8-1-2011

Using Growth Rate of Reading Fluency to Predict Performance on Statewide Achievement Tests

Rachelle Whittaker Hinkle
Indiana State University

USING GROWTH RATE OF READING FLUENCY TO PREDICT PERFORMANCE ON STATEWIDE ACHIEVEMENT TESTS

A dissertation

Presented to

The College of Graduate and Professional Studies

Department of Communication Disorders, Counseling, School, and Educational Psychology

Indiana State University

Terre Haute, Indiana

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy in Guidance and Psychological Service

---

by

Rachelle Whittaker Hinkle

August 2011

Keywords: reading fluency, curriculum-based measurement, statewide achievement

COMMITTEE MEMBERS

Committee Co-Chair: Linda Sperry, Ph.D.

Professor of Communication Disorders and Counseling, School, and Educational Psychology

Indiana State University

Committee Co-Chair: Damon Krug, Ph.D.

Assistant Professor of Communication Disorders and Counseling, School, and Educational Psychology

Indiana State University

Committee Member: Eric Hampton, Ph.D.

Associate Professor of Communication Disorders, and Counseling, School, and Educational Psychology

Indiana State University

Committee Member: Ronald J. Carlini, Ph.D.

Adjunct Assistant Professor of School Psychology

The University of Tennessee, Knoxville

ABSTRACT

Federal legislation has prescribed the increased use of statewide achievement tests as the culmination of a student's knowledge and ability at the end of a grade level; however, schools need to be able to predict those who are at-risk of performing poorly on these high-stakes tests. Three studies served to identify a means of predicting statewide achievement test scores in either third or eighth grade based on CBM reading scores and rates of improvement at first, second, and third grades or third, fourth, and fifth grades using readily available statistical procedures. One-half of the third-grade data was used in Study 1, while the prediction equation generated in Study 1 was validated on the second half in Study 2. The results of Study 1 indicated that, of the sample of over 1,200 third-grade students who took the third-grade statewide achievement test, the second- and third-grade spring CBM reading scores explained the highest amount of variability in third-grade reading scores; however, reading rate of improvement was also significant. The prediction equation from Study 1 was cross-validated in Study 2 on over 1,200 third-grade students, which indicated that there was more than 95 percent concordance that those who were predicted to pass the third-grade statewide test did pass. However, when using the second-grade spring cut score of 90 words read correctly per minute, the accuracy of prediction was diminished. In Study 3, using nearly 250 eighth-grade students' scores, reading fluency scores in third, fourth and fifth grades explained approximately 30 percent of statewide achievement test scores; however, rate of improvement was not significant in any of the grades.

ACKNOWLEDGMENTS

Since I began my graduate education at Indiana State University, I have been influenced by so many different people and circumstances. I want to thank the people who, over the past ten years, have changed and improved me.

I want to thank my husband, Allen, for his love, friendship, and sacrifice in order for me to pursue this degree. Through your encouragement, patience, and strength, I have been able to endure late nights and long hours of work that have taken place in order to achieve my career and educational goals. I have truly been blessed to have your steadfast presence in my life over the past fifteen years, and you will always remain my best friend and life partner. With you I am more than I ever could have been, and I am so glad that I get to share the ups and downs of life. And, because of you, I have hope in a fulfilling future.

My parents, George and Kathryn Whittaker, are truly the reason why I had the confidence to set my goals high and to have the work ethic to complete them. Your encouragement to pursue any dream is one of the greatest gifts that you could have ever bestowed upon me. When I was discouraged, you listened and provided insight, and when I achieved goals, you celebrated my accomplishments. Because of you I am not afraid to try.

I also want to send my heart to my dearest friend, Sherri Eaton-Bin Daar. The memories that I have of Indiana State University and of this degree I share with you. I may have achieved success, but it was only with your friendship and guidance that I could do it. We did it all together, and you will always be a sister to me. Because of you I got more than a just a piece of paper; I got a smart, funny, and compassionate friend with whom I can commiserate and laugh.

I also want to acknowledge those who have joined me in this journey in the past six years, my friends and colleagues in Tennessee. With your help I learned how to do what I was trained to do better and more confidently. I also learned from you how to think and apply my knowledge to be a positive force in the world. It is also because of you that I changed my dissertation topic to be something that can be of benefit to the kids, to the community, and to other educators. Because of you I am a better coworker, friend, and school psychologist.

Finally, I want to express sincere gratitude to my committee for the unique skills that you have brought to the educational experience of completing a dissertation. Without your time, patience, guidance, experience, and encouragement, I would not have been able to complete the dissertation, nor would I have been inspired to consider further research in my post-graduate activities. I want to give special thanks to Drs. Krug and Sperry. Although I did not spend any time with you in the classroom, I know that you did more than any long-lost student could ever have hoped for in an adviser. You never seemed to give up on me and went beyond any expectations that Indiana State University could have of you. I am truly lucky for what you have sacrificed to help me push past my fear of research. Because of you, I am a better researcher and a well-rounded professional.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Within the last decade, the American educational system has witnessed dramatic changes in federal legislation that affect general and special education funding and accountability procedures. State and local educational agencies have been forced to make difficult decisions in order to meet the demands of the federal government and their renewed interest in education. Many educators and educational systems are slow to change, while others have embraced the opportunity to improve student outcomes and rise to the challenge of meeting the demands of federal legislation.

The changes that must occur in order to meet federal mandates include improved data-driven decision making and improved student outcomes on high-stakes testing. The rationale behind these mandates appears to be multilayered. School systems must improve student outcomes by providing better services through better-trained staff, early intervention, and research-based interventions. Interventions alone are not sufficient; they must be measured using validated and reliable assessments in order to make real-time, in-the-moment changes that would lead to a swift change in interventions and thus, improve success (Linn, 2005; Mellard, 2004).

The role of school psychologists in assisting schools who wish to meet the requirements of new federal laws varies; however, school psychologists are specifically trained to identify research-based interventions, collaborate with other school personnel to implement interventions, collect data on intervention effectiveness, and analyze the data using proven research methods. School professionals rely heavily upon the expertise of school psychologists in order to administer many aspects of federal law that dictate the identification of students with academic difficulties. While some students require very little above the standard curriculum in order to be successful, others require much more intensive support. School psychologists can assist school professionals in making decisions about improving student outcomes.

Federal Legislation Drives Educational Decision-Making

Objectivity in educational accountability is an important goal in the federal legislation. In 2001, President George W. Bush signed into law the No Child Left Behind Act (NCLB), which required states to make specific actions to "close racial achievement gaps" and bring "all students to proficiency within the next eight years" (Lee, 2006, p. 5). States are required to do this by conducting annual testing in third through eighth grades in both English/language arts and mathematics. States are also required to use these data to establish goals for adequate yearly progress (AYP) between the years 2002 and 2014, where the end goal is to have all students at or above proficiency by 2014.

School systems in many states measure AYP through standardized testing in core subject areas. Additionally, many school systems have begun to use students' performance on these high-stakes tests as a measure of teacher effectiveness, adding additional pressures to individual teachers and groups of students to have adequate scores (Linn, 2005). Teachers and school administrators need a means of improving scores on standardized tests, and they look to many different models in order to bring students to proficiency in the math and reading curriculum.

