Using Student Performance During a Reading Intervention to Predict Student

Outcomes and Performance on Accountability Measures of Reading

A DISSERTATION SUBMITTED TO THE FACULTY OF

UNIVERISTY OF MINNESOTA

BY

Lorien Parson

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Matthew K. Burns

July, 2012

Acknowledgements

I would like to acknowledge my family, in particular Aaron and Elias for their patience

and support while I completed this academic journey.

I would also like to acknowledge my advisor, Dr. Matt Burns for his constant support and

guidance.

Finally, I would like to acknowledge the people that are committed to the challenge of

helping at-risk students by providing high quality intervention services, with very little

monetary compensation, every day.

Abstract

An existing data set for a sample of 3rd grade students was used to determine the relationship between performance during a reading intervention and short-term achievement test outcomes, and long-term risk status. Students participated in a reading intervention, one-on-one practice with a trained adult, during which weekly curriculum based measurement-reading data and dosage of intervention service time were collected. The sample was comprised of reading curriculum based measurement progress monitoring data for 99 students from 8 different schools. The recommendation was for students to receive 60 minutes per week of fluency sessions, the average student received a dosage of 51 minutes per week. Student scores on district and state reading tests from the spring of third grade, and student participation in pre-referral or special education programming 1 and 2 years post participation were also examined. Reading test data were provided by the schools including student performance on the spring 3rd grade No Child Left Behind state reading test, and for 6 of the schools, a district reading test by the Northwest Evaluation Association called the Measures of Academic Progress. Schools also provided existing information about free/reduced lunch eligibility and risk-status, for each of the students.

The following research questions guided the study: (a) how does performance during a third-grade reading intervention predict future performance on district and state reading tests; and (b) how does performance during a third-grade reading intervention predict future risk-status in 4th and 5th grades?

Multiple linear regression models were used to determine the amount of variance accounted for in state and district reading test performance, using the predictor variables related to participation in the 3rd grade reading fluency intervention. Twenty-three percent of the variance in state test scores, and 44% of the variance in district test scores could be explained using multiple linear regression models with the predictor variables: lunch status, slope, baseline and end levels, and dosage. There were different significant coefficients in each model, and 2 of the schools did not administer the MAP.

Risk-status 1 and 2 years following participation in the 3rd grade reading intervention also was determined, using the predictors: performance during the intervention, dosage, and lunch status. Logistic regression models used to predict placement in special education in 4th or 5th grades were not significant. The logistic regression model used to predict student participation in pre-referral programming in 4th grade was not significant, but the model was significant in 5th grade. Variables were backwards eliminated to further determine which variable(s) in the model was/were significant in predicting pre-referral participation in 5th grade, and baseline level at the onset of the 3rd grade reading intervention was the significant predictor. An estimated probability plot illustrated that students with a higher baseline level were less likely to participate in pre-referral programming in 5th grade, versus students with lower baseline level performance.

The findings from the first part of this study further support the existing research that reading curriculum based measurement does predict performance on other reading tests (Baker, et al., 2008; Chard, Vaughn & Tyler, 2002; Lennon & Slesinski, 1999) and

state reading tests (Stage & Jacobsen, 2001; Good, Simmons & Kame'enui, 2001; Crawford, Tindal, Stieber, 2001; McGlinchey & Hixson, 2004; Keller-Margios, Shaprio & Hintz, 2008). The current study is unique in its application of progress monitoring data for students that participated in a reading intervention, versus seasonal benchmark scores. The second part of the study examined risk-status in 4th and 5th grades for the students who participated in the reading intervention service in 3rd grade. Results supported the findings of other studies, that some intervention participants struggle to maintain gains in subsequent years, (Byrne & Fielding-Barnsley, 1993; Bus & Van Ijzendoorn, 1999; O'Connor, Notari-Syverson, & Vadasy, 1996, 1998) and that curriculum-based measurement baseline level, can be a strong predictor of growth (Silberglitt and Hintz, 2007). The current study was unique in that much of prior research has examined results for students that participate in interventions higher in dosage, diverse in skill areas and prior to 3rd grade. The findings from this study suggest that progress-monitoring data at the on-set, and during an intervention, do relate to performance on other reading assessments, and may predict future risk-status.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Statement of the Problem

Previous research found that multi-skill reading interventions can lead to long term growth and reduction in students being identified with a disability (Torgesen & Davis, 1996; Lennon & Slesinski, 1999) , but few researchers have taken the approach of Simmons et al. (2008) and implemented reading interventions for 3 to 4 years within a Response to Intervention (RTI) tiered system. The current study is intended to add to the growing body of research around implementation of RTI in schools and explores the degree to which student progress during an intervention predicts some selected academic outcomes. Moreover, research about the effects of RTI implementation and reduced placement in special education is promising (Burns, Appleton & Stehouwer, 2005; Hughes & Dexter, 2010), but more research is needed about student performance in later elementary school years, and if there is a relationship with weekly progress monitoring data. The current study included long term follow-up of students who participated in a tier-2 reading intervention, as recommended by Hughes and Dexter (2010), and examined performance on state and district reading tests, and referrals to intervention teams or prevalence of special education placement.

The research questions are: (1) How does student performance during a tier 2 intervention, predict performance on group accountability measures of reading? (2) How well do level, slope, lunch status and dosage, combined, predict future risk status 1 and 2 years after intervention.

Overview of Methods

Third grade students in 8 different elementary schools participated in a reading fluency intervention delivered in a one-on-one format by trained adults. Curriculum based measurement-reading (CBM-R) was administered on a weekly basis for 99 students that participated in the intervention. The progress monitoring data were used to determine baseline level and end level scores for each student record, and the slope values calculated by the graphing software were also used. Tutoring session time was recorded by tutors and converted into a dosage variable, to determine the average number of minutes per week of intervention time for each student. Schools provided state reading test scores, district reading test scores (Measures of Academic Progress), special education status and pre-referral participation as available, for each student.

Overview of Results

Multiple linear regression models were used to determine the percentage of variance accounted for in student performance on the state and district reading tests. Student performance during the intervention was significant in predicting student performance on district and state reading tests. The model using progress monitoring baseline level, slope and end level was significant and explained 23% of the variance in state test scores, slope was the significant predictor. A logistic regression model was then applied to determine the odds of meeting proficiency on the state reading exam given progress monitoring baseline level scores, slope, and end level. Estimated probability plots were then generated to illustrate the probability of meeting proficiency on the state

reading exam, given the mean baseline level, the range of end level scores and strong or weak slope values. Strong slope values, even with high end level scores, resulted in an increased probability of meeting proficiency on the state exam, compared to low slope values with high end level scores. For the district reading test, 44% of the variance in student test scores could be explained with a regression model. The model included: baseline level, dosage, end level, and lunch status; with dosage and end level as the significant coefficients in the model.

The other outcomes investigated in this study related to student risk-status in 4th and 5th grades. The four logistic regression models used the 5 predictors of 3rd grade intervention curriculum based measurement baseline, slope, end level, eligibility for free/reduced price lunch, and dosage of the intervention. The four dependent variables included: pre-referral participation in 4th grade, pre-referral in 5th grade, special education in 4th grade, and special education in 5th grade.

The two logistic regression models with participation in special education or pre-referral programming in 4th grade as the outcomes, were not significant. The model predicting special education participation in 5th grade was not significant, but approached significance, and the model predicting pre-referral participation in 5th grade was significant. In the model predicting 5th grade participation in pre-referral programming, baseline level was the only significant predictor. To further examine the practical significance of baseline level, an estimated probability plot was generated. The plot illustrates how a baseline level score below 20 words per minute results in the maximum

probability, a greater than 60% probability of participation in pre-referral programming in 5th grade, compared to a much lower probability for higher baseline level scores.

Implications for Research

Researchers have found that curriculum based measurement-reading (CBM-R) seasonal benchmark scores predicted performance on state reading tests (Stage & Jacobsen, 2001; Good, Simmons & Kame'enui , 2001; Crawford, Tindal & Stieber, 2001; McGlinchey & Hixson, 2004; Keller-Margios, Shapiro & Hintze, 2008). Studies have also reported that progress during a reading intervention can also predict performance on standardized reading achievement tests (Baker, et al., 2008; Chard, Vaughn & Tyler, 2002; Lennon & Slesinski, 1999). The current study extends the research by demonstrating that progress monitoring data during an intervention, predict performance on state and district, large scale reading assessments. However, different predictor variables were significant in the two regression models. The slope variable was a significant coefficient in the model predicting performance on the state exam, whereas baseline and end level variables were not. Student mean end level scores were below the winter and spring target scores for 3rd grade, yet the average state reading test score met the state proficiency standard. Similar to the current study, others (Simmons, et al. 2008; Shaw & Shaw, 2002) reported that students with low CBM-R scores in the spring, have successfully passed state reading exams.

The current study found that for students participating in a reading intervention, slope, or progress via and during an intervention, may be an important factor related to performance on state reading exams. Dosage and end level were significant predictors

related to the percent of variance accounted for on the district assessment results. These findings suggest that level and slope progress monitoring variables are both important for students that participate in reading interventions, and that variance in reading test scores can be accounted for based on student progress monitoring data. Using benchmark scores for entire grade levels, Young-Suk, Petscher, Schatschneider and Floorman (2010) reported that both were important, curriculum based measurement level and slope, in predicting reading performance of 3rd grade students.

The second part of this study examined how progress monitoring performance, lunch status, and dosage of the reading intervention predicted risk-status in subsequent grades. The logistic regression models were not significant in predicting special education participation in 1 or 2 years post-intervention participation. The model predicting pre-referral participation in 4th grade was also not significant, but the model predicting pre-referral participation in 5th grade was significant. Baseline level was the significant predictor in the model. Silberglitt and Hintz (2007) reported that students with low baseline level scores had weaker growth than students with higher baseline level scores. Weaker growth could have been only a slight concern in 4th grade, but may have become more serious for these students, in 5th grade, hence the significant logistic regression model predicting participation in pre-referral programming in 5th grade. Furthermore, intervention follow-up studies have shown that some students fail to maintain gains in years post intervention (Byrne & Fielding-Barnsley, 1993; Bus & Van Ijzendoorn, 1999; O'Connor, Notari-Syverson & Vadasy, 1996, 1998).

Implications for Practice

Collecting and interpreting progress monitoring data for students that receive interventions was related to student performance on reading tests and risk for pre-referral and/or special education participation. Previous research also demonstrated a relationship between implementing tiered interventions with progress monitoring and referrals to special education (Burns et al., 2005). Thus, school-based personnel should consider facilitating tiered interventions and collecting progress monitoring data, as such practices can impact student performance on state and district reading exams. Progress monitoring data and the relationship to future student participation in pre-referral and special education was explored in this study, and results indicate it is a topic that warrants further research in the field.

<div align="center">

Limitations

</div>

Limitations of this study relate to external validity such as: the sample itself and related record keeping by tutors and the schools, local decision-making by supervising teachers including the use of target scores for determining student participation, less than the recommended dosage, and factors related to progress monitoring. Future research should consider replication with a larger and more diverse sample, further exploration of progress monitoring data prior to participation in an intervention, the application of more rigorous decision-making criteria for exiting students from the intervention, dosage of tier 2 interventions and examination of processes of pre-referral teams. The slope variable was a value reported by the progress monitoring software and leptokurtic in its raw data form, according to descriptive statistics. It was therefore changed to a dichotomous

variable, using percentile rank cut-offs. Changing this variable resulted in honoring statistical assumptions, but may have resulted in an increase in type 2 error.

Delimitations

Delimitations include the data provided in the existing data set, local decisions made at the schools, the fluency focus of the reading intervention services, and the school records of student test performance, dosage, and risk-status. The study used an existing progress-monitoring data set of students that participated in a reading fluency intervention in 3rd grade, and only records that had 6 or more data points. The dosage records revealed that the average student received less than the recommended amount of service time per week. Supervising teachers at each of the schools selected students for participation and also determined when students exited tutoring services. Although guidelines for student participation were provided, data suggest that supervising teachers did not consistently apply the guidelines, for example students participated that were above or extremely below seasonal benchmark targets as indicated by the range of baseline level scores. District assessment scores for the Measures of Academic Progress were not available for 2 schools because the districts did not administer the test. All tutors that worked with the students delivered fluency interventions, as that was the training provided to the tutors and the focus of their role.

Definitions

Curriculum-based measurement (CBM): An assessment approach developed by Deno (1985) that uses brief, technically adequate measures of academic performance that can be administered repeatedly across time.

<u>Criterion referenced target scores:</u> CBM scores for various grade levels, calculated using

cohort student data from seasons and grade levels and their corresponding state

test scores to determine which CBM scores predict, with a specific degree of

probability, the likelihood of achieving a particular score on a reading

achievement test (Bollman, Silberglitt & Gibbons, 2007; Hintze & Silberglitt,

2005).

<u>Estimated probability plots</u>: Used to provide a visual representation of the model

illustrating the probability of success, via results from a logistic regression

analysis. The plots are in the form of a sigmoid, or s-shaped curve, constrained to

values between 0.0 and 1.0 (Agresti, 2007).

<u>Rasch unIT (RIT) scores</u>: are equal interval scores generated via completion of the

computer adaptive tests, the Measures of Academic Progress, (NWEA, 2011),

allowing for measurement of student performance and growth over time.

<u>Backward elimination</u>: Beginning with a full regression model, terms are sequentially

removed beginning with the term with the highest *P*-value, elimination stops

when additional removal results in a significantly poorer fit (Agresti, 2007).

<u>Ordinary least squares (OLS):</u> OLS is a form of statistical regression, that can be applied

to formative curriculum based measurement data to calculate weekly growth rates

(Deno, Fuchs, Marston, & Shin, 2001; Fuchs, Fuchs, Hamlett, Walz & Germann,

1993) and trend/slope lines used for decision-making (Good & Shinn, 1990;

Shinn, Good, & Stein, 1989; Christ, Pike, & Monaghen, 2011).

Chapters in the Dissertation

There are four additional chapters that comprise this dissertation. Chapter 2 is a literature review including (a) longitudinal reading intervention research, (b) reading interventions as related to RTI, and (c) RTI decision making and assessment as applied to reading. Chapter 3 is a description of the methodology applied in the dissertation. It includes participants, the assessment and risk-status predictor and response variables, and the data analysis procedure. Chapter 4 is the summation of the results, as related to the research questions, including tables and figures with the inter-correlations, multiple and logistic regression models, and estimated probability plots. Chapter 5 is the discussion of the findings within the context of the research literature and current practice in the field, and also includes study limitations and suggestions for future research.

CHAPTER 2

REVIEW OF THE LITERATURE

Response to Intervention (RTI) involves using student assessment data to drive intervention efforts for individual and groups of students. Better understanding of the short- and long-term effects of reading interventions on student achievement and placement in special education could help schools in their selection and implementation of interventions for at-risk students. The current study examined potential implications of participating in a one-on-one reading intervention as part of an RTI model on student achievement and special education placement.

The purpose of the current study was to examine the long-term student outcomes for 99 students that participated in a reading intervention in 3rd grade. Student achievement, progress, and state accountability test data, along with categorical data regarding participation in tier 3 interventions or placement in special education 1 and 2 years after services, were used to determine (a) how pre and post-intervention level, dosage and slope during a tier 2 intervention predicted performance on group accountability measures of reading, and (b) how well level, dosage, slope and free/reduced lunch status, combined, predicted future risk status 1 and 2 years after intervention services.

The next portion is a literature review of the findings from longitudinal reading intervention research related to the areas of basic literacy skills including: phonemic awareness, phonics and fluency. This is followed by a summary of reading interventions

within the context of RTI. The final topic addresses the related decision-making models and assessments related to RTI as applied to the area of reading.

<center>**Longitudinal Reading Intervention Research**</center>

Intervening early with at-risk students can be effective. Some interventions have been found to provide improvements in performance during and immediately after services, whereas others have had mixed results semesters and years post-intervention. Below is a brief summary of reading interventions and their subsequent effects on student learning.

Phonemic and Phonological Awareness

Intervention research with kindergarten students often focuses on phonemic awareness, which is the area of literacy that involves the ability to manipulate phonemes in words, it helps prime the connection between sounds and print, and is thought to be one of the pre-literacy skills that can help most children learn to decode (National Institute of Child Health and Human Development [NICHD], 2000; Snowling & Hulme, 2007). It is thought to lead to alphabetic coding, which enables children to recognize words (Stanovich, 2000). Phonemic awareness is a sub-category of the larger skill area of phonological awareness which encompasses identification and manipulation of larger aspects of spoken words, beyond phonemes, common examples include rhyming and alliteration (NICHD, 2000). Both phonemic and phonological awareness are teachable and can be promoted by specific explicit attention to instructional variables (Smith, Simmons & Kame'enui, 1998). Instruction and practice in phonological awareness with kindergarten students for 3 and 6 months resulted in significantly improved skills for

students (Schneider, Ennemoser & Roth, 1999; Torgesen & Davis, 1996). In a synthesis of 20 years of research, phonological awareness was a common and effective aspect of reading interventions used with kindergarten students, resulting in moderate to high effect sizes for students with and without disabilities (Cavanaugh, Kim, Wanzek, Vaughn, 2004).

Children with strong phonemic awareness have an easier time learning to read than children with weaker skills (Cunningham, 1999), and interventions regarding phonemic or phonological awareness can mutually strengthening each other (Shaywitz, 2003). Longitudinal studies of intervening to support phonological skill development in young children, have demonstrated some positive outcomes in subsequent years. Preschoolers that received phonemic awareness interventions had higher skills in first sound fluency, (Koutsoftas, Harmon & Gray, 2009) spelling, reading words, and applying the alphabetic principle for decoding than a control group during kindergarten of formal schooling (Byrne & Fielding-Barnsley, 1993). Kindergarten students that participated in a phonological awareness intervention had higher end of school year scores than a control group, and at the beginning of first grade the students still performed higher than the control group (O'Connor, et al., 1996, 1998).

In follow-up studies of the students that participated in the Byrne and Fielding-Barnsley, (1993, 1995) intervention, a combination of phonological and phonemic awareness, the treatment group exceeded the control group in pseudoword reading, but not real word reading at the end of 1st grade. In 2nd grade, treatment students exceeded controls in reading comprehension scores (Byrne & Fielding-Barnsley, 1995). In third

grade, the treatment and control groups did not differ in reading or listening

comprehension, nor in the area of word identification, however, for pseudo word reading,

the treatment group performed better than the control group. Additional follow-up was

conducted when students were in 5th grade (Byrne, Fielding-Barnsley & Ashley, 2000)

and again the treatment exceeded the control group in word attack, and reading of regular

and irregular words. Moreover, meta-analytic research found that phonological

awareness interventions with young children led to small long-term effects for reading

comprehension and negligible effects for overall reading achievement (Bus & Van

Ijzendoorn, 1999). Thus, early literacy interventions may need to address more than

phonemic awareness to have strong long-term impact on students' overall reading

achievement.

Phonics

Phonics is a common area of reading instruction that typically follows or

accompanies phonemic awareness, and is the aspect of instruction that addresses using

letter-sound correspondences to identify words (NICHD, 2000). Phonics instruction

increases a student's ability to decode words, which is the largest difference of various

reading skills between strong and weak readers (Juel, 1988).

Phonics interventions in kindergarten. A combination of phonemic awareness

and phonics with kindergarten students, can be a powerful intervention. In studies by

Ball and Blachman (1988, 1991) such an approach resulted in significantly better

phoneme segmenting, letter-sound knowledge, and applied phonics skills for students in

the treatment group. Fuchs, et al. (2001), also examined the outcomes of intervening

with kindergarten students in multiple areas of phonological awareness and phonics and reported that treatment groups exceeded the control group on assessments of phonological awareness, phonics skills, sight-word reading, connected-text reading and spelling.

