

Pages in Study: 109 Candidate for the Degree of Doctor of Philosophy

Major Field: Educational Psychology

Scope and Method of Study: The purpose of this study was to evaluate whether the word level composition of curriculum-based measurement reading passages significantly impacted the oral reading fluency performance of at-risk readers. Specifically, the impact of high-frequency and decodable words was evaluated on first-, second-, and third-grade students' absolute reading rate and growth profiles.

In the current study, passages containing significantly different levels of high-frequency and decodable words were administered to two groups of students eight times over the course of four weeks. Students read aloud from the passages for one minute and words correct per minute were calculated.

Findings and Conclusions: Results from a mixed-model repeated-measures analysis of variance indicated that the percentage of high-frequency and decodable words did not significantly impact either absolute reading rate or reading growth profiles. Visual inspection of grade level data suggested that differences might exist in the reading performance of students who read from the high-frequency and those who read from decodable passages; however, insufficient power prevented statistical analyses of these grade-level differences. Implications for practice and future research are discussed

Advisor's Approval: _____

VITA

Staci Kanoelani Cumming

Candidate for the Degree of

Doctorate of Philosophy

Dissertation: THE EFFECTS OF DECODABILITY AND HIGH-FREQUENCY WORDS ON CURRICULUM-BASED MEASUREMENT ORAL READING FLUENCY OUTCOMES FOR AT-RISK STUDENTS

Major Field: Educational Psychology
Option: School Psychology

Biographical:

Personal Data:

Born in Wailuku, Hawaii, on January 15, 1980, the daughter of John and Maude Cumming.

Education:

Graduated from Kaʻahumanu Hou High School, Puʻunene, Hawaii in May 1998; received Bachelor of Arts degree with a major in Psychology and a minor in Sociology from Oral Roberts University, Tulsa, Oklahoma in May 2002; received Master of Science degree in Applied School Psychometrics from Oklahoma State University, Stillwater, Oklahoma in July, 2004. Completed the requirements for the Doctorate of Philosophy degree with a major in Educational Psychology and option in School Psychology at Oklahoma State University in July, 2008.

Experience:

Employed by Oklahoma State University, Department of Educational Psychology as a graduate teaching assistant; completed school-based and clinic-based practicums in consultation, intervention, and assessment at Oklahoma State University; employed by Oklahoma State University, Department of Education as a school-based literacy coach; completed pre-doctoral internship at Cypress-Fairbanks Independent School District.

Professional Membership:

American Psychological Association, National Association of School Psychologists, Texas Association of School Psychologists.