One such model, response-to-intervention (RtI), has begun to identify students at risk for non-proficiency as early as possible (D. Fuchs & Fuchs, 2006). The RtI methodology of identification was made popular with the reauthorization of the Individuals with Disabilities Education Improvement Act (IDEIA, 2004), in which RtI was permitted as a means of identifying students with learning disabilities as opposed to the more traditionally known discrepancy model. The discrepancy model was formalized by federal regulations published in 1977 (U.S. Office of Education) and was used in nearly every state. It involved comparing a student's IQ standard score against the student's achievement. If there was a significant discrepancy between intelligence and achievement (to be specifically determined at the state level), then the student would be qualified to receive special education services. With the passage of IDEIA, educators saw RtI as a way not only to identify students who may perform poorly on standardized, high-stakes testing but also to provide options for improving the scores of such students who may not have been previously identified for special education with the discrepancy model (D. Fuchs & Fuchs, 2006). Many school psychologists are also proponents of RtI because it is a means for reducing the need for special education referrals, thus freeing up their time to perform other underutilized skills within the school setting, including consultation, and direct intervention (Hale, Kaufman, Naglieri, & Kavale, 2006).

Reading Skill Determines Future Academic Success

Because reading is a foundational skill upon which many other skills and areas of academic instruction are based, schools are extremely concerned with increasing the reading abilities of their students as a means to meet their AYP goals. Especially since the implementation of IDEIA (2004) and RtI, many students are identified with reading difficulties before they even take their first statewide assessment test in third grade. The utility of identifying at-risk students early and providing the necessary intervention or remediation in reading skills deficits is obvious to many educators.

The purposes of RtI models are to provide interventions to children with academic difficulties and to identify those who do not respond to general education instruction. This is especially important when applied to reading instruction. Reading has been proposed as a primary skill on which success in other academic skills areas is based. Reading problems comprise over three-fourths of all special education referrals (Nelson & Machek, 2007). Because reading difficulties are so prevalent, the prevention of reading difficulties is of great interest to school psychologists as well as educators and parents. Therefore, RtI may also be a point of interest for educational stakeholders.

The most common form of learning disability is reading disability; as many as 80% of children with learning disabilities have reading difficulties (Aaron, 1997). The most significant deficit in students with a reading disability is word recognition skills (Aaron & Joshi, 1992). Word recognition skills are associated with improved comprehension (National Reading Panel, 2000). It is among the most crucial of skills in reading fluency. Word recognition can be broken into two parts: the ability to automatically decode words and the ability to recognize words accurately (National Reading Panel, 2000).

Of the ways that an individual can demonstrate reading difficulties, slow reading speed is the one that is the most frequently assessed (National Reading Panel, 2000). Children who have difficulty decoding words also struggle with reading speed because it takes more time to identify, or sound out, words, regardless of how well they can comprehend the material they read (Aaron & Joshi, 1992). One of the most widely-accepted RtI models utilizes two criteria for making decisions: fluency (calculated in words read correctly per minute below a goal score) and rate of growth (calculated in terms of increase in words read correctly per minute divided by the amount of time between assessments) in comparison to the child's grade level peers (L. S. Fuchs & Fuchs, 1997).

Reading speed and other measures of reading fluency impact reading comprehension skills (Good, Simmons, & Kame'enui, 2001; Marston, 1989). As many as 50% of students with learning disabilities do not pass statewide achievement tests, which is one of the most critical reading comprehension assessments impacting AYP (Schulte, Villwock, Whichard, & Stallings, 2001). Over a period of several years, this lack of proficiency in grade-level curriculum standards could produce significant decreases in meeting graduation requirements. Fewer students meeting graduation requirements could lead to lower graduation rates from high school, another key component of measuring AYP in a school district. Additionally, students with learning disabilities are twice as likely to drop out of high school as their peers (U. S. Department of Education, 2002).

Need for Current Study

The purpose of this study is to examine possible techniques regarding how local educational agencies can improve student outcomes by using readily available, efficient, cost-effective data. Additionally, this research seeks to help educational agencies with early identification of students who may need support in order to perform well on standardized assessments, which is a priority for schools concerned with meeting the demands of AYP. Because reading difficulty has been identified as one of the most significant barriers to success on standardized tests, there is a strong need to use cost- and time-effective assessments of reading to identify those who may not perform well on standardized tests in sufficient time to intervene and remove that potential for failure. While school psychologists have unique training to help local educational agencies to identify students at risk and prevent further reading difficulty, many have access only to basic research analysis methods. This problem addresses a secondary aim for this research study, which is to use statistical methods that can be employed by school psychologists, such as multiple regression, to determine the correlation of reading variables with statewide achievement tests.

There are several questions that are raised in the current climate of education in regard to reaching the needs of students and still remaining in compliance with legislation. For example, is it possible to accurately identify students who are at risk for failure on standardized tests, and is it also possible to achieve this identification early enough to intervene and influence their performance? By examining reading fluency scores and growth rates longitudinally using less-complex statistical procedures to predict performance in primary grades as well as in intermediate grades, this study will assist school personnel to make use of cost-effective data to make high-impact decisions about student performance.

CHAPTER 2

REVIEW OF LITERATURE

School psychologists provide direct and indirect support to students with a variety of needs, although they are most known for their role in helping schools to identify students with educational disabilities. About 75% of a school psychologist's time is spent in the activities necessary to determine if a student is eligible for special education services (Reschly & Wilson, 1995), more than half of which is often activities needed to identify learning disabilities. Changes in federal legislation are primarily responsible for the changing role of school psychology in public education.

Federal Legislation Drives Educational Decisions

Public education in the United States has been a state- and local-government funded institution since the mid-1800s. Federal and state funding has been tied to several key court decisions and laws modeled after the goal of visionary leaders, including Thomas Jefferson and Horace Mann, to educate members of society sufficiently enough to govern themselves (Kober, 2007; Thattai, 2001). Prior to the early 1900s, education was primarily provided to those who could afford to pay for it. The needs of society and economic hardships had a tendency to lead to changes in public education. The most current legislation driving education reform now is No Child Left Behind (NCLB, 2001), which has sprung from the need for Americans to compete within a global society. Whereas previous legislation, namely the Rehabilitation Act of 1973, Section 504, the Individuals with Disabilities Education Act of 1975, the Americans with Disabilities Act of 1990, and Title I, has attempted to address meeting the educational needs of those with disabilities, NCLB has attempted to address public education's failure to educate all children to a level of proficiency adequate to be competitive with other leading countries of the world (Lee, 2006).

No Child Left Behind

NCLB (2001) has become the driving force behind accountability in the education system of the United States. With the goal for all students to become proficient in curriculum standards at every grade level in math and reading, the face of education has changed dramatically. School systems, and ultimately teachers as implementers of the curriculum, are under pressure to become more effective in students' retention of skills and knowledge so that the annual student assessments reflect favorably upon them. Statewide testing and the goals for AYP have created a data-driven decision-making model for education, with reading and math as the crux of all educational skills. The amount of time educators spend in these subject areas on average has increased by 43% since 2002 and now far exceeds that of all other subject areas, which the majority of districts have reduced by at least 150 minutes per week (Center on Education Policy, 2008).