Intervening intensively with at-risk kindergarten students, can result in many students maintaining gains into later grades (Coyne, Kame'enui, Simmons & Harn, 2004) especially when the intervention focused on both phonological awareness and the alphabetic code (Vadasy, Sanders & Peyton, 2006). For example, 62% of at-risk second graders that participated in reading intervention services in kindergarten or kindergarten and first grade, scored at grade level benchmarks, as compared to 45% of not at-risk kindergarten students who did not receive intervention (Cartledge, Yurick, Sing, Keyes, & Kourea, 2011). Some at-risk kindergarten students that made rapid growth during reading intervention services maintained their skills into the spring of 3rd grade, as did students who made slower progress and received services in kindergarten and first grade (Vellutino, et al., 1996; Vellutino, Scanlon & Zhang, 2008). Lennon and Slesinski (1999) delivered daily intervention in multiple areas of reading to kindergarten students, which resulted in improved short-term reading skills for the treatment group, but also reduced the likelihood of later requiring special education services as compared to the control group. However, phonological awareness and phonics interventions with kindergarten students consistently led to long-term gains on measures of phonics there were less direct benefits for vocabulary and reading comprehension outcomes (Torgesen, Wagner, & Roshette, 1997).

Phonics interventions in later elementary grades. Some students do not make sufficient gains in kindergarten and continue to struggle in first grade (Vellutino, Scanlon, & Sipay, 1997), but reading interventions with first-grade students has also led to long-term gains 2 years later (Vadasy, Sanders, & Abbott, 2008). Torgesen, et al. (2001) trained staff to work with at-risk first grade students for an additional 2 hours per day in the areas of phonological awareness, phonics and written language. Standard scores on reading tests improved, from 1 standard deviation below the mean, to scores in the average to above average range. Torgesen, et al. (2001) also found that the strongest impact was, "on the children's ability to apply phonemic decoding strategies to unknown words" (Al Otaiba & Torgesen, 2007). Moreover, interventions with the First-Grade PALS (Mathes, Howard, Allen & Fuchs, 1998) resulted in improved outcomes for low-achieving students, and non-significant or short-term gains for average to high-achieving students (Mathes & Babyak, 2001; Mathes, Torgesen & Allor, 2001).

Reading Fluency

Fluency is another important reading skill area noted by the National Reading Panel report (NICHD, 2000). Fluency involves reading letter names, letter sounds, words, or connected text accurately and automatically (Vaughn & Linan-Thompson, 2004) thereby easing working memory and supporting comprehension (Chard, et al., 2002; LaBerge & Samuels, 1974). Logan (1997) described automatic reading as occurring via repeated encounters, or instances, therefore resulting in memory retrieval that has speed, is effortlessness, uses autonomous processing and does not require conscious awareness. When automaticity is applied to lower level reading skills such as

letter patterns, and at the word level, higher processing can be applied in the form of comprehension (Perfetti, 1985; Samuels & Flor, 1997)

Fluency is commonly improved by using a technique called repeated readings, which involves reading the same text multiple times (Samuels, 1997; Nathan & Stanovich, 1991). Common components of a repeated reading intervention include having students read chorally, pairing students to read and reread text to one another, and having students read while listening to audio recordings of text (Vaughn & Linan-Thompson, 2004). The fluency and comprehension of students with and without learning disabilities, can be improved with re-reading practice via adult modeling, goal setting, cuing and corrective feedback (Chard, et al., 2002; Therrien, 2004).

Therrien (2004) reviewed the existing literature about repeated reading, and found 18 studies which fell into the areas of transfer or non-transfer for students with and without disabilities. Transfer was described as how well a reader transfers the reading skill to an unstudied text. Components of transfer studies included delivery of the intervention by peers or adults, the use of modeling, corrective feedback, performance criteria, comprehension, and charting. Transfer interventions at the bottom end of the range included fluency practice without modeling, resulting in an effect size of .30. At the high end, transfer fluency with performance criteria resulted in an effect size of 1.70 for fluency. Transfer applied to comprehension resulted in the weakest effect size when modeling was used, and with the greatest effect when the tutor was an adult. For students without disabilities, transfer studies resulted in fluency and comprehension effect sizes of .59 and .18 respectively, and higher effect sizes, .79 and .41 for students with disabilities.

Non-transfer meant how well a student's comprehension and fluency improved as applied to the text used for the repeated reading. The author categorized non-transfer studies according to intervention components including: cuing students to read for rate, answering questions, a combination of the two (rate and answering questions), corrective feedback, and re-reading a text 2 to 4 times. For non-transfer fluency, the lowest effect size was attributed to studies where students read passages 2 times, and strongest when cuing focused on fluency and comprehension and passages were read 4 times For comprehension, non-transfer fluency studies effect sizes were weakest when passages were read 3 times, and cuing focused on fluency. Comprehension for non-transfer was strongest when fluency cueing focused on comprehension and passages were read 4 times. Non-transfer repeated reading resulted in fluency and comprehension effect sizes of .85 and .64 for students without disabilities, and .75 and .73, respectively, for students with disabilities.

The Therrien (2004) meta-analysis described above, deconstructed findings via component level aspects of transfer and non-transfer forms of repeated reading interventions. From a macro perspective, the impact of repeated reading interventions, result in students with higher fluency and better story recall skills than students who only read a passage one time (O'Shea, Sindelar, & O'Shea, 1987; Sindelar, Monda & O'Shea, 1990). Moreover, students who participated in fluency practice with adults in daily 5 minute sessions (Mercer, Campbell, Miller, Mercer & Lane, 2000), or as little as 10 minute sessions every 4 or 5 days, demonstrated improved fluency and comprehension. Adults who delivered one-on-one fluency practice with at-risk 2nd grade students resulted

in improved fluency and comprehension, compared to control groups (Begeny, et al., 2010; Begeny, Mitchell, Whitehouse, Samuels, & Stage, 2011). Fluency practice with pairs and small groups of at-risk and very low achieving students, have also been shown to be effective (Begeny, Krouse, Ross, & Mitchell, 2009; Vadasy & Sanders, 2008).

Long-Term Effects

Long-term follow-up research focused on the impact of fluency interventions, is limited, most interventions involve a combination of phonics and fluency. Blachman, et al. (2004) delivered 8 months of 50 minute phonological, orthographic, and fluency focused reading intervention sessions to second- and third-grade students, and then conducted a follow-up study 1 year later. During the intervention, the reading skills of the treatment groups exceeded that of the control groups. At 1 year follow-up, the treatment group's rate of progress had slowed, matching the rate of the control group. Intervention with a cohort of kindergarten and first-grade students at-risk for reading problems that combined phonemic awareness, phonics, sight-word reading, and fluency led to 84% of the students scoring in the average range on a test of basic skills by the end of the first, second, and third-grades (Vellutino, Scanlon, & Zhang, 2008), which is an effect that has been successfully replicated multiple times (Burns, Senesac & Silberglitt, 2008; O'Connor, et al., 1996; 1998).

These reading intervention follow-up studies indicate positive immediate, short-term, and 2 year maintenance of skills for many students, although not all. Some kindergarten and first grade students were keeping pace with their peers, at the time of a 1 year follow-up (Vaughn, Wanzek, Linan-Thompson & Murray, 2007; Vellutino, et al.,

1996) or were performing better than at-risk peers 2 years later (Burns et al., 2008). However, small numbers of older students struggled to maintain gains 1 year post intervention, (Blachman, et al., 2004) and some older students failed to maintain the same rate of improvement in 3rd grade and beyond without intervention (Vellutino, et al., 2008).

Growth measures from the winter of kindergarten with students who participated in reading intervention accurately predicted 75% of the children who would continue to make adequate progress in the future or who would continue to need intervention (Vellutino, et al., 2008). Vaughn, et al., (2007) worked with struggling students in first grade, providing daily 30 minute sessions in multiple areas of reading. Students who had responded slowly during first grade received intervention in second grade as well. The rapid responders during 1st grade maintained their skills through 2nd grade, without intervention, scoring higher than the slow responders who received 2 years of intervention. Results indicated that 1st grade slope of growth was a strong predictor of future performance.

Synthesis

Research demonstrates that interventions can help most students improve reading skills, but some students do not demonstrate measurable gains, or fail to maintain gains over time (e.g., 2% to 5% Al Otaiba & Torgesen, 2007, 2% Simmons et al., 2008). Thus, more follow-up research is needed to know how older students that participated in reading interventions perform 2 or more years post intervention (Al Otaiba & Torgesen, 2007). Moreover, research that found that a small percentage of students will continue to

need intensive intervention suggests the need for a systematic RTI model to meet the needs of all students.

Reading Interventions and RTI

Organizing school-wide assessment data into three tiers has become the basis for most RTI models (Shaprio & Clemens, 2009; Sugai & Horner, 2009; Marston, 2003). RTI involves using on-going progress information beginning in kindergarten, or earlier, to implement interventions as student progress indicates a need. Research about the effects of RTI implementation on student performance and placement in special education is promising. In a meta-analysis of field and research-based implementation of RTI models, Burns, et al., (2005) found that the RTI model improved student performance and systemic variables (e.g., fewer students referred for special education, fewer students retained in a grade, etc.). Hughes and Dexter (2010), summarized 11 field studies of RTI that were published in peer reviewed journals, 4 of the studies reported higher student reading outcomes as a result of RTI models.

The success of RTI models, greatly depends on data-based decision making. The three-tier model is a fundamental aspect of RTI that provides a structure for organizing data and intervention services to improve student outcomes. Each of the three tiers has a role to play in responding to student needs via instruction and intervention.

The base of the triangle is tier 1 and represents the 85% to 90% of students who demonstrate adequate reading skills. Thus, the majority of the students at each grade level should be performing at national or criterion-referenced academic standards in the fall, winter and spring of every school year. If less than 80% of the students are

performing at grade level, then a universal class-wide or school-wide intervention is warranted (Witt & VanDerHeyden, 2007; Batsche, et al., 2006). Strengthening the core instruction by providing more explicit instruction in phonics, has led to improvements in student achievement. At one school, the percentage of students scoring below the 25th percentile, fell from 31.8% of first graders to 3.8% after 4 years (King & Torgesen, 2000). A comprehensive RTI prevention model that employed flexible grouping and grade-level data team meetings led to an increase in the percentage of students who met grade-level curriculum-based measurement targets, from 38% to 70%, and the number of students that met proficiency on state reading tests improved from 51% to 80% (Bollman, Silberglitt & Gibbons, 2007).

Tier 2 consists of selected interventions for the 5% to 15% of students at-risk for academic failure (Vaughn & Linan-Thompson, 2004). Tier 3 targeted interventions are for 1% to 7% of the student body with the greatest concerns (Sugai, Sprague, Horner, & Walker, 2000). The goal is for schools to improve student performance so that ideally, 100% of students are performing in tier 1, and tiers 2 and 3 are not needed.

The effectiveness of tier 2 and tier 3 interventions are typically assessed by the growth of student progress via intervention, and reduction in the number of students found eligible for special education services (Marston, 2003). In terms of growth, about 70% of students that receive tier 2 interventions make gains to a level indicating that they are no longer at-risk, and maintain these skills 1 year post intervention, whereas students that receive tier 3 make more rapid growth with intervention than peers that do not indicate a need for intervention (Vaughn, et al., 2007; Vaughn, et al., 2008; Vaughn, et

al., 2009). Research syntheses have found that for students in kindergarten through 3ʳᵈ grade: students in interventions perform better than control groups, a daily fluency and phonics combination is most effective, smaller group sizes are desirable, and effect sizes for students in kindergarten and first grade are greater than students in 2ⁿᵈ and 3ʳᵈ grades (Scammacca, Vaughn, Roberts, Wanzek, Torgesen, 2007; Wanzek & Vaughn, 2007; Wanzek & Vaughn, 2010). With regards to special education eligibility, findings from RTI school implementation of tiers 2 and 3, has resulted in gradual reductions in the numbers of students identified for special education as cohorts of students experienced multiple years of RTI via professional development for their teachers (O'Connor, Fulmer, Harty, & Bell, 2005; Wanzek & Vaughn, 2011).

Decision making within RTI

Although in-depth data analysis and individualized interventions are needed for students with the greatest need (i.e., tiers 2 and 3; Burns, Christ, Boice, & Szadokierski, 2009), previous research found that resource decisions for significantly struggling students were based on data that did not distinguish students labeled as learning disabled and low achieving (Ysseldyke, Algozzine, Shinn & McGue, 1982). Teacher decisions were biased toward referral information, and used limited test data when recommending students for special education (Algozzine & Ysseldyke, 1981; Algozzine, Ysseldyke, Hill, 1982). Research has consistently questioned traditional approaches to identifying students with a learning disability (LD) because of resulting inconsistent LD prevalence rates across states and poor predictions regarding student response to interventions (Reschly, Kicklighter, McKee, 1988b; Vellutino, Scanlon & Lyon, 2000).

The data from the compilation of studies conducted by the Institute for Research on Learning Disabilities (Ysseldyke, et al., 1983), litigation (Reschly, Kicklighter & McKee, 1988a, 1988b, 1988c), concerns about overrepresentation of minority students in special education (Reschly, 1988) and the problem solving/consultation movement (Sheridan, Kratochwill & Bergan, 1996), led to a paradigm shift. Pre-referral intervention teams became common in most states (Carter & Sugai, 1989; Buck, Polloway, Smith-Thomas, & Cook, 2003) to assist teachers in implementing interventions with at-risk students before special education assessments were conducted (Graden, Casey & Christenson, 1985), and relied heavily on consultation and measuring student progress (Carter & Sugai, 1989). Yet another evolution of the pre-referral process has occurred, moving to the more current RTI model. Research on different types of pre-referral intervention teams found that using the model resulted in fewer students being referred for special education eligibility evaluations (Burns et al., 2005; Burns & Symington, 2002; Bollman, et al., 2007) and a higher proportion of students who were evaluated being identified with a special education disability probably because of the implementation of research-based interventions as a screener (McNamara & Hollinger, 2003).

An RTI model analyzes student academic progress with regards to level and slope to determine which students should receive special education services due to a LD (Fuchs & Fuchs, 1998; Fuchs, Fuchs & Speece, 2002). Students are dually discrepant if they score below an accepted standard for demonstrating adequate skill and for their rate of growth (Fuchs, 2003). Speece, Case and Molloy (2003) found that across multiple

studies, the application of the dual discrepancy model has more reliably resulted in identifying a population of students who do not respond to interventions, than the intellectual/achievement discrepancy model. Moreover, students who were identified as dually discrepant after intervention demonstrated significantly lower reading skills on an independent measure than a comparable at-risk population who was not discrepant (Burns & Senesac, 2002).

Early identification decisions

A body of research has been growing in an effort to determine how early and with which assessments, one can identify students at-risk for reading disabilities, and which intervention approaches effectively improve student outcomes. The follow-up pre-school, phonemic awareness study by Byrne, et al., (2000), reported that the poor readers in 5th grade, were the students that had made slow progress during the intervention in pre-school, and had low phonemic awareness test scores at the end of the pre-school intervention. Al Otaiba and Fuchs (2002) reviewed the research regarding the characteristics of non-responders, and found that low phonological awareness skills were common. Correlational studies reviewed by Nelson, Benner and Gonzalez (2003) also indicated phonological awareness, but also identified short-term memory, rapid automatic naming, inattentive behavior, poor acquisition of the alphabetic principal, and intellectual functioning. Going beyond the correlation approach used by Nelson, et al.(2003) research has found that intellectual functioning is not a predictive characteristic when determining who will respond to interventions (Burns & Scholin, in press; Vellutino, Scanlon & Zhang, 2007).

Speece, et al. (2011) determined that a combination of a teacher rating scale of reading problems, a measure of sight word reading efficiency, and word identification reading fluency produced a valid and efficient model for predicting reading risk status at the end of first grade, resulting in an Area Under the Curve, value of .96. Meta-analytic research found that slope pre-intervention level, as applied to RTI, was a strong predictor of future intervention effectiveness (Burns & Scholin, in press; Tran, Sanchez, Arellano, & Swanson, 2011). The studies reported correlations of .70 (Burns & Scholin, in press) and .72 (Tran et al., 2011) between pre-intervention and post-intervention levels, but the correlation between pre-intervention reading fluency and growth as measured by slope fell to .37 (Burns & Scholin, in press).

With the growing knowledge of the profile of skill deficits common in students that continue to struggle in later years, comes the need for more research regarding efficient and early identification of at-risk students. Early identification can lead to effective interventions that may need to be applied over several years, continuing into 3rd grade for some students, but potentially resulting in fewer students in special education. Moreover, Hughes and Dexter (2010) noted the need for more RTI related studies, one area being longitudinal work, tracking the performance of students who were identified early in their academic progress and receive early intervention.

Many have advocated for research focused on a combination of level and slope to provide a more accurate representation of student performance relative to interventions, and to do so with curriculum-based measurement (Al Otaiba & Fuchs, 2002; Burns & Senesac, 2005; Fuchs & Fuchs, 2004). Oral reading fluency is a commonly used

curriculum-based measure (CBM; Deno, 1985), and consists of having students read leveled passages for 1 minute and using established scoring rules to determine the number of words read correctly per minute. Many schools administer CBM-Reading (CBM-R) oral reading fluency measures 3 times per year, as a benchmarking screening tool, for implementing RTI. Deno (2003) stated that CBM-R can be used for many purposes that support an RTI model, including: (1) universal screening, (2) providing teachers with formative assessment data to guide decisions, (3) predicting student performance pre-kindergarten through school years, relative to important criteria such as target scores, growth targets, and performance on state exams, (4) evaluation of pre-referral and special education interventions, and (5) special education identification. Because CBM-R is being used to make important decisions for schools and individual students, researchers have been addressing the assessment's aspects of level (through screening) and slope, via progress monitoring. These points will be discussed below.

Screening

CBM- Reading (CBM-R) is used in many schools with RTI models, in conjunction with target scores, to measure student performance prior to state accountability exams, (No Child Left Behind Act, 2002), and improve instruction for students (Batsche et al., 2006). Bias associated with CBM-R is an important topic that has been researched using CBM assessment results and student demographics. CBM-R was shown to not be biased when using common demographics of race, age, gender and socio-economic status to predict reading achievement (Hintze, Calahan, Matthews, Williams & Tobin, 2002),

Although psychometric data for CBM are important, the goal of reading interventions is to improve student skills to a point at which the student can continue to progress independently at a rate similar to or greater than an established criteria or the rate of their peers (Fuchs & Fuchs, 2007). In a large longitudinal study, (Young-Suk, Petscher, Schatschneider, & Foorman, 2010) students were followed from kindergarten to 3rd grade and assessed frequently in reading skills via the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) oral reading fluency and multiple standardized reading comprehension, tests to determine growth and what different assessments contributed to a student's performance on reading achievement tests. When examining first grade assessments and growth, oral reading fluency slope in 1st grade was a strong variable, as were baseline level oral reading scores in 2nd and 3rd in relation to a student's reading comprehension in 3rd grade.

Because state accountability testing is the criterion to which student reading performance is compared, several researchers have studied the relationship between CBM data and state accountability test scores. Stage and Jacobsen (2001) were first to study the use of CBM-R rates to predict student performance on a state accountability exam. Fourth grade students were tested at the beginning, middle and end of the school year using curriculum-based oral reading fluency measures. Multiple regression analysis found that student oral reading fluency levels at the three points during the school year, better predicted performance on the state exam than did the slope. Specifically, the researchers found that student seasonal fall scores in oral reading fluency (level) better

predicted student performance on the state exam in the spring, than their oral reading fluency slope across the school year.

