CORRELATIONS AND PREDICTIVE ABILITY OF ORAL READING FLUENCY AND THE WILSON READING SYSTEM ON END OF YEAR ASSESSMENTS

A Dissertation
Submitted
to the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

By
Kristin Zielinski
August, 2010

Examining Committee Members:

Catherine Fiorello, Advisory Chair, School Psychology
Joseph Ducette, Psychological Studies in Education
Erin Rotheram-Fuller, School Psychology
James Connell, School Psychology
Ken Thurman, Special Education Curriculum, Instruction, and Technology in Education

ABSTRACT

Title: Correlations and Predictive Ability of Oral Reading Fluency and the Wilson Reading System on End of Year Assessments

Candidate's Name: Kristin Zielinski

Degree: Doctor of Philosophy

Temple University, 2010

Doctoral Advisory Committee Chair: Dr. Catherine Fiorello

To insure academic success for students with learning disabilities, it is critical that educators are able to make timely and effective instructional decisions. The focus of this study was to evaluate the relationship and effectiveness of two different progress monitoring tools, a measure of oral reading fluency and the Wilson Reading System's level documentation against measures of reading achievement. Reading achievement was specifically, examined using cluster scores from the Woodcock Johnson Tests of Achievement – Third Edition (WJ-III) Broad and Brief Reading domains. The current study investigated archival data of 51 students, grades four through eight, enrolled in a small, private school for students with learning disabilities during the 2008-2009 school year. Oral reading fluency benchmark and regular progress monitoring scores were collected from students and the progress monitoring scores were converted into an overall rate of improvement. Data from students receiving instruction in the Wilson Reading System were also collected as levels mastered throughout the school year. All students received pre- and post-testing on the WJ-III Reading domain. No significant relationships were found between oral reading fluency rate of improvement and Wilson level. Additionally, only the oral reading fluency spring benchmark significantly predicted end of year WJ-III performance after pre-test scores were taken into account. Supplemental

analyses did find that students who completed levels seven and eight in the 2008 – 2009 school year consistently performed better than peers who completed levels two through six at all oral reading fluency benchmark periods. Limitations to the study and implications for future research and practice are discussed.

ACKNOWLEDGMENTS

A fellow graduate student shared an anonymous quote with me once, “Wisdom is not measured in test scores but rather in resiliency, compassion, and courage.” This has continued to resonate with me through the years. I would like to first thank my family for instilling in my these and other qualities that I needed to complete my graduate program. My mother and late father, Barbara and Michael, have provided me with continual love and support throughout my entire life. I owe so much to both of them. My smart and successful brother, Michael, who has kept me driven to complete while laughing along the way. Jill, my sister, has always been a role model at the ready with love and encouragement. My late grandmother, Sophie, whose fighting spirit is worthy of emulation. Without my family, I would not have accomplished what I have in my life and these accomplishments would have little meaning without my family to share them with.

I would also like to thank my friends and fellow runners who have offered me support, respite, and laughter through the years. They have all helped me develop a healthly balance of work and play, distraction and focus, as well as mind and body. I have been lucky to establish so many lasting friendships through all stages of my life. May we all continue to adventure together.

Finally, I would like to thank those that have supported me through this process. Temple University’s School Psychology faculty are incredibly intelligent, caring, and supportive. Dr. Erin Rotheram-Fuller for her guidance and support, Dr. Patricia Feuerstien for her confidence in me and teaching me what it is to think on my feet, Dr. Catherine Fiorello for all her time and energy as my committee chair. Additionally, I have been fortunate to work for the past two years with very talented teachers and staff. My colleagues have helped make completing this dissertation possible. Thank you Dr. Grace Ashton and Dr. Nicole Schwartz for all the support and understanding.

TABLE OF CONTENTS

PAGE

ABSTRACT iii

ACKNOWLEDGMENTS v

LIST OF TABLES vii

CHAPTER

1. INTRODUCTION 1
2. REVIEW OF LITERATURE 8
3. METHODS 40
4. RESULTS 48
5. DISCUSSION 58

REFERENCES CITED 69

LIST OF TABLES

Table Page

1. *Table 4 -1. Descriptive Statistics for Demographic Variables.* 48

2. *Table 4-2. Totals and Percentages of Highest Wilson Level by Group and Grade..* 49

3. *Table 4-3. ANOVA results for Highest Wilson Level Achieved by Oral Reading Fluency Rate of Improvement.* 50

4. *Table 4-4. Means and Standard Deviations of Group Differences in Oral Reading Fluency Rate of Improvement.* 50

5. *Table 4-5. Means and Standard Deviations of WJ Pre and Post-test W Scores by Highest Wilson Level Achieved..* 51

6. *Table 4-6. ANCOVA results for Highest Wilson Level Achieved by WJ Basic Reading Skills Cluster w Scores.* 52

7. *Table 4-7. ANCOVA results for Highest Wilson Level Achieved by WJ Broad Reading Skills Cluster w Scores.* 52

8. *Table 4-8. Partial Correlations for Basic Reading Skills Cluster and Highest Wilson Level Achieved Controlling for Pre-test Basic Reading Skills Cluster.* 53

9. *Table 4-9. Regression Table for Broad Reading Skills Cluster and ORF Spring Benchmark.* 54

10. *Table 4-10. Means and Standard Deviations of Oral Reading Fluency Benchmarks* 55

11. *Table 4-11. Highest Wilson Level Achieved by Oral Reading Fluency Benchmark Period MANOVA Results.* 56

12. *Table 4-12. Means and Standard Deviations of Wilson Group Differences in Slope of Oral Reading Fluency Benchmark Scores.* 57

13. *Table 4-13. Highest Wilson Level Achieved by Oral Reading Fluency Benchmark Slope ANOVA Results* 57

14. *Table 5-1. Steps of the Wilson Reading System's Scope and Sequence and Corresponding Word Attack Items on the Woodcock Johnson Tests of Achievement – Third Edition.* 62

CHAPTER 1

INTRODUCTION

For the past 20 years, reading difficulty has been considered a major public health concern (Sweet, 2004). The ability to read is a critical skill not only through a student's educational career, but through his or her entire life. Reading instruction begins with a child's first school experience and is a primary focus in the elementary grades. Cunningham and Stanovich (1997) conducted a longitudinal study that followed a group of students from 1st through 11th grade and found that even students that lag in first grade but catch up in third or fifth have a good prognosis for their level of future reading engagement. Good, Simmons, and Kame'enui (2001) have found that one foundational reading skill that differentiates successful from unsuccessful readers is the ability to read connected text accurately and fluently, also called oral reading fluency. Recent research that has focused on improving student achievement has identified the importance of tools that routinely and efficiently monitor student progress toward goals of reading success and can be used as formative evaluations (Busch & Reschly, 2007; Jenkins et al., 2009; Vaughn et al., 2008). The formal assessment of curriculum-based Oral Reading Fluency (ORF) provides educators with valuable information as to student progress toward their reading goals. Oral Reading Fluency is a set of standardized procedures used to monitor student reading progress over time.

During the early elementary years, a number of students are identified as having a learning disability, which is a neurological disorder characterized by persistent difficulties learning specific skills among which are basic reading skills. These students often require intensive intervention over and above their typically developing peers to achieve reading success. Curriculum-based measurement, originally created for students receiving special education services, is an important progress monitoring tool for all. It is commonly used to improve individual instructional programs and, recently, to predict performance on high stakes assessments (Deno, 2003). Models of Response to Intervention (RtI) call for continual progress monitoring so educators can make informed, data-based decisions regarding a student's educational programming. For students that have been moved through the tiers of RtI and qualify for special education services, progress monitoring is an important tool in the development and implementation of their Individual Educational Plans (IEP).

While there are many interventions that exist that seek to remediate reading difficulties, one such program, the Wilson Reading System, is an explicit, highly intensive method that requires extensive teacher training to execute properly. Students are moved through different syllable patterns of the English language sequentially through 12 different steps. Thus, the steps in the Wilson Reading System do not correspond with grade level performance and initial steps of this program may not translate to immediate improvement as measured by oral reading fluency.

This study sought to contribute to the current state of research and understanding of the role of oral reading fluency as students with learning disabilities progress through the Wilson Reading System. Specifically, the relationship between oral reading fluency

and Wilson level attained along with the contribution each makes to predicting end-of-year performance on a norm-referenced assessment were evaluated.

Statement of Purpose

The current study sought to contribute to the understanding of what measures of oral reading fluency can and can not tell educators about reading achievement in an accuracy-based reading intervention. Fluency-based progress monitoring is critical information often used in practice to make important educational decisions. The purpose of this study is to evaluate the appropriateness of a measure of oral reading fluency when students require instruction in the Wilson Reading Program, an intensive, accuracy-based reading intervention.

Context

This study evaluated the relationship of oral reading fluency by the levels achieved in an explicit, structured, supplemental reading instruction (Wilson) as well as student reading performance on a norm-referenced end-of-year assessment. The data were collected to monitor student progress in a small, independent school serving students with learning disabilities. The school served 105 students with learning disabilities in grades one through nine by providing them with intensive interventions, one of which is the Wilson Reading System during the 2008-2009 school year.

Research Questions and Hypotheses

This study addresses three research questions, as follows:

Question 1: Does Rate of Improvement on measures of oral reading fluency differ based on the Wilson level?

Hypothesis 1: Due to the nature of the Wilson Reading System's scope and sequence, the mastery of initial levels of word study may not be significantly related to increased performance on measures of oral reading fluency.

Question 2: Is there a difference in the prediction of WJ scores across Wilson levels, after accounting for last year's WJ sores?

Hypothesis 2: Due to the nature of the Wilson Reading System, it is hypothesized that students that have reached higher Wilson levels will have significantly higher performance on the WJ Basic Reading Skills Cluster compared to students who have not progressed as far through the Wilson Reading System. The same holds true for the Broad Reading Skills Cluster; however, not to the same magnitude since the Broad Skills Cluster includes measures of reading fluency and passage comprehension.

Question 3: Does Wilson end-level for the 08-09 school year predict Woodcock Johnson Tests of Achievement – Third Edition (WJ-III) w-scores in Basic Reading Cluster administered in spring, 2009 scores over and above the prior year's WJ performance?

Hypothesis 3: Since the Wilson Reading program focuses heavily on word reading skills, it is hypothesized that improved decoding skills will have a direct impact on the WJ-III Basic Skills Reading Cluster. Thus, the Wilson level a student achieves should predict his or her reading performance on the WJ-III.

Question 4: For this population, is oral reading fluency benchmark scores or rate of improvement a better predictor of WJ-III w-score in Reading skills over and above the previous year's WJ w-scores?

Hypothesis 4: Prior research suggests that benchmark oral reading fluency performance will be the single best predictor. However, in the present population, it is hypothesized that rate of improvement through instructional level material will be a better predictor because it may be more sensitive to changes in performance.

This paper will review the characteristics of students with learning disabilities that impede successful development of reading skill, reading theory and reading difficulty, response to intervention, Wilson Reading System, development of curriculum-based measurement and progress monitoring assessments, and curriculum-based measurement's ability to predict student performance.

Definitions

Basic Reading Skills – Student ability to phonetically decode words in the English language as well as identify non-phonetic sight words. Within basic reading skills the term decoding refers to the ability to convert letters and combinations of letters into sounds associated with the English language.

Wilson Reading System **–** Is a supplemental reading curriculum designed to teach decoding and encoding (spelling) skills to upper elementary students through adults. While the primary focus of the Wilson program is at the word level, its developers state that it also includes sight word instruction, fluency, vocabulary, oral expressive language development, and comprehension. Instruction is delivered in a sequential and multisensory manner that requires mastery at each step. For more specific information as to the scope and sequence of the Wilson steps, please refer to the appendix.

Curriculum-Based Measurement (CBM) – This term refers to a specific set of standardized administration and scoring procedures for measuring oral reading fluency or MAZE performance. Additional characteristics of CBM include multiple equivalent passages, technical adequacy in terms of reliability and validity, and time efficiency.

Oral Reading Fluency (ORF) **–** The number of words a student can correctly read in connected text with a given time limit when given a grade-level appropriate passage.

Aimsweb Oral Reading Fluency Passages - A specific CBM procedure for assessing the acquisition of reading skills. They are designed to be short (one minute) fluency measures used to regularly monitor the development of reading skills. The text of this report refers to the passages from the Oral Reading Fluency subtest from

Edformation, Inc. (Howe and Shinn, 2002) as Aimsweb passages, which is the term used broadly in the reading practice community.

Progress Monitoring – A process by which data are collected and used to gauge reading growth over time. Progress monitoring tools need to be sensitive to student growth over a brief period of time, have multiple forms, as well as to be quick and easy to administer. Curriculum-based measurement is one measure used for progress monitoring.