Proficiency levels in reading and math are set by each state; however, in fourth and eighth grades the National Assessment of Education Progress (NAEP) must be given in order to evaluate national progress in meeting the proficiency goals set forth by NCLB. This assessment also evaluates the equity of the NAEP and the individual state's academic assessments and proficiency levels, as there is a relatively strong relationship between the majority of state assessments and the NAEP (Bandeira de Mello, Blankenship, & McLaughlin, 2009). At the same time, because proficiency levels are set by each state, there are differences among the difficulty levels of each state's performance expectations in order for a student to be considered "proficient." Individual schools and districts are being held accountable for their state goals under the threat of being taken over by federal or state agencies to ensure that those schools would meet AYP standards.

According to the federal NCLB (2001), all students in all states must be assessed annually in reading and mathematics in third through eighth grades. Each test must also be linked to that state's curriculum standards and should be used to determine if a student is proficiently meeting the state's curriculum standards for the child's respective grade level. Additionally, the students in each district and school must make AYP, meaning that the school must either maintain or increase by a certain degree the number of students who are considered proficient on the statewide achievement test each year. This is measured primarily by each state's annual achievement testing as well as by attendance and graduation rates; additionally, there are often documented performance gaps between the general population average and the average of children with racial and socioeconomic disadvantages (Lee, 2006; Linn, 2005). Schools who fail to make AYP may be allowed provisions in their annual improvements through what NCLB calls "safe harbor," in which schools can still meet AYP if there are 10% fewer students who fall below proficiency than the year before and there is improvement among the disadvantaged subgroups, which is a considerably difficult goal to achieve as well (Linn, 2005).

High-stakes tests, as they have come to be known, are designed to mirror the state curriculum in each of the major subject areas of that state. The focus of NCLB on English/language arts and mathematics has caused there to be more emphasis and instructional time in those subject areas, and whether intended or not, more emphasis in presentation of the information that will be on a test (Center on Educational Policy, 2008). Accountability standards that are built into NCLB, which include student performance on these high-stakes tests, may also affect educational resources, time, and funding, as education becomes more driven by productivity (Linn, 2005).

A by-product of NCLB and other federal laws applicable to public education, including the IDEIA (2004), is the requirement for educators to be "highly qualified" in the areas they teach and to use "evidence-based instruction," adding more strain to educators' already full work-load and leading many to seek more training (Reeves, Bishop, & Filce, 2010). Federal lawmakers are increasingly turning to objective, business-like data-keeping, measurement, and accountability in productivity as they continue to reduce spending in all facets of government, including education. Federal and state spending is highly tied to IDEIA; legislators have taken a renewed interest in producing successful citizens, including those with disabilities.

Individuals with Disabilities Education Improvement Act

IDEIA (2004), which is the current revision of the original IDEA of 1975, continues to mandate that all children be able to access a free appropriate public education within the least restrictive environment just as the original IDEA; however, the IDEIA was reframed in light of the NCLB (2001). NCLB created more emphasis on "highly qualified" teachers, AYP, and evidence-based practices. IDEIA's reauthorization extended many statutes from NCLB and applied them to each school's responsibilities to children with educational disabilities. Just as in the AYP goals asserted under NCLB, IDEIA requires highly qualified instructors to be licensed and proficient in the subject areas they teach and to use evidence-based instructional practices for all children with disabilities (Yell, Shriner, & Katsiyannis, 2006).

Schools are federally mandated to implement IDEIA (2004) and yet meet the requirements of NCLB. Changes to improve student outcomes have led to the use of the high-quality instruction mandated by IDEIA to help children who have learning difficulties, but who have not yet been identified with educational disabilities. Mandated high-quality instruction helped to resolve the dispute over the best way to identify students with disabilities that began as soon as it was placed into federal law in 1975 with the first authorization of IDEA.

The discrepancy model, which was formally adopted as the primary means of identifying learning disability by the U.S. Office of Education in 1977, was defined as a discrepancy between IQ and achievement (U.S. Office of Education, 1977). At the time that it went into effect, it had very little research to support its prescription (Restori, Katz, & Lee, 2009). School professionals and researchers quickly realized some of the downfalls of its use, including poor ability to identify learning disabilities early, inconsistencies in implementation, questionable use of time-intensive intelligence testing to aid in helping students' difficulties, high false negatives with students who demonstrate low cognitive skills, and high false positives with students who may have lacked appropriate instruction (Fletcher, Coulter, Reschly, & Vaughn, 2004; Hale et al., 2010; Reschly, 2002; Restori et al., 2009; Willis & Dumont, 2006). Additionally the discrepancy model was criticized by Restori et al. (2009) for its tendency not to identify struggling students until they reached higher grades when the critical window of instructional opportunity had passed.

Response-to-intervention (RtI) is a marked change from the IQ-achievement discrepancy model (Aaron, 1997; Ardoin, Witt, Connell, & Koenig, 2005). Stemming from the twin pressures deriving from the use of school consultation models and from the growing dissatisfaction with the discrepancy mode, the educational setting was ripe for RtI to begin its realization in the public schools nearly 30 years later. The goals of RtI are to improve student academic and behavioral skills and to provide data for identifying learning disabilities (Fletcher & Vaughn, 2009; D. Fuchs & Fuchs, 2006).

Response to Intervention (RtI)

RtI has been researched since the 1980s. The National Research Center for Learning Disabilities (Mellard, 2004) recommended that specific features be present in all RtI programs: (a) high quality classroom instruction as demonstrated through assessments across classrooms within each grade level, (b) research-based instruction that has been proven to be valid in making academic gains for students, and (c) classroom performance. Classroom performance can be measured through assessments such as statewide achievement tests, universal screening of students not meeting curriculum standards, continuous progress monitoring of both behavior and academic performance of students in the general curriculum, research-based interventions provided in addition to instruction in the general curriculum, progress monitoring of target behaviors or academic skills during interventions, and fidelity measures to ensure consistency of instruction (Mellard, 2004). Additionally, Mellard (2004) described common attributes of RtI which include the concept of multiple tiers of increasingly intense student interventions; implementation of a differentiated curriculum; instruction delivered by staff other than the classroom teacher; varied duration, time, and frequency of interventions; and categorical and non-categorical placement decisions.

Researchers have theorized a few different approaches for RtI implementation, the two primary approaches being either the standard treatment protocol or the problem-solving model (Christ, Burns, & Ysseldyke, 2005; D. Fuchs, Mock, Morgan, & Young, 2004). Christ et al. (2005) described both RtI approaches and summarized that there is little difference in practical application between the two. Specifically, they asserted that the primary difference between the standard protocol approach and problem-solving approach is the level of individualization, including the depth of problem analysis that occurs before selecting an intervention, the design of an intervention, and methods of implementing the chosen intervention. Christ et al. also concluded that in many cases the methods are blended when implemented at the school or district level.

The purposes of RtI models are to provide interventions to children with academic difficulties and to identify those who do not respond to appropriate instruction. Just as in the case of the IQ-achievement discrepancy model, RtI has been met with criticism from the school psychology and educational communities, including variation in implementation across school districts, variation in RtI protocols, lack of empiricism of the assumptions influencing RtI procedures, potential for false positives, and difficulty of intervention implementation (Hale et al., 2010; Marston, 2005; Reynolds & Shaywitz, 2009; Wanzek & Vaughn, 2007). However, IDEIA (2004) regulation has essentially permitted the use of RtI as an equally valid option for the identification of learning disabilities.