Correlations between CBM oral reading fluency assessments and state test scores, have been reported by a variety of researchers (Barger, 2003; Buck and Torgesen, 2003; Good, Simmons, & Kame'enui, 2001; Shaw & Shaw, 2002; Vander Meer, Lentz & Stollar, 2005; Wilson, 2005). Correlations range from .67 for a sample of 364 third grade students (Good, et al., 2001), to .81 for fourth-grade students (McGlinchey & Hixson, 2004). With correlation established, receiver –operating characteristic (ROC) curves have been used to determine cut or target scores, for seasonal benchmarking, relative to student performance on state exams (Silberglitt & Hintze, 2007; Roehrig, Petscher, Nettles, Hudson & Torgesen, 2008). For example, in Florida, benchmark assessments were conducted with 3rd grade students in the September (Fall), December (Winter 1) and February/March, (Winter 2), the Winter 2 score had the highest correlation with performance on the state reading exam, administered in March (Roehrig, at al., 2008). Criterion reference targets were then determined, the Winter 2 target for low risk was selected as 98 words correct per minute (wcpm), which resulted in 81% of the students meeting the state test goal. In a study by Crawford, Tindal and Stieber (2001) 94% of students with oral reading fluency scores of 119 wcpm in 3rd grade passed the third grade state reading test.

CBM-R data are reliable and predict performance on a state reading exam. However, studies varied in findings as to which seasonal level score, spring (Keller-Margulis, Shapiro, & Hintze, 2008; Good, et al., 2001), or fall (Stage & Jacobson, 2001),

may be the better predictor of achievement on state tests. However, progress monitoring baseline level and end level scores were not included in these studies, and is an area for further research. Slope was also inconsistent as an independent variable. Stage and Jacobsen (2001) found that individual fall scores were a better predictor than slope using 3 data points, and Keller-Margulis and colleagues (2008) reported that slope after 2 years did predict students' state reading test scores, but 1 year's worth of slope data did not. The studies discussed used slope from benchmark assessments and not progress monitoring data, indicating a need for investigating slope within progress monitoring as a predictive variable.

Monitoring Progress

Monitoring progress using CBM-R results in measures of both level and slope. Level is a score obtained during 1 sitting, either an individual score of 1 probe (Ardoin et al., 2004) or a score resulting from a median of 3 probes, whereas slope is growth between scores across time (Shinn, 2002). Research has found that slopes derived from CBM-R cannot be predicted by student demographic variables alone, indicating low assessment bias and strong validity (Chard, et al., 2008). Recent research supported these prior findings, CBM-R slope was not biased with regards to socio-economic status, but special education status was a significant demographic predictor for 3rd grade students (Yeo, Fearrington & Christ, 2011).

Deno, Fuchs, Marston and Shin (2001) used linear growth via ordinary least squares regression, to estimate weekly growth rates, a higher rate for students in general education of 2 words per week, and a lower rate for students in special education, less

than 1 word per week. Silberglitt and Hintze (2007) used hierarchical linear modeling with fall, winter, spring data for a very large sample of student scores (7,544) in 2nd through 6th grade and found that the fall score was the critical variable related to student growth rates. Low fall scores had corresponding lower growth rates, and were significantly different from students with higher fall scores, which had higher growth rates.

The distribution of slopes derived from CBM-R has been extensively studied, but with different results. One study collected data from a sample of students weekly for an entire year and determined that average student growth was linear (Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993), but later research found that growth was most rapid during early elementary grades, and particularly from fall to winter versus winter to spring (Ardoin & Christ, 2008), when applying a piece-wise model, (Christ, Silberglitt, Yeo, & Cormier, 2010), and at a faster rate for general education students compared to special education students (Christ, et al., 2010). Research by Katz, Stone, Carlisle, and Corey (2008), demonstrated that students with similar academic and demographic characteristics as students with high incidence disabilities, made significantly slower growth than non-disabled peers.

Given that CBM-R progress monitoring data are frequently used in RTI models, to determine if instruction or intervention is effective for students, error associated CBM-R data has become a topic of research. The standard error of measurement (SEM) associated with CBM – R is high with small numbers of data points, and improves with a larger number of samples, making it important to collect an adequate amount of data

prior to making high stakes decisions (Poncy, Skinner & Axtell, 2005; Shinn, Good, & Stein, 1990; Christ & Ardoin, 2006). Christ (2006) examined the technical adequacy of using oral reading fluency progress monitoring data to determine a reliable and stable slope, and reported a SE*b* range of 9.19 for 2 data points, to 1.0 for 20 data points. A SEM range of 5 to 15 words, was determined using tri-annual data consisting of 8,200 student records (Christ & Silberglitt, 2007).

Due to this summation of findings regarding growth rates, SEM, and performance of students receiving special education, high stakes decisions using CBM-R should be thoughtfully considered. Optimum testing conditions should be applied to reduce chances for error, and sufficient data need to be collected. The best method for calculating growth/slope for benchmarking, versus progress monitoring data, is an area that warrants further research.

Synthesis

The findings from this literature review indicate that early intervention can improve outcomes for at-risk students, resulting in immediate and long-term impact on student reading skills and decreased likelihood of placement in special education. Not all researchers who intensively worked with students in kindergarten and first grade have published findings related to student performance in later elementary years, which suggests an area that needs further research. Intervening with at-risk students beginning in 3rd grade is not a common focus in the literature, yet state accountability exams are commonly administered at the end of 3rd grade or in 4th grade. RTI is a paradigm that is driving school staff and administrators to use data to measure student progress via core

instruction and interventions. Schools employing systemic school-wide RTI models are applying universal screening commonly using CBM, and intervening with all at-risk students in the school regardless of grade level. CBM has been found to have strong predictive validity of student performance on achievement tests, including state reading exams, and can be used to predict which students will continue to struggle if schools do not provide on-going intervention support.

Study Purpose

Previous research found that multi-skill reading interventions can lead to long term growth and reduction in students being identified with a disability (Torgesen & Davis, 1996; Lennon & Slesinski, 1999) , but few researchers have taken the approach of Simmons et al. (2008) and implemented reading interventions for 3 to 4 years within an RTI tiered system. The current study is intended to add to the growing body of research around implementation of RTI in schools and explores the degree to which student progress during an intervention predicts some selected academic outcomes. Moreover, research about the effects of RTI implementation and reduced placement in special education is promising (Burns et al., 2005; Hughes & Dexter, 2010), but more research is needed about student performance in later elementary school years, and if there is a relationship with weekly progress monitoring data. The current study will include long term follow-up 1 and 2 years later, of students who participated in a tier-2 reading intervention, as recommended by Hughes and Dexter (2010), and will examine performance on state and district reading tests, and referrals to intervention teams or prevalence of special education placement.

The research questions are: (1) How does student performance during a tier 2 intervention, predict performance on group accountability measures of reading? (2) How well do level, slope, lunch status and dosage, combined, predict future risk status 1 and 2 years after intervention.

CHAPTER 3

METHOD

The purpose of the current study was to examine the outcomes of a sample of third-grade students who were at-risk third for reading problems after participating in a systematic reading intervention. Specifically, outcomes related to student performance on the state reading exam, and a standardized reading achievement test, as well as placement in pre-referral intervention programming and special education. The research questions were: (1) How does student performance before and during a tier 2 intervention, predict performance on group accountability measures of reading? (2) How well does student performance and dosage during a tier 2 intervention predict future risk status 1 and 2 years after intervention services? The first hypothesis was that student progress during a reading intervention would predict student performance on state accountability tests and other reading achievement tests. The second hypothesis was that student performance during a reading intervention would predict student participation in pre-referral and special education programming.

Participants

The current study used existing data for students that participated in a reading intervention service in 8 rural elementary schools in one Midwestern state. The size of the average community was 3,300 residents (U.S. Census Bureau, 2000). Students were all in 3rd grade. The average elementary school enrollment was 586 students. Most of the schools, 6, served students in kindergarten through 6th grade. However, one school had

students from kindergarten through and 3rd grade, and another enrolled students from 3rd through 5th grade.

The percentage of students in 3rd grade that met proficiency on the state reading exam was gathered from the state department of education website, to further describe the schools. Test results regarding the average score for students at each of the schools, were not found on the data reporting portion of the website. Data was not available for the year the student participants completed the test, nor for the years prior. Results were available in a graphic representation, in 20 unit increments along the 'x' axis for 3rd grade students in the cohort that followed the students in the current study. The state average for that year was 79%, the 8 schools had the following rates of meeting proficiency: 82%, 85%, 77%, 81%, 84%, 78%; 86%, and 81%; the range was 77 – 86%, resulting in an average of 81.75% of students having met proficiency on the exam.

Students

The original sample contained 112 students, but 13 students were removed because they had 5 or fewer total progress monitoring data points. Thus, 99 students comprised the final sample. Due to absenteeism or students leaving the school, test scores and records of participation in pre-referral interventions or special education were missing for some students. Results from the state reading test were missing for 6 student records. Six of the eight schools used the Measures of Academic Progress (MAP; Northwest Evaluation Association [NWEA], 2005) as their district reading test, resulting in scores for 72 students. Records of special education and pre-referral participation 1 year and 2 years post intervention were missing for 6 and 12 students, respectively. It was

assumed that those data were missing at random and they were excluded from the final analysis.

The racial make-up of the group of students was homogenous, in that 93% were white, the remaining 7% were African-American, Hispanic, American Indian/Alaska Native, and Asian. Demographic data were missing for 6 students. None of the students were identified as receiving services for English language learning needs, and 38% of the students were eligible for a free or reduced price lunch (data were missing for 6 students).

Tutors

The students were served by eight full-time tutors and one part-time tutor, all of whom were recruited over the summer and hired prior to the start of the school year. The nine tutors worked in eight different schools, (one school had both a full- and part-time tutor). All tutors had high school diplomas, one had an associate's degree, two had some college, and four had undergraduate degrees.

Local decision-making by instructional coaches varied, regarding the number of students serviced by tutors at each grade level. There was a large range in the number of 3rd grade students served by the tutors at the different schools. One tutor had only two third-grade students, another had four, two tutors had seven, and the remaining worked with 9, 10, 11, 13, and 36 students. The tutor that serviced 36 students over the course of the school year, worked at the 3rd – 5th grade elementary school, servicing only students in 3rd grade for the entire school year.

Variables and Measures

Three quantitative measures of reading performance were gathered from student records: (1) CBM oral reading fluency scores, (2) state reading exam scores, and (3) the MAP reading achievement test results. Tutors maintained service logs, recording tutoring session durations with students. Tutor logs were used to calculate an estimated dosage variable. Information regarding student placement in pre-referral and special education programming was gathered from the schools.

The research questions were addressed using multiple and logistic regression analyses. The predictor variables included: lunch status, student reading progress monitoring scores during service, and intervention dosage. The outcome variables were student performance on state and district reading tests at the end of the school year, student participation in the special education pre-referral processes 1 or 2 years post 3rd grade, and placement in special education services 1 and 2 years post service.

Predictor Variables

This study was comprised of categorical and continuous predictor variables. The list of predictor variables included: (1) free/reduced price lunch status, (2) oral reading fluency level at the beginning of the intervention (3) oral reading fluency level at the end of the intervention, (4) slope during the intervention, and (5) dosage of intervention for each student.

Lunch status. There was large variation within the sample, and across the schools, with regards to free/reduced price lunch status of the students. According to the Department of Education (1996) free and reduced price school meal status is as an

indicator of socioeconomic status used by the government to determine eligibility for federal and state education funding and programs such as Title 1. The National Center for Education Statistics (2007) and the federal education law, No Child Left Behind (2002) use lunch status as an indicator of poverty. Free/reduced price lunch was included as a predictor variable. Some students moved after intervention services, and therefore free/reduced lunch records were missing for 6 students. Students not eligible for free or reduced price lunch were coded a 0, students eligible for free or reduced lunch status were coded a 1.

Oral reading fluency level. Tutors were trained in assessment of oral reading fluency with curriculum based measurement-reading (CBM-R) and fluency interventions. Tutors were trained in CBM-R using the AIMSweb procedures and the accompanying third-grade oral reading fluency probes (AIMSweb, 2002; Shinn & Shinn, 2002; Howe & Shinn, 2002).

All students were administered three grade-level oral reading fluency passages, and the median number of words read correctly was used to determine student eligibility for participation in the service. Students with median scores below grade level target scores were eligible to receive the service. The fall target was 70 words read correctly per minute (wrcm) and the winter target was 91 wrcm. This initial score was used as the student's beginning level score in the current study. When intervention services ceased, an end median score was also derived from three oral reading fluency samples. The term target scores refers to grade level criterion reference fall, winter or spring oral reading fluency scores found to have strong predictive validity with regards to student

performance on the state third grade reading test (Hintz & Silberglitt, 2005). Therefore, a student with a score of 85 wrcm in the fall, which is above the fall target score of 70, would not be considered in need, but a student with a score of 85 wrcm in the winter, below the winter target of 91 wrcm, would be considered in need.

The AIMSweb oral reading passages for 1st through 8th grades had alternate form reliability ranging from .81 to .90 (Howe & Shinn, 2002). In a study examining parallel forms administered to 4th grade students, the correlation was .94 (Tindal, Marston & Deno, 1983). The reliability of test-retest administrations 1, 5 and 10 weeks apart for 3rd through 6th grade students were .90, .90 and .82 respectively (Marston, 1982; Shinn, 1981).

Another aspect of CBM-R reliability is administration and scoring. The existing data set used in this current study did not include inter-rater reliability scores. Assessment reliability was conducted informally during tutor training, with participants sharing their scores with their colleagues and the trainer, but specific inter-observer agreement scores were not recorded. According to research, oral reading fluency inter-rater reliability for trainees is usually quite high (e.g., 99%; Marston, 1989).

Slope. Weekly 3rd grade level parallel form CBM-R reading passages were administered to each student to monitor student progress. A resulting slope value was generated by AIMSweb for all progress monitoring graphs, using the ordinary least squares regression method (OLS). The OLS method has been used to calculate weekly growth rates for general and special education students in elementary grades (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, Fuchs, Hamlett, Walz & Germann, 1993). OLS is

used to estimate progress monitoring growth, resulting in trend/slope lines for guiding instructional decision-making (Good & Shinn, 1990; Shinn, Good, & Stein, 1989; Christ, Pike, & Monaghen, 2011). The standard error of the slope is strongly influenced by the number of data points, and factors related to the testing materials and environment (Christ, 2006). Participants in the current study had as few as 6 data points, an average of 18, and a maximum of 30.

One can examine the standard error of slope estimate for the scores for the current study within the context of Christ's (2006) research. For example, according to Christ (2006), six progress monitoring data points, the minimum in the current study, produced a standard error of estimate of between .97 and 8.71; comparatively a record with 18 data points, the average in the current study, had a much smaller standard error of slope, ranging between .18 and 1.64. A total of 30 data points had an even smaller slope range of .08 to .76. Due to the large standard error associated with slope, percentile ranks were determined for each of the 99 slope values, which were then coded as dichotomous, those below the 49th percentile were coded 0, and slope values at the 50th percentile and higher were coded 1. There were 49 slope values coded 0, and 50 slope values coded 1. When a student met the exit criteria, the intervention would stop and the tutor would begin to monitor the student's progress. A new student would then be assessed. If the new student scored below target, or if they were assessed between benchmark seasons, and scores indicated they were at-risk of reaching target, they would begin receiving services. The tutor's supervising teacher was involved in reviewing assessment results and the selection of students. Tutors and supervising teachers were trained to select students just

below target, referred to as tier 2 students, versus students with scores far below target, as that would indicate a need for more intensive instruction or intervention to be provided by a teacher. No bottom cut score was provided to tutors, rather on-going consultation was provided by the supervising teachers throughout the school year. Selected students were progress monitored 1 time per week during service, by the tutors.

Dosage. Another predictor variable is the 'dosage' of intervention service time each student received. Students were scheduled to receive the reading fluency service from trained tutors for a minimum of 60 minutes per week, in multiple sessions per week of 10, 15 or 20 minutes each. Tutors were trained to maintain written records, or logs, of each session conducted with each student, which were entered into an Excel spreadsheet as the number of minutes per week. Tutoring time was available for 95 student records, tutoring time was missing for 4 students. To calculate dosage of service for each student, the total number of minutes of service per student was determined. The total number of minutes of service was divided by the number of weeks of service to determine an estimated dosage for each student that was an average number of minutes per week of service. The estimate of the number of weeks of service for each graph was determined via the number of progress monitoring data points between the baseline level score and the end level score for each graph. There were as few as 3, (because the baseline level variable's data points and end data points comprised the baseline level and end level values) and at the maximum, as many as 30, with an average of 16 data points, across the progress monitoring graphs.

Response Variables

There were four response variables for this study including (1) performance on the third grade state reading achievement test, (2) the 3rd grade MAP for reading, (3) documentation of student involvement in pre-referral intervention processes and, (4) records of placement in special education.

State reading assessment. Students were administered the state reading achievement test in the spring of the 3rd grade. The test was administered to students at their respective schools, in paper/pencil format and was comprised of multiple choice and constructed response items representing three areas. The 44 multiple choice items were distributed across three areas aligned with the state standards for reading. The area of vocabulary was the smallest, with 6 to 8 test items comprising 14% to 18% of the test; comprehension had 9 to 18 test items or 20% to 41% of the test; and literature had the largest number of items with 15 to 24 which comprised 34% to 55% of the test. There were between 5 and 6 passages total, comprised of fiction, non-fiction and poetry, with a targeted word count range of 1800-2000, and a degrees of reading power range of 40-56 (Minnesota Department of Education, n.d.).

Student raw scores were converted to three-digit scale scores. Scale scores corresponded with 1 of 4 qualitative levels of proficiency. Scale scores at or below 339 *do not meet* state standards, scores between 340 and 349 *partially meet* state standards, those between 350 and 364 are considered to *meet* state proficiency standards, and scores at or above 365 *exceed* state standards (Minnesota Department of Education, n.d.).

District reading test. All but one of the school districts administered the MAP in the spring of 3rd grade. Scores for students who completed the MAP were provided by school administrators. The MAP is a computer adaptive test aligned with state standards, intended to provide schools with rapid results to aid in instruction. The third grade MAP test for Reading includes a broad reading score that is comprised of four sub-tests: Word Recognition, Literal Comprehension, Interpretive Comprehension, and Evaluative Comprehension. The MAP produces Rasch unIT (RIT) scores, which are equal interval scores calculated based on each individual test item's difficulty (NWEA, 2011). The score represents a level of difficulty at which the student was correct 50% of the time. A lower RIT range means a student could answer more than 50% of the items correctly, and a higher RIT range indicated they could answer fewer than 50% of the items correctly. Students with the same scores are at the same level, a 2nd grader with a score of 190 has the same meaning as a 7th grader with a score of 190 (NWEA, 2011). A RIT score less than 180 is equal to performing at a level on the state exam of *not meeting* state proficiency standards. Students with scores in this low range have limited reading skills including the association between single words and pictures, sight words, and recognition of initial consonants. The RIT score of 180, is at the 12th percentile, and is equivalent to *partially meeting* state standards. The RIT score of 190, is at the 26th percentile, and is equivalent with scoring in the *meets standards* range on the state exam. The RIT score of 202, is at the 58th percentile, and corresponds with the state exam's score range titled *exceeds standards* (Cronin, 2007).

District reading test. All but one of the school districts administered the MAP in the spring of 3rd grade. Scores for students who completed the MAP were provided by school administrators. The MAP is a computer adaptive test aligned with state standards, intended to provide schools with rapid results to aid in instruction. The third grade MAP test for Reading includes a broad reading score that is comprised of four sub-tests: Word Recognition, Literal Comprehension, Interpretive Comprehension, and Evaluative Comprehension. The MAP produces Rasch unIT (RIT) scores, which are equal interval scores calculated based on each individual test item's difficulty (NWEA, 2011). The score represents a level of difficulty at which the student was correct 50% of the time. A lower RIT range means a student could answer more than 50% of the items correctly, and a higher RIT range indicated they could answer fewer than 50% of the items correctly. Students with the same scores are at the same level, a 2nd grader with a score of 190 has the same meaning as a 7th grader with a score of 190 (NWEA, 2011). A RIT score less than 180 is equal to performing at a level on the state exam of *not meeting* state proficiency standards. Students with scores in this low range have limited reading skills including the association between single words and pictures, sight words, and recognition of initial consonants. The RIT score of 180, is at the 12th percentile, and is equivalent to *partially meeting* state standards. The RIT score of 190, is at the 26th percentile, and is equivalent with scoring in the *meets standards* range on the state exam. The RIT score of 202, is at the 58th percentile, and corresponds with the state exam's score range titled *exceeds standards* (Cronin, 2007).