Learning Disabilities - The regulations for Public Law (P.L.) 101-476, the Individuals with Disabilities Education Act (IDEA), formerly P.L. 94-142, the Education of the Handicapped Act (EHA), define a learning disability as a "disorder in one or more of the basic psychological processes involved in understanding or in using spoken or written language, which may manifest itself in an imperfect ability to listen, think, speak, read, write, spell or to do mathematical calculations."

Woodcock Johnson Tests of Achievement – Third Edition (WJ-III) – This is an individually administered, norm-referenced measure of oral language and general academic achievement. It was designed to provide methods for determining areas of student strength and need. For the present study, this measure represents the criterion variable. While the Woodcock-Johnson Tests of Achievement assess student performance over a wide range of skills, only the Broad Reading and Basic Reading Skills will be used in the present study. The Broad Reading Cluster is compromised of the Letter-Word Identification, Reading Fluency, and Passage Comprehension subtests. The Word Attack and Letter-Word Identification subtests are used to derive the Basic Reading Skills Cluster.

CHAPTER 2

LITERATURE REVIEW

Learning Disabilities

The Individuals with Disabilities Education Act (IDEA) is the main federal program authorizing state and local funding for special education and related services for children with disabilities, including learning disabilities. In 2004, President Bush signed the Individuals with Disabilities Education Improvement Act (P.L. 108-446), a major reauthorization and revision of IDEA. The new law preserves the basic structure and rights guarantees of IDEA but also includes new provisions regarding how schools can determine whether a child has a specific learning disability and may receive special education services. The legal definition of a learning disability in IDEA 2004 reads as follows:

> "The term 'specific learning disability' means a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which disorder may manifest itself in the imperfect ability to listen, think, speak, read, write, spell, or do mathematical calculations. Such term includes such conditions as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia. Such term does not include a learning problem that is primarily the result of visual, hearing, or motor disabilities, of mental retardation, of emotional disturbance, or of environmental, cultural, or economic disadvantage."

The field of special education is now facing a policy debate about the identification procedures used for learning disabilities (Kavale et al., 2008). With the passing of the Individual with Disabilities Education Improvement Act (IDEIA 2004),

school districts are now able to adopt an IQ-discrepancy or Response to Intervention model for diagnosing learning disabilities. Critics of the IQ-discrepancy model typically state that the model fails to identify students early enough, fails to differentiate between poor students with a discrepancy and poor readers without, and lacks instructional implications (Case, Speece, & Molloy, 2003). Fuchs and Fuchs (1998) conceptualized how Response to Intervention (RtI) could be used to identify learning disabilities using the Dual Discrepancy (DD) model. Students would be identified as learning disabled if the child's level of performance and rate of improvement were both below classmates following an intervention. However, many questions still remain if districts were to use a response to intervention approach, such as how to define and measure responsiveness, whether gains be maintained in regular education, and whether a general education or an intensive intervention be used (Brueggemann et al., 2008).

The majority of the population to be used in the present study include students with learning disabilities that impact their ability to acquire age-appropriate basic reading skills, thus making the Wilson Reading System an appropriate intervention. Students with reading difficulties typically have deficits in one or more of three cognitive domains: language comprehension, phonological processing, or fluency/naming speed (Fletcher, Lyon, Fuchs & Barnes, 2007). Students that have difficulties in language comprehension struggle to build connections among sounds, words, meanings, and ideas while students with fluency deficits may be accurate in their understandings of sounds or word knowledge, but are extremely slow. The final deficit is in phonological awareness, and these students have weakness in their underlying phonological skills. Students that have a

'double deficit' in fluency and phonological awareness are typically the most difficult to remediate (Wolf & Bowers, 2000).

Models of the Reading Process

Models of reading have been developed and explored by members of the education, psychology, and neuroscience communities that have reflected views of reading through the respective lenses of their disciplines. Below is a review of select reading models from three different perspectives: top-down information processing, bottom-up information processing, and interactive processing.

Some contemporary models of reading processes stem from constructionist perspectives that find that learning occurs when new information is integrated with existing knowledge, meaning that reading is more driven by what is in the reader's head than on the page. This is also referred to as 'top-down' information processing. Two of the most popular theories that have come from this perspective are Psycholinguistic and Whole Language theory. Both theories see the role of language and background knowledge as crucial to successful reading and feel that it is an educator's role to immerse children in authentic experiences, rich language, and quality literature (Tracey & Morrow, 2006). Whole Language theory is a very broad approach to language development and has been heavily influenced by Dewey (Goodman & Goodman, 1981). Reading is considered a natural extension of language development and the whole language approach de-emphasizes reading readiness skills, because they are viewed as abstract minutia, and focuses on comprehension (Goodman & Goodman, 1981). Psycholinguistics, on the other hand, focuses more closely on language's role in reading.

Psycholinguists find that readers use a series of internalized cueing systems to guide them towards reading comprehension. The cues are processed by syntactic, semantic, and graphophonic systems that utilize grammar, vocabulary, and letter-sound knowledge respectively to guide reading. Graphophonic cuing, or decoding, is found to be a final cueing system after predictions and background knowledge have been ineffective. A study by Goodman (1965) compared the ability of 100 students in grades first through third to read isolated wordlists and the same target words in connected text. He presented descriptive statistics that indicated 69% of first grade students assessed were able to correctly read two thirds of the words initially missed when presented in isolated wordlists compared to 81% of second graders and 91% of third graders. Goodman hypothesizes that the improved student performance when presented with connected text is the students' ability to utilize language cues that were 'outside of words,' which excludes decoding. Criticism of these theories includes that such learning is internal and difficult to measure, and is rooted in prediction and hypothesis testing, which is not always applicable to all types of text.

Another set of models has been the product of cognitive psychology's Information Processing Theory, which views learning as the processing, storage, and retrieval of knowledge (Slavin, 1997). Also known as 'bottom-up' models, these models posit that reading takes place as information is processed from lower to higher-order stages (Tracey & Morrow, 2006). While there are many variations of models based on information processing theory, two of the most prominent are Atkinson and Shiffrin's 1968 Information Processing Model and LaBerge and Samuels 1974 Automatic Information Processing Model (Tracey & Morrow, 2006). The Information Processing

Model describes the reading process as beginning with a reader's exposure to print. Information is first perceived in the Sensory Register or Sensory Memory. This is the stage that visual stimuli, in the form of letters and words, are interpreted. Next, information is temporarily held in Working Memory and combined with knowledge from long-term memory. Information that receives sufficient attention will be encoded into Long-Term Memory. Slavin (1997) states that Long-Term Memory can be understood as "the component of memory in which large amounts of information can be stored for long periods of time" (p.191). This entire process is facilitated by executive control processes.

LaBerge and Samuels developed their Automatic Information Processing model of reading in the mid-70s. Prior to 1974, the majority of these theories were composed of two interdependent elements, decoding and comprehension (Shinn et al., 1992), but then LaBerge and Samuels (1974) added in the third element of automaticity. Their model introduced five separate information processors: visual, phonological, episodic, semantic, and attention. The concept of automaticity in this model develops when the processors function proficiently and free up limited cognitive resources to be focused elsewhere. Both theories assume that words must be decoded automatically before they can be understood or comprehended.

The two previous categories of reading models can be criticized as being linear and one-directional. Rumelhat (1977) created Rumelhart's Interactive Model. Similar to 'bottom up' information processing models, the reading process begins with visual input; however, from that point, information is simultaneously processed by a variety of processors that include word order (syntax), message construction (semantics), visual

input (orthographics), and word knowledge (lexical). Stanovich (1980) later added that when one processor is disabled or overloaded, the other processors will compensate.

Another contemporary model is the Parallel Distributed Model. It has been regularly updated with the most recent version by Plaut and McClelland (1993). Three of the four reading processors are organized in a triad where information interacts through an orthographic (print recognition), phonological (sounds associated with words), and meaning (vocabulary) processor (Moats & Davidson, 2008; Tracey & Morrow, 2006). Extending off the meaning processor is a context processor that constructs and monitors meaning at the phrase, sentence, paragraph and full-text levels. The context processor is highly dependent on an individual's prior, or background, knowledge. The model is based on the notion that information is stored as connections between the different processors and these connections become stronger and faster with repeated connections (Moats & Davidson, 2008; Tracey & Morrow, 2006).

A Framework of Reading Instruction

While the select theories of the cognitive process of reading have been discussed, all include processing of the words in print (decoding) and the ability to derive meaning from these words (comprehension). Since LaBerge and Samuels (1974) added the element of automaticity, other models of reading, such as the Parallel Distributed Model, have included similar concepts. The natural extension of a discussion of the models of reading is how these processors and skills are developed through instruction. The elements of decoding, fluency, and comprehension are reviewed and presented through the framework of the instructional hierarchy developed by Haring, Lovitt, Eaton, and Hansen (1978). This instructional hierarchy presents learning as a series of linear steps.

First, a student must acquire basic skills, then perform these skills quickly and easily in order to reach the highest step, which is the ability to use the skills in the widest possible range.

There is remarkable consistency across theories of reading that find overall reading achievement is causally influenced by proficiency in foundational skills (Good, Simmons, & Kame'enui, 2001). The importance of foundational skills is fairly intuitive; if a student is unable to identify words in print, there is little chance of reading success. These foundational skills typically include phonological awareness, decoding, and a developmentally appropriate sight word bank.

Phonological Awareness is the ability to identify, think about, and manipulate parts of words including syllables, onsets and rimes, and phonemes (Torgesen & Hudson, 2006). Decoding skills typically refer to an individual's ability to assign letters in print to sounds and ultimately letter patterns to words. The primary difference between tasks of phonological awareness and decoding is that phonological tasks are entirely auditory in nature; decoding tasks require the representations of sound in letters, words, and print.

A sufficient sight word bank is critical in acquiring reading skill in the English language. The English language is orthographically complex because it is rooted in multiple languages, which results in sounds of our language represented in multiple ways. Consider the /f/ sound as a prime example because it can be orthographically represented as an f, ff, ph, and sometimes gh. The different spellings of words are often related to their Anglo-Saxon, Latin, or Greek roots. Sight words do not always follow conventional spelling rules and cannot be decoded (e.g., the, what, or island).

Haring and colleagues' (1978) next step is fluency, which the National Reading Panel (2000) describes as the ability to read orally with accuracy, speed, and with proper expression or prosody. Accuracy builds from the instructional hierarchy and speed refers to the rate individual words are recognized as well as fluidity moving through text (Torgesen & Hudson, 2006). Prosody is used to describe the rhythmic and tonal aspects of speech applied to oral reading (Torgesen & Hudson, 2006). The role of fluency is best understood through the Cognitive Processing Theory, which suggests that when students are able to read fluently they are able to free up cognitive resources and concentrate more on deriving meaning from a passage. Torgesen and Hudson (2006) find that a fluent reader can maintain the performance of accurate reading at a conversational rate with appropriate prosody for long periods of time, retain the skill after long periods of no practice, and can generalize across texts.

There is general consistency among theories regarding the role of fluency in reading achievement. There is even a substantial body of evidence generated from research involving children with reading disabilities to support the importance of reading fluency. In 2003, de Jong and van der Leij studied characteristics of Dutch children with dyslexia. Their research determined that in a language with a transparent orthography, these children could acquire high levels of phonological awareness, but regardless of the transparency of the language, impairments in fluency are a main characteristic of dyslexia. However, the assessment of fluency is a source of great contention for both researchers and educators. Prosody has been difficult to assess, leading to a dearth of evidence of its role in comprehension, and the few research studies focusing on prosody have not been able to define the role of reading with prosody in reading comprehension.

In a study by Schwanenflugel et al. (2004) children's oral reading was recorded, reformatted into a digital wave file, and analyzed for characteristics of prosodic speech compared to adult controls. The findings of this study did not support the role of prosody as either an assist to or a reflection of adequate reading comprehension over the relationship between decoding speed and comprehension.

The more easily assessed and empirically-supported aspect of fluency is reading rate and accuracy, which the National Reading Panel (2000) describes as automaticity. Automaticity is "the processing of information that ordinarily requires long periods of training before the behavior can be executed with little effort or attention. (National Reading Panel, p. 3-7, 2000)." This definition stresses the role of practice in learning to read. Further analysis by the National Reading Panel found that data reveal gradual, continuous improvement over time.

The final steps of the instructional hierarchy are generalization and adaptation, which can be understood as the abilities to use foundational skills in the widest possible range and modify and adapt skills to new situations for the purpose of comprehension (Haring et al., 1978). Davis (1944) conducted the first psychometric analysis of comprehension skills – word meanings, word meanings in context, follow passage organization, main thought, answer specific text-based questions, text-based questions with paraphrasing, draw inferences from context, literary devices, and author's purpose. These skills assist the reader in building the intended representation of information or ideas that the author intended.