Reading Skills and Reading Disability

Because reading is a foundational skill upon which many other skills and areas of academic instruction are based, schools are extremely concerned with increasing the reading abilities of their students as a means to gaining ground toward their AYP goals. This focus has been assisted by IDEIA (2004) and RtI (Mellard, 2004), which allow for many students to be identified with reading difficulties before they even take their first statewide assessment test in third grade. Identifying students who are at risk of reading difficulties and providing interventions for reading deficiencies as early as possible has become a significant emphasis in education. Further, more than half of all special education students have a learning disability, mostly in reading, which is a disability category that has grown by more than 300% since 1976 (President's Commission on Excellence in Special Education, 2002). Thus, the importance of reading is being emphasized by general education and AYP and special education in the form of students with learning disabilities.

Importance of Reading Fluency

According to the National Reading Panel (2000), reading is composed of the successful acquisition of five basic skills: phonemic awareness, phonics and blending, reading fluency, comprehension, and vocabulary. Reading fluency is defined as "the ability to read a text quickly, accurately, and with proper expression" (National Reading Panel, 2000, p. 11). Fluency is important because effective reading requires word recognition skills at a rate fast enough that the reader is able to focus attention on the meaning of a passage. Skilled reading fluency also requires the recognition of punctuation and determination of where to place emphasis or pause within a passage in order to make sense of the text. Summarily, fluency is a recognized aspect of successful comprehension because it allows the reader to expend less effort on the decoding of words and focus more on the meaning of words and word phrases (Aaron & Joshi, 1992; National Reading Panel, 2000).

The premise of RtI as it relates to reading is to conduct frequent measures in order to document progress in reading, which are used to help in identification of students with reading disabilities. Fluency is an area that has empirical evidence of repeated assessment forms that can be conducted frequently, namely curriculum-based measures, which will be discussed further in this document. Because RtI effectiveness is often assessed using measures of reading fluency, it is difficult to conceptualize the identification of students with reading disabilities by means of RtI without discussing reading fluency.

Students with Reading Disabilities

In regard to the use of RtI in the identification of reading disabilities, it is important to review the RtI process. In a standard-treatment protocol, which appears to have a more stringent research-base, Marston (2005) described three-tiered methods of RtI in which Tier 1 is core reading instruction, Tier 2 is supplemental instruction to the core curriculum, and Tier 3 is core instruction plus intensive interventions.

Following the early intervention phases, Vellutino et al. (1996) reported that there was a notable (67%) reduction in the number of students who scored below average on measures of reading fluency. Additionally, in a review of RtI programs, D. Fuchs, Mock, et al. (2004) reported that participation in interventions resulted in as much as a 20% reduction in the number of students with below average reading skills and that 98% of schools who implemented RtI provided anecdotal evidence demonstrating a reduction in special education referrals. Marston (2005) also conducted a review of RtI reports, indicating that students who participated in Tier 1 alone resulted in a 3% reduction of special education referrals; however, students who participated in Tier 2 resulted in as many as 75% no longer requiring intervention. The most remarkable outcome of Marston's report is that over 90 percent of students who participated in all three tiers of intervention responded to the interventions. Another impact of RtI was that placements in special education decreased in all grade levels by at least 20%, thus decreasing the need for special education.

In most commonly used RtI models, decisions are made at the end of each tier. Although they vary, these decisions are often based on normative data of reading fluency scores, including scores or rate of improvement over the course of the intervention being above or below the 25th percentile (McMaster & Wagner, 2007). L. S. Fuchs, Fuchs, and Speece (2002) argued that the use of both a single score and growth rate over a specified period of time is warranted in the decision-making process of RtI.

Reading fluency growth rates differ for students who have been identified with a learning disability (Deno, Fuchs, Marston, & Shin, 2001). For example, in first grade, when comparing students with and without a learning disability, Deno et al. (2001) found that students with a learning disability had approximately one-half the growth rate of typically developing students; however, this difference appeared to be less dramatic in higher grades. Students with disabilities continued to demonstrate a rate of growth that had essentially reached a plateau after second grade, even with instruction, while their same-aged peers' rates of improvement did not reach a plateau until fifth or sixth grade. The rate of increase in both groups also tended to be the same once a relative plateau had occurred (.60 for general education students and .58 for special education students). Once a student falls behind his or her normally developing peers, it continues to be difficult for that student to "catch up." This lag is described as the Matthew effect by Stanovich (1986), who suggested that students who start out as poor readers are more likely to remain poor readers because they read less than good readers.

Curriculum-Based Measurements of Reading

According to the National Reading Panel (2000), there are several informal procedures that can be used in the classroom to assess fluency: (a) informal reading inventories, (b) miscue analysis, (c) pausing indices, (d) running records, and (e) reading speed calculations. All of these procedures require oral reading of text and can be used to provide an index of fluency. Additionally, oral reading fluency can also be assessed in a manner that emphasizes prosody, which "requires examiners to describe the pitch, stress, and duration with which children express text" (L. S. Fuchs, Fuchs, Hosp, & Jenkins, 2001, p. 251), but this is difficult to assess reliably and efficiently. Currently, the most widely used method to assess reading fluency is curriculum-based measurement (CBM), because it is a low-cost method for assessing reading on a frequent basis. The development and application of CBM is generally credited to Stanley Deno at the University of Minnesota's Institute for Research on Learning Disabilities during the mid-1970s for the purpose of aiding special educators' use of reliable, objective data in making decisions about student progress and improving the quality of instruction (Deno, 1985; Deno et al., 2001; Stecker, Fuchs, & Fuchs, 2005).

Since its beginnings over 20 years ago, CBM of reading fluency has been used by educators and researchers to provide easily accessible data regarding a child's performance and growth (Deno, 1985). Growth, which can be calculated using several data points over time and displayed as a graphical line of growth, has been shown to differ at varying ages and grade levels (Silberglitt & Hintze, 2007). For example, children who are identified as beginning readers have higher growth rates as compared to older students. Curriculum-based measurement scores and growth rates are also being used by educators to make decisions regarding educational placement and are used in RtI models (D. Fuchs, Fuchs, & Compton, 2004).

Curriculum-based measurement is cost-effective for repeated sampling because there are alternative test forms, which can be either a part of a commercially created series of CBM passages (e.g., Dynamic Indicators of Basic Early Literacy Skills [DIBELS] or AIMSweb), or created from classroom texts using resources available on the world-wide web (e.g., http://www.interventioncentral.org). Deno (1985), in his review of CBM, observed that the advantages of CBM were improved communication due to graphical representations of the CBM scores over time, increased sensitivity to changes in student ability and skill, an improved data base for use in making educational decisions, peer referencing in order to choose the most appropriate comparisons of proficiency for a student, and cost effectiveness when compared to standardized assessments.

The most widely used method of CBM for reading fluency (R-CBM) primarily involves counting the number of words read aloud correctly by a student from a passage of text in one minute. Another form of R-CBM that is widely used is isolated word lists. However, Jenkins, Fuchs, van den Broek, Espin, and Deno (2003) reported that text fluency explained a significantly greater portion of the variance in reading comprehension than did list fluency.