The MAP is a computer adaptive test that uses Item Response Theory in its construction design. Reliability was reported by the publisher in the form of marginal reliability (Samejima, 1994). Marginal reliability applies the test information function to determine correlation values between two hypothetical scores for a given student, and can be used to calculate reliability for parallel test forms (NWEA, 2003). The marginal reliability for the MAP reading exam in 3rd grade, was .95 for fall and .95 for spring (NWEA). Test-retest reliability was also calculated for students that had taken the MAP multiple times. For students who completed fall and spring administrations, the test-retest correlation value was .87; the correlation test-retest reliability for students that completed the test in two consecutive spring seasons was .89 (NWEA).

Student MAP test results are provided to teachers in the form of individual student score reports and class reports that further disaggregate test results including descriptions of particular skill strands that need improvement. Third grade spring MAP student scores were provided to the researcher by school or district testing administrators or building administrators, the broad reading score was the single score available for the largest number of students, 72.

Pre-referral and special education status. Pre-referral and special education information was maintained by the schools in electronic and written format, respectively. During the pre-referral process, teams of school staff developed and implemented interventions to help struggling students improve their performance in school. Building administrators were active participants on each of the building pre-referral teams, therefore, they provided documentation to the researcher of students' involvement in pre-

referral support. Documentation included written forms in student files, and handwritten administrator notes on meeting agendas. For each student, administrative and individual student records were reviewed to determine if there was written documentation of pre-referral interventions during either the 1st or 2nd year post intervention.

All schools maintained electronic records of which students were receiving special education services. Using these records, school or district special education administrators reported to the researcher which students were eligible for special education in the 1st or 2nd year post participating in the reading intervention.

Procedure

Tutors were trained to serve students individually in kindergarten through third grade. During the first week of school, tutors were provided 3 days of training in kindergarten through third grade CBM, AIMSweb data entry, four early literacy interventions used with younger students, as well as six scripted connected text fluency interventions used with the 2nd and 3rd grade students. An instructional coach at each building was provided a 2-hour orientation of the program, and then collaborated with tutors regarding selection of students, development of schedules, and communication with teachers regarding implementation of the program throughout the year.

 All students were to be scheduled for 60 minutes of 1 on 1 tutoring service per week, delivered in either 10, 15 or 20 minute sessions. Students were serviced every week, progress monitoring data were collected 1 time per week, and students were exited from service once scores indicated consistent performance above the aim-line (more

detail is provided later, about this aspect of the program). Students began receiving

services in September at the earliest, and in April at the latest.

Students received services until their scores were consistently above the aim-line,

as illustrated on their progress monitoring graph. AIMSweb generates an aim-line from

the student's initial median score to the target score entered by users. Tutors entered the

third grade level spring target score of 107 wrcm, for the last day of the school year.

Intervention services were supposed to cease when student graphs contained three

consecutive weekly data points that were above the aim-line. After services ended,

progress monitoring was to be conducted by tutors for 3 consecutive weeks, and scores

were added to the progress monitoring graph. The average number of data points per

student was 18.

Descriptive information about the participants (i.e., lunch status, race/ethnicity,

pre-referral and special education involvement, and language) was provided by school

administrators. Some of the independent variables used in this study existed, and other

independent variables were derived. The independent variable baseline level was present

for every participant and was the first data point for every student's progress monitoring

record, which was a median score from three oral reading probes. When administering

three probes, six scores are generated, 3 words read correct scores, and 3 error scores.

The words read correct score with the median numeric value and the error score with the

median value comprised the final median score. For example, if a student read 45 words

correct and 3 errors, 51 words correct and 7 errors, and 43 words correct and 5 errors, the

median score would be 45 words correct and 5 errors (Howe & Shinn, 2002).

Determining a median reading score, uses the students scores from reading 3 different texts to produce a measure of central tendency that represents a student's actual performance, (as opposed to a mean which would involve calculation of a student's reading score, using the scores from the 3 different texts).

Design and Analysis

A descriptive research design was applied to answer the research questions, by using existing records from the intervention program and school records. The independent variable lunch, was coded 1 for students eligible for the free or reduced lunch program, and 0 for students ineligible. The first research question inquired about student performance during a tier 2 intervention predicting performance on group accountability measures of reading. The first part of this question was answered by conducting binary correlations with the predictor and response variables. Next, a series of regressions with each independent variable and the state reading exam as the dependent variable were conducted to examine the variance in test scores accounted for by the predictor variables. A final regression analysis was conducted with selected variables to determine the model that best predicted performance on the state reading exam. Finally, a logistic regression was conducted to determine the log-odds related to student performance during the intervention and performance on the state reading exam. For the logistic regression, student scores of 350 or higher, thus meeting the state's proficiency standard, were coded as a 1, and scores below 350 were coded as 0.

The second part of the first question was also answered using regression analyses, utilizing the same independent variables, but with the MAP data as the dependent

variable. A multiple regression analysis was conducted to determine how well the model related to student performance on the district administered test.

The second research question inquired about oral reading fluency and dosage during a tier 2 intervention predicting future risk status for students 1 and 2 years after intervention services. For this second question, a series of four binary logistic regression analyses were conducted, using lunch status and slope as binary covariates. The predictors for all the logistic regressions included: oral reading fluency baseline level, oral reading fluency end level, slope and dosage. The response variables were dichotomous, 1 for participation in pre-referral or special education programming, 0 for no participation. The first binary logistic regression analysis used pre-referral status 1 year post intervention as the dependent variable, the second analysis used pre-referral status 2 years post intervention as the dependent variable. The third and fourth analyses used placement in special education programming as the dependent variable 1 and 2 years post intervention, respectively. Results from the four analyses indicated how well the regression models estimated the odds that a student would participate in pre-referral or special education programming 1 and 2 years post participation in a reading intervention.

CHAPTER 4

RESULTS

The first research question examined whether, student performance during a tier 2 intervention predicted performance on group accountability measures of reading. The second research question inquired how well level, slope, dosage, and lunch status predicted future risk status for students 1 and 2 years after intervention services.

Variables

The sample sizes, means, standard deviations, ranges and confidence intervals for the continuous predictor variables appear in Table 1. Notable aspects of each variable are discussed in turn below.

Table 1

Descriptive Statistics for Predictors and Reading Test Response Variables

Variable	*n*	*M (SD)*	Range	95% CI
Predictor				
Baseline Level	99	53.88 (24.14)	5 to 107 wrcm	49.12 to 58.64
Slope	99	1.81 (2.75)	-.50 to 15.58 wrcm	1.27 to 2.35
End Level	99	87.85 (18.49)	10 to 131 wrcm	84.21 to 91.49
Dosage	95	51.04 (14.67)	15 to 95 min	48.09 to 53.99
Response				
District Test	72	191.21 (11.23)	159 to 214 RIT	188.62 to 193.8
State Test	91	354.48 (17.44)	304 to 398 scale scores	350.85 to 358.11

Note. CI = Confidence interval, wrcm = words read correctly per minute.

The variable lunch status was dichotomous, a 1 indicated the school records stated the student was eligible for free or reduced price lunch, and a 0 meant the student was not eligible. Lunch status was available for 93 records, 58 students were ineligible, and 35 were eligible.

Baseline level score, a student's oral reading fluency eligibility score, was another predictor and was determined immediately prior to the start of intervention services. The mean student baseline level score was below both the fall and winter target scores ($M =$ 53.88; $SD = 24.14$) of 70 and 91 words per minute, respectively. When displayed on a histogram, the distribution of baseline level scores appeared normal (skewness = .18, kurtosis = -.89).

An end level median score was determined for each student, using the last 3 progress monitoring scores of each student's record. The mean end level score ($M =$ 87.85, $SD = 18.49$) was below the winter and spring target scores of 91 and 107, respectively (ServeMN, 2005), indicating students may still have been at-risk for not meeting proficiency on the state reading test. The end level variable had a distribution that appeared normal (skewness = -1.06, kurtosis = 2.78).

The mean slope value was positive, ($M = 1.81$, $SD = 2.75$) approaching a rate of 2 words per minute per week. The distribution for the slope variable was positively skewed, with a sharp peak (skewness = 2.41, kurtosis = 7.36). The shape of the distribution was not normal, which would have violated one of the assumptions required for regression. Due to the leptokurtic shape of the distribution for the slope variable, and the concerns regarding error associated with slope values calculated with fewer data

points (Christ, 2006), slope was made a dichotomous variable. The slope values were first assigned percentile ranks within the sample distribution. Next, the slope values ranging from the 1st to 49th percentiles, were coded 0, and slope values at the 50th percentile and higher were coded 1.

A histogram of the dosage variable indicated a normal distribution (skewness = .194, kurtosis = .219). Students received an average of 51 minutes (M = 51.04, SD = 14.67) per week of intervention, which was less than the recommended amount of 60 minutes per week (ServeMN, 2005). Only 24% of the students received the recommended program dosage of 60 minutes or more per week.

The descriptive statistics for the response variables show a smaller standard deviation and range for district assessment scores compared to the state reading exam results. There were 72 MAP scores, the average was a RIT of 191, (M = 191.21, SD = 11.23) which is close to Cronin's (2007) target score of 190 and equal to the 26th percentile. When displayed on a histogram the distribution of MAP scores appeared normal (skewness = -.23, kurtosis = -.10).

A score of 350 or higher on the state reading test is considered meeting state proficiency standards. The mean score (M = 351.12, SD = 28.81) met proficiency standards. State reading test scores were available for 93 of the 99 participant records. The distribution of state reading test scores were negatively skewed (skewness = -3.41) and sharply peaked (kurtosis = 18.46), resulting in a non-linear S-like pattern associated with distributions with short or long tails that are commonly influenced by outliers.

The shape of the distribution for this variable would violate the assumption of normality, required for regression. Using a box and whisker chart, there were 2 obvious outlier scores at the lower end. The district reading test scores for both these students was reviewed. The district reading test scores for both participants were extremely high, above the 202 cut score, (203 and 214) indicating that they were very likely to perform in the highest bracket called, 'Exceeding Standards' on the state reading test (Cronin, 2007). Due to the extremely high performance of these students on the MAP, compared to their low performance the state exam, the validity of their scores on the state reading test were determined to be questionable and the outliers were removed. With the 2 outliers removed, ($n = 91$), the distribution of state reading test scores appeared normal (skewness = .10, kurtosis = .51). The mean became slightly higher, and the standard deviation was smaller ($M = 354.48$, $SD = 17.44$). In the final sample ($n = 91$), there were 53 student scores that *met* or *exceeded* the state's proficiency standards and 38 student scores that *did not meet* or only *partially met* state standards.

There were four dependent variables for the research questions related to risk-status: special education 1-year post intervention, special education 2-years post, pre-referral participation 1 year after intervention, and pre-referral 2 years after participation. One year after intervention 93 of 99 records were available, 14 of the records indicated students were receiving special education services, and 19 of the records indicated students were participating in pre-referral interventions. Two years after intervention, records were available for 87 of the 99 students. At that time, 16 students were receiving special education, 12 of which had been receiving services the previous year, 4 students

were new recipients.   Pre-referral participation 2 years later was at 25, an increase of 11

new students, 14 students were recipients the previous year.

Inter-correlations Between Variables

Inter-correlations were the first step in analyzing the relationship between the

predictor variables (baseline level, end level, slope, dosage, and lunch status) with the

outcome variables of student performance on district and state reading examinations.

The results from the inter-correlations appear in Table 2.

Table 2

Intercorrelations Among Predictor Variables and State and District Reading Tests

Variables	Baseline Level	End Level	Dosage	Lunch Status$_1$	Slope$_1$	State Exam	District Exam
Predictors							
Baseline level	-						
End level	.64*	-					
Dosage	.16	.02	-				
Lunch status$_1$	-.03	-.02	.18	-			
Slope$_1$	-.10	-.19	-.21*	-.06	-		
Responses							
State Reading	.37*	.39*	.07	-.11	-.29*	-	
District Reading	.34*	.49*	-.33*	-.28*	-.24*	.81*	-

1 = correlations for lunch and slope were determined using Spearman's rho

* $p < .05$.

Each pair of correlations was calculated using Pearson correlations, except for the dichotomous variables slope, and lunch, which were calculated using Spearman's rho.

Three of the five predictors, slope, baseline level score, and end level scores, were significantly correlated with the state reading test. The other two predictors, dosage and lunch, did not correlate significantly with the state reading test scores. Four of the five predictors correlated significantly with the district reading test, baseline level, end baseline level, dosage and lunch status, but slope did not. Slope and dosage correlated significantly with one another, as did the two variables baseline level and end level. The other predictors correlated moderately with one another. The district and state reading tests also correlated significantly.

Predicting Reading Test Performance

The first part of this study was examining how the independent variables (baseline level, end level, slope, dosage, and lunch status) predicted student performance on two different reading exams. Multiple linear regressions were conducted, using an alpha level of p < .05 to determine significance.

To determine which variables to enter into a multiple linear regression, Pearson correlations between the independent variables and the reading tests were examined. The variables dosage and lunch did not significantly correlate with state test scores. Slope, baseline level and end level all were significantly correlated with scores on the state reading exam, and were entered into a linear multiple regression model.

The full regression model used slope, baseline level, and end level as predictors, and was significant, $R^2 = .23$, $F(3, 87) = 8.73$, $p < .05$. The model explained 23% of the

variance in student performance on the state reading test. Each individual coefficient in the model was further examined. Slope had a statistically significant effect on state reading test scores. The unstandardized regression coefficient (*b*) for slope was -7.86, *t* (87) = -2.37, *p* < .05, meaning that for slope values below the 50th percentile, student performance on the state reading exam decreases by -7.86 scale score points and state reading test scores increased by 7.86 scale score points for slope values at or above the 50th percentile, after controlling for baseline level and end level scores. The unstandardized regression coefficients for baseline level and end level were not significant. For baseline level the *b* was 0.16, *t* (87) = 1.88, *p* >.05, and for end level, it was *b* = 0.23, *t* (87) = 1.86, *p* >.05.

Cook's Distance (Cook's D) was applied post-hoc to the final model to determine the impact of deleting the two state test outlier observations from the regression analysis. Cook's D is a measure of the influence of deleted observations on the results of a regression analysis. The cut-off value of $D_i > 4/n$ (Bollen & Jackman, 1990) was used, and suggested that values > 0.043 indicated a strong influence on the estimated regression coefficients in the final model. The outlier test scores of 175 and 214 resulted in Cook's D values of *D* = 0.39, and *D* = 0.23 respectively, which demonstrated that removing the outlying data did not significantly affect the results.

To determine how well a model combining slope, baseline level, and end level predicted student passing on the state reading exam, a logistic regression was conducted. State test scores were converted into a dichotomous variable. Specifically, scores at the level of meeting state proficiency standards, 350 or higher, were coded 1, and scores at

349 or lower were coded 0. The overall logistic regression model was significant, χ^2 (3, $n = 91) = 21.36 \, p < .05$, meaning that at least 1 of the independent variables was significant at predicting the odds of meeting proficiency on the state exam.

Logit values from this logistic regression model were used to construct estimated probability plots. These figures illustrate the estimated probabilities of meeting proficiency on the state reading exam given slope values coded 1, referred to as strong slope values, those at or above the 50th percentile or 0, and weak, for slope values below the 50th percentile.
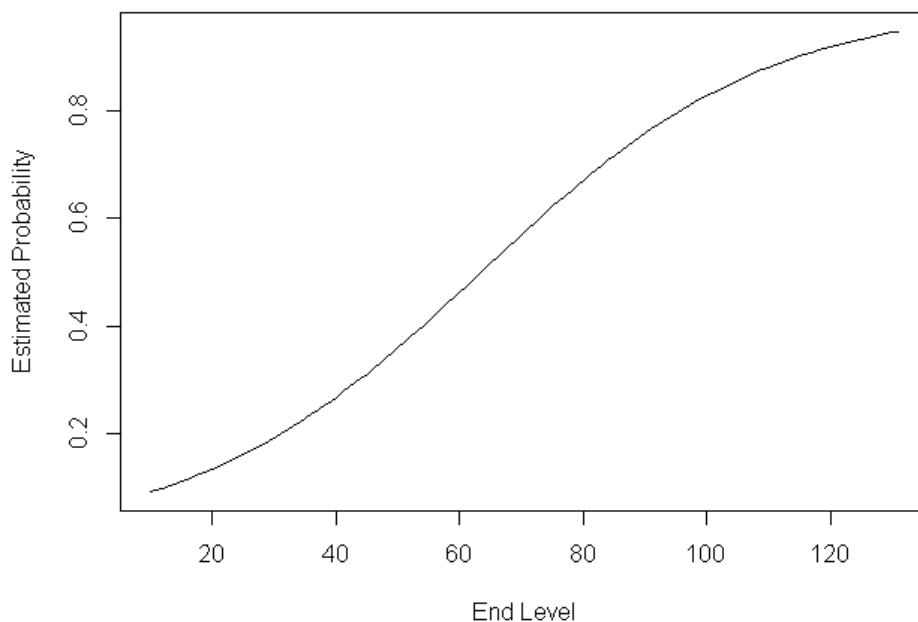

Figure 1: Estimated probabilities of meeting proficiency on the state reading test given the mean CBM-R baseline level, strong slope and all CBM-R end level scores
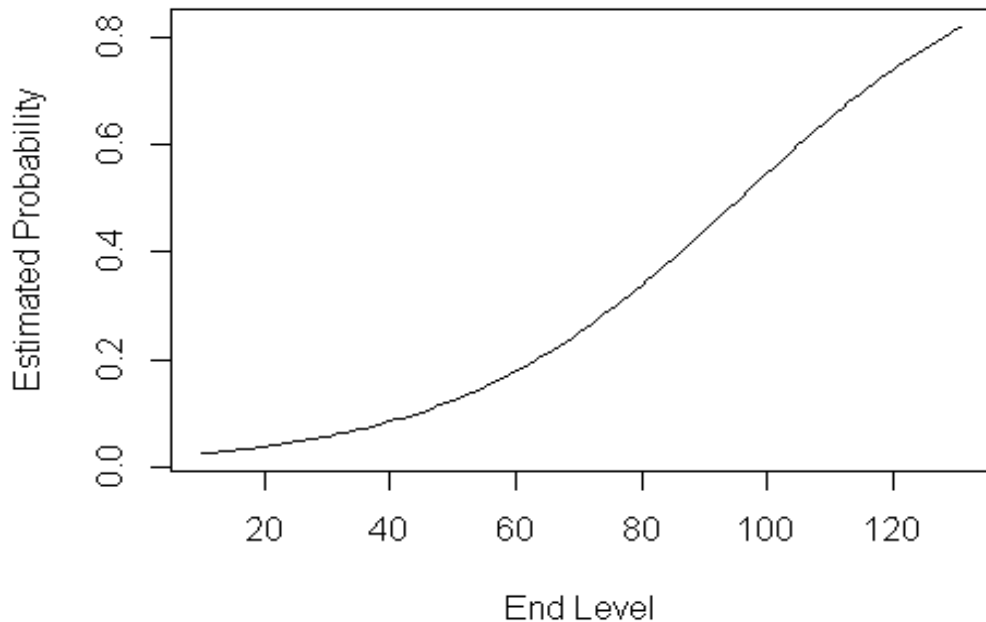
Figure 2: Estimated probabilities of meeting proficiency on the state reading test given the mean CBM-R baseline level, weak slope and CBM-R end level scores

Because baseline CBM level was not a significant coefficient in the logistic model, the mean value of 53.88 words per minute was used in both plots. Because end level was a significant predictor, all the CBM end level values (10 – 131 words read correctly per minute) were used in both plots.

The two sigmoid, or s-curve, plots in Figures 1 and 2 are visually different, Figure 1 has fairly equally short tails at both ends, whereas Figure 2 has a longer and more gradually inclined tail at the lower end of the X axis, and a short tail with virtually no curve at the high end of the X axis. Using Figures 1 and 2 as visual representations, of

the findings, given an end level score of 80 words per minute (the sample mean was

87.85), and a strong slope, the estimated probability of meeting proficiency on the state

reading exam is approximately 60%; however, with a weak slope, it is less likely,

approximately 30%.