Additional Reading Factors and Assessment

Differences in developmental reading trajectories can be explained by a predictable and consequential series of reading-related activities that begin with difficulty in foundational skills, progress to fewer encounters with and exposure to print, and culminate in decreased motivation and desire to read (Stanovich, 1986; 2000). The amplification of any initial advantage (SES, health, early literacy) that leads to cumulative differences that widen pre-existing gaps is also known as the Matthew Effect (Ceci & Papierno, 2005). This mechanism is embedded in the social and cognitive contexts of schooling (Cunningham & Stanovich, 1997). The instructional hierarchy does not incorporate the role of additional factors such as language and cognitive skills as major contributors to reading comprehension. Reading experience has a large impact on comprehension ability, but is it the ability to decode words or the exposure to language structures that contributes more?

Aspects of language that are critical to reading comprehension as well as listening comprehension including vocabulary knowledge (semantics) and an understanding of grammar and sentence structure (syntactics) (Cain & Oakhill, 2007). After word reading, language skills were found to be the second strongest predictors of reading comprehension (Goff, Pratt, & Ong, 2005). These findings support the bottom-up and interactive reading models presented earlier that highlight the role and importance foundational skills have to support reading comprehension. Along the same lines, Schatsneider and colleagues (2004) sought to identify the major reading, cognitive, and linguistic skills that contribute to individual differences in performance on the reading portion of the Florida statewide assessment, called the FCAT, at third, seventh, and tenth grades. Using a dominant analysis, which is a method for comparing the importance of

predictors in a multiple regression, standardized measures of oral reading fluency were most dominant and able to account for 55% of the variance in third grader's Florida statewide assessment. Verbal knowledge and reasoning was the second most dominant, followed by nonverbal reasoning and memory factors.

Where members of the research and education communities have diverged is the definition and influence fluency, as currently assessed, has on comprehension. The efficacy of oral reading fluency as an appropriate measure of reading proficiency and comprehension has strong support (Shinn, Good, Knutson, Tilly, & Collins, 1992). Good et al. (2001) examined the relation of CBM oral reading fluency to the reading process from a theoretical perspective in grades three and five. Oral reading fluency correlated positively with standardized measures of literal ($\underline{r} = 0.64$) and inferential ($\underline{r} = 0.63$) reading comprehension. Using confirmatory factor analysis, evidence indicated oral reading fluency fits into current theories of reading and that foundational skills, including the ability to read with accurately and fluently with connected text, can differentiate between successful and unsuccessful readers.

Spear-Swerling (2006) evaluated the contributions of oral language, word accuracy, and single word reading speed to passage comprehension at third grade. The researchers found that third grade oral reading fluency was significantly related to reading comprehension ($\underline{r} = 0.57$), but with the other two variables accounted for, oral language explained the largest amount of variance in passage comprehension (24%). When third grade variables were placed into a regression analysis to predict student performance on fourth grade statewide assessments, all variables combined accounted for 60% of the variance ($R^2 = .62$, adjusted $R^2 = .574$, $\underline{F}(6, 50) = 13.6$, $\underline{p} < .001$). However,

only two variables had significant final beta weights, and only two variables accounted for unique variance in fourth-grade reading with the other variables controlled: Passage Comprehension (B= 0.401, $\underline{p}$ <0.01, $\Delta R2$=0.067, $\underline{p}$ <0.01) and Oral Reading Fluency (B=0.407, $\underline{p}$ <0.05, $\Delta R2$=0.041, $\underline{p}$ <0.05).

Recent research has been establishing an empirical base supporting measures of oral reading fluency as moderately to highly correlated with state assessment measures in Arizona (Wilson, 2005), Colorado (Shaw & Shaw, 2002), Florida (Buck & Torgesen, 2003), North Carolina (Barger, 2003), Ohio (Vander, Meer, Lentz, & Stollar, 2005), Pennsylvania (Shapiro et al., 2006), and Washington (Stage & Jacobsen, 2001). All of these studies found significant correlations between oral reading fluency and overall reading performance ranging from 0.6 to 0.8, which is considered a moderate to high relationship. These studies are discussed in greater detail in the CBM and High Stakes Assessment section of this literature review.

A recent study by Shapiro et al. (2008) looked at the predictive ability of oral reading fluency and a group-administered measure of reading comprehension called the 4Sight Benchmark Assessments, and the Pennsylvania System of School Assessment (PSSA) on 1,000 students in grade three through five. The 4Sight Benchmark Assessments are one hour, group reading comprehension assessments. They are designed to resemble the PSSA reading assessment in format and results are delivered by assigning students to similar categories as the PSSA (Advanced, Proficient, Basic, and Below Basic). A very strong relationship was observed between fall and winter ORF across all grades ($\underline{r}$=0.92–0.94), while a relatively weaker but strong correlation existed on the 4Sight measure between both times ($\underline{r}$ =0.72–0.80). Analysis of the relationship between

ORF and 4Sight and the statewide achievement test indicated that 4Sight was more strongly correlated with PSSA ($\underline{r}$ = 0.71-0.78) than ORF ($\underline{r}$ = 0.64-0.75) across all grade levels. The combination of ORF and the 4Sight measure yielded the best predictive strength for students whose performance on both measures was strong enough to be categorized as low risk or proficient respectively. Individually, the 4Sight measure reportedly performed as well as or better than the measure of oral reading fluency for students in grades three and four. There was no difference between the Area Under Curve (AUC) for oral reading fluency and 4Sight Assessments for students in the fifth grade. The Area Under Curve is an indicator of the overall performance of a diagnostic test's ability to accurately discriminate true positives and negatives. The data indicate that oral reading fluency alone is able to do a reasonable job at identifying pass/fail rates for students who are at either end of the spectrum (at-risk or low risk), but a poor job with students whose performance falls in between.

Samuels (2007), on the other hand, has concerns about the misuse of the term fluency and feels that ORF are measures of speed and not fluency. In his commentary, Samuels is critical of the widespread use of brief measures of oral reading fluency, where instructional decisions can be based on student performance on tasks where no comprehension is required. He expresses feelings shared by some in the research community that by focusing so much on speed students are sent the wrong message and, ultimately, measures of ORF push students to read at a rate past what could have possibly aided in reading comprehension. LaBerge and Samuels (1974) in their description of the automaticity theory of reading note that it is the simultaneity of decoding and comprehension that is the essential characteristic of fluency. Secondary characteristics of

fluency such as speed, accuracy, and expression are indicators, but not essential. Samuels suggests that presence of comprehension questions following measures of oral reading fluency would decrease students' reading speed, but would be more representative of current reading achievement.

Research by Yovanoff and colleagues (2005) evaluated the importance of vocabulary and oral reading fluency in reading comprehension for five cross-sectional samples of approximately 6,000 students in grades four through eight. Overall, researchers concluded that fluency was less important as grade levels increase, particularly after fourth grade. However, data were collected from two different samples that reflected students taking two different forms of the district's fluency and comprehension assessment. For students in the district, oral reading fluency was a significantly more important predictor for one sample than the other. The samples had been matched for each school in the district based on gender, ethnicity, SES, and previous reading achievement, which suggest differences in passage difficulty and remains an area in need of further study.

Evidence suggests that students with difficulty decoding words also struggle with comprehension, but is fluency with foundational skills sufficient to guarantee reading success? Research by Reuzel and Hollingsworth (1993) found that intervention designed to increase oral reading fluency increased performance on three of four comprehension measures used over a round-robin reading control group.

While oral reading fluency does not directly relate to content and performance standards used to design state-wide assessments, there is mounting evidence that indicates oral reading fluency can be used as an identifier of students at-risk for poor

performance and is unbiased to ethnicity (Hintze et al., 2002). This study evaluated the performance of 136 students in grades two through five. The sample consisted of 65 African-American students and 71 Caucasian students. The socioeconomic status was 47% low income, 33% middle, and 20% in the high-income range. A study conducted by Stage and Jacobsen (2001) used Hierarchical Linear Modeling (HLM) to determine the predictive ability ORF had on 173 fourth graders' Washington state performance assessment. This study used curriculum-based measures derived from basal texts. There were overall medium correlations between the Washington state performance assessment and oral reading fluency scores across fall, winter, and spring scores, that were statistically significant and ranged from 0.43 to 0.44. In addition to evaluating the difference between different benchmark performances, the researchers also posited that a student's overall slope was more indicative than at any particular time point. The slope represents the rate of improvement across benchmark assessments over time. Contrary to their hypothesis, the slope of benchmark scores was less predictive than an individual benchmark time period ($\underline{r} = 0.26$). Additional studies that specifically review this topic with oral reading fluency measures are discussed later in this review.

Response to Intervention

Response to Intervention (RtI) is a means of providing early intervention services for all students at risk for reading failure using continuous progress monitoring such as curriculum-based measures as one aspect of identification. Many of the same policy makers behind Response to Intervention initiatives were also responsible for Reading First Initiative of NCLB 2002 (Fuchs & Fuchs, 2006). The basic tenants of RtI are identification, monitoring, multi-tiered instruction, and the notion of an intervention as a

test (Fuchs & Fuchs, 2006). By identifying and supporting students prior to experiencing marked and sustained reading failure, contemporary educators and researchers hope to avoid the Matthew Effect described earlier by Ceci and Papierno (2005) and Stanovich (1986, 2000).

Underlying these tenants is the need for assessment instruments that have the ability to accurately and reliably identify those students who will be most likely to experience reading difficulties in the future based on their performance on critical indicators (Good, Simmons, & Kame'enui, 2001), must be sensitive to small changes, and must not confound future data with practice or carryover effects.

Different models of RtI have anywhere from two to four tiers. The tiers signify an increase in the 'intensity' of intervention the students are provided whether in the explicitness, frequency, or duration of instruction; group size; or practitioner qualifications. Instructional decisions for students identified as at-risk can be made using a standard protocol or problem-solving approach. By using a standard protocol, students move through tiers based on their response to previously established, more intensive, instruction. The critique of a standard protocol approach is that judgments are made as to a student's response to instruction that may have not been appropriate to their instructional needs (Daly, Martens, Barnett, Witt, & Olsen, 2007). A problem-solving approach, on the other hand, is more tailored to an individual student's areas of need. This is typically carried out by the school-based intervention system who must utilize available resources to carry out specific interventions chosen with a student's academic needs in mind. Marston (2005) reviews a number of studies utilizing standard protocol

and problem solving models and reports decreases in special education referrals in both models.

Proper implementation of RtI calls for the assessment of every student with a screener that demonstrates utility for predicting end-of-year performance on a high-stakes test (Fuchs & Fuchs, 2006). This can be accomplished through previous year's high stakes test performance (lowest 25th percentile), a different norm-referenced measure, or a criterion-referenced measure (like Aimsweb). Identified at-risk students are then monitored on a more frequent basis and provided with more intensive instruction. Proponents of Response to Intervention stress the importance of multiple sources of information when evaluating progress. Curriculum-based measurement is designed so that students are exposed to material that will be instructed over the school year with different skills mixed in. The strength of this method of assessment is that measures are quick and easy to administer (Cusumano, 2007) and provide information as to the maintenance of skills (Singer-Dudek & Greer, 2005). Norms have been developed by Hasbrouck and Tindal (1992) among others. Also, with the help of web-based curriculum-based measurement websites such as Aimsweb or DIBELS, aggregate growth norms can be collected and updated annually.

Another source of information can be norm-referenced achievement tests. These measures are individually administered, broad indicators designed to measure students' relative position compared to their same-aged or same-grade peers. One example of this type of assessment is the Woodcock Johnson Tests of Achievement. These measures compare a student's performance to a national average, which is valuable information, but also means a lack of overlap with local or classroom curriculum (Wright, 2001). One

positive aspect of norm-referenced assessments is that they provide an ideal index of individual differences or can function as a criterion for gauging capacity to recognize such differences in other measures. However, pre or post-test gains from achievement tests are not satisfactory criteria as measures of growth (Jenkins et al., 2009). Also, these types of assessments are not sensitive to small skill developments or gains across short time spans, which makes it more difficult to detect and react to inadequate growth.

Assessment practices contribute to increased levels of reading achievement ONLY when they can answer important questions for teachers and schools as well as enable informed instructional decision-making (Coyne & Harn, 2006). Data-based decision making is the cornerstone of Response to Intervention and best practice for serving students receiving special education. Data identify students that may be at risk for reading failure, the impact of an intervention, movement through tiers, as well as students that may not fall within the proficient range as measured by statewide assessments. However, even though data evaluation is so critical to the ongoing success of a response to intervention framework, more research is needed.