Not all CBMs are created alike. One argument against the use of passages from the curriculum was described by Shapiro (2004). If passages are chosen directly from the curriculum, there are often differences in the readability level among samples of one grade level's texts. To determine the grade level of a passage and to ensure the similarities between passages chosen for progress-monitoring, or comparison over time, the readability of passages must be measured. Readability can be determined using reading probe generators found on the internet (e.g., http://interventioncentral.org and http://www.readability-software.com) that offer formulas that are vocabulary-based, count the number of syllables within words, and/or compare the length of sentences. AIMSweb passages have been graded based on the Lexile formula (http://www.lexile.com). Hintze and Christ (2004) found that controlled passages (not taken directly from the child's current curriculum) reduced measurement error, leading to increased sensitivity and reliability of the CBM.

Curriculum-based measurement can also be used qualitatively to provide diagnostically useful descriptions of performance. For example, a teacher or examiner using R-CBM may not only count the number of words read correctly, but may also note the types of decoding errors made, the kinds of decoding strategies used on unknown words; the reliance upon specific features of language; the prosody of the student's reading; and the types of self-corrections, pacing, and scanning strategies used. Additionally, Stecker et al. (2005) summarized preceding research on CBM in this way:

> Over the past 25 years, numerous investigations have emerged that utilized CBM in a variety of ways including, but not limited to: (a) establishing norms for screening and identifying students in need of special education services (Shinn, 1989b), (b) evaluating the effectiveness of educational programs (Tindal, 1992), (c) reintegrating students with disabilities into general education classrooms (D. Fuchs, Fernstrom, Reeder, Bowers, & Gilman, 1992), (d) monitoring progress and planning instruction within general education classrooms (L.S. Fuchs, D. Fuchs, Hamlett, Phillips, & Bentz, 1994), and (e) identifying potential candidates for special education evaluation using a dual-discrepancy model of low level of performance and inadequate rate of improvement (L.S. Fuchs et al., 2002). (p. 765)

Data that are gathered as individuals practice reading over time may reveal gradual and incremental improvement (National Reading Panel, 2000). These data, often in the form of CBM, may be used to calculate a rate of improvement (ROI) or measure of student growth. There is, however, more than one way to estimate ROI, or a model of student growth.

Deno et al. (2001) described both a simple linear relationship to model student progress within one academic year and a quadratic analysis of the slope to demonstrate a negatively accelerating pattern of student performance improving over the course of a year but decreasing gradually across years. In other words, a typically developing student is expected to make some progress each year. A second assumption is that the degree of improvement will decline as the child grows older. Following from these two assumptions, a comparison of the scores within each year using the simple linear relationship would provide one with the ability to observe this process. The quadratic analysis provides the information needed to know exactly how much decline there is across the years for each child.

Additionally, Yeo, Fearrington, and Christ (2011) utilized oral reading fluency and maze (reading comprehension fluency) scores from over 1,700 students to generate linear growth models for third through eighth grades. What the authors found was that slope was generally consistent among students regardless of demographic characteristics but that oral reading fluency linear slope in third and seventh grades was not consistent with the other grades. Yeo et al. suggested that the slope of oral reading fluency scores can be used as a predictor of reading performance in fourth, fifth, sixth, and eighth grades.

Silberglitt and Hintze (2007) described estimation of growth rates using hierarchical linear modeling (HLM), the results of which indicated that the slopes of oral reading fluency rates of improvement were significantly lower for students in the bottom- and upper-most deciles than for students within the average-level performance deciles. HLM is an example of a linear mixed model and is useful for the analysis of longitudinal data because it allows for missing data. It also assumes independence between subjects, that there is not independence between measurements, that the measurements are nested within students, and that the response variable is normally distributed. The use of HLM in assessing growth with CBM scores was also supported by Shin, Espin, Deno, and McConnell (2004) to evaluate growth patterns, growth rates, developmental lag, and instructional factors facilitating growth when measuring the relationship between instruction and growth.

Additionally, Christ (2006) indicated that when linear models are used to predict or summarize CBM performance, the calculation of the standard error of estimation, a measure of the accuracy for the point predictions of the regression line, is needed. For this reason, the author used a monotonic linear model to estimate the "standard error of slope across multiple progress-monitoring durations" (Christ, 2006, p. 130). This is in contrast to the HLM used by Silberglitt and Hintze (2007) because HLM does not provide this type of error measurement alone. The commercially produced CBM series, AIMSweb, also produces an estimate of ROI over time for each student as well as an estimated norms table for a group of students, a school, or an entire school district. The statistical procedure for the calculation of the ROI is not published by AIMSweb; however, L. S. Fuchs and Shinn (1989) and L. S. Fuchs, Fuchs, Hamlett, Walz, and Germann (1993) reported that it is based on a simple linear growth model between the fall and spring data points. In contrast to the AIMSweb fall-to-spring method of calculating ROI, Wiley and Deno (2005) found that fall scores are much lower than winter and spring scores, indicating a falsely elevated growth rate estimate. This indicates that there are differing conceptualizations of estimating growth rates within one year.

This limited review indicates that various methods for calculating rates of improvement are used. The theoretical reasoning for using HLM is that it allows for analysis of variables with nested relationships or variables that would break the assumption of multicollinearity required in other statistical procedures. On the other hand, HLM is impractical for use within schools because many researchers are not well-versed in its applications or use beyond theoretical knowledge. Simpler models including linear regression provide a practical solution to the problems faced by school personnel who are seeking to improve the application of research to drive decisions regarding service delivery and improvement of student outcomes.

Predictors of Statewide Achievement Tests

Curriculum-Based Measurement and Statewide Achievement Tests

Because of the changes in federal law that accompanied the IDEIA (2004), many school systems have switched to an RtI system of identifying children with learning disabilities. A significant aspect of RtI programs for academic problems is the use of CBM to monitor progress and to make decisions about the presence of a disability. Additionally, following the authorization of NCLB (2001), school systems are under pressure to provide documentation of academic progress to the state educational agencies that govern them. This documentation is often in the form of statewide achievement test scores. The use of CBM data to predict statewide achievement test proficiency would help schools make decisions at both the individual and system levels regarding programming and prospective educational outcomes for students.

There have been several studies to examine the predictive power of CBM scores on statewide achievement tests. Several of these studies, for the sake of standardization, have tested the validity of the cut scores generated by Good, Kaminski, and Dill (2002) in the manual for the DIBELS, or have used CBM norms that were developed from a sample of over 9,000 students by Hasbrouck and Tindal (1992). For example, according to Good et al. and Hasbrouck and Tindal, the cut score for students in the spring of 3rd grade, "low risk," is reading at least 110 words correctly per minute, the cut score for "some risk" is reading between 80 and 110 words correctly per minute, and the cut score for "at risk" is reading below 80 words correctly per minute. There are different cut scores, or end of year goals, for each grade. Unless otherwise stated, these are the cut scores primarily used in the literature reviewed below.

McGlinchey and Hixson (2004) used R-CBM cut scores to determine how reliably they could predict meeting a proficiency score on statewide reading achievement tests in fourth grade

(*N* = 1,362). Using a cut score of 100 or more words read correctly per minute, the researchers found that they could reliably identify in nearly 75% of the children whether or not they would pass the state reading assessment with few false positives or negatives. Similar results were found in third grade by Good et al. (2001) in four cohorts of approximately 300 students, except with much better predictive power, as 96% of students were correctly identified as passing based on a cut score.