Predicting student performance on the district reading exam involved a multiple

linear regression. The independent variable with the weakest correlation with the district

reading test, slope, was entered into a simple regression model with district MAP test

results as the dependent variable, and found not to be significant. The remaining four

independent variables, were simultaneously entered into a multiple regression model,

which was significant, $R^2 = .44$, $F(4, 61) = 11.951$, $p < .05$. For students that participated

in the reading intervention, 44% of the variance in their MAP scores could be explained

based on their scores at the beginning of the intervention (baseline level), the dosage of

intervention they received, lunch status, and their performance at the end of the

intervention (end level). Two of the four independent variables, end level and dosage,

had a statistically significant effect on MAP scores. The unstandardized regression

coefficient (*b*) for end level was significant, .25, $t(57) = 3.192$, $p < .05$, meaning that for

each 1 additional word read correct in 1 minute at the end of intervention participation,

students' MAP scores increase by .25 RIT points. In other words, for every 4 additional

words read correct in 1 minute at the end of intervention services, students' MAP scores

increase by 1 RIT point, controlling for baseline level, dosage, and lunch. The *b*

associated with the variable dosage, was also significant $b = -.23$, $-.32_{65}$ $t = -3.20$, $p < .05$.

These results suggest that for each minute less of dosage of intervention service time,

students' average MAP scores were reduced by .23 RIT points, or for every 5 minutes less of average weekly intervention time, students' MAP scores decreased by 1.15 RIT points, controlling for baseline level, and lunch. The b for baseline level was not significant $b = .11$, $t = 1.76$, $p > .05$, nor was lunch $b = -2.47$, $t = -1.04$, $p > .05$. In summation, the full multiple linear regression model with all four predictors was significant at predicting student performance on the district reading test, but only two unstandardized regression coefficients were significant, end level and dosage.

Predicting Risk-Status

The second part of this study was examining student risk-status 1 and 2 years post intervention. Risk status was defined as placement either in pre-referral programming or special education 1 and 2 years following the intervention. The specific research question was, how does performance during an intervention predict pre-referral or special education placement 1 and 2 years post intervention?

Four binary logistic regression analyses were conducted with the covariates: baseline level, end level, dosage, lunch status, and slope. Logistic regression models were used to determine how the models correctly classified student risk status 1 and 2 years post intervention, and to determine whether the independent variables as a whole significantly affected the dependent variables. There were two dichotomous categorical variables in the models, lunch status and slope. Lunch status was coded 1 if the student was eligible to receive free/reduced lunch, and 0 if the student was not eligible. Slope was coded 0 if the slope value was equal to a percentile rank value of 49th percentile or less, and 1 if the slope value was equal to the 50th percentile or higher. The dependent

variables were also coded 1 or 0. The dichotomous dependent variable, special education status, was coded 1 if the student was receiving special education according to school records, and coded 0 if they were not. The dependent variable pre-referral participation, was coded 1 when there was documentation of participation in pre-referral, or coded 0 for no pre-referral documentation. In logistic regression, when any cases are missing covariates, they are removed from the model. Therefore, the models analyzing 1 year post intervention outcomes included 86 cases, and the models examining outcomes 2 years post intervention included 82 cases, the results are listed in Tables 3 and 4.

The first and second logistic regression analyses examined the outcomes for students receiving or not receiving Special Education 1 and 2 years post intervention, results appear in Table 3. The chi-squared statistic was performed to determine whether special education status is more likely for some students, given a model with the independent variables of baseline level, end level, dosage, lunch, and slope. One year after intervention, special education status was equally possible for all students, χ^2 (5, $n=$ 86) = 8.34, $p >.05$. The overall model two years after intervention was also not significant, χ^2 (5, $n = 82$) = 10.22, $p >.05$. As a whole, the independent variables in the models did not contribute significantly to predicting student placement in special education 1 and 2 years after participating in an intervention. The models did correctly classify 83% and 78% of the students, respectively, labeled special education 1 and 2 year's post-intervention.

The third and fourth logistic regression analyses examined the outcomes for students that participated in a pre-referral program 1 and 2 years post intervention, results

appear in Table 4. The dichotomous dependent variable was pre-referral status, which was coded 1 if documentation at the school indicated the student received pre-referral interventions and 0 if there was not any documentation. Again slope and lunch status were dichotomous covariates. The other predictors included baseline level, end level and dosage.

One year after intervention, pre-referral status was equally possible for all students, χ^2 (5, $n = 86$) $= 7.12$, $p > .05$, the independent variables as a whole did not significantly predict pre-referral status. Two years after intervention, the model to determine the odds of students being in pre-referral programming was significant, χ^2 (5, $n = 82$) $= 12.79$, $p < .05$.

Table 3

Summary of Logistic Regression Analysis for Variables Predicting Special Education

Predictor	1 Year					2 Years				
	b	*SEb*	*B*	*P value*	*Classification Accuracy Rate Overall*	*b*	*SEb*	*B*	*P value*	*Classification Accuracy Rate Overall*
Special Education					82.6					78.0
Baseline Level	- 0.01	0.02	0.99	0.50		- 0.01	0.02	0.99	0.42	
End Baseline Level	0.02	0.02	1.02	0.44		0.01	0.02	1.01	0.73	
Slope	-1.45	0.72	0.24	0.04*		-1.74	0.69	0.18	0.01*	
Dosage	0.06	0.03	1.06	0.03*		0.04	0.02	1.04	0.11	
Lunch	0.33	0.71	1.39	0.64		- 0.14	0.66	0.87	0.84	
Constant	-5.38					-2.60				
χ^2		8.34		0.14			10.22		0.07	
df		5					5			

Note: Lunch predictor coded as 1 for *yes free/reduced* and 0 for *no free/reduced*. Slope coded 0 for values from 1st to the 49th percentile values, 1 for 50th percentile and higher   *p <.05

Table 4

Summary of Logistic Regression Analysis for Variables Predicting Pre-referral Participation

Predictor	1 Year					2 Years				
	b	*SEb*	*B*	*P value*	*Classification Accuracy Rate Overall*	*b*	*SEb*	*B*	*P value*	*Classification Accuracy Rate Overall*
Pre-referral					76.7					69.5
Baseline Level	- 0.01	0.02	0.99	0.51		-0.04	0.02	0.97	0.02*	
End Baseline Level	- 0.03	0.02	0.97	0.15		0.01	0.02	1.01	0.63	
Slope	0.41	0.56	1.51	0.46		- 0.29	0.53	0.75	0.58	
Dosage	- 0.02	0.02	0.98	0.33		- 0.03	0.02	0.97	0.08	
Lunch	0.13	0.60	1.13	0.84		- 0.54	0.57	0.58	0.34	
Constant	2.06					2.40				
χ^2		7.12		0.21			12.79		0.03*	
df		5					5			

Note: Lunch predictor coded as 1 for *yes free/reduced* and 0 for *no free/reduced*. Slope coded 0 for values from 1st to the 49th percentile values, 1 for 50th percentile and higher   *$p < .05$

Backwards elimination was performed to determine the significant coefficient(s) contributing to predicting student participation in pre-referral programming in 5th grade, and any multi-colinearity. This was performed by systematically removing the least significant coefficient in the full logistic regression model, retaining all the other variables, and applying logistic regression. End baseline level was the first variable removed from the regression, as it was the least significant coefficient in the full model. In the subsequent models, the following variables were removed in the following order: slope, lunch, then dosage. No co-linearity was found between variables, as CBM-R baseline level was the only significant coefficient in every model. Following this process resulted in CBM-R baseline level being the only remaining significant coefficient, χ^2 (1, $n= 82) = 7.87$, $p < .05$. The logit from this final model, was then used to determine the probability of participation in pref-referral programming in 5th grade using the following formula and given the range of baseline scores.

$$p = \frac{e^{\log it}}{1 + e^{\log it}}$$

Figure 3 shows that probability of participation in pre-referral programming in 5th grade is higher with lower CBM-R baseline level scores.
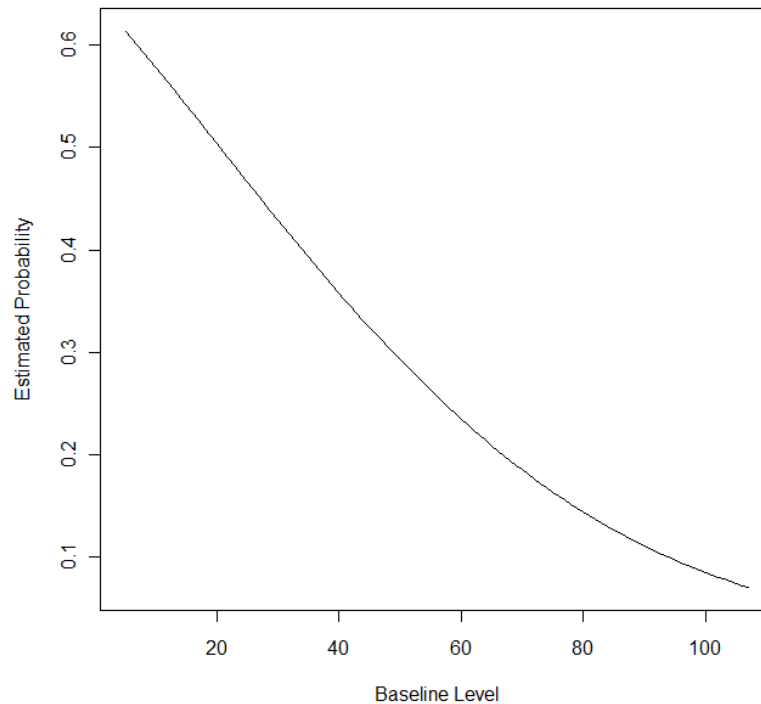
Figure 3: Estimated probability of participation in pre-referral programming in 5th grade given CBM-R Baseline Level scores

In summation, the models used to determine the odds of special education participation 1 and 2 years after 3rd grade, were not significant. The logistic regression model used to determine the odds of pre-referral involvement 1 year after intervention, was not significant, but the model analyzing pre-referral participation 2 years after intervention, was significant. The estimated probability of participation in pre-referral programming in 5th grade, was an inverse relationship, the lower the CBM-R baseline level score, the higher the probability of participation.

CHAPTER 5

DISCUSSION

The research questions for this dissertation included: (1) How does student performance during a tier 2 intervention, predict performance on group accountability measures of reading? (2) How well do level, slope, lunch status and dosage, combined, predict future risk status 1 and 2 years after intervention? Student performance during an intervention predicted student performance on state and district reading exams, and student placement in pre-referral programming 2 years post intervention. Linear regression modeling found that 44% of the variance in student performance on district reading assessments could be accounted for when using the predictors baseline level, end level, dosage and lunch. The variables end level and dosage, were significant predictors in the model. Student performance on the state reading exam was predicted using baseline level, end level and slope. It was determined that 23% of the variance in student test scores could be accounted for by the model, and baseline level was the significant coefficient. A logistic regression model using the variables baseline level, slope and end level to classify student performance as passing or failing the state reading exam, was significant.

Risk-status, defined as placement in special education or pre-referral programming post intervention, were also examined. Logistic regression analyses were used to determine the significance of the models and correct classification of student participation in pre-referral programming and special education 1 and 2 years post intervention. Student performance during this particular reading intervention did not predict student placement in special education, nor did student participation in pre-

referral programming 1 year after intervention participation. However, the logistic regression model using the outcome variable placement in pre-referral programming 2 years after intervention participation was significant in predicting student participation.

In the following discussion, the results are considered within the context of the research literature, and implications for practice are considered. Next, the limitations of the study are described, and suggestions for future research are proposed.

Predicting Student Performance

The regression model using baseline level, end level and slope to predict student performance on the state reading test, was significant, explaining 23% of the variance in student test scores. A logistic regression model was then applied, using the same variables baseline level, end level and slope to determine the odds of meeting proficiency on the state exam. This overall model was also significant, with slope and CBM-R end level being the significant predictors. The results support prior research demonstrating that oral reading fluency scores predict performance on state and district reading tests (Stage and Jacobsen, 2001; Good, Simmons & Kame'enui, 2001; Crawford, Tindal & Stieber 2001; McGlinchey & Hixson, 2004; Keller-Margios, Shaprio & Hintze, 2008). However, the current study extends this research in that it applied CBM-R progress monitoring data for students that participated in a tier 2 reading intervention.

Predicting student performance on the state reading test was further examined via estimated probability plots. Two estimated probability plots were developed, using the mean baseline level CBM-R score for the sample, all end level CBM-R scores, and categorical slope values of 1 or 0. The plots, Figures 1 and 2, provide a visual representation of estimated probabilities of meeting proficiency on the state exam given

the mean baseline level score of the sample, strong or weak slope and all end level values. Figure 1 shows that progress monitoring graphs with a strong slope can still meet proficiency on the state exam, even with a lower end level CBM-R score. Figure 2 shows that with a weak slope, one needs a higher end level CBM-R score to meet proficiency on the state exam. The figures illustrate the important role that these two variables: slope and CBM-R end level, played in predicting student performance on the state test.

The average end level score in the data set was 88 words per minute, below the winter and spring oral reading fluency target scores for 3rd grade, and only 12% of the CBM-R end level scores were at the spring criterion reference target score of 107. Over half the students met proficiency on the state exam, which could be another indication of the importance of slope, or progress via the intervention. These results are similar to Simmons, et al. (2008) and Shaw and Shaw (2002) who reported that students performed better than expected on a standardized achievement test and the state reading test based on spring oral reading fluency performance. These studies and the findings in the current study could suggest that lower target scores may be sufficient in predicting student performance on reading achievement tests.

Similar to other studies, the current study also found a relationship between progress monitoring performance during a reading intervention and performance on comprehensive measures of reading skills (Baker, et al., 2008; Chard, Vaughn & Tyler, 2002; Lennon & Slesinski, 1999; Torgesen et al., 2001) as measured by the Measures of Academic Progress (MAP; Northwest Evaluation Association [NWEA], 2003). Performance during the reading intervention did predict performance on the MAP. The multiple linear regression model with four predictor variables: baseline level, end level,

dosage, and lunch status was significant and explained 44% of the variance in student test scores on the MAP. The findings from the current study suggest that CBM-R progress monitoring data from participants in an intervention can predict reading achievement.

Young-Suk, Petscher, Schatschneider, and Foorman's (2010) longitudinal study found benchmark level and slope in 1st through 3rd grade to be strong predictors of future performance, including that of 3rd grade reading achievement. In the current study, level and slope were also significant predictors. The linear regression models predicting performance on the reading achievement tests, CBM-R end baseline level was significant in the model with MAP test scores as the dependent variable, and approached significance in the model with state reading test as the dependent variable. CBM-R baseline level approached significance in both models. Slope was a significant coefficient in the model predicting performance on the state reading test, and dosage was significant in the model with MAP as the outcome. This suggests that improvement with a tier-2 fluency intervention may support success on the state exam, and the amount of intervention time per week is important for performance on the district test.

Progress monitoring data captures student growth, holistically, via instruction and intervention. The findings in the current study are similar to other studies of fluency interventions demonstrating that student progress via reading fluency interventions, and instruction in the classroom are a strong combination, (Begeny, 2011; Begeny, et al., 2010; Begeny, Mitchell, Whitehouse, Samuels, & Stage, 2011; Mercer, Campbell, Miller, Mercer & Lane, 2000) contributing to student performance on reading exams. The progress the students made via fluency interventions, helped the students improve reading skills, as found in other studies, (O'Shea, Sindelar & O'Shea, 1987; Sindelar, Monda,

O'Shea, 1990), enough that more than half met proficiency on the state exam. The role

of instruction also likely contributed to student performance, and third grade teachers

may have provided explicit instruction in comprehension skills (Gersten et al., 1998;

Mastropieri, Scruggs, Bakken & Whedon, 1996) that worked well in conjunction with the

fluency practice provided by tutoring sessions, and the combination of both the tier 1

instruction and tier 2 intervention helped student reading skills overall.

Predicting Risk-Status

The second part of the study examined student risk-status 1 and 2 years after

participation in the reading intervention. The dual discrepancy model uses student data

via implementation of interventions to evaluate student progress, those making very weak

to no gains as indicated by level and slope, may be considered candidates for special

education (Fuchs & Fuchs, 1998; Fuchs, Fuchs & Speece, 2002). Speece, Case &

Molloy. (2003), found that a dual discrepancy model, more accurately identified students

who made poor gains, than the intellectual/achievement discrepancy model.

There are degrees to which students may make gains relative to level and slope, as

a result of intervention, and therefore conceptual degrees of risk when evaluating

progress via level and slope. For example, students that begin an intervention at-risk, and

make gains, may be considered no longer at-risk or always responsive, versus students

who participate in interventions but make weaker to no gains, have been categorized as

sometimes responsive, poor readers, or non-responsive (Al Otaiba & Fuchs, 2006;

Vellutino, Scanlon, & Zhang, 2007). The current study applied two categories of risk-

status 1 and 2 years post intervention, pre-referral participation and special education.

The hypothesis was that student performance during a reading intervention would predict

student placement in pre-referral interventions and special education in one or two years following participation.

One year post intervention, in 4th grade, the logistic regression models with baseline level, end level, slope, lunch status and dosage as predictors and either pre-referral or special education participation as the dependent variables were not significant. Two years post intervention, when students were in 5th grade, using the same predictors, the logistic regression model with pre-referral participation as the outcome was significant, and the model with special education participation as the outcome approached significance, $p = .069$. This may indicate a gradation difference in risk status between the categories of pre-referral and special education, and the impact of time. Two year's post intervention many students were failing to make sufficient progress, and hence were participating in pre-referral programming, but were not so far behind that they were participating in special education. Hence some of the students may have still been in tier 2 in 4th grade, (which may or may not have included intervention participation) but were further behind in 5th grade demonstrated by participation in pre-referral programming, but not so severely behind that they were predicted to be in special education. These findings suggest that these at-risk students were likely making poorer growth than general education students, a phenomenon that has been demonstrated via growth models comparing general and special education students (Christ, Silberglitt, Yeo & Cormier, 2010). Long-term studies have also indicated that some students struggled to maintain gains in years post intervention (Byrne & Fielding-Barnsley, 1993; Bus & Van Ijzendoorn, 1999; O'Connor, Notari-Syverson, Vadasy, 1996; 1998).

CBM-R baseline level was the only significant coefficient in the model with 5 variables, used to predict the odds of pre-referral participation in 5th grade. The estimated probability plot, figure 3, shows that baseline scores below 20 have a 50 – 60% ceiling probability of pre-referral participation in 5th grade, where as the probability is smaller for higher baseline level scores. Tran, Shanchez, Arellano and Swanson, (2011) reported that early baseline level CBM scores are strong predictors of risk status in 3rd grade. Silberglitt and Hintz's (2007) study also reported the importance of baseline level CBM-R scores for students in 2nd through 6th grades in predicting growth, reporting that students in the lowest decile group, based on baseline level score, demonstrated significantly weaker CBM-R growth than students with baseline level scores in the median level. Although it was not the purpose of the current study to explore all the reasons why students may have participated in pre-referral intervention in 5th grade, it's possible that classroom teachers were noticing the Silberglitt and Hintze (2007) findings, that some of the students that had low baseline level scores in 3rd grade, were demonstrating low fall benchmark scores and growth in 4th and 5th grades, resulting in pre-referral interventions in 5th grade.