When oral reading fluency data are being evaluated, information is typically presented in a line graph with words correctly read over time. Students are given target performances based on either a norm-based rate of improvement or the benchmark normative performance for the student's grade level. Using a student's baseline performance, a line is drawn from their initial performance to their end goal. This line is called the target line. Data evaluation occurs through the visual inspection of this graph where the student's actual performance is compared to this target line.

Many articles that review Response to Intervention procedures highlight the importance of the periodic comparison of ORF scores to the goal line (Stecker, Lembke, & Foegen, 2008) as well as decision rules about whether to extend or discontinue a diagnostic trial period (Vaughn & Fuchs, 2003). Wright (2001) describes decision rules as "firm guidelines for interpreting the 'pattern' of data points on a graph (p. 3-5)." However, different decision-making frameworks exist and suggest that educators should not evaluate progress until a program has been in place for three to four weeks (Fuchs & Vaughn, 2005; Stecker, Lembke, & Foegen, 2008; Wright, 2001) or as many as seven to ten data points (Shinn, Good, & Stein, 1989). The difficulty with such decision-making frameworks that call for fixed data points before decisions are made is that these models do not specify frequency of administration. For example, Shinn (2007) suggests collecting data points one to two times per week with decisions made after seven to ten data points. Based on the above suggestions, a student could be in an intervention for a range three and a half weeks to as long as two and a half months before the decision-making framework calls for evaluation.

In a highly cited article, Vaughn, Linan-Thompson, and Hickman (2003) explored a response to treatment model that provided supplemental reading instruction to 45 regular education, second grade students identified as at-risk. The student's progress was reviewed every 10 weeks and students were exited from the intervention if they were able to meet criteria of 50 words per minute on a curriculum-based oral reading measure, a median score of 55 on the Test of Oral Reading Fluency, and a passing score on the Texas Primary Reading Inventory Screener. Students were provided with 35 minutes of daily intervention that covered writing, instructional level reading, fluency, word

analysis, and phonemic awareness. The results of the study found that 10 of the students were able to meet exit criteria after 10 weeks, 14 after 20 weeks, and 10 after 30 weeks. Eleven students were not able to meet exit criteria. All of the students that exited after 10 weeks continued to make appropriate gains in oral reading fluency and nine of the 14 students that exited after 20 weeks continued to have an acceptable trajectory in oral reading. There was no information provided as to the gains these students met early in the intervention; however, it begins to shed light on the length of time required to evaluate if a student is benefiting from a comprehensive reading intervention.

Additionally, many of these decision frameworks do not account for differences in interventions. Following this section is a review of the Wilson Reading System, its scope and sequence as well as research base. With an in-depth understanding of the Wilson Reading System, it will become apparent that either the goal setting or data evaluation stages must account for the specific intervention in a problem-solving model, especially for students with learning disabilities.

Wilson Reading System

The Wilson Reading System was developed by Barbara Wilson and published in 1988 (Wilson Language Training Corp., 2004). It is a direct, multi-sensory, supplemental language instruction designed for students in the upper elementary grades through adulthood. The program is supplemental to regular reading instruction and is organized around the six syllable types found in the English language and teaches the structure of words in a systematic and cumulative manner (Wilson Language Training Corp., 2004). In addition to direct reading instruction, the students have extensive practice with controlled vocabulary for decoding and spelling application. Comprehension is addressed

with vocabulary work and through the development of visualization skills with the reading of literature to students (Wilson Language Training Corp., 2004).

The Wilson Reading System teaches a specific sequence of skills. At each step, students learn skills to mastery for both reading and spelling before progressing to the next step. Steps 1-2 emphasize phonological awareness. These steps teach letter/sound correspondence for closed syllables, the identification of sound units, phoneme segmentation, and blending. At the end of step 2, up to 6 sounds are blended and sequenced in a syllable. Step 3 introduces multisyllabic work, combining closed syllables. Steps 4-6 teach vowel-consonant-"e": open; and consonant "le" syllables. In Step 6, Anglo-Saxon suffix endings are taught also. Steps 7-12 teach higher level word structure. Words containing sound options are introduced for reading and spelling. "R" controlled and diphthong/vowel digraph syllables are taught, as well as additional work with spelling rules and suffixes (O'Connor & Wilson, 1995). Throughout the progression through Wilson steps, instruction delivery reserves time in each lesson for vocabulary and comprehension development.

Evidence of effectiveness is reported in multiple sources. However, the majority of these publications are books or book chapters describing the conceptualization, the process of delivering the Wilson Reading Program, or offer case study evidence. There is a dearth of evidence published in peer-reviewed journals. Currently, there is one study out of nine reviewed that met the evidence standards of the What Works Clearinghouse (WWC), an initiative of the U.S. Department of Education's Institute of Education Sciences to provide clear and concise information on instructional programs and interventions. Studies that provide strong evidence for an intervention's effectiveness are

characterized as *Meet Evidence Standards*. In order to meet evidence standards, a study has to be a randomized controlled trial or a quasi-experiment with one of the following three designs: quasi-experiment with equating, regression discontinuity designs, or single-case designs (IES, 2008).

Torgeson et al.'s (2006) study was conducted to evaluate the effectiveness of four different interventions (Wilson Reading, Failure Free Reading, Spell Read, and Corrective Reading) for upper elementary school students with reading problems. The study used data collected during the 2003 –2004 school year on 772 students in third and fifth grade across 27 school districts in western Pennsylvania. Of the larger sample, 162 students received instruction in a Wilson Reading program that was modified to include only word reading instruction. The students were selected based on screening criteria that identified students as 'struggling readers.'

Sample information provided found that 33% of students in the sample, evenly distributed across third and fifth grade, were diagnosed with a learning or another disability. The study did not evaluate these students as a subgroup, but instead dissected the sample based on high and low performance on Word Attack and receptive vocabulary as measured by the Peabody Picture Vocabulary Test – Third Edition (PPVT-III). The results found a significant impact of the modified Wilson Reading program for students in the third grade as well as students with very high or very low receptive language PPVT-III scores at baseline. Lastly, the effects of the Wilson Reading program were not significant for students that had both low Word Attack and low PPVT-III scores at baseline.

Overall, evidence indicated that the Wilson Reading produced statistically significant improvement on performance on the Woodcock Reading Mastery Test – Revised (WRMT-R), with effect sizes at $r^2 = 0.59$ and $r^2 = 0.29$ for third and fifth grade respectively. No significant improvements were found on Word Identification or Passage Comprehension subtests on the WRMT-R as well as on the Group Reading Assessment and Diagnostic Evaluation (GRADE) for students that received the Wilson Reading Program. Additionally, even though Torgeson and colleagues (2006) met the evidence standards of the WWC, it is important to mention the study used a modified Wilson program that focused exclusively on word-level intervention.

Other studies of the efficacy of the Wilson Reading System that did not meet requirements of the What Works Clearinghouse include two studies that utilized data collected in the Wilson Teacher Training process. The first study, by O'Connor and Wilson (1995) evaluated data on 220 students in grades three through twelve across schools in Massachusetts, New Jersey, and Maine. Data collected in this study included student's raw score performance on pre- and post-tests on the Woodcock Reading Mastery Test and the Woodcock Reading Mastery Test – Revised Assessment's Word Attack, Word Identification, Passage Comprehension, and Total Reading. Pre and post t-tests on raw scores revealed significant gains on the Word Attack and Passage Comprehension subtests. The magnitude of the change was reported for all four subtests as improvement in grade level. While not explicitly stated, it is assumed that the researchers used Grade Equivalents of raw scores and made statements such as "Significant gains in Word Attack were obtained; the average gain was 4.6 grade

Levels… The average word attack gain of 4.6 grade levels indicates that the Wilson Reading System greatly improved students' decoding abilities (O'Connor & Wilson, 1995, pg 250)." Unfortunately, grade equivalent scores lack the statistical properties to make these comparisons meaningful.

The second study also analyzed data collected on 375 students in grades three to eight through the Wilson Teacher Training process during the 1999-2000 and 2000-2001 school years (Wood, 2002). Woodcock Reading Mastery age-based standard scores were used as a pre- and post-test. Results of the entire sample mean performance on subtests and cluster scores on the Woodcock Reading Mastery Assessment yielded significant gains in Word Identification, t (350) = -12.35, $p < .0001$; Word Attack, t (348) = -22.56, $p < .0001$; Passage Comprehension, t(348) = -13.72, $p < .0001$; and Total Reading Cluster, t(348) = -15.69, $p < .0001$.

In sum, the Wilson Reading System is a direct, multi-sensory, supplemental language instruction program of reading and spelling skills. The collection of unpublished articles available on the Wilson Language website do not use procedures that can stand up to statistical scrutiny, although the absence of evidence-based research is not an indicator that the Wilson Reading System is ineffective. The Wilson Reading System has a wealth of clinical observations to support its claims of effectiveness even though it still lacks empirical support published in peer-reviewed, scholarly journals.

Curriculum-Based Measurement (CBM)

Academic success is conceptualized as an interaction of ability, motivation, and instruction (Walberg & Tsai, 1983). While passionate, this conceptualization is difficult to quantify for the purpose of formative or summative assessment of student

achievement. The ‘essence of reading’ is generally considered reading comprehension, or the ability to derive meaning from text (Durkin, 1993; Shinn et al., 1992). Students with learning disabilities often have difficulty achieving age-appropriate reading achievement without supplemental support and instruction. To allow schools the opportunity to provide timely remediation to students identified with learning disabilities, it is essential to understand the components of reading achievement as well as to identify tools that facilitate the judgment of a students’ response to intervention or those that require a change in instruction. Curriculum-based measurement, originally developed in the 1970s and 1980s by Deno and associates in the Institute for Reasearch and Learning Disabilities at the University of Minnesota, is a vital tool to this end.

Curriculum-based measures, which are a set of specific standardized procedures, were initially designed to measure the effectiveness of special education interventions (Deno, 2003). Currently, the use of curriculum-based measurement has expanded to include screening, pre-referral evaluation, placement in remedial or special education programming (Deno, 2003), improving individual instructional programs, and predicting performance on high-stakes assessment (Good, Simmons, & Kame’enui, 2001).

Curriculum-Based Measurement has steadily gained attention since its inception in the 1970s. Originally, CBM was developed by special educators for special education students (Deno, 2003). The central purpose of monitoring student progress is to gauge reading growth, which is a critical statistic in judging response to intervention or the effectiveness of an intervention. One of the many advantages of these procedures is that it is quick and readily repeated. Jenkins et al. (2009) state the importance of these characteristics when working with students with learning disabilities, “Students with

learning disabilities struggle with learning to a degree that outstrips special education resources, a fact that makes special education teachers sensitive to the tension between using lesson time for teaching versus assessing" (p. 160). Since CBM tools also have multiple, equivalent forms it allows for more frequent monitoring. In theory, this enables inadequate growth to be detected and reacted to more quickly. However, judgments of adequate versus inadequate growth must be considered in the context of the intervention. To draw conclusions too prematurely may lead to an effective intervention being abandoned before the student has time to respond.

Deno, Fuchs, Marston, and Shin, (2001) illustrated how curriculum-based measurement can be used to establish growth standards for students with disabilities. Initially, researchers took a convenience sample of 2,675 regular education and 324 special education students in grades one through six and evaluated their growth rate in oral reading fluency. All students were evaluated on grade-level appropriate passages three times throughout the school year; the median score at each time point was used in the analyses. There was no information provided as to the specific disability status of the special education students.

The first analysis calculated average weekly increase for students by grade and special education status. Differences in growth rates existed at first grade; however, the difference progressively decreased until grades five and six, where regular and special education students having virtually identical growth rates. The non-special education first grade student population had an average growth rate of 1.8 words per week, then decreased to 1.66, 1.18, 1.01, .58, and .66 for grades two through six respectively. Special education students, on the other hand, averaged .83 words per week, with .57 in second

grade, .58 in grades three through five, and .62 in grade six. These rates remained similar even when initial performance was controlled. However, the researchers were not content to declare such unsatisfactory growth rates as the norm for students with disabilities.

In the second portion of the article, the researchers analyzed existing research literature for studies that used CBM data to identify how much progress special education students made when provided with effective treatments. Five criteria were adopted for use in selection appropriate studies: the treatment had to last at least 10 weeks, the study had to report pre and post-test data using CBM, report separate outcomes for students with disabilities, utilize a control or contrast design, and report statistically significant improvement for at least one treatment condition. Only five studies met the selection criteria, and the researchers reported that the mean improvement with an effective intervention was at 1.39 words per week for students in grade two. By no means definitive, this line of research begins to establish that with effective intervention, students with disabilities may be able to achieve similar growth rates to their typically developing peers.

Curriculum Based Assessment and High Stakes Testing

Measures of oral reading fluency have been used extensively in research as well as practice and are often an outcome measure in program evaluation. Below is a review of research on different curriculum-based measurement instruments relevant to predictive power on various statewide assessments.