Shapiro, Keller, Lutz, Santoro, and Hintze (2006) also compared R-CBM scores with the state reading assessments in over 2,200 second through fifth graders. The researchers found significant correlations of .62 and .69 between oral reading fluency and the statewide achievement test. Additionally, the winter R-CBM demonstrated the strongest correlation. Cut points were generated using a statistical procedure called receiver operating characteristic (ROC) curves in which multiple models were generated to identify the point at which sensitivity and specificity between the R-CBM with the statewide achievement test is optimum. The results indicated a correct prediction that a student who was below the R-CBM cut score of 125 or 126 words read correctly per minute in the winter would be below the proficiency cut score on the statewide achievement tests with 84-94% accuracy, while the prediction that a student who was above the R-CBM cut score would be above the statewide test cut score was only accurate just over half of the time (58-68%). Shapiro et al. asserted that these levels suggest that R-CBM would provide useful screening information regarding whether a student will pass or fail statewide testing. However, it should also be noted that the cut scores that Shapiro et al. used were generated by the data rather than being drawn from the assessment manual, as in Good et al. (2002). Actually, the cut scores (125-126 words read correctly per minute) that were generated by the data were much higher than the at-risk range used by Good et al., but Shapiro et al. indicated that these cut scores generated optimum levels of specificity and sensitivity.

Cut scores of R-CBM and statewide achievement tests were also examined by Crawford, Tindal, and Stieber (2001), who found that in a sample of 51 third graders assessed across a two-year period, 81% whose R-CBM scores were at or above the 50th percentile (based on oral reading fluency norms generated by Hasbrouck and Tindal, 1992) passed the statewide test in reading. They also found that 94% of children who read above a cut score of 119 words read correctly per minute passed the statewide reading test. In expanding the study to longitudinal analysis, 78% of the children whose second grade R-CBM scores were above the 25th percentile passed the statewide test in third grade. Additionally, 29% of those who scored below the 25th percentile also passed the statewide achievement test. The results also indicated that 100% of students who were reading at least 72 words read correctly per minute in the second grade passed the statewide achievement test in third grade. The difference in predictive power was not statistically significant; however, the across-year correlation was stronger between second grade oral reading rate and success on the reading test (.66) than the same-year third-grade oral reading rate and the statewide test proficiency (.60). The authors also found that there was no significant difference between the rates of improvement of children whose initial scores were low compared with those whose initial scores were high. This research supports the use of not only CBM to predict proficiency on high-stakes tests, but also the use of those data across years. An obvious limitation of the Crawford et al. study, however, was the small sample size. Similar correlations using a slightly larger sample size were found by Wood (2006) for fourth grade (n = 101) same-year comparison of reading fluency and statewide test score (.64) and fifth grade (n = 98) same-year comparisons (.68).

Goffreda, Diperna, and Pedersen (2009) examined more longitudinal data when they attempted to predict second-grade TerraNova proficiency and third-grade statewide achievement tests using 67 students' first-grade letter-naming fluency, phoneme sound fluency, nonsense word fluency, and oral reading fluency scores. What they found after using logistic regression was that first-grade oral reading fluency was the only significant predictor of second-grade TerraNova proficiency and third-grade statewide achievement tests proficiency. However, a significant limitation of this study was the low number of students in the sample.

Hosp, Hosp, and Dole (2011) also examined the utility of first-, second-, and third-grade oral reading fluency scores to predict same-year administration of first-, second-, and third-grade statewide achievement tests. Of the nearly 4,000 students who were included in the study, predictive validity was good for oral reading fluency across all grade levels, but validity was not maintained when applied to ethnic, gender, disability, and socioeconomic status subgroups, as the effects that occurred indicated predictive validity bias against traditionally underrepresented subgroups. However, the statistical analyses (ROC curves and quantile regression) used in this study are complex and were chosen because they were more concerned with examining the predictions generated for a variety of demographic characteristics than on the predictive validity of oral reading fluency on statewide achievement tests. While the complex statistical analyses described in this study may be helpful for making rate of improvement predictions, it is likely that few school psychologists, and even fewer teachers or administrators would know how to conduct them.

Silberglitt, Burns, Madyun, and Lail (2006) examined the relationship between statewide testing scores in reading and R-CBM in over 5,400 third through eighth graders and found that R-CBM scores at lower grades had higher correlations to achievement test scores than did R-

CBM scores at higher grades. In their study, Silberglitt et al. determined that R-CBM scores increased at a much greater degree in the lower grades than in the higher ones. Further, they established that the relationship between state test scores and R-CBM scores was much stronger at third grade than at either fifth or eighth grades. No comparison was made across grades in this study.

Merino and Beckman (2010) took a different approach to predicting statewide achievement tests, using both oral reading fluency and maze scores. In a study of over 375 students in second through fifth grades, the authors used multiple regression techniques to predict same-year, spring statewide achievement tests. The results indicated that maze scores alone did not significantly predict statewide achievement tests in any of the grades, but oral reading fluency scores alone and the combination of oral reading fluency and maze scores did in all of the grades. Of further interest was that these variables behaved differently in predicting the fall statewide achievement tests: oral reading fluency alone and the combination of maze and oral reading fluency again predicted scores in all of the grades, but maze scores also were positive predictors of fourth-grade fall statewide achievement test scores. In summary, oral reading fluency was the better predictor of statewide achievement tests and was the strongest in second grade. The authors indicated that the results may not be generalizable to the U.S. population of students because over a third of the sample students were English language learners. In addition, the results were found on such a small sample.

Additionally, Silberglitt and Hintze (2005) also found a significant relationship between R-CBM and state achievement tests in a sample of over 2,100 first through third graders, but they pointed out that the correlations were higher for those CBM assessments temporally closer to the administration of the statewide achievement test. What is of more interest, particularly in relationship to the current study, is that the R-CBM scores as far back as first grade were able to predict with 80% accuracy those who were likely to pass the state achievement test in reading in third grade. Hintze and Silberglitt (2005) confirmed these results with over 1,700 children, but stated that when several statistical methods were used to calculate the predictive value of CBM on statewide achievement tests, there were no significant differences in the predictability of the methodology used. The authors also noted that using benchmark cut scores to make these predictions is problematic in that the predictions can vary within one school year, so for this reason it may be valid to consider another means of utilizing benchmark data within one year, namely, rate of growth.

Barger (2003) sought to identify a specific cut score that would be the best able to predict statewide achievement test scores. In a brief technical report, he examined 38 third-grade students' spring oral reading fluency scores and correlated them with the third-grade statewide achievement test. Using the 40th percentile cut scores prescribed by Good et al. (2002), Barger found that 92% of the students who scored above 110 words read correctly on the oral reading fluency assessments also gained a high classification on the statewide achievement test. The problem concerned predicting who would fail the statewide achievement test. Barger hypothesized that the small number of students who actually failed the high-stakes assessment contributed to his low ability to predict who these students were.

Similar results were found by Shaw and Shaw (2002) on a sample of 52 third-grade students, indicating that 91% of students who read at least 90 words correctly in one minute scored within the proficient or advanced ranges. Additionally, 27% of students who scored fewer than 90 words read correctly in one minute on the spring oral reading fluency measure demonstrated a proficient score. It should be noted that the discrepancy in these figures does not indicate a calculation error; the authors did not comment on those whose scores fell within the "partially proficient" range, which is a category not consistently present in all states (Bandeira de Mello et al., 2009).