Potential Implications for Practice

The primary implications of the current study's results for practice are that progress monitoring data during a reading intervention does predict student performance and risk-status for pre-referral participation. The findings support the Tran, et al., (2011) meta-analysis that CBM level and slope are strong predictors of performance and risk in the early grades. The results from this study indicate positive results for intervention initiatives, as student CBM-R performance related to receiving an intervention, did

predict their performance on other reading assessments, with many students meeting proficiency, even with CBM-R scores below oral reading fluency targets. CBM-R end baseline level was a significant coefficient for state test scores and approached significance for district scores. CBM-R baseline level approached, but was not a significant coefficient in either model. This suggests that how a student ends intervention services and improves skills as a result of intervention, may be more importantly related to their performance on reading exams in the short-term, than CBM-R scores at the beginning of an intervention. These findings suggest the importance of CBM-R data-based decision making throughout tier-2 intervention delivery.

Although CBM-R end level score was an important coefficient in the models predicting test performance at the end of 3rd grade, CBM-R baseline level was a significant coefficient in predicting pre-referral participation 2 years later in 5th grade. These findings likely indicate that CBM-R performance at the end of an intervention in 3rd grade may not be sustained into 4th and 5th grades. This is evidence that some students may need additional intervention in 4th and 5th grades to maintain progress made or to make additional progress. Practitioners supporting RTI implementation within schools should assist in the facilitation of tier-2 interventions in middle elementary school years. For example, reading fluency interventions after 3rd grade have been shown to be effective (Neddenriep, Fritz, & Carrier, 2011; Musti-Rao, Hawkins, & Barkely, 2009). From a prevention and early intervention perspective, practitioners should work to intensify kindergarten through 3rd grade instruction and intervention within an RTI framework, so fewer students are continuing to struggle in 3rd grade and beyond.

When considering implications for practice, several factors related to local decision-making should be considered. Supervising teachers at each of the schools determined which students participated, which fluency interventions students received, when students received their 60 minutes per week of tutoring sessions and when students stopped receiving the intervention. The recommended participants were students who scored below grade level oral reading fluency targets. There were two baseline level scores in the sample that did not meet the suggested eligibility criteria, one record with a score of 106 words read correctly per minute (wrcm) in November, which was above the upcoming Winter target of 91 wrcm, (and almost at the spring target score of 107) and another record with a baseline level score of 107 wrcm in March. This indicates that teachers selected students for participation based on factors other than students' CBM-R scores. School psychologists collaborating with instruction coaches and teams in making decisions about student participation in tier 2 interventions may provide literature about CBM assessments, criterion reference target scores, and general outcome measures (Fuchs & Fuchs, 2004; Shinn, 1989). School psychologists can collaborate and support other instructional coaches, (see Denton & Hasbrouck, 2009) to ensure the RTI model within the school results in growth for all students, including those with scores at grade level CBM-R targets.

Dosage was also a variable influenced by local decisions and circumstances, and a strong predictor of student performance on the MAP, but not the state exam. Dosage relates to time and consistency of student reading practice. Brenner and Heibert (2010) found that the average 90 minute reading instruction block includes only 15 minutes of actual reading time. If daily reading time is so small, the dosage of a tier-2 fluency

intervention is an important factor. Unfortunately, local decisions about scheduling

sessions, or interruptions, meant the average student did not receive the full

recommended dosage of 60 minutes per week. This could be easily remedied by school

psychologists helping to schedule daily, staggered intervention time for entire grade

levels, when teachers and tutors work with students, consisting of 30 to 60 minutes

(Miller, 2010). It is possible that if student dosage had been consistently at 60 minutes

per week or higher, then more of the variance in student performance on the state and

MAP exams would have been explained with the models, and the probability of

participation in pre-referral programming in 5th grade would have been lower. Dosage

could be made less of a concern, potentially, if principles of brief experimental analyses

were applied to interventions being delivered to students making weak progress. School

psychologists or other trained staff, could tailor interventions for individual students,

such as adding motivational incentives, or intensifying forms of error correction, thereby

more rapidly improving the reading fluency performance of students (McComas et al.,

2009).

Local decisions were made regarding when to exit students. The guideline

provided was to dismiss students when sores were above the aim-line. Although some

students were making good gains, the fact that the average student's end level score was

below winter and spring target scores may have meant that some students were exited too

early from services. Some of these students likely continued to score below grade level

targets in subsequent years, as the logistic regression model predicting participation in

pre-referral programming 2 years after 3rd grade, was significant. The exit criteria could

also be related to the outcomes, and a more rigorous standard such as scoring at spring

target score prior to dismissal, could have improved student outcomes, by resulting in more weeks of sessions.

Despite low oral reading fluency scores and less than the recommended amount of intervention time, the average student met proficiency on the state exam, scoring 1 point above the minimum score required by the state. These results indicate that the average student made just enough gains to pass the state exam in 3rd grade, even though they failed to meet grade level oral reading fluency targets. Because the average student did not achieve the third grade spring target for oral reading fluency, and just barely met proficiency on the state exam, many probably continued to score below grade level targets in 4th grade, and in 5th grade. Potentially, students may have had oral reading fluency scores so far below grade level targets that the schools began providing pre-referral intervention services in 5th grade.

Decision-making at the school, teacher, and individual student level are important factors influencing the format in which RTI is conducted. School psychologists can assist with implementation of comprehensive RTI assessment and intervention programming to ensure all students are making progress, and play a central role in the coordination of interventions that are provided to students in need.

Limitations

The primary limitations of this study relate to external validity. The contributing factors include: lack of information regarding quality core instruction in tier 1, the size and homogeneity of the sample, missing school records, progress monitoring, and the inconsistent administration of the recommended dosage of intervention.

Tier 1

Students were provided their tier 2 interventin service sessions in addition to their core reading instruction occuring within tier 1. Tier 1 instruction was not the focus of the current study, but is a factor that contributes to a student's overall reading progress (King & Torgesen, 2000; O'Connor, et al., 1996, 1998), and likely contributed to the progress of the students in the current study. Students were in schools that used a variety of curriculua for tier 1, i.e., Reading Mastery, Houghton Mifflin, Scott Forseman, and models for aligning assessment and instruction resources and techniques such as prior participation in Reading First and RTI. The average student met proficiency on the state reading exam, and yet the average student ended tutoring failing to achieve the winter target of 90 words per minute. These findings suggest a limitation to the study, the contribution of tier 1 instruction to student progress, performance, and risk-status.

Sample

Students were in schools in relatively small communities, with economic diversity, but limited racial and language diversity. The sample represented the typical population composition in small communities in the state. Performance of a larger and more racially, and linguistically diverse sample would have lent more support for the social validity of the findings relating to the performance of students in urban and suburban settings. The size of the sample in the current study was small, compared to the sample of 2,400 used by Baker, et al. (2008) in which a larger amount of variance (51%) of 3rd grade state test scores was explained using benchmark level and slope. Baker et al. (2008) had samples of entire grade levels, where as the current study had a sub-group of

at-risk students. However, if the study was replicated with a larger sample, more variance may be explained.

Students were selected by the supervising teachers, and told that students selected for participation should be those just below grade level CBM-R seasonal targets, that could use the fluency focused reading practice provided by the tutors. Therefore the sample was a selective at-risk group. If pre-referral participation is considered a less severe category of risk than special education this may explain why the logistic regression models did not predict student participation in special education 1 and 2 years post intervention.

The standard deviation for baseline level scores was 24 wrcm and the mean baseline level score was 54 wrcm. It is possible that the majority of the students that teachers selected were perceived to be moderately at-risk readers. This may partially explain why pre-referral participation could be predicted in 5th grade, but not in 4th grade, as this group of students may have been slowly falling behind their peers, and not becoming extremely discrepant until 5th grade.

An existing data set and school records were used from which data were missing for some students with regards to test scores and the risk status variables. State and district test scores came from school records 2 years after students had been in 3rd grade. Two of the schools did not administer the MAP to its students, and therefore scores were not available. At several of the schools, electronic records of student performance on the state reading test were not maintained. Principals provided scores to the researcher by checking for a test score sticker located in a student's file, in some cases, this was missing. The unavailability of MAP scores for some student records was another

limitation. Two of the school districts where students had received the reading

intervention did not administer the MAP test to any of their students at the time. For the

schools that did administer the MAP, some scores were missing. A larger sample size for

the MAP results would have improved the validity of the conclusions drawn from the

analyses.

Assessment Prior to 3rd Grade

One of the purposes for the study was to determine if progress monitoring

assessment variables related to intervention participation in 3rd grade could predict

special education placement 1 and 2 years later. Student placement in special education

was not predicted based on the independent variables used in this study. This finding

supports the research that additional or other assessment data, in conjunction with

progress monitoring data, may be necessary to predict which students are most at-risk for

severe difficulties reading. In a recent assessment study by Fuchs, et al. (2012) first

grade cognitive measures and Word Identification Fluency in December and May,

explained 50.3 and 62.1%, respectively, of the variance in 5th grade reading

comprehension test scores. This suggests that a particular battery of assessments in the

early years may be a powerful predictor of risk-status in future years. In the longitudinal

study of outcomes for students that participated in a pre-school phonemic awareness

intervention (Byrne, Fielding-Barnsley & Ashley, 2000), those later found eligible for

special education in 5th grade made the same progress as others in phoneme identification

in kindergarten and first grade, but made poorer progress than peers in letter

identification and all subsequent follow-up reading testing in kindergarten through 5th

grades. Recent research has found that teacher ratings, in addition to student reading

fluency skills can predict at-risk readers in 1st grade (Speece, et al., 2011) and 4th grade (Speece et al., 2010). Research has found that phonological awareness in young students predicts phonics development and is an area of skills that may be particularly challenging for young struggling readers (Torgesen & Wagner, 1998). AlOtaiba and Fuchs (2002) found that students with low phonological awareness, and weak rapid word naming and phonological memory in the early grades, were the students who failed to make progress during interventions, and subsequently became likely candidates for special education.

Records of Risk Status and Performance Indicators

The records for the risk status variables, participation in special education and pre-referral programming 1 and 2 years post intervention were a combination of electronic and paper records. All schools maintained an electronic data base of which students were receiving special education services. However, records of student participation in pre-referral programming varied, and principals provided information based on their reports and the existence of documentation of student participation. Five of the districts were starting to use a new electronic system, but due to software errors, were not able to access all electronic records. Paper records did exist across all schools, but varied widely between schools and districts, from as little as meeting agendas listing the student's name, to detailed intervention planning forms completed by teachers. Due to the differences in school policies and procedures regarding record keeping on pre-referral participation, a decision was made to have principals provide the data to the researcher regarding their judgment of student participation in pre-referral programming. A limitation was that pre-referral status was based on the principal's subjective opinion regarding what they considered to be existing documentation.

The reading tests completed by students were different in administration format, and not all schools administered the MAP. Six of the eight schools used the MAP, which may have related to the outcomes of the regression models used to predict test performance. The model used to predict student performance on the state test explained 23% of the variance in scores, where as the model used to predict scores on the MAP explained 44% of the variance. The state test was administered as a paper/pencil fixed length exam, and the MAP is a computer adaptive test. Because the tests differ in administration modalities such as the difficulty of items presented to test takers, the number of items completed by students (usually fewer MAP items, than state test items) and the ability/inability to change a prior response (National Center for Educational Statistics, 2009), it is possible the differences between the tests related to student performance and consequently the percentage of variance accounted for by the two models.

Progress Monitoring

Another limitation to the study was the number of progress monitoring data points for each student record and related issues regarding the slope variable. Not all student records included a large number of progress monitoring data points, due to the fact that some students participated in the intervention toward the end of the school year, or exited when they had 3 to 5 consecutive data points above their aim-line. The oral reading fluency variables of baseline level, slope and end level scores were available for 99 students that participated in the reading intervention during some portion of their 3rd grade year, and had at least 6 data points. However, six progress monitoring data points can result in a large standard error of slope, a range of 0.97 to 8.71(Christ, 2006). When

progress monitoring data points are entered into AIMSweb individual progress graphs, the software generates a slope value for each graph, calculated using ordinary least squares, which assumes linear progress. The predictor variable slope, in the original form of ordinary least squares values, had a lepturkotic distribution, violating the assumption of normality, and prompting the assignment of percentile ranks and dummy coding into a dichotomous variable. Recently, some researchers have examined fall, winter, spring CBM growth using various methodologies that may produce more reliable estimates for slope (Christ, et al., 2010). More research is needed to know if these methodologies could apply to progress monitoring slope data.

Seasonal Affect and Dosage

Students who participated in the reading intervention entered when there was availability in tutor schedules, as other students were successfully dismissed. This meant students started at different times throughout the school year, in the fall when growth is typically very rapid, as well as in the spring when it is much slower (Christ, et al., 2010). Furthermore, students were eligible because they were below target scores, which increased in rigor from fall, to winter, to spring. A student may have been above fall target in September, and therefore was not considered a candidate in the fall, but then scored below winter target and became a candidate for participation. The variables baseline level and end level occurred at different times during the school year. It's possible that the season of the school year may have impacted some of the slope values derived from progress monitoring scores. However, season of the year may not have been as important as the regularity with which students received sessions. Another limitation of this study was the fact that the average student received 51 minutes per

week of intervention service, less than the recommended 60 minutes per week of intervention. This made it challenging to know the true impact of the fluency intervention services, because not all students received the recommended amount of time. Dosage was a significant regression coefficient in the model used to predict student performance on the district assessment, but not for predicting student performance on the state reading test. If students had received the recommended 60 minutes per week, or daily sessions, it is possible dosage may have been a stronger predictor, impacting student performance on the state test, and contributing to the predictions of risk-status 1 and 2 years later with regards to special education or pre-referral participation.

<div align="center">**Directions for Future Research**</div>

Future research could include: (a) replication with a larger and more diverse sample, (b) further examination of factors related to CBM-R such as baseline level, slope and end level, (c) progress monitoring, (d) criteria used for discontinuing fluency intervention sessions, (e) dosage for tier 2, and, (f) examination of pre-referral teams.

Sample Size

It was acknowledged that the sample was limited in its language and racial diversity. This study could be replicated with a more language diverse sample. Research has demonstrated that English Language Learners (ELL) respond well to interventions in the early grades, achieving grade level criteria and higher (Lesaux & Siegel, 2003; Vaughn et al., 2006). Furthermore, oral reading fluency leads to valid progress monitoring decisions with ELL students because it has been found to predict the performance of ELL students on reading achievement tests (Baker & Good, 1995; Wiley & Deno, 2005). Future research should include data for ELL students to determine how

the progress of ELL students during this particular reading fluency intervention relates to performance on reading achievement and future risk-status.

CBM-R Baseline, Slope, and End Level

Future research could consider strengthening the reliability and validity of some of the predictor variables used in the models. The slope variable could be improved as a predictor in models analyzing student performance on the state and district reading tests. A larger sample of student records with at least 18 progress monitoring data points each could improve the validity of the slope value (Christ, 2006). Due to the quadratic growth of oral reading fluency over a typical school year, with the greatest amount of progress occurring from fall to winter (Ardoin & Christ, 2008), future studies could examine the impact that initial intervention baseline scores have with regards to slope depending on the point in time of the school year when intervention services begin. Future research could include analyze the progress monitoring data for students receiving reading intervention services and apply different growth analyses, such as those used by Christ, et al., (2010). Although Christ, et al., (2010) reported weaker seasonal effects for special education students, compared to general education students, the results of a third group of students, those receiving interventions such as the sample used in the current study, were not specifically addressed. It would be useful information to know if there is also steepest growth from fall to winter for students receiving an intervention, and weaker growth from winter to spring, or if students receiving intervention services continue to make strong perhaps linear growth from fall to spring. Such results would indicate that students 'catch-up' if provided this particular intervention model. Future research could compare progress monitoring growth by creating groups of records with similar baseline

level scores, with the distinction being various points throughout the school year, and comparing growth across the school year.

Progress Monitoring

Three predictor variables in the current study: baseline level, end level and slope were all based on scores collected via the administration of AIMSweb oral reading fluency probes. Recent research has focused on improving the quality of oral reading fluency probes used for formative assessment. The current study used the AIMSweb probes, which were systematically tested on 20 students each, to eliminate probes with substantial error (Shinn & Howe, 2002). Predicting future student reading performance and risk-status could be improved with probes that produce progress monitoring levels and slopes associated with less error. Future research could examine the outcomes for reading intervention students using different progress monitoring probes and their ability to predict student performance on the state exam (Ardoin & Christ, 2009).

Over half the students met proficiency on the state test, scored below winter and spring target scores at the end of services. These results indicate that some students may not need to reach grade level oral reading fluency target scores to perform well on reading comprehension measures, similar to Simmons' et al., (2008) findings. Another consideration is the role that classroom instruction has on student performance. Future research could include classroom observations to determine how teachers teach comprehension and test strategies, in a manner that supports the learning of at-risk students, such as the use of explicit instruction (Archer & Hughes, 2011).

Exit Criteria

The guideline for when to stop providing intervention services to students could be further explored. Students were exited when their progress monitoring scores were consistently above the aim-line, otherwise students remained in service. To date, there have been 5 methods used by researchers for determining student responsiveness/nonresponsiveness to interventions: median split, normalization, benchmark, dual discrepancy, and slope discrepancy (Fuchs, and Deshler, 2007). The dual discrepancy model, most closely resembles the guidelines used by the supervising teachers in the current study. Dual discrepancy uses CBM level and slope values, to determining student movement between tiers, particularly if students are failing to make progress and may require more intensive tier 2 or tier 3 intervention (Fuchs & Fuchs, 1998; Fuchs, Mock, Morgan & Young, 2003). For example, Fuchs and Fuchs (2007) suggest that a 3rd grade student reading less than 50 words per minute (level), and making less than .50 words per week growth (slope), would be identified for tier 2 intervention, and would remain in intervention for 10 to 20 weeks, with weekly data collection, at which point level and slope would be reviewed to determine if the student should receive tier 1 only, (due to adequate progress), be changed to a different tier 2 intervention, or receive tier 3 interventions. More research is needed to determine which model has the highest degree of efficacy, and can be applied by practitioners.

Dosage for Tier 2

In addition to the application of a grade level exit target score requirement, future research could include assigning students to varying degrees of intervention 'dosage'. In the current study, the average student received 51 minutes per week, yet students were to

be scheduled for 60 minutes per week. Dosage as little as 5 minutes daily, was found to improve reading fluency scores for students with disabilities (Mercer, et al., 2000). In another study of a fluency intervention and dosage, Begney (2011) delivered his Helping Early Literacy with Practice Strategies (HELPS) intervention to students for either 15 or 30 minutes per week, 1.5 to 3 sessions, and found that there was not a significant difference between these groups on a fluency and comprehension measure, but both groups exceeded the control group on the assessments.

From an RTI perspective, there doesn't appear to be consistent agreement as to the number of sessions and session durations (dosage) defining tier 2 interventions. McCook (2006) stated that tier 2 interventions should occur 2 to 3 times per week for 30 minutes a session, resulting in 60 to 90 minutes per week. Fuchs and Fuchs (2007) recommend that tier 2 intervention services occur 3 to 4 times per week for 30 to 60 minute sessions, resulting in a minimum of 90 minutes per week or a maximum of 240 minutes per week. McMaster and Wagner (2007), describe a case study of a student in tier 2 intervention that received 35 minute sessions 3 times per week, for a total of 105 minutes per week. Vaughn (n.d.) recommends that ELLs receive tier 2 intervention services for 30 minutes, 5 days per week, resulting in 150 minutes per week. These examples, indicate that the RTI three-tier model itself, is a framework, and that tier 2, as part of the model, has a framework in itself that may be developed at the school district or school level (Vaughn & Wanzek, 2007). The range of recommendations, 2 – 5 sessions per week, for a total of 60 to 240 minutes per week, is an area of study that could be further explored. In the meta-analysis by Tran, et al., (2011), hierarchical linear modeling showed that dosage (length and number of sessions) had no significant

moderating effects on posttest effect sizes, for reading interventions, and pre-intervention performance was the strongest variable. This meta-analysis included a variety of reading interventions. Directly researching the impact of dosage by randomly assigning students to the same intervention for different amounts of time, would help determine if dosage does relate to student growth, performance, and risk status.