AIMSweb is a progress monitoring system based on direct, frequent and continuous student assessment. The results can be reported to educators via a web-based data management and reporting system to determine response to intervention (Pearson

Ed. Inc, 2008). Research on the relationship of oral reading fluency measures and statewide assessments has been repeatedly conducted in field studies since 2001. Many of the studies use Pearson Correlations to analyze their data, which looks at the linear relationship between the two measures. Overall findings across studies report a moderate to strong positive correlation (0.60 - 0.74) between oral reading fluency and third grade statewide standardized assessments in North Carolina, Florida, Pennsylvania, Ohio, Colorado, and Arizona (Barger, 2003; Buck & Torgeson, 2003; Shapiro, 2006; Vandeer Meer, Lentz, & Stollar, 2005; Wilson, 2005; Wood, 2006). These results indicate that higher scores on these measures are associated with higher scores on the statewide assessment. Additionally, oral reading fluency also provides reading targets called cut scores that indicate students with low risk, some risk, or are at risk for reaching subsequent literacy goals (Good et al., 2002). In addition to overall correlations at third grade, certain researchers have extended their analyses to include the evaluation of oral reading fluency's correlations among different grades, benchmark assessments, and subgroups that will be discussed below.

Shaw and Shaw (2002) conducted analyses to correlate Colorado state assessments with fall, winter, and spring oral reading fluency benchmark assessments. The researchers found in Colorado that all three scores correlated significantly with student performance on the third grade state assessment given in the spring, with the spring oral reading fluency benchmark having the highest correlation of 0.8. Additionally, the three time-periods were significantly correlated with each other, ranging from 0.89 - 0.93. Additional analyses indicated that cut scores above and below 90 words correct per minute were able to provide a high probability of correct prediction of success

or failure on the statewide assessment. In other words, 91% of students that were able to read over 90 words per minute in the spring assessment had passed the Colorado statewide exam. Only 27% of students that read below 90 words per minute passed the exam.

Barger (2003) also correlated oral reading fluency measures with the North Carolina End of Grade Reading Assessment. Data from this analysis were from the spring benchmark assessment and the End of Grade Reading Assessment given one week later. Results found a high correlation of 0.73, which is consistent with other statewide assessments. When Barger (2003) looked at student performances at different cut scores, which are scores that determine if a student is low, some, or at-risk, he found that 100% of students scoring 100 words per minute or more performed at the proficient level on the statewide exam and 50% of students scoring 70-99 reached the proficient level. This indicates that the strength of prediction of students identified as some risk is not as clear.

Wilson (2005) collected oral reading fluency on 241 third grade students and correlated the data with their performance on the reading portion of the Arizona state assessment. There was a strong correlation (.74) between spring benchmark ORF and overall reading performance. Wilson (2005) also analyzed subgroup performance based on ethnicity, gender, free/reduced lunch status, and English language learners. While all subgroups had positive and significant correlations at $\underline{p}<.01$, the strongest correlations were among Hispanic students ($\underline{r} = 0.78$), ELL students ($\underline{r} = 0.78$), students receiving free lunch ($\underline{r} = 0.74$), and males ($\underline{r} = 0.72$).

Buck and Torgeson (2003) also evaluated subgroup performances when comparing the relationship between 1,102 third graders' oral reading fluency and the

Florida statewide assessment. The ORF data collected were gathered a month after the Florida statewide exam and had a significant correlation of 0.70. The correlation was also significant and similar among students of different ethnicities with correlations of 0.70, 0.62, and 0.74 for white, African American, and Hispanic students respectively. However, typical cut scores of 110 words per minute were not as predictive for African American and Hispanic students. In other words, minority student performance relative to cut scores was not as predictive of success on the Florida statewide assessment. In another study from Florida, Roehrig and colleagues (2008) found exactly the same overall correlations between oral reading fluency and the Florida statewide assessment ($\underline{r}$ = 0.70). Using revised cut scores, the researchers were able to predict performance on the Florida assessment equally well regardless of student socioeconomic, language, or ethnicity.

Wood (2006) conducted a concurrent validity study that evaluated oral reading fluency winter benchmark assessments and the performance of 281 students in grades three, four, and five on Colorado's statewide assessment. Overall correlations between ORF and statewide assessment performance were strong for third ($\underline{r}$ = 0.70), fourth ($\underline{r}$ = 0.67), and fifth grades ($\underline{r}$ = 0.75). Additionally, he also found that previous year's statewide assessment data and current year ORF accounted for 62% of the variance in fourth grade statewide performance and 70% of the variance for fifth grade student performance. Wood did not evaluate subgroup difference primarily due to limited variability in the sample.

A large N study was conducted by Baker and colleagues (2008) that evaluated data from approximately 2,400 students in grades one through three in Oregon. Using

Growth Curve Analysis, data were analyzed by collapsing cohorts to determine the consistency of oral reading fluency over time, as well as using second grade ORF and standardized assessment to predict performance on a third grade standardized assessment. Results relevant to the present study found that oral reading fluency scores from six different time points from fall of second grade to spring of third grade were moderately correlated with a norm-referenced standardized assessment in the spring of third grade (r = 0.58, 063, 063, 0.65, 0.68, 0.67). In the best fitting Growth Curve Model, the oral reading fluency intercept, slope, and second grade standardized assessment accounted for 59% of the variance in scores.

Shapiro and colleagues (2006) conducted correlations of data from Pennsylvania's System of School Assessment (PSSA) and oral reading fluency data in a small suburban and moderately sized urban/suburban district. The overall findings of the study found statistically significant correlations between oral reading fluency and Total Reading Performance across districts with correlation coefficients approaching 0.70. Correlations were also provided for students at the third and fifth grade at each of the three benchmark time-periods for each district. The highest correlation for third grade students was in the spring benchmark assessment (r = 0.67), which is the closest time point to the administration of the PSSA. However, when the same analyses were run for fifth graders, the highest correlation was the winter benchmark score which ranged across districts from .64 to .69. The correlations between fifth grade oral reading fluency and the PSSA varied across districts and time-periods with the lowest being fifth graders from district 2 in the fall (r = 0.25) to the winter scores for district 1 (r = 0.69).

Oral reading fluency measures are technically adequate and have moderate to strong positive correlations with multiple statewide assessments. Across the studies, a number of patterns emerge. While different studies used different benchmark periods in their analyses, typically the highest correlation was with the winter or spring scores. Also, researchers evaluated the predictive strength of cut scores at each benchmark period finding that overall established cut scores are fairly predictive for students with either very high or low reading fluency; however, the strength diminishes with students identified as 'some risk.' Lastly, in Shapiro's (2006) article, the correlation between 5th graders' oral reading fluency and PSSA performance was more variable across time and school districts.

In sum, Curriculum-based measurement can play a critical role in RtI and fits theoretically into models of reading and reading instruction. The quick and easy detection of progress on a measure that highly correlates with high-stakes achievement measures and assists educators in providing timely and effective intervention. However, special populations of students, namely those with disabilities, often require a level of intensity and specificity that may not translate into immediate change in rate of improvement on oral reading fluency measures. In fact, the Wilson Reading System typically requires three years to move students through the entire program. Decision rules should consider type and intensity of intervention before changes are made. The present study hopes to explore patterns in student's rate of improvement across steps of the Wilson Reading System to assist in decisions regarding a student's instructional programming.

CHAPTER 3

METHOD

Participants

The participants in this study consisted of a subsample (N=51) of 115 students enrolled in a small private school for students with learning disabilities. The data used in this project are historical data collected by the school to monitor progress and assist in educational decision-making. Participants were selected based three different criteria: instruction in the Wilson Reading System during the 2008 – 2009 school year, regular progress monitoring using the Aimsweb progress monitoring system, and pre and post-test scores from the Woodcock Johnson Tests of Achievement. The final sample was 51 students in grades four through eight with 73% male. Wilson instruction was supplemental to the school's core literature-based reading instruction. In addition, some students also participated in the Read Naturally program.

The target school is a 1st through 11th grade private school servicing students with learning disabilities located in an urban setting in Pennsylvania that serves students from surrounding urban and suburban school districts. Admission to the school requires a documented learning disability diagnosed through a psychoeducational evaluation. The school currently employs over 50 teachers and staff and reports a 1: 6 student teacher ratio.

Measures

Aimsweb

The Aimsweb oral reading fluency passages are a set of standardized, individually administered measures designed to assess accuracy with grade level and instructional level passages. These are research supported, strong predictors of later reading achievement. Aimsweb Oral Reading Fluency (ORF) assesses how many words a student can accurately and fluently read in one minute when presented with connected text.

Oral Reading Fluency Benchmark Assessments: Oral Reading Fluency (ORF) was measured by evaluating the number of correct words read per minute on a grade level passage from AIMSWEB. These assessments took place three times per year; September, January, and May. The examiner in the present study took each student out of the classroom to a quiet area to read three passages for one minute each. If a student did not know a word or hesitated for more than three seconds, the examiner would say the word and mark it as an error. If the student did not correctly read a word, omitted a word, or substituted a word it was also marked as an error on the administrator copy of the passage. After reading the three passages, the student was sent back to the classroom and the examiner calculated the words correct and errors for each passage. However, only words correct per minute were recorded as a student's benchmark score. The median words correct score across the three passages was documented as the final reading rate.

These assessments would typically take place over a two-week period. There were no integrity checks on the examiner's procedure and, therefore, there is no information on the accuracy of AIMSWEB data collection. The examiner is a Masters level Certified

School Psychologist that received specific training in benchmark and curriculum-based measurement through graduate coursework and the Aimsweb training workbook.

Oral Reading Fluency Survey Level Assessment: At the time of the September benchmark period students performing below the 25th percentile according to AIMSWEB aggregate norms were also administered a survey level assessment. Per AIMSWEB survey level assessment procedures, students were administered single passages below their grade level and downward until their performance fell within the average range on a passage. When student performance fell above the 25th percentile of the national aggregate norms for the passage level given, that passage level became the level at which they were monitored for the remainder of the school year. Error and scoring criteria remained the same as the benchmarking procedure.

Oral Reading Fluency Progress Monitoring Passages: All students were monitored either weekly or every two weeks at the survey level or grade appropriate level regularly from September through May. The level at which students were monitored was dependent on student performance at the fall benchmark and survey level assessment. Data were collected by the examiner or the students' reading teachers, all of whom hold special education teaching degrees. Teachers were trained by the examiner and the procedures for training are described in the fidelity section (see below). These scores will be used to calculate the slope of the end-of-year rate of improvement variable as well as the slope of rate of improvement through each Wilson level achieved.

Woodcock Johnson Tests of Achievement – Third Edition assessment: The Woodcock Johnson Tests of Achievement – Third Edition (WJ-III) is a standardized achievement test developed by Riverside Publishing Company. The primary purpose is

to assess the achievement level on various academic skills of individual students as they compare to other students their chronological age.

Students in the current study take the Woodcock Johnson Tests of Achievement – Third Edition annually throughout their career at the school. Each assessment is administered individually and takes approximately one hour. Students are given between 12 and 15 subtests depending on grade level. W-scores represent both a person's ability in addition to the task difficulty and are generally considered to be a growth scale. They can be interpreted as an 'absolute' rather than relative measure of academic performance (e.g., changes in standard scores reflect changes in group position, while changes in w-scores reflect criterion-referenced changes). For the purpose of the present study, w-scores from the Letter Word Identification, Word Attack, Reading Fluency, and Passage Comprehension subtests, as well as the Broad Reading and Basic Reading Skills Cluster w-scores will be used for analysis. The Broad Reading Cluster is composed of the Letter-Word Identification, Reading Fluency, and Passage Comprehension subtests. The Word Attack and Letter-Word Identification subtests are used to derive the Basic Reading Skills Cluster.

Standardized assessment procedures were carried out as designed in the administrator's manual by the examiner or a clinical psychology post-doctorate/school psychology intern, both of whom have the required graduate-level training in educational assessment and diagnostic decision making. Assessments took place individually in a quiet room and were administered individually.

The standardization of the WJ-III used a large, nationally representative sample of 8,818 subjects in over 100 geographically diverse U.S. communities, with 4,783 in the K through 12th sample.

During the norming process, the reliability was calculated for all of the WJ Achievement subtests except for speeded tests using the split-half procedure. The split-half procedure compared the consistency between odd and even test items for each individual subtest. Review of median statistics across each technical age level find a high level of reliability with all reading subtest coefficients at or above r = 0.80 (McGrew & Woodcock, 2001). For speeded subtests such as reading fluency, a special test-retest study was completed. Reliability coefficients for speeded reading subtests fell between r = 0.76 and 0.96 across age groups (McGrew & Woodcock, 2001).