Wilson (2005) also reported on nearly 250 third-graders that the same cut scores used by Good et al. (2002) which resulted in showing that 93% of students falling in the at-risk category (less than 80 words read correctly per minute) were not proficient on the statewide achievement test, and approximately 82% of students who fell within the low risk category (reading at least 110 words correctly per minute) demonstrated proficient statewide achievement test scores. Vander Meer, Lentz, and Stollar (2005) reported similar findings with a sample of over 360 third- and fourth-grade students, indicating that 96% of students who scored at-risk, or read below 80 words read correctly per minute, on the third-grade spring oral reading fluency measure scored below proficient on the fourth-grade fall statewide achievement test. They also indicated that 97% of fourth-grade students who scored within the at-risk range, or reading less than 71 words correct per minute on fourth-grade fall oral reading fluency measures (within the at risk range), were also not proficient on the fourth-grade fall statewide achievement test. Conversely, 72% of those whose cut scores were above a 93 (within the low-risk range) on either oral reading fluency administration demonstrated proficient scores on the fourth-grade assessment. The findings of Vander Meer et al. are reassuring, as they were able to predict, with relative stability over time, students who would score poorly on the fourth-grade reading assessment.

Roehrig, Petscher, Nettles, Hudson, and Torgesen (2008) also attempted to identify a specific cut score to help with decision making on identifying those at risk for performing poorly on the statewide achievement test. In an examination of over 16,500 students and cross-validation upon nearly 17,000 additional students, Roehrig et al. attempted to use third-grade fall and winter oral reading fluency scores to predict both the statewide achievement test and the Stanford Achievement Test (SAT-10). Using the cut scores recommended by Good et al. (2002), the authors identified students who demonstrated levels of risk on the oral reading fluency scores; however, they adjusted the cut scores because the Good et al. cut scores were not as sensitive and specific so as to accurately identify risk in the sample. Therefore, these adjusted cut scores were shifted to the following ranges: third-grade fall oral reading fluency was adjusted from ≥ 77 to ≥ 76 being low risk and from ≤ 53 to ≤ 45 being at risk, and third-grade winter oral reading fluency was adjusted from ≥ 87 to ≥ 84 being low risk and from ≤ 62 to ≤ 55 being at risk. Using the adjusted cut scores, the cut scores accurately sorted the statewide achievement test scores as well; students whose oral reading fluency scores indicated risk proportionately indicated risk on the statewide achievement test, though there was a lowered risk of false-positives but not false-negatives. The study also indicated that the winter oral reading fluency score demonstrated the strongest relationship with the statewide achievement test and the SAT-10. However, further analysis was not conducted, as the focus of the research was in identifying the predictive bias of oral reading fluency predictions upon ethnic- and income-related factors, not necessarily validity or relationship with statewide achievement tests. The authors stated that there was no predictive bias across demographic factors.

Rate of Growth and Statewide Achievement Tests

Stage and Jacobsen (2001) examined the relationship between the growth rate of several R-CBM benchmark scores and the statewide achievement test in fourth grade (N = 173) and found that although the ROI was a significant predictor of statewide achievement test scores, a single R-CBM score was a better predictor. They noted the deceleration of growth rate by fourth grade may have also been a confounding factor in the poor predictive value. However, the researchers' sample was small and they cautioned against discounting the use of growth rates as a predictor of statewide achievement test scores until further research with a wider population could be conducted.

A second study that examined the utility of both oral reading fluency and reading growth to predict a statewide achievement test score was conducted on nearly 12,000 first- through third-grade students over a two-year period (Baker et al., 2008). Using the 40th percentile oral reading fluency benchmarks recommended by Good et al. (2002), Baker et al. found that first-grade oral reading fluency correlated well with the first-grade SAT-10, and both first- and second-grade oral reading fluency scores correlated well with the second-grade SAT-10. Using growth curve analyses, a model of best fit was developed in which various possible combinations of the oral reading fluency intercept (point at which the slope of the rate of improvement is zero), oral reading fluency rate of improvement (slope), and first-grade spring SAT-10 were used to predict second-grade SAT-10 performance. Of all of the possible combinations of variables, the combination of second-grade spring oral reading fluency score and second-grade slope explained 70% of the variance in the second-grade SAT-10 score. Additionally, the best fit growth model combination of intercept, first-through-second grade slope, and first-grade SAT-10 provided the best fit and explained 76% of the variance in second-grade SAT-10 scores. However, the first-grade SAT-10 score was the best single predictor of second-grade SAT-10 scores, closely followed by first-grade oral reading fluency slope. The prediction of third-grade statewide achievement tests held similar results, as second-grade oral reading fluency intercept and slope accounted for 52% of the variance in the third-grade test, and that the best-fitting model included intercept, slope, and the second-grade SAT-10; all three predictors accounted for 59% of the variance in third-grade statewide achievement test scores. However, the second-grade oral reading fluency intercept predicted more variance in third-grade achievement test scores than the oral reading fluency rate of improvement. It was noted that all of the predictor variables demonstrated high correlations with each other, violating the assumption of multicollinearity. However, this study helped to demonstrate that rate of improvement is a valid and strong predictor of statewide reading tests and in combination with other factors, including intercept and previous achievement test scores, can be used to develop a prediction model.

Finally, Chard et al. (2008) followed 668 kindergarten and first-grade students through third grade and utilized oral reading fluency growth curve, SAT-10 vocabulary, SAT-10 comprehension, first-grade letter-naming fluency, first-grade phonemic segmentation fluency, first-grade nonsense-word decoding, first-grade Woodcock Reading Mastery Test-Revised (WRMT-R), and the Social Skills Rating Scale academic competence, problem behavior, and social skills subscales to predict third-grade achievement test scores. Chard et al. used path analysis to create a prediction model. Significant predictors of oral reading fluency initial status included letter-naming fluency, alphabetic principle, and academic competence, which accounted for 75% of the variance in initial oral reading fluency. Significant predictors of oral reading fluency rate of improvement were alphabetic principle and academic competence, which accounted for 11% of the variance in oral reading fluency rate of improvement. Significant predictors of third-grade SAT-10 scores were first-grade passage comprehension; oral reading fluency rate of improvement; academic competence; first-grade phonemic segmentation fluency rate of improvement; low problem behavior; and not being African American. This last factor may indicate a predictive bias of CBM, such as was addressed by Roehrig et al. (2008). The most notable predictors of the third-grade SAT-10 were first-grade passage comprehension from the WRMT-R and that the effect of oral reading fluency rate of improvement was not significant. Chard et al. indicated that the reason that slope was not significant was that it was mediated by other factors in the path analysis, including alphabetic principle and passage comprehension skills.

Curriculum-based measurement data, because of their cost-effectiveness, utility of multiple assessments within relatively short periods of time, and sensitivity to student growth, provide optimum conditions for research applications. In light of NCLB, the application of CBM in predicting student success on statewide achievement tests provides schools with the capability of identifying children at risk of failure and of providing the appropriate instruction at just the right time to influence the greatest impact on statewide achievement tests in the future. Silberglitt and Hintze (2005) provided substantial reason to believe that this application of CBM is within the grasp of schools to positively impact AYP and student achievement. Stage and Jacobsen (2001), Baker et al. (2008), and Chard et al. (2008) also indicated that the examination of rate of growth on CBMs continues to be warranted, especially in predicting success on statewide achievement tests.