Pre-Referral Teams

The current study found that when using the reading intervention progress variables from third grade the probability of student participation in pre-referral interventions 2-years later was significant. Further examination of the functioning of the pre-referral teams in these grades may indicate why students were predicted to be participants.   The use of RTI models and pre-referral consultation teams in schools usually results in fewer students being assessed for special education, compared to the classic discrepancy model (Burns & Symington, 2002). Although it was not a focus of the current study, all of the schools were implementing aspects of systemic RTI models, such as data-based decision making, interventions, and pre-referral consultation teams. Burns and Symington (2002) found that treatment integrity of pre-referral intervention teams likely correlated strongly with student outcomes. Future research could include evaluating the effectiveness of the pre-referral intervention teams in 4th and 5th grades.

Some of the students recommended for pre-referral interventions in 5th grade may have demonstrated low literacy skills and weak response to interventions as early as kindergarten or first grade (Al Otaiba & Fuchs, 2006; Vellutino, et al. 1996; Fuchs et al., 2012). Future research could include reviewing student records from kindergarten and first grade, including assessment and intervention information. Assessment data from

those grade levels could be included in regression models, used to predict pre-referral participation in future elementary years.

Supervising teachers were to select students for the reading services that were scoring below grade level targets. This study found that CBM-R progress monitoring data for students receiving a fluency intervention in 3rd grade, does predict student performance on state and district reading tests at the end of the year, and student risk for participation in pre-referral programming 2 years later. Local decisions were made to enroll some students in the intervention services based on information other than oral reading fluency baseline level score. Third grade students with baseline level CBM-R scores below 10 words per minute, and two students with baseline levels above the spring target score were serviced. Having supervising teachers self-report why they select students for a reading fluency intervention service, would be informative. There could be behavioral reasons related to a performance deficit, a skill deficit, or a combination of the two, (Duhon et al., 2004; Daly, Persampieri, McCurdy, Gortmaker, 2005; Eckert, Ardoin, Daly & Martens, 2002).   Burns and Coolong-Chaffin (2006) suggest that school psychologists be involved with administrators and teachers in the building-wide RTI and pre-referral intervention process from assessment to the interpretation of results, and selection of students and interventions. Future research could include evaluating the effectiveness of a collaborative model of selecting students for interventions, monitoring their progress, and making intervention decisions based on student performance.

Synthesis

This discussion reviewed the findings from the current study within the context of past and on-going research related to predicting student performance and risk status, and

considered the implications for practice within school settings as districts and buildings further develop RTI assessment systems and interventions. The study used predictor variables from progress monitoring data, lunch status, and dosage to predict performance on state and district reading test scores and participation in pre-referral or special education programming 1 and 2 years later. The multiple regression models were significant, with regards to the amount of variance explained in state and district reading scores, 23 and 44% respectively. The logistic regression models predicting participation in special education in 4th and 5th grades were not significant. The logistic regression model predicting documentation of pre-referral participation in 4th grade was not significant, but the model for 5th grades was significant. Proposed limitations included factors related to: the sample, assessment, school records, progress monitoring, the topic of growth as related to the season of the school year, and dosage. Directions for future research were proposed, including: a larger and more diverse sample with a high number of progress monitoring data points for each record, a revised exit criteria requiring specific scores rather than a positive slope, more investigation of dosage as related to student progress, and further examination of teacher and staff decision-making processes and pre-referral teams.

References

Agresti, A. (2007). *An introduction to categorical data analysis, second edition.*

Hoboken, NY: John Wiley & Sons, Inc.

AIMSweb. (2002). *Reading curriculum-based measurement (R-CBM) standard oral*

reading fluency assessment passages. Retrieved from http://www.aimsweb.com

Algozinne, R., & Ysseldyke, J. (1981). Special education services for normal children.

Better safe than sorry. *Exceptional Children, 48,* 238 – 242.

Algozinne, R., & Ysseldyke, J., Hill, C. (1982). Psychoeducational decision making as a

function of the amount of information reviewed. *Journal of Psychology in the*

Schools, 19, 328-334.

Al Otaiba, S., & Fuchs, D. (2002). Characteristics of children who are unresponsive to

early literacy intervention: A review of the literature. *Remedial and Special*

Education, 23(5), 300-316.

Al Otaiba, S., & Fuchs, D. (2006). Who are the young children for whom best practices

in reading are ineffective? An experimental and longitudinal study. *Journal of*

Learning Disabilities, 39 (5), 414-431.

Al Otaiba, S., & Torgesen, J. (2007). Effects from intensive standardized kindergarten

and first-grade interventions for the prevention of reading difficulties. In S.R.

Jimerson, M.K. Burns & VanDerHeyden, A.M. (Eds.), *Handbook of Response to*

Intervention: The science and practice of Assessment and Intervention (212-222),

New York, NY: Springer.

Archer, A.L., & Hughes, C.A. (2011). *Explicit instruction: Effective and efficient*

teaching. New York, NY: Guilford Press.

Ardoin, S.P., Witt, J.C., Suldo, S.M., Connel, J.E., Koenig, J.L., Resetar, J.L., et al. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probe when conducting universal screening. *School Psychology Review, 33,* 218 – 233.

Ardoin, S.P., & Christ, T.J. (2008). Evaluating curriculum based measurement slope estimate using data from tri-annual universal screenings. *School Psychology Review, 37*, 109-125.

Ardoin, S. P., & Christ, T. J. (2009). Curriculum-Based Measurement of Oral Reading: Standard Errors Associated with Progress Monitoring Outcomes from DIBELS, AIMSweb, and an Experimental Passage Set. *School Psychology Review, 38* (2), 226-283.

Baker, S.K., & Good, R.H. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: A validation study with second-grade students. *School Psychology Review, 24,* 561-578.

Baker, S.K., Smolkowski, K., Katz, R., Fien, H., Seeley, J.R., Kame'enui, E.J., & Thomas Beck, C. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37* (1), 18-37.

Ball, E.W., & Blachman, B.A. (1988). Phoneme segmentation training: Effect on reading readiness. *Annals of Dyslexia 38,* 208 – 225.

Ball, E.W., & Blachman, B.A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26,* 49-66.

Barger, J. (2003). Comparing the DIBELS Oral Reading Fluency indicatora nd the North

    Carolina end of grade reading assessment (technical report). North Carolina

    Teacher Academy, Ashville, NC (2003).

Batsche, G.M., Elliott, J., Graden, J., Grimes, J., Kovaleski, J.F., & Prasse, D. et al.

    (2006). *Response to Intervention: Policy Considerations and Implementation.*

    Alexandria, VA: National Association of State Directors of Special Education.

Begeny, J.C. (2011). Research Brief: Effects of the Helping Early Literacy with Practice

    Strategies (HELPS) reading fluency program when implemented at different

    frequencies. *School Psychology Review, 40* (1), 149 – 157.

Begeny, J.C., Krouse, H.E., Ross, S.G., & Mitchell, R.C. (2009). Increasing elementary-

    aged students' reading fluency with small-group interventions: A comparison of

    repeated reading, listen passage preview, and listening only strategies. *Journal of*

    Behavioral Education, 18 (3), 211-228.

Begeny, J.C., Laugle, K.M., Krouse, H.E., Lynn, A.E., Tayrose, M.P., & Stage, S.A.

    (2010). A control-group comparison of two reading fluency programs: The

    Helping Early Literacy with Practice Strategies (HELPS) program and the Great

    Leaps K – 2 reading program. *School Psychology Review, 39* (1), 137 – 155.

Begeny, J.C., Mitchell, R.C., Whitehouse, M.H., Samuels, F.H., & Stage, S.A. (2011).

    Effects of the HELPS reading fluency program when implemented by classroom

    teachers with low-performing second grade students. *Learning Disabilities*

    Research & Practice, 26 (3), 122-133.

Blachman, B. A., Fletcher, J.M., Schatschneider, C., Francis, D.J., Shaywitz, B.A.,

    Shaywitz, S.E. (2004). Effects of intensive reading remediation for second and

third graders and 1-year follow-up. *Journal of Educational Psychology, 96* (3), 444-461.

Bollen, K.A. & Jackman, R. (1990). Regression diagnostics: An expository treatment of outliers and influential cases. In: J. Fox & J. Scott Long (Eds.), *Modern Methods of Data Analysis* (pp. 257-91). Newbury Park: Sage.

Bollman, K., Silberglitt, B., & Gibbons, K. (2007). The St. Croix River Education District Model: Incorporating Systems-Level Organization and a Multi-Tiered Problem-solving Process for Intervention Delivery. In S.R. Jimerson, M.K. Burns & VanDerHeyden, A.M. (Eds.), *Handbook of Response to Intervention: The science and practice of Assessment and Intervention* (319 - 330), New York, NY: Springer.

Brenner, D., & Hiebert, E.H. (2010). If I follow the teachers' editions, isn't that enough? Analyzing reading volume in six core reading programs. *Elementary School Journal, 11* (3), 347 – 363.

Buck, G.H., Polloway, E.A., Smith-Thomas, A., & Cook, K.W. (2003). Pre-referral intervention processes: A survey of state practices. *Exceptional Children, 69*(3), 349 – 60.

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 Tech Rep. No. 1) Tallahassee, FL: Florida Center for Reading Research.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research

implemented models. *Journal of Psychoeducational Assessment, 23*, 381-394.

Burns, M., K., Christ, T.J., Boice, C.H., & Szadakierski, I. (2009) Special Education in an RTI Model: Addressing Unique Learning Needs. In: T. Glober & S. Vaughn (Eds.), *The Promise of Response to Intervention: Evaluating Current Science and Practice* (pp. 267 - 285) Guilford Press.

Burns, M.K., & Coolong-Chaffin (2006). Response to Intervention: The role of and effect on school psychology. *School Psychology Forum: Research in Practice, 1* (1), 3 - 15.

Burns, M.K. & Scholin, S. (in press). Relationship between pre-intervention data and post-intervention reading fluency and growth: A meta-analysis of assessment data for individual students. *Psychology in the Schools*.

Burns, M.K., & Senesac, B.V. (2005). Comparison of dual discrepancy criteria to assess response to intervention. *Journal of School Psychology, 43*, 393-406.

Burns, M.K., Senesac, B.J., & Silberglitt, B. (2008). Longitudinal effect of a volunteer tutoring program on reading skills of students identified as at-risk for reading failure: A two-year follow-up study. *Literacy Research and Instruction, 47,* 27-37.

Burns, M.K., & Symington, T. (2002). A meta-analysis of pre-referral intervention teams: Student and systemic outcomes. *Journal of School Psychology, 40,* 437 – 447.

Burns, M.K., & Wagner, D. (2008). Determining an effective intervention within a brief experimental analysis for reading: a meta-analytic review. *School Psychology Review,* 37 (1), 126 – 136.

Bus, A. & Van Ijzendoorn, M. (1999). Phonological awareness and early reading: A

meta-analysis of experimental studies. *Journal of Educational Psychology, 91,*

403-414.

Byrne, B. & Fielding-Barnsley, R. (1993). Evaluation of a program to teach phonemic

awareness to young children: a 1-year follow-up. *Journal of Educational*

Psychology, 85(1), 104-111.

Byrne, B. & Fielding-Barnsley, R. (1995). Evaluation of a program to teach phonemic

awareness in children: a 2- and 3-year follow-up and a new preschool trial.

Journal of Educational Psychology, *87*(3), 488-503.

Byrne, B., Fielding-Barnsley, R., Ashley, L. (2000). Effects of preschool phoneme

identity training after six years: outcome level distinguished from rate of

response. *Journal of Educational Psychology, 92* (4), 659 – 667.

Carter, J., & Sugai, G. (1989). Survey on prereferral practices: Responses from state

departments of education. *Exceptional Children, 55*, 293-302.

Cartledge, G., Yurick, A., Singh, A.H., Keyes, S.E., & Kourea, L. (2011). Follow-up

study of the effects of a supplemental early reading intervention on the

reading/disability risk of urban primary learners. *Exceptionality, 19,* 140-159.

Cavanaugh, C.L., Kim, A., Wanzek, J., & Vaughn, S. (2004). Reading Interventions for

At-Risk Students: Twenty Years of Research. *Learning Disabilities: A*

Contemporary Journal 2(1), 9 – 21.

Chard, D.J., Stoolmiller, M., Harn, B.A., Wanzek, J., Vaughn, S., Linan-Thompson, S., &

Kame'enui, E.J. (2008). Predicting reading success in a multilevel schoolwide

reading model: A retrospective analysis. *Journal of Learning Disabilities, 41,* 174 – 188.

Chard, D.J., Vaughn, S., & Tyler, B. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities*, 35, 386-406.

Christ, T.J. (2006). Short-term estimates of growth using curriculum based measurement of oral reading fluency: estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35*(1), 128-133.

Christ, T.J., Pike, A., & Monaghen, B. (2011). CBM-R Progress Monitoring Outcomes: Reliability and validity of slope estimates compared to 3, 4, and 5 point decision rules. Retrieved from

http://www.mspaonline.net/Conference2011/Christ%20Power%20Point.pdf

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.

Christ, T.J., Silberglitt, B., Yeo, S., & Cormier, D. (2010). Curriculum-Based Measurement of Oral Reading: An Evaluation of Growth Rates and Seasonal Effects Among Students Served in General and Special Education. *School Psychology Review, 39* (3), 447-462.

Coyne, M., Kame'enui, E.J., Simmonos, D.C., Harn, B.A. (2004). Beginning reading intervention: inoculation or insulin: First-grade reading performance of strong responders to kindergarten intervention. *Journal of Learning Disabilities, 37* (2), 90 - 104

Crawford, L., Tindal, G., Stieber, S. (2001). Using oral reading fluency rate to predict

    student performance on statewide achievement tests. *Educational Assessment, 7*

    (4), 303-323.

Cronin, J. (2007). *A study of the alignment of the NWEA RIT Scale with the Minnesota*

    Assessment System. Lake Oswego, OR: Northwest Evaluation Association.

    Retrieved from http://www.isd518.net/district/curr/MN%20Alignment%

    20to%20NWEA%20RIT.pdf

Cunningham, P.M. (1999). *Phonics they use: Words for reading and writing* (3rd ed.).

    Boston, MA: Addison, Wesley.

Daly, E.J. III, Persampieri, M., McCurdy, M., Gortmaker, V. (2005). Generating reading

    interventions through experimental analysis of academic skills: Demonstration

    and empirical evaluation. *School Psychology Review, 34* (3), 395 – 414.

Daly, E.J. III, Witt, J.C., Martens, B.K., & Dool, E.J. (1997). A model for conducting a

    functional analysis of academic performance problems. *School Psychology*

    Review, 26, 554 -574.

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

    Children, 52, 219-232.

Deno, S.L. (2003). Developments in curriculum-based measurement. *Journal of Special*

    Education, 37 (3), 184 - 192

Deno, S.L., Fuchs, L.S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement

    to establish growth standards for students with learning disabilities. *School Psychology*

    Review, 30, 507-524.

Denton, C.A., & Hasbrouck, J. (2009). A description of instructional coaching and its relationship to consultation. *Journal of Educational and Psychological Consultation, 19,* 150 – 175.

Duhon, G.J., Noell, G.H., Witt, J.C., Freeland, J.T., Dufrene, B.A., & Gilberson, D.N. (2004). Identifying academic skills and performance deficits: The experimental analysis of brief assessments of academic skills. *School Psychology Review, 33,* 429 – 443.

Eckert, T.L., Ardoin, S.P., Daly, E.J., III. & Martens, B. K. (2002). Improving oral reading fluency: An examination of the efficacy of combining skill-based and performance-based interventions. *Journal of Applied Behavior Analysis, 35,* 271-281.

Fuchs, D., Compton, D.L., Fuchs, L.S., Bryant, V.J., Hamlett, C.L., & Lambert, W. (2012). First-grade cognitive abilities as long-term predictors of reading comprehension and disability status. *Journal of Learning Disabilities, 45* (3), 217 – 231.

Fuchs, D. & Deschler, D. (2007). What we need to know about responsiveness to intervention (and shouldn't be afraid to ask). *Learning Disabilities Research and Practice, 22* (2), 129 – 136.

Fuchs, D., Fuchs, L.S., McMaster, K.L., Yen, Loulee, Svenson, E. (2004). Nonresponders: How to find them? How to help them? What do they mean for special education? *Teaching Exceptional Children, 36* (6), 72-77.

Fuchs, D.J., Fuchs, L.S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N.J., Braun, M., &

O'Connor, R.E. (2001). Is reading important in reading-readiness programs? A randomized field trial with teachers as program implementers. *Journal of Educational Psychology, 93*(2) 251-267.

Fuchs, D., Fuchs, L.S., Thompson, A., Al Otaiba, S., Yen, L., Yang, N.J., & O'Connor, R.E. (2002). Exploring the importance of reading programs for kindergartners with disabilities in mainstream classrooms. *Exceptional Children, 68,* 295 – 310.

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

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.

Fuchs, L.S., & Fuchs, D. (2004). Using CBM for progress monitoring. Retrieved from http://www.studentprogress.org.

Fuchs, L.S., & Fuchs, D. (2007*) Using curriculum-based measurement to determine Response to Intervention (RTI), Progress monitoring in the context of responsiveness-to-intervention, manual..* Retrieved from National Center on Student Progress Monitoring: http://www.studentprogress.org/weblibrary.asp#research.

Fuchs, L.S., Fuchs, D., Hamlett, C.L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review, 22,* 27-48.

Fuchs, L.S., Fuchs, D., & Speece, D. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disability Quarterly, 25,* 33-45.

Gersten, R., Williams, J. P., Fuchs, L., Baker, S., Koppenhaver, D., & Spadorcia, S. et al. (1998). *Improving reading comprehension for children with disabilities: A review of research.* Final report. Washington, D.C: Special Education Programs (ED/OSERS). (ERIC Document Reproduction Service No. ED451650).

Good, R.H., & Kaminski, R. A. (2002). Dynamic Indicators of Basic Early Literacy Skills (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement Available: http://dibels.uorego.edu

Good, R. H., & Shinn, M.R. (1990). Forecasting accuracy of slope estimates for reading curriculum-based measurement: Empirical evidence. *Behavioral Assessment, 12,* 179-193.

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

Graden, J.L., Casey, A., & Christenson, S.L. (1985). Implementing a prereferral intervention system: Part 1: The model. *Exceptional Children, 51*, 377-384.

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

Hintze, J.M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high stakes testing. *School Psychology Review, 34,* 372-386.

Howe, K.B. & Shinn, M.M. (2002). *Standard reading assessment passages (RAPs) for use in general outcome measurement: A manual for describing development and technical features*. Eden Prairie, MN: Edformation. Retrieved from http://www.aimsweb.com

Hughes, C., & Dexter, D. D. (2010.) *Field Studies of RTI programs.* Retrieved from RTI Action Network: http://www.rtinetwork.org/learn/research/field-studies-rti-programs

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first to fourth grades. *Journal of Educational Psychology, 80* (4), 437 – 447.

Katz, L.A., Stone, C. A., Carlisle, J.F., Corey, D.L., Zeng, J. (2008). Initial progress of children identified with disabilities in Michigan's Reading First schools. *Exceptional Children, 74* (2), 235-256.

Keller-Margulis, M.A., Shapiro, E.S., Hintze, J.M.(2008). Long-term Diagnostic Accuracy of Curriculum-Based Measures in Reading and Mathematics. *School Psychology Review, 37* (3), 374 – 390.

King, R., & Torgesen, J.K. (2000). *Improving the Effectiveness of Reading Instruction in One Elementary School: A description of the Process.* (Technical Report #3) *Retrieved from* the Florida Center for Reading Research http://www.fcrr.org/TechnicalReports/Hartsfieldnew.pdf

Koutsoftas, A.D., Harmon, M.T., Gray, S. (2009). The effect of a tier 2 intervention in phonemic awareness in a response to intervention model in low-income preschool classrooms. *Language, Speech, and Hearing Services in Schools, 40*, 115 – 130.

LaBerge, D., & Samuels, S.J. (1974). Toward a theory of automatic information

processing in   reading. *Cognitive Psychology, 6,* 293 – 323.