McGrew and Woodcock (2001) recently published a normative update for the scoring of WJ-III subtests. It provides age-based norms by month from ages 24 months to 19 years and by year from ages 2 to 90+ years. The technical manual presents a considerable amount of evidence supporting the validity of scores from the test, noting that the earlier versions of the battery have also been shown to have validity. The WJ III ACH content is similar to other achievement tests in subject areas and to established practice in schools (McGrew & Woodcock, 2001). A concurrent validity study by Martelle and Smith (1994) evaluated the similar WJ-Revised edition and the Wechsler Individual Achievement Test for students identified with learning disabilities. The broad reading domain scores between the two assessments were moderately to highly correlated (r = 0.70 to r = 0.81). These results suggest that the WIAT and WJ-R measure reading

skills in similar ways as evidenced by strong correlations between the global reading scales of both tests.

Procedures

Data for this project were collected within the school for the purpose of progress monitoring. The first source of data was individual student oral reading fluency assessments that were conducted by the examiner. These assessments were conducted in September, January, and May of the 2008-2009 school year as a standard practice within the school to determine students' level of achievement and improvement in their individual instructional programs. The second data source, the Woodcock Johnson Tests of Achievement – Third Edition, were collected by the examiner or a clinical psychology post-doctorate student. Finally, information regarding progress through the Wilson Reading System was collected from the students' Wilson Instructors' lesson plans and attendance records. Once all the data were aggregated and de-identified, they were released for the purpose of this secondary analysis.

Wilson Reading System – The school from which data were collected requires all teachers to document student progress through the program. All students new to the Wilson Reading program begin at substep 1.3 (please refer to the appendix for specific substep information). Before students are permitted to move onto the next substep, they must complete an exit mastery assessment, scoring 100% accuracy in reading a list of real words and a list of nonsense words. Data generated for the present study included variables of the starting and ending Wilson level for the 2008-2009 school year and average days spent in each level.

Data Collection: Data for analysis were assembled by the examiner as part of typical job duties. Data were checked for consistency on multiple occasions by the examiner.

Fidelity

The fidelity of administration is the responsibility of the examiner and Director of Psychological Services. Special education teachers conducting regular progress monitoring were individually trained by the examiner. Teachers were initially given instruction in the administration, scoring, and data reporting procedures. Teachers were also observed scoring passages presented as part of the AIMSWEB training materials, until inter-rater agreement between the examiner and teacher was at or better than 98%.

Missing Data

All students received benchmark and Woodcock-Johnson – III assessments. Missing data occurred when students transferred in or out of the school. Students transferring in after benchmark assessments have been administered will not have records until the subsequent benchmark period, either winter 2009 or spring 2009. Transfer students received their WJ-III assessments upon entering the school and, therefore, will not be included in analyses that seek to evaluate the predictive ability of oral reading fluency or Wilson level on the WJ-III. All students will be included in analyses for which all data are present.

Design and Data Analysis

The focus of this study was to evaluate the relationship and effectiveness of two different progress monitoring tools, a measure of oral reading fluency and the Wilson Reading System's level documentation against measures of reading achievement,

specifically, cluster w scores from the Woodcock Johnson Tests of Achievement – Third Edition (WJ-III) Broad and Brief Reading domains. The current study examined the following questions: (1) Is change in the Rate of Improvement on measures of oral reading fluency based on the Wilson level? (2) Is there a difference in the prediction of Woodcock Johnson Tests of Achievement – Third Edition (WJ-III) scores across Wilson levels after accounting for last year's WJ-III scores? (3) Does Wilson end-level for the 08-09 school year predict WJ-III w-scores in Basic and Broad Reading Cluster administered in spring, 2009 scores over and above prior year's WJ performance? (3) For this population, is oral reading fluency benchmark scores or rate of improvement a better predictor of WJ-III w-score in Reading skills over and above previous year's WJ w-scores?

CHAPTER 4

RESULTS

A combination of analytic procedures was designed to address three research questions. The first research question examined the relationship between the oral reading fluency rate of improvement and progression through the Wilson Reading System. To address this question, correlations and ANOVAs were used. The second and third research question examined both evaluation methods' ability to predict end of year achievement on a norm-referenced measure of academic performance. The three research questions were addressed and the results are presented below. The data were evaluated for and met basic statistical assumptions such as Test for Normality and Levene's Tests of Homogeneity.

As an initial analysis, descriptive information was computed. These data are contained in Table 4-1 and Table 4.2.

Table 4 – 1. Descriptive Statistics for Demographic Variables.

	N	Percent
Gender		
Female	14	27
Male	37	73

Table 4 - 1. (continued)

	N	Percent
Grade Level		
4	11	21.6
5	14	27.5
6	13	25.5
7	8	15.7
8	5	9.8

Table 4-2. *Totals and Percentages of Highest Wilson Level by Group and Grade.*

	Highest Wilson Level by Group					
	Levels 2 – 4		Levels 5 – 6		Levels 7 - 8	
Grade	N	Percent	N	Percent	N	Percent
4	9	47.3	1	5.9	1	7.7
5	5	31.6	5	29.4	2	15.4
6	2	10.5	8	47.1	2	15.4
7	1	5.2	2	11.7	5	38.5
8	1	5.2	1	5.9	3	23
Total	19	38.8	17	34.7	13	26.5

Rate of Improvement and Wilson Reading Levels

To answer the first research question, whether Rate of Improvement on measures of oral reading fluency differed based on Wilson level, a one-way analysis of variance

was conducted. First a low, medium, and high Wilson level achieved variable was created: early Wilson levels (1-4), middle Wilson levels (5-6), and higher Wilson levels (7-8). Next the examiner computed the Total Rate of Improvement variable by collapsing the oral reading fluency scores from the regular progress monitoring into a single slope score. The dependent variable was the rate of improvement in oral reading fluency over the 2008-2009 school year. The ANOVA was not found to be significant (Table 4-3) and scores are reported in Table 4-4.

Table 4-3. ANOVA results for Highest Wilson Level Achieved by Oral Reading Fluency Rate of Improvement.

Source	*df*	*F*	*η*	*p*
Wilson Level by Group	2	.671	.028	.516
Error	46			
Total	49			

Table 4-4. Means and Standard Deviations of Group Differences in Oral Reading Fluency Rate of Improvement.

Wilson Level by Group	Mean	SD	N
Levels 2 – 4	.10	.26	19
Levels 5 – 6	.16	.47	17
Levels 7 - 8	.24	.35	13

Wilson Achievement and End of Year Reading Assessments

The second research question examined, differences in the prediction of WJ-III scores across Wilson levels, after accounting for last year's WJ scores. This question was analyzed using two, one-way Analysis of Covariance (ANCOVA). Table 4 - 5 outlines means and standard deviations of WJ scores by highest Wilson level achieved.

Table 4 – 5. *Means and Standard Deviations of WJ Pre and Post-test W Scores by Highest Wilson Level Achieved.*

	Highest Wilson Level by Group					
	Levels 2 – 4		Levels 5 – 6		Levels 7 - 8	
Variable	X	SD	X	SD	X	SD
WJ Pre-test Broad w score	481.80	16.39	481.85	8.90	502.92	9.65
WJ Post-test Broad w score	492.87	15.54	490.77	8.27	509.23	7.33
WJ Pre-test Basic w score	481.47	17.78	481.08	11.39	504.54	11.32
WJ Post-test Basic w score	495.80	20.49	491.31	10.68	514.85	8.22

The first ANCOVA evaluated outcomes in the Broad Reading Cluster which includes word reading, fluency, and comprehension, while the second ANCOVA evaluated the Basic Reading Skills Cluster which included sight word and phonetic

decoding skills. The independent variable for both analyses, Wilson level, includes three levels: early Wilson levels (1-4), middle Wilson (5-6), and higher Wilson levels (7-8). The outcome variable for the first ANCOVA was the 2009 WJ-III Basic Reading Skills Cluster w score and the 2009 WJ-III Broad Reading Skills Cluster w score for the second.

Table 4 – 6. ANCOVA results for Highest Wilson Level Achieved by WJ Basic Reading Skills Cluster w Scores.

Source	*df*	*F*	*η*	*p*
WJ Pre-test Basic	1	100.046	.72	<.001
Highest Wilson Level	2	.862	.042	.430
Error	39			
Total	43			

Table 4 – 7. ANCOVA results for Highest Wilson Level Achieved by WJ Broad Reading Skills Cluster w Scores.

Source	*df*	*F*	*η*	*p*
WJ Pre-test Broad Reading w Score	1	62.016	.62	<.001
Highest Wilson Level	2	.801	.04	.456
Error	38			
Total	42			

The covariate was the 2008 WJ-III Basic Reading Skills Cluster w score and 2008 WJ-III Broad Reading Skills Cluster w scores respectively. A preliminary analysis evaluated the homogeneity-of-slopes assumption and indicated that the error variance of the dependent variable was equal across groups. The results for both ANCOVAs were not significant. Please refer to tables 4 – 6 and 4 –7 for results of these analyses.

The third research question looked at whether Wilson end-level for the 08-09 school year predicted Woodcock Johnson Tests of Achievement – Third Edition (WJ-III) w-scores in the Basic Reading Cluster administered in spring, 2009 scores over and above prior year's WJ performance. A partial correlation was conducted among the highest Wilson level achieved with the current WJ-III Basic Skills performance holding the previous year's WJ-III Basic Skills Reading Cluster w scores constant. The results of this analysis were not significant ($p = 0.32$). Please refer to tables 4 - 8 for partial correlations.

Table 4 –8. Partial Correlations for Basic Reading Skills Cluster and Highest Wilson Level Achieved Controlling for Pre-test Basic Reading Skills Cluster.

	Post-Test WJ Basic Reading Skills Cluster
Highest Wilson Level Achieved	-0.16

Oral Reading Fluency and End of Year Assessments

The final analysis examined whether the spring oral reading fluency benchmark or rate of improvement was a better predictor of WJ-III w-score Broad Reading over and above previous year's WJ-III performance for the current population. A regression was used with previous year's WJ-III w-scores entered first as a covariate, ORF benchmark

and ROI added in stepwise fashion, with current WJ-III performance as the criterion variable. As previous research has suggested, spring ORF benchmark scores were related to performance on WJ Broad Reading Cluster w scores. Rate of Improvement on regular progress monitoring over the school year was not a significant predictor. The regression equation predicting Broad Reading skills was significant (p = 0.001). The WJ Pre-test Broad Reading score accounted for 81% of the variance (R^2 = .818), with spring ORF associated with an increase in variance of 7%. The Total Rate of Improvement did not account for any additional variance. Table 4 - 9 outlines the results of this analysis.

Table 4 – 9. Regression Table for Broad Reading Skills Cluster and ORF Spring Benchmark.

		Model 2	
Variable	Model 1 B	B	95% CI
Constant	125.9	259.317	164.185 – 354.448
WJ Pre-Broad Reading Score	.760	.452	.242 - .663
Spring ORF Benchmark		.172	.078 - .267
Total Rate of Improvement		-3.487	8.995 – 2.021
R^2	.747	.818	
F	117.835	57.039	
ΔR^2		.072	
ΔF		7.498	

Post Hoc Analyses

Since the preceding analyses did not support the hypotheses, the final hierarchical regression was not conducted. However, supplemental, post hoc analyses were performed to better understand the dataset. A one-way multivariate analysis of variance (MANOVA) was conducted to explore differences in the highest Wilson level achieved determined group differences on the fall, winter, and spring benchmark scores. The independent variable, highest Wilson level achieved included three levels: early Wilson levels (1-4), middle Wilson levels (5-6), and higher Wilson levels (7-8). The dependent variables were the fall, winter, and spring oral reading fluency benchmarks. The MANOVA was found to be significant, Wilks' $\Lambda = .677$, F (6, 82) = 2.948, $p = .012$. The highest Wilson group level achieved accounted for 17.7% of the variance. The means are reported in table 4-10 and the MANOVA results are presented in table 4 – 11.

Table 4 – 10. Means and Standard Deviations of Oral Reading Fluency Benchmarks.

	Highest Wilson Level by Group					
	Levels 2 – 4		Levels 5 – 6		Levels 7 - 8	
Variable	X	SD	X	SD	X	SD
ORF Fall Benchmark	72.56	41.028	59.27	19.815	104.31	25.840
ORF Winter Benchmark	86.44	41.767	73.27	22.496	123.23	24.806
ORF Spring Benchmark	91.94	40.342	80.60	21.784	130.31	23.211

Table 4 –11. Highest Wilson Level Achieved by Oral Reading Fluency Benchmark Period MANOVA Results.