Advantages of This Study

Several researchers have found CBM to have significant predictive value on student proficiency on statewide achievement tests; however, these have primarily focused on third- or fourth-grade assessments and few have expanded the longitudinal studies to first grade. Additionally, few of the studies have examined the longitudinal benchmark data in elementary school to predict success in eighth grade. Only three studies have examined ROI as a predictor of statewide test success (Baker et al., 2008; Chard et al., 2008; Stage & Jacobsen, 2001), and one of which was done on a very small sample size (Stage & Jacobsen, 2001). The current study seeks to add to the literature in this area by replicating research regarding third-grade state tests, while adding to the literature through a downward extension of prediction to data from earlier grades. The current study also seeks to predict middle-school (eighth grade) success in elementary school and readdress the theory of Stage and Jacobsen (2001) to use growth rates to predict success in a sample large enough to detect significant correlations. Finally, the current study also will utilize the benefit of its large sample to cross-validate the prediction equations on the third-grade statewide achievement test.

CHAPTER 3

METHOD

The current research examined this topic through three studies. The specific research questions for each study were as follows:

1. Do reading fluency scores in third grade predict statewide achievement test scores in third grade (within the same year)? Several studies have previously established this in other samples; however, will those findings generalize to a larger sample? It was expected that third-grade R-CBM scores would be a significant predictor of third-grade statewide achievement test scores.

2. Do reading fluency scores in earlier grade levels predict statewide achievement test scores at different grade levels (i.e., reading fluency in first and second grades to predict third grade, and reading fluency at third, fourth, and fifth grades to predict eighth grade statewide achievement test scores)? While a few studies have calculated positive predictive power of first- and second- grade reading fluency scores on third-grade statewide tests, only one has attempted to predict eighth-grade test scores. This study attempted to provide information regarding how early a child's eighth-grade success can be predicted. While the relationship between all grades' R-CBM scores will be significant predictors of statewide achievement test scores, it was likely that the R-CBM scores that are temporally closest to the statewide achievement test would demonstrate the highest predictive value.

3. Do reading fluency rates of improvement in third grade predict statewide achievement test scores in third grade? Only three studies (Baker et al., 2008; Chard et al., 2008; Stage & Jacobsen, 2001) have attempted to use rates of improvement to predict statewide test scores at third grade, one of which (Stage & Jacobsen, 2001) with a very small sample size. This study expanded upon those findings. It was anticipated that the reading fluency rate of improvement (ROI), like that generated by the AIMSweb program, would be a significant predictor of statewide achievement test scores at third grade.

4. Do reading fluency rates of improvement at primary grade levels (first and second grades) predict future achievement test scores (third grade)? Similar to the question above, how soon can reading rates of improvement predict success on the third-grade test? It was expected that the second-grade ROI would provide a higher predictive value with third-grade achievement test scores than first-grade ROI, but both would be significant.

5. Do reading fluency rates of improvement at intermediate grade levels (third, fourth, and fifth grades) predict future achievement test scores (eighth grade)? No known studies have attempted the upward extension of Stage and Jacobsen (2001), Baker et al. (2008), and Chard et al. (2008). This study attempted to demonstrate that reading rates of improvement can be used to predict success in eighth grade and how soon one can determine when to intervene. Fifth-grade rates of improvement were expected to likely have the highest predictive value, but it was predicted that third-and fourth-grade reading fluency rates of improvement would also be significant.

6. Do reading fluency rates of improvement in third grade, following the winter-spring linear regression model proposed by Wiley and Deno (2005), predict the statewide achievement test scores in third grade? The simple ROI proposed by Wiley and Deno would be applied to the

current sample for comparison. A winter-spring ROI was expcted to provide a significant prediction of third-grade achievement test scores.

7. Do reading fluency rates of improvement, following the Wiley and Deno (2005) model, at earlier grade levels (first and second grades, or third, fourth, and fifth grades) predict future achievement test scores (third or eighth grades)? Alternatively, the use of differently calculated ROI scores in predicting success in third and eighth grades were examined. The rates of improvement that were temporally closest to the achievement test sample were expected to demonstrate the greatest amount of predictive value.

8. Which form of reading fluency rate of improvement is a better predictor of statewide achievement test scores? Following the research regarding how rates of improvement were calculated, two practical forms of ROI were used to determine if there is a statistical difference between them in predictive value toward statewide achievement tests. It was anticipated that the two forms of ROI would be similar in their ability to predict achievement test scores.

9. Which is a better predictor of statewide achievement test scores: reading fluency scores or rates of improvement? Finally, the comparison must be made regarding which is better in predicting statewide achievement test scores: the reading fluency scores or the rates of improvement. While both were expected to be significant predictors of statewide achievement tests, the oral reading fluency scores would provide the greatest degree of explained variance.

Study 1

Research questions 1, 2, 3, 4, 6, 7, 8, and 9, which apply to the children who completed the third-grade statewide achievement test in spring 2009 are answered in Study 1. These included the prediction models pertaining to the third-grade statewide achievement test using the first-, second-, and third-grade reading fluency measures and the subsequent rates of improvement.

Participants

Students in a large-sized Southeastern city from grades 1 through 3 were given R-CBM probes three times annually (September, January, and April) between the years of 2003 and 2009. Students having Individualized Education Plans (IEPs) were also administered the R-CBM probes, with the exception of students with severe cognitive or adaptive delays who behaviorally could not attempt the work. The children included in this study were those who took the third-grade statewide achievement test in April 2009 and who had complete R-CBM benchmark data from first, second, and third grades.

Tthe school system had approximately 56,000 children in kindergarten through twelfth grades. There were 3,276 students who took the third-grade statewide achievement test in 2009; however, 846 of these children were missing benchmark assessment data. Therefore, children with incomplete data were omitted from the study. One-half of the children who had complete data sets were randomly assigned using SPSS to the second study for cross-validation and were not included in the current study.

Study 1 included 1,214 students. Demographic information that was retained included age at the time of the 2009 statewide achievement test, gender, ethnicity, socioeconomic status as measured by participation in the free and reduced lunch program, and special education and Section 504 certification. All other identifying information was removed from the data.

Measures

Curriculum-based measurement. One primary type of R-CBM is oral reading fluency (ORF), in which the student reads aloud for one minute and the examiner counts the number of words read correctly. Curriculum-based measures that were downloaded from AIMSweb (Edformation, 2005) were used as the reading fluency benchmark assessments. Three passages were administered at each benchmark, and the median words read correctly (WRC) and the median errors were used as the benchmark score. The same three passages were administered at all three benchmark periods each year for a given student, all of which were on the grade level of the student. Children who were in the third grade during the 2008-2009 school year participated in the benchmark assessments in 2006-2009 (grades 1-3).

Reliability and validity of CBM. Curriculum-based measurements have long been studied for their practical use in educational decision making. The criterion-related validity and reliability of CBM probes were described through a comprehensive literature review and summary by Marston (1989) and updated by Good and Jefferson (1998), Shinn and Shinn (2002), and Wayman, Wallace, Wiley, Tichá, and Espin (2007). The criteria described in Wayman et al. regarding the strength of correlations were used: ≥.70