Lennon, J.E., & Slesinski, C. (1999). Early intervention in reading: Results of a

screening and intervention program for kindergarten students. *School*

Psychology Review, 28(3), 353-364.

Lesaux, N.K., & Siegel, L.S. (2003). The development of reading in children who speak

English as a second language. *Developmental Psychology, 39,* 1005-1019.

Logan, G.D. (1997). Automaticity and reading: Perspectives from the instance theory of

automatization. *Reading and Writing Quarterly: Over-coming Learning*

Difficulties, 13, 123-146.

Marston, D.B. (1982).The technical adequacy of direct, repeated measurement of

academic skills in low achieving elementary students. (Doctoral   dissertation,

University of Minnesota, 1982). *Dissertation Abstracts International*, 43, AAT

8301966.

Marston, D.B. (1989). A curriculum-based measurement approach to assessing academic

performance: What it is and why do it. In M.R. Shinn (Ed.), *Curriculum-based*

measurement: Assessing special children (pp.18-78). New York: Guilford.

Marston, D.B. (2003). Tiers of intervention in responsiveness to intervention: prevention

outcomes and learning disabilities identification patterns. *Journal of Learning*

Disabilities, 38 (6) 539-544.

Mastropieri, M.A., Scruggs, T. E., Bakken, J. P. & Whedon, C. (1996). Reading

Comprehension: A synthesis of research in learning disabilities. In T.E. Scruggs &

M.A. Mastropieri (Eds.), *Advances in learning and behavioral disabilities* (Vol. 10B, pp. 201-227). Greenwich, CT: JAI Press.

Mathes, P.G., & Babyak, A.E. (2001). The effects of peer-assisted literacy strategies for first-grade readers with and without additional mini-skills lessons. *Learning Disabilities Research & Practice, 16,* 1, 28 – 44.

Mathes, P.G., Howard, J.K., Allen, S.H., & Fuchs, D. (1998). Peer-assisted learning strategies for first grade readers: Responding to the needs of diverse learners. *Reading Research Quarterly, 33* (1), 62 – 94.

Mathes, P.G., Torgesen, J.K., & Allor, J.H. (2001). The effects of peer-assisted learning strategies for first-grade readers with and without additional computer-assisted instruction in phonological awareness. *American Educational Research Journal, 38* (2), 371-410.

McComas, J., Wagner, D., Chaffin, M., Holton, E., McDonnell, M., & Monn, E. (2009). Prescriptive analysis: further individualization of hypothesis testing in brief experimental analysis of reading fluency. *Journal of Behavioral Education, 18* (1), 56 – 70.

McCook, J.E. (2006). *The RTI Guide: Developing and implementing a model in your schools*. Horsham, PE: LRP Publications.

McMaster, K. L., & Wagner, D. (2007). Monitoring response to general education instruction. In S.R. Jimerson, M.K. Burns & VanDerHeyden, A.M. (Eds.), *Handbook of Response to Intervention: The science and practice of Assessment and Intervention* (223-233), New York, NY: Springer.

Mercer, C.D., Campbell, K.U., Miller, M.D., Mercer, K.D., & Lane, H.B. (2000). Effects

of a reading fluency intervention for middle schoolers with specific learning

disabilities.*Learning Disabilities: Research & Practice, 15* (4), 179 – 189.

Miller, D. (2010). Scheduling considerations for RTI at the elementary school level.

Retrieved from http://www.rtinetwork.org/rti-blog/entry/1/99

Minnesota Department of Education (n.d.), *Minnesota Comprehensive Assessments-*

Series II (MCA-II). Retrieved from

http://education.state.mn.us/mdeprod/groups/Assessment/documents/Spreadsheet/0

17874__.pdf

Musti-Rao, S., Hawkins, R.O., Barkley, E. (2009). Effects of repeated readings on the

oral reading fluency of urban fourth-grader students: Implications for practice.

Preventing School Failure, 54 (1), 12 – 23.

Nathan, R.G., & Stanovich, K.E. (1991). The causes and consequences of differences in

reading fluency. *Theory into Practice, 30* (3), 176 – 185.

National Center for Education Statistics (2007). *The condition of education, Note 1:*

Commonly used variables 2007. Retrieved from

http://nces.ed.gov/programs/coe/supnotes/2007-n01.asp

National Institute of Child Health and Human Development. (2000). Report of the

National Reading Panel. Teaching children to read: An evidence-based

assessment of the scientific research literature on reading and its implications for

reading instruction(NIH Publication No. 00-4769) Washington, DC: U.S.

Government Printing Office.

NCLB (2002). *No Child Left Behind Act.* P.L. No. 107-110, 115 Stat. 1425 (2002).

Retrieved from http://www.ed.gov/policy/elsec/leg/esea02/index.html

Neddenriep, C.E., Fritz, A.M., Carrier, M.E. (2011). Assessing for generalized

improvements in reading comprehension by intervening to improve reading

fluency. *Psychology in the Schools, 48* (1), 14 – 27.

Nelson, R.J., Benner, G.J., & Gonzalez, J. (2003). Learner characteristics that influence

the treatment effectiveness of early literacy interventions: A meta-analytic review.

Learning Disabilities Research and Practice, 18, 255 – 267.

Northwest Evaluation Association (NWEA) (2011). *The Rit Scale.* Retrieved from

http://www.nwea.org/support/article/532/rit-scale

Northwest Evaluation Association (NWEA) (n.d.). *Standard Error of Measurement.*

Retrieved from http://www.nwea.org/support/article/555/standard-error-measure

Northwest Evaluation Association (NWEA) (2003). *Technical Manual for the NWEA*

Measures of Academic Progress and Achievement Level Tests. Retrieved from

http://www.scribd.com/doc/41812304/NWEA-Technical-Manual-2003

Northwest Evaluation Association. (2005). *2005 Normative Data: Monitoring Growth in*

Student Achievement. Retrieved from http://www.nwea.org

O'Connor, R.E., Fulmer, D., Harty, K., & Bell (2005). Layers of reading intervention in

kindergarten through third grade: Changes in teaching and student outcomes.

Journal of Learning Disabilities, 38 (5), 440 – 455.

O'Connor, R.E., Notari-Syverson, A., & Vadasy, P.F. (1996). Ladders to literacy: The

effects of teacher-led phonological activities for kindergarten children with and

without disabilities. *Exceptional Children, 63* (1), 117-130.

O'Connor, R.E., Notari-Syverson, A., & Vadasy, P.F. (1998). First-grade effects of teacher-lead phonological activities in kindergarten for children with mild disabilities: A follow-up study. *Learning Disabilities Research & Practice, 13* (1), 43-52.

O'Shea,L.J., Sindelar, P.T., & O'Shea, D.J. (1987). The effects of repeated readings and attentional cues on the reading fluency and comprehension of learning disabled readers. *Learning Disabilities Research, 2,* 103-109.

Perfetti, C.A. (1985). *Reading Ability*. New York: Oxford University Press.

Reschly, D.J. (1988). Minority mild mental retardation overrepresentation and special education reform. *Exceptional Children*, *54*, 316-323.

Reschly, D., Kicklighter, R.& McKee, P. (1988a). Recent placement litigation, part I: Regular education grouping: comparison of Marshall (1984, 1985) and Hobson (1967, 1969) *School Psychology Review,* 17(1), 9-21.

Reschly, D., Kicklighter, R.& McKee, P. (1988b). Recent placement litigation, part II, minority EMR overrepresentation: comparison of Larry P. (1979, 1984, 1986) with Marshall (1984, 1985 and S1 (1986). *School Psychology Review, 17*(1), 22-38.

Reschly, D., Kicklighter, R.& McKee, P. (1988c). Recent placement litigation, part III, Analysis of differences in Larry P., Marshall and S-1 and implications for future practices. *School Psychology Review, 17*(1), 39-50.

Roehrig, A.D., Petscher, Y., Nettles, S.M., Hudson, R.F., Torgesen, J.K. (2008). Accuracy of DIBELS Oral Reading Fluency Measure for Predicting Third Grade

Reading Comprehension Outcomes. *Journal of School Psychology, 46* (3), 343-366.

Samuels, S. J. (1997). The method of repeated readings. *The Reading Teacher, 50* (5), 376-381.

Samuels, S.J. & Flor, R. (1997). The importance of automaticity for developing expertise in reading. *Reading and Writing Quarterly, 13* (2), 107-122.

Samejima, F. (1994). Estimation of reliability coefficients using the test information function and its modifications. *Applied Psychological Measurement, 18* (3), 229-244.

Scammacca, N., Vaughn, S., Roberts, G., Wanzek, J., Torgesen, J. (2007). *Extensive reading interventions in grades K – 3: From research to practice.* Portsmouth, NH: RMC Research Corporation, Center on Instruction.

Schneider, W., Ennemoser, M., & Roth, E. (1999). Kindergarten prevention of dyslexia: Does training in phonological awareness work for everybody. *Journal of Learning Disabilities, 32* (5), 429 – 436.

ServeMN (2005). *Minnesota Reading Corps Institute, September 2005.*

Shaw, R., & Shaw, D. (2002). DIBELS Oral Reading fluency-Based Indicators of the third-grade reading skills for Colorado State Assessment Program (CSAP) (Technical Report). Eugene, OR: University of Oregon. Retrieved from https://dibels.uoregon.edu/techreports/shaw_csap_technical_report.pdf

Shaywitz, S. (2003). *Overcoming Dyslexia: A new and complete science-based program for reading problems at any level.* New York, NY: Knopf.

Sheridan, S., Kratochwill, T., & Bergan, J. (1996). *Conjoint Behavioral Consultation: A Procedural Manual.* New York, NY: Plenum Publishing Corporation.

Shinn, M.R.(Ed.). (1989). *Curriculum-based measurement: Assessing special children.* New York, NY: Guilford.

Shinn, M. R. (2002). Best practices in using curriculum-based measurement in a problem-solving model. In A. Thomas, & J. Grimes (Eds.), *Best practices in school psychology, IV (*pp. 671-198). Bethesda, MD: National Association of School Psychologists.

Shinn, M.R., Good, R.H., & Stein, S. (1990). Summarizing trend in student achievement: A comparison of methods. *School Psychology Review, 18,* 356-370.

Shinn, M.R., & Shinn, M.M. (2002). *Administration and scoring of Reading-Curriculum Based Measurement for use in general outcome measurement* (Technical Manual). Eden Prairie, MN: Edformation.

Silberglitt, B., & Hintze, J. (2007). How much growth can we expect? A conditional analysis of R-CBM Growth rates by level of performance. *Exceptional Children, 74* (1), 71 – 84.

Simmons, D.C., Coyne, M.D., Kwok, O., McDonagh, S., Harn, B.A., Kame'enui, E.J. (2008). Indexing Response to Intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41* (2), 158 – 173.

Simmons, D.C., Fuchs, L.S., Fuchs, D., Mathes, P.G., Hodge, J.P. (1995). Effects of explicit teaching and peer tutoring on the reading achievement of learning disabled

and low-performing students in regular classrooms. *The Elementary School Journal, 95* (5), 387-408.

Simmons, D.C., Coyne, M.D., Kwok, O., McDonagh, S., Harn, B.A., Kame'enui, E.J. (2008). Indexing response to intervention: A longitudinal study of reading risk from kindergarten through third grade. *Journal of Learning Disabilities, 41* (2), 158 – 173.

Sindelar, P.T., Monda, L.E., & O'Shea, L.J. (1990). The effects of repeated readings on instructional and mastery level readers. *Journal of Educational Research, 83* (4), 220-226.

Smith, S.B., Simmons, D.C., & Kame'enui, E.J. (1998). Phonological awareness: Instructional and curricular basics and implications. In D.C. Simons & E.J. Kame'enui (Eds.) *What reading research tells us about children with diverse learning needs: Bases and basics.* Mahwah, NJ: Lawrence Erlbaum Associates.

Snowling, M.J., Hulme, C. (2007). *The science of reading: A handbook.* Malden, MA: Blackwell Publishing

Speece, D. L., Case, L. P., & Molloy, D. E. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research and Practice, 18,* 147-156.

Speece, D.L., Ritchey, K.D., Silverman, R., Schatschneider, C., Walker, C.Y., Andrusik, K.N. (2010). Identifying children in middle childhood who are at risk for reading problems. *School Psychology Review, 39* (2), 258 – 276.

Speece, D.L., Schastschneider, C., Silverman, R., Pericola Case, L., Cooper, D.H., & Jacobs, D.M. (2011). Identification of Reading Problems in First Grade Within a

Response-to-Intervention Framework. *The Elementary School Journal, 111*(4), 585 – 607.

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.(2000). *Progress in understanding reading: Scientific foundations and new frontiers.* New York: Guilford.

Sugai, G., & Horner, R.H. (2009). Responsiveness-to-intervention and school-wide positive behavior supports: Integration of multi-tiered system approaches. *Exceptionality, 17* (4), 223 – 237.

Sugai, G., Sprague, J.R., Horner, R.H., & Walker, H.M. (2000). Preventing School Violence: The Use of Office Discipline Referrals to Assess and Monitor School-Wide Discipline Interventions. *Journal of Emotional and Behavioral Disorders*, *8* (2), 94-101.

Therrien, W. (2004). Fluency and comprehension gains as a result of repeated reading. *Remedial and Special Education, 25* (4), 252-261.

Tindal, G., Marston, D. B., & Deno, S.L. (1983). *The reliability of direct and repeated measurement* (Research Reprot No. 109). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Torgesen, J.K. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice*, *15*(1), 55-64.

Torgesen, J.K., Alexander, A.W., Wagner, R.K., Rashotte, C.A., Voeller, K.S., & Conway, T. (2001). Intensive remedial instruction for children with severe reading disabilities: immediate and long-term outcomes from two instructional approaches. *Journal of Learning Disabilities, 34,* (1), 33 – 58.

Torgesen, J.K., & Davis, C. (1996). Individual difference variables that predict response to training in phonological awareness. *Journal of Experimental Child Psychology, 63*(1), 1-21.

Torgesen, J.K., & Wagner, R.K. (1998). Alternative diagnostic approaches for specific developmental reading disabilities. *Learning Disabilities Research and Practice, 13* (4), 220-232.

Torgesen, J.K., Wagner, R.K., & Rashotte, C.A., (1997). Prevention and remediation of severe reading disabilities: Keeping the end in mind. *Scientific Studies of Reading, 1* (3), 217 – 234.

Tran, L., Sanchez, T., Arellano, B., & Swanson, H.L. (2011). A meta-analysis of the RTI literature for children at risk for reading disabilities. *Journal of Learning Disabilities, 44* (3), 283 – 295.

U.S. Census Bureau (2000). *American Fact Finder*. Retrieved from http://factfinder.census.gov/servlet/SAFFPopulation

United States Department of Education (1996). *Archived Information: Policy Guidance for Title 1, Part A – Improving Basic Programs Operated by Local Educational Agencies*. Retrieved from http://ed.gov/legislation/ESEA/Title_I/memofree.html

Vadasy, P.F.& Sanders, E.A. (2008) Repeated reading intervention: outcomes and interactions with readers' skills and classroom instruction. *Journal of Educational Psychology*, *100*(2), 272-290.

Vadasy, P.F., Sanders, E.A., & Abbott, R.D. (2008). Effects of supplemental early reading intervention at 2-year follow up: Reading skill growth patters and predictors. *Scientific Studies of Reading, 12* (1), 51 – 89.

Vadasy, P.F., Sanders, E.A., & Peyton, J. (2006). Code-oriented instruction for kindergarten students at risk for reading difficulties: A randomized field trail with paraeducator implementers. *Journal of Educational Psychology, 98* (3), 508-528.

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. (n.d.). *Response to Intervention in reading for English language learners.* Retrieved from http://www.rtinetwork.org/learn/diversity/englishlanguagelearners

Vaughn, S., Mathes, P., Linan-Thompson, S., Cirino, P., Carlson, C., & Pollard-Duradola, S., et al. (2006). Effectiveness of an English intervention for first-grade English language learners at risk for reading problems. *The Elementary School Journal, 107,* 153-180.

Vaughn, S., & Linan-Thompson, S. (2003). What is special about special education for students with learning disabilities? *Journal of Special Education, 69,* 391 – 409.

Vaughn, S., & Linan-Thompson, S. (2004). *Research-based methods of reading instruction grades K-3.* Association for Supervision and Curriculum Development: Alexandria, VA.

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

Vaughn, S., Linan-Thompson, S., Woodruff, T., Murray, C., Wanzek, J., Scannaca, N., et al. (2008). Effects of professional development on improving at-risk students' performance in reading. In C.R. Greenwood, R.R., Kratochwill, & M. Clements (Eds.), *Schoolwide prevention models: Lessons learned in elementary schools* (pp. 115 – 142). New York: Guilford.

Vaughn, S., & Wanzek, J. (2007). *Reading and Response to Intervention (RTI): How students benefit from multi-tiered instruction and intervention*. Retrieved from http://www.ncldtalks.org/content/interview/detail/1204/

Vaughn, S., Wanzek, J., Linan-Thompson, S., & Murray, C.S. (2007). Monitoring response to supplemental services for students at risk for reading difficulties: High and low responders. In S.R. Jimerson, M.K. Burns & VanDerHeyden, A.M. (Eds.), *Handbook of Response to Intervention: The science and practice of Assessment and Intervention* (234 - 243), New York, NY: Springer.

Vaughn, S., Wanzek, J., Murray, C., Scammaca, N., Linan-Thompson, S., & Woodruff, T.L. (2009). High and low responders to early reading intervention: Response to intervention. *Exceptional Children, 75,* 165 – 183.

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

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

Vellutino, F.R., Scanlon, D.M., & Zhang, H. (2007). Identifying reading disability based on Response to Intervention: Evidence from early intervention research. In S.R. Jimerson, M.K. Burns & VanDerHeyden, A.M. (Eds.), *Handbook of Response to Intervention: The science and practice of Assessment and Intervention* (185 – 211). New York, NY: Springer.

Vellutino, F.R., Scanlon, D.M., & Zhang, H. (2008). Using response to kindergarten and first grade intervention to identify children at-risk for long-term reading difficulties. *Reading and Writing: An Interdisciplinary Journal, 21* (4), 437 – 480.

Wanzek, J. & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36* (4), 541 – 561.

Wanzek, J. & Vaughn, S. (2010). Tier 3 interventions for students with significant reading problems. *Theory Into Practice, 49,* 305 – 314.

Wanzek, J., & Vaughn, S. (2011). Is a three-tier reading intervention model associated with reduced placement in special education? *Remedial and Special Education, 32* (2), 167-175.

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

Wilson, J. (2005). *The relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and oral reading fluency to performance on Arizona Instrument to Measure Standards (AIMS)* (Technical Report). Tempe, AZ: Assessment and Evaluation Department, Tempe School District No. 3.

Witt, J.C., VanDerHeyden, A.M. (2007). The system to enhance educational performance (STEEP): Using science to improve achievement. In S.R. Jimerson, M.K. Burns & VanDerHeyden, A.M. (Eds.), *Handbook of Response to Intervention: The science and practice of Assessment and Intervention* (223-233), New York, NY: Springer.

Yeo, S., Fearrington, J., & Christ, T.J. (2011). An investigation of Gender, Income, and Special Education Status Bias on Curriculum-Based Measurement Slope in Reading. *School Psychology Quarterly, (26)* 2, 119 – 130.

Young-Suk, K., Petscher, Y., Schatschneider, C., Foorman, B. (2010). Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *Journal of Educational Psychology, 102* (3), 652 – 667.

Ysseldyke, J.E., Algozzine, B., Shinn, M. R. & McGue, M. (1982). Similarities and differences between low achievers and students labeled learning disabled. *The Journal of Special Education, 16,* 73 – 85.

Ysseldyke, J.E., Thurlow, M., Graden, J., Wesson, C., Algozzine, B., & Deno, S.L. (1983). Generalizations from five years of research on assessment and decision-

making: The University of Minnesota Institute. *Exceptional Education Quarterly, 4,* 75 – 93.