Source	*df*	*F*	*η*	*p*
Highest Wilson Level	6	2.948	.177	.012
Fall ORF	2	7.536	.260	.002
Winter ORF	2	9.007	.295	.001
Spring ORF	2	9.869	.315	<.001
Error	43			
Total	46			

Follow-up tests were conducted to evaluate pairwise differences among the means. Analyses of variance (ANOVA) on each dependent variable were conducted as follow-up tests to the MANOVA, using the Bonferroni method. The high Wilson level group performed significantly higher on oral reading fluency benchmarks at all three time periods than the low or middle groups, which did not differ (means and SDs in Table 4-10).

A second post hoc analysis was completed that evaluated the relationship between highest Wilson level achieved and the change in benchmark scores across the school year. The independent variable, highest Wilson level achieved included three levels: early Wilson levels (1-4), middle Wilson (5-6), and higher Wilson levels (7-8). The dependent variable was obtained by transforming the fall, winter, and spring oral reading fluency benchmark scores into a single change score by calculating the slope of the three benchmark points across the school year. The ANOVA was not found to be significant.

The means and standard deviations are presented in Table 4 – 12 while ANOVA results are presented in table 4 – 13.

Table 4 - 12. Means and Standard Deviations of Wilson Group Differences in Slope of Oral Reading Fluency Benchmark Scores.

Wilson Level by Group	Mean	SD
Levels 2 – 4	.08953	.07597
Levels 5 – 6	.08569	.04877
Levels 7 - 8	.10757	.06861

Table 4 - 13. Highest Wilson Level Achieved by Oral Reading Fluency Benchmark Slope ANOVA Results

Source	*df*	*F*	*η*	*p*
Between Groups	2	.452	.019	.639
Error	46			
Total	49			

CHAPTER 5

DISCUSSION

This research hoped to assess two different methods for evaluating reading progress, oral reading fluency and movement through a highly structured reading program. The literature in this field indicates the utility of measures of oral reading fluency as a data-based decision making tool (Shinn, 2007; Vaughn & Fuchs, 2003; Vaughn, Linan-Thompson, & Hickman, 2003) as well as the effectiveness of the Wilson Reading System (O'Conner & Wilson, 1995; Torgeson et al., 2006). However, data based decision-making research has focused on developing an acceptable interval of time for evaluating improvement. Many decision-making models do not consider that interventions might take different amounts of time to show results. As a result, this study intended to answer three research questions: (1) Does the rate of improvement on measures of oral reading fluency differ based on Wilson level? (2) Does Wilson end-level for the 2008-2009 school year predict WJ-III w – scores in Basic and Broad Reading Clusters? (3) For this population, is oral reading fluency benchmark scores or rate of improvement a better predictor of WJ-III w – score in reading skills?

Overall results were not as expected, but they still provide valuable information regarding judgments of reading progress. Analysis began with a number of correlations. Interestingly, a student's rate of improvement was not correlated with spring benchmark, post-test Broad and Basic reading cluster scores, or student grade level.

(1) Rate of Improvement and Wilson Reading Level

The relationship between rate of improvement in oral reading fluency and highest Wilson level achieved. The results showed that the total slope of regular progress monitoring across the school year was not significantly correlated to the highest Wilson level achieved. This is consistent with the original hypothesis that, due to the nature of the Wilson Reading System's scope and sequence, progress through Wilson levels may not directly translate into progress as measured by oral reading fluency. Further analysis took a different perspective and evaluated the difference in rate of improvement by grouping Wilson levels into beginning, middle, and high groups. The results were not significant, which indicates that rates of improvement were not different among groupings of highest Wilson level achieved.

While the findings corroborated the original hypothesis that progress through the Wilson Reading levels would not be significantly related to rate of improvement in oral reading fluency, the reason remains unclear. Research suggests that reading fluency is less important as grade levels increase, especially after fourth grade (Yovanoff et al., 2005), which is the lowest grade in the current study. Furthermore, research has not yet established the most effective method for calculating the rate of improvement. Stage and Jacobsen (2001) found that the slope from the three benchmark periods were less predictive than individual benchmark periods in students' performance on standardized assessments, but Baker et al (2008) found a moderate correlation between the rate of improvement and standardized assessments by calculating the slope from six different timepoints throughout the school year. The nature of the Wilson Reading System's scope and sequence can intuitively be another probable explanation for the lack of relationship

between rate of improvement and Wilson level achieved. As mentioned before, the Wilson Reading System scope and sequence were designed to expose students to the syllable patterns in the English language, beginning with the most common, rather than following a grade level (Wilson Language Training Corp, 2004).

(2) Wilson Achievement and End of Year Reading Assessments

Preliminary correlations found a significant, positive relationship between highest Wilson level achieved and WJ post Broad and Basic reading cluster w scores. This finding suggested that the higher Wilson levels achieved are associated with higher Broad and Basic reading skills clusters. However, the results of the ANCOVA for both Broad and Basic reading cluster w scores indicate that pre-test WJ performance accounted for 62% and 73% of the variance, respectively while the Highest Wilson Level Achieved only accounted for 4% in both analyses. This means that once pre-test performance on the WJ is taken into account, there are no differences between the three Wilson score groups. Most likely, the statistical strength of the pre-test in a measure as robust as the WJ does not leave much more variance to be explained by progress through the Wilson Reading program.

These findings are surprising none the less based on published and unpublished Wilson efficacy data (O'Conner & Wilson, 1995; Torgeson et al., 2006; Wood, 2002). Previous efficacy research on the Wilson Reading System used similar timelines (seven to ten month intervention periods) and found significant improvement in reading achievement on WJ subtest and cluster scores; however, standard scores or grade equivalents were typically used instead of w scores. Additionally, many of these studies used pre-Wilson certification teachers while the present study used data generated from

certified Wilson teachers. In the case of the Basic Reading Skills cluster score, the original hypothesis suggested that Wilson levels would be a significant predictor of end-of-year WJ performance in basic reading skills. The basic skills cluster score was a combination of student performance on Letter-Word Identification, a sight word activity, and Word Attack, which directly assesses decoding skills (McGrew & Woodcock, 2001). These are surprising results since the Wilson Reading System heavily focuses on word reading skills. Unlike previous research that utilized age-based standard scores and found a significant improvement in reading performance, the present study used w scores, which is a statistic from which standard scores are derived. W scores provide an understanding of performance free from an individual student's relative position among peers. The current research found that once pre-test WJ scores are accounted for, Wilson level is not a significant predictor. The most likely explanation is that the Woodcock Johnson is, by nature, too broad of a measure to register small increases in reading skills in a calendar year.

Another possible explanation for the lack of predictive ability of Wilson group level is that there is no evidence to suggest that the WJ's evaluation of decoding skills maps onto the scope and sequence of the Wilson Reading System. In other words, the order of syllable types instructed through the Wilson Reading System may not be consistent with the manner in which the WJ evaluates decoding skills. An item analysis of the WJ Word Attack Subtest against the Wilson Reading System's scope and sequence yielded an interesting finding. Of the 12 Wilson steps, nine steps are represented in one of the 29 items in the A form of the Word Attack subtest and, of the nine steps represented, five have fewer than two items on the entire subtest. In administering the WJ

Word Attack subtest, most school age children are asked to decode nonsense words. The WJ is designed to initially offer easier items and then progress to more difficult items; however, when comparing the Word Attack subtest to the Wilson scope and sequence there is little agreement. Table 5 – 1 outlines these findings.

Table 5 – 1. *Steps of the Wilson Reading System's Scope and Sequence and Corresponding Word Attack Items on the Woodcock Johnson Tests of Achievement – Third Edition.*

Step Number	Number of Items	Item Number
Step 1	6	4, 6 – 9, 15
Step 2	4	12, 17, 22 , 23
Step 3	0	NA
Step 4	1	19
Step 5	0	NA
Step 6	1	21
Step 7	1	29
Step 8	1	14
Step 9	8	5, 11, 13, 16, 18, 20, 26, 27
Step 10	0	NA
Step 11	5	10, 24, 28, 30, 31
Step 12	2	25, 32

Note. Number of Items was based on highest Wilson level required to decode word.

The WJ Broad Reading Cluster is derived from three subtests: Letter-Word Identification, Reading Fluency, and Passage Comprehension (McGrew & Woodcock, 2001). One would not intuitively expect significant change in Reading Fluency or Passage Comprehension since the Wilson Reading System is primarily a decoding/encoding reading intervention. This is supported by research from Torgeson et al (2006), whose study did not find that an adapted Wilson Reading program had any significant impact on the Letter Word Identification or Passage Comprehension subtests of the Woodcock Johnson. However, in an unpublished article and using standard scores instead of w scores, Wood (2002) found significant gains in the Woodcock Reading Mastery Passage Comprehension, Letter Word Identification, Word Attack subtests, as well as the Total Reading Cluster. As highlighted in the literature review, the unpublished Wilson articles do not use procedures that stand up to statistical scrutiny. While initial correlations suggested a positive relationship, once the previous years' WJ performance was taken into account, the Wilson level achieved was not a significant predictor of end of the year WJ scores. This may be support for the use of an alternative measure in evaluating program effectiveness or the use of scores other than Grade Equivalent or standard score.

Oral Reading Fluency and End of Year Reading Assessments

Results showed that WJ pre-test performance was the single best predictor of post-test WJ performance accounting for 74% of the variance. The Spring ORF Benchmark only explained an additional 7% of the variance and the rate of improvement on measures of oral reading fluency across the school year did not contribute at all. A small but significant contribution of spring benchmark performance is consistent with

research by Yovanoff et. al (2005) which argues that fluency is less important as grade levels increase. Much of the current research on oral reading fluency focused on the ability of the benchmark assessment to predict performance found a moderate to high ability on statewide standardized assessments (Barger, 2003; Buck & Torgeson, 2003; Shapiro, 2006; Vandeer Meer, Lentz, & Stollar, 2005; Wilson, 2005; & Wood, 2006). However, statewide assessments vary in both form and function from the Woodcock Johnson Tests of Achievement. The Woodcock Johnson Tests of Achievement is meant to parse out specific reading skills and then aggregate performance statistically while state assessments evaluate student performance against grade level standards.

However, the author also posited that rate of improvement would have had some predictive power, which was not the case. Given the previous analyses, this finding is not surprising as the rate of improvement did not correlate with Wilson, WJ-III, or benchmark performance. This is contrary to findings by Baker and colleagues (2008) that found the rate of improvement in oral reading fluency scores from six different time periods throughout two school years was moderately correlated on third grade state assessments. There is a developing body of research looking into the Standard Error of Measurement (SEM) of oral reading fluency rate of improvement. Christ and Silberglitt (2007) found that for students in grades four through five, there was a standard error of measurement between six and fourteen words per minute and a median of 10 words per minute for every time point. With such variability in any single score and as many as 36 data points per student, the potential SEM increases exponentially. This has major implications for data-based decision making.

Post Hoc Analyses

In the first post hoc analysis, it was found that the group who completed Wilson levels 7 – 8 in the 2008 – 2009 school year performed better than students who completed levels 2 – 6 across the school year. In other words, students who completed the highest Wilson levels that year demonstrated a stronger oral reading fluency performance in the fall, winter, and spring benchmark periods. This is an interesting finding in that, while intuitive, did not manifest itself in other analyses conducted in this study. There was no significant difference in Wilson level achieved by rate of improvement or WJ reading cluster performance. Possibly, by completing levels six and seven, students have acquired sufficient decoding skills to be able to demonstrate real improvement. Since students in the middle group typically reached levels five and six by the end of the school year, there was not sufficient time to demonstrate progress in overall rate of improvement.

The second post hoc analysis evaluated the relationship of the change in improvement across benchmark periods by highest Wilson level achieved. The results of the ANOVA were not significant, which is somewhat consistent with research by Stage and Jacobsen (2001), who found benchmark scores are more predictive than the rate of change in benchmark scores over the school year. Due to the small N or the sample's characteristics, the present study did not find any group differences. This provides additional support for the notion that, while the Wilson Reading program is an effective reading intervention, progress through the program does not necessarily translate into progress in external measures of reading proficiency such as oral reading fluency.

Limitations

One of the biggest limitations to the current study is the small sample size and short time frame. With a total of 51 students over four grades it is difficult to control or co-vary for grade. The small sample size also required Wilson levels to be collapsed in order to be compared, which may have impacted analyses. The three Wilson level groups had only 13 – 19 students each. The brief time frame for data collection and the relatively new school where data were collected also limited analysis of Wilson levels one through eight. It is recommended that future analyses collect cross-sectional data over multiple years.

Another potential confound is the presence of concurrent reading instruction for all students in the sample. While the Wilson Reading program is designed to cover many aspects of reading, it is primarily a word reading intervention and, thus, considered supplemental instruction. As mentioned previously, students also receive direct instruction in oral reading as well as comprehension.

Implications for Future Research

Future research could investigate the longitudinal relationship between the Wilson Reading Program and oral reading fluency benchmark performance. One of the limitations of the current study was availability of only a single year's worth of Wilson Reading data. Due to the nature of the population, progress through the Wilson Reading program tends to be slower than the targeted four levels in one school year. Utilizing a larger sample, exploration of student progress through the entire program would be a valuable contribution to the knowledge base for interventions for students with learning

disabilities. This information would provide information to assist in determining the level of frequency or intensity of the Wilson Reading program as an intervention.

Another avenue to explore is the use of w scores to evaluate program effectiveness, especially in the Wilson Reading program, which lacks a sufficient body of empirical support at this time. Unpublished Wilson efficacy articles typically use grade equivalents or standard scores, neither of which are appropriate statistics for measuring acquisition of skills.

A final research area that requires more support is using oral reading fluency rate of improvement in data based decision-making. The Wilson Reading program by desgin does not have a scope and sequence consistent with grade level material. Research finds that oral reading fluency has support as an appropriate measure of reading proficiency (Shinn et. al, 1992); however, more recent research by Christ and Silberglitt (2007) cite the large standard error of measurement for every data point used. For students with disabilities, especially in the acquisition of basic reading skills, progress can often be slow, which may compound the issue.

Implication for Practice

The first research question sought to determine the relationship between the highest Wilson level achieved and rate of improvement on a measure of oral reading fluency. The second and third research questions also used a different evaluation measure, the Woodcock Johnson Tests of Achievement, to see if progress through the Wilson Reading program translated into significant progress on a broad reading assessment. It appears that progress through the Wilson Reading program does not become immediately evident in sensitive measures such as oral reading fluency or more

robust measures of reading achievement (i.e., Woodcock Johnson Tests of Achievement). Additionally and most importantly, the use of a fluency-based progress monitoring tool is not always appropriate when students are receiving an accuracy-based reading intervention. Therefore, it is important for educators to remember that any decision regarding a student's education or educational programming should not be based on a single source of information but rather a synthesis of all relevant educational data, including norm-referenced assessment, curriculum-based measurement, and criterion based mastery measurement.

REFERENCES CITED

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

Barger, J. (2003). Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina end of grade reading assessment. (Technical Report). Asheville, NC: North Carolina Teacher Academy.

Brueggemann, A. E., Kamphaus, R. W., & Dombrowski, S. C. (2008). An impairment model of learning disability diagnosis. *Professional Psychologist: Research and practice*, *39*(4), 424-430.

Buck, J., & Torgesen, J. (2003). The Relationship Between Performance on a Measure of Oral Reading Fluency and Performance on the Florida Comprehensive Assessment Test. (FCRR Technical Report #1) Tallahassee, FL: Florida Center for Reading Research.

Busch, T.W., & Reschly, A.L. (2007). Progress Monitoring in Reading: Using curriculum-based measurement in a response-to-intervention model. *Assessment for Effective Intervention, 32*(4), 223-230.

Cain, K., & Oakhill, J. (Eds.) (2007). Children's comprehension problems in oral and written language: A cognitive perspective. New York, NY, Guilford Press.

Case, L. P., Speece, D. L., & Molloy, D. E. (2003). The validity of a response to instruction paradigm to identify reading disabilities: A longitudinal analysis of individual differences and contextual factors. *School Psychology Review*, *32*(4), 557-582.

Ceci, S. J., & Papierno, P. B. (2005). The rhetoric and reality of gap closing: When the 'have-nots' gain but the 'haves' gain even more. *American Psychologist*, *60*(2) 149-160.

Christ, T.J., Silberglitt, B. (2007). Estimates of the Standard Error of Measurement for Curriculum-Based Measures of Oral Reading Fluency. *School Psychology Review*, *36*(1), 130-146.

Coyne, M. D., & Harn, B. A. (2006). Promoting beginning reading success through meaningful assessment of early literacy skills. *Psychology in the Schools*, *43*(1), 33-43.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability ten years later. *Developmental Psychology*, *33*(6), 934-945.

Cusumano, D. L. (2007). Is it working? An overview of CBM and its uses for assessing instructional, intervention, or program effectiveness. *The Behavior Analyst Today*, *8*(1), 24-34.

Daly, E. J., Martens, B. K.; Barnett, D.; Witt, J. C.; & Olson, S. C. (2007) Varying

Intervention Delivery in Response to Intervention: Confronting and resolving challenges with measurement, instruction, and intensity. *School Psychology Review, 36* (4), 562-581.

De Jong, P. F., & Van der Leij, A. (2003). Developmental changes in the manifestation of a phonological deficit in dyslexic children learning to read a regular orthography. *Journal of Educational Psychology*, *95*(1), 22-40.

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

Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, *37*(3), 184-192.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). Learning Disabilities: From identification to intervention. New York, NY: Guilford Press.

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

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

Goff, D. A., Pratt, C., & Ong, B. (2005). The relations between children's reading

comprehension, working memory, language skills, and components of reading decoding in a normal sample. *Reading and Writing, 18,* 583-616.

Good, R. H., Kaminski, R. A., Smith, S., & Bratten, J. (2001). Technical Adequacy of Second Grade DIBELS Oral Reading Fluency Passages (Technical Report No. 8). Eugene, OR: University of Oregon.

Good, R. H., Simmons, D. C., & Kame'enui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for 3rd grade high-stakes outcomes. *Scientific Studies of Reading*, *5*(3), 257-288.

Good, R. H., Simmons, D., Kame'enui, E., 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.

Goodman, K. S. (1965). A linguistic study of cues and miscues in reading. *Elementary English, 42,* 639-643.

Goodman, K., & Goodman, Y. (1981). A whole-language, comprehension-centered reading program. *Program in Language and Literacy Occasional* (Paper No. 1). Tucson, AZ: Arizona University.

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

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

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

Howe, K. B., & Shinn, M. M. (2002). Standard reading assessment passages for use in general outcome assessment: A manual describing development and technical features. Eden Prairie: MN: Edformation Inc.

Institute of Educational Sciences (2008). Evidence standards for reviewing studies. Retrieved from the What Works Clearinghouse website: http://ies.ed.gov/ncee/wwc/pdf/wwc_version1_standards.pdf

Jenkins, J. R., Graff, J. J., & Miglioretti, D. L. (2009). Estimating reading growth using intermittent CBM. *Exceptional Children, 75* (2), 151-165.

Kavale, K. A., Kauffman, J. M., Bachmeier, R. J., & LeFever, G. B. (2008). Response-to-intervention: Separating the rhetoric of self-congratulation from the reality of specific learning disability identification. *Learning Disability Quarterly, 31*(3), 135-150.

Moats, L. C., & Davidson, M. (2008). Language Essentials for Teachers of Reading and Spelling. Sopris West: Boston, MA.

Marston, D. (2005). Tiers of intervention in response to intervention: Prevention outcomes and learning disabilities identification patterns. *Journal of Learning Disabilities, 38* (6), 539-544.

Martelle, Y., & Smith, D. K. (1994). The relationship of the WIAT and WJ-R tests of achievement in a sample of students referred for learning disabilities. Paper presented at the annual meeting of the National Associated of School Psychologists, Seattle, WA.

McGrew, K. S., & Woodcock, R. W. (2001) Woodcock Johnson Technical Manual: Normative Update. Riverside Publishing Company: Itasca, Illinois.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidenced-based assessment of scientific research literature on reading and its implications for reading instruction*. (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

O'Connor, J., & Wilson, B. (1995). Effectiveness of the Wilson Reading System used in public school training. In McIntyre and Pickering (eds.) 1995. *Clinical Studies of Multisensory Structured Language Education* (pp. 247 – 253). Salem, OR: International Multisensory Structured Language Education Council.

Reuzel, D. R., & Hollingsworth, P. M. (1993). Effects of fluency training on second graders' reading comprehension. *Journal of Educational Research, 86* (6), 325-331.

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

Samuels, S. J. (2007). The DIBELS test: Is speed of barking at print what we mean by reading fluency? *Reading Research Quarterly, 42*(4), 563-566.

Schatschneider, C., Buck, J., Torgesen, J., Wagner, R., Hassler, L., Hecht, S., & Powell-Smith, K. (2004). A Multivariate Study of Factors That Contribute to Individual Differences in Performance on the Florida Comprehensive Reading Assessment Test. Technical Report 5, Tallahassee, FL: Florida Center for Reading Research.

Schwanenflugel, P. J., Hamilton, A. M., Kuhn, M. R., Wisenbaker, J. M., & Stalh, S. A. (2004). Becoming a fluent reader: Reading skills and prosodic features in the oral reading of young readers. *Journal of Educational Psychology, 96*(1), 119-129.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measurement and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*(1), 19-34.

Shapiro, E. S., Solari, E., & Petscher, Y. (2008). Use of a measure of reading comprehension to enhance prediction on the state high stakes assessment. *Learning and Individual Differences, 18,* 316 - 328.

Shaw, R. & Shaw, D. (2002). DIBELS Oral Reading Fluency-Based Indicators of Third

Grade Reading Skills for Colorado State Assessment Program (CSAP). (Technical Report) Eugene, OR: University of Oregon.

Shinn, M. R., Good, R. H., & Stein, S. (1989). Summarizing trend in student achievement: A comparison of methods. *School Psychology Review*, *18*(3), 356-370.

Shinn, M. R.; Good, R., Knutson, N., Tilly, W. D., & Collins, V. L. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review*, *21*(3), 459-479.

Shinn, M. R. (2007). Identifying students at risk, monitoring performance, and determining eligibility within response to intervention: Research on Education and Benefit from academic intervention. *School Psychology Review, 36* (4), 601 – 617.

Singer-Dudek, J., & Greer, R. D. (2005) A long-term analysis of the relationship between fluency and the training and maintenance of complex math skills. *The Psychological Record*, *55*, 361-376.

Slavin, R. E. (1997). Educational psychology: Theory and practice. Boston, MA: Allyn & Bacon

Spear-Swerling, L. (2006). Children's reading comprehension and oral reading fluency in easy text. *Reading and Writing*, *19*, 199-220.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated

performance-based assessment using oral reading fluency. *School Psychology Review*, *30*(3), 407-419.

Stanovich, K. E. (1986). Matthew Effects in Reading: Some Consequences of Individual Differences in Acquisition of Literacy. *Reading Research Quarterly*, *21*, 360-407.

Stecker, P. M., Lembke, E. S., & Foegen, A. (2008). Using Progress monitoring data to improve instructional decision making. *Preventing School Failure*, *52* (2), 48-60.

Sweet, R. W. (2004). The big picture: Where we are nationally on the reading front and how we got there. In P.McCardle & V.Chhabra (Eds.), *The voice of evidence in reading research*. Baltimore: Paul H. Brookes.

Torgesen, J. K. & Hudson, R. (2006). Reading fluency: critical issues for struggling readers. In United States Department of Education, Office of the Under Secretary (2003). No Child Left Behind: A toolkit for teachers. Washington, DC.

Torgesen, J., Myers, D., Schirm, A., Stuart, E., Vartivarian, S., & Mansfield, W., et al. (2006). National assessment of Title I interim report—Volume II: Closing the reading gap: First year findings from a randomized trial of four reading interventions for striving readers. Retrieved from Institute of Education Sciences, U.S. Department of Education Web site: http://www.ed.gov/rschstat/eval/disadv/title1interimreport/index.html.

Tracey, D. H., & Morrow, L. M. (2006). Lenses of reading: An introduction to theories and models. New York, NY: Guilford Press.

Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). The relationship between Oral Reading Fluency and Ohio Proficiency Testing in Reading (Technical Report). Eugene, OR: University of Oregon.

Vaughn, S., & Fuchs, L. S. (2003). Redefining learning disabilities as inadequate response to instruction: The promise and potential problems. *Learning Disabilities Research and Practice, 18*(3), 137-146.

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

Walberg, H. J., & Tsai, S. L. (1983). Matthew effects in education. *American Education Research Journal, 20*(3), 359-373.

Wilson Language Training (2002). Wilson literacy solutions: Evidence of effectiveness. Retrieved from Wilson Language Trainging Corporation website http://www.wilsonlanguage.com/ w_wrs.htm.

Wilson Language Training Corporation (2004). The Wilson Reading System. Retrieved from Wilson Language Trainging Corporation website: http://www.wilsonlanguage.com/w_wrs.htm.

Wilson, J. (2005). The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency to performance on Arizona Instrument to Measure Standards (AIMS). Tempe, AZ: Tempe School District No. 3.

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11*(2), 85 – 104.

Wright, J. (2001). Curriculum-based measurement: A manual for teachers. Syracuse, NY: Syracuse City Schools. Retrieved from Intervention Central website: www.interventioncentral.org.

Yovanoff, P., Duesbery, L., Alonzo, A., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice*, *24*(3), 4-12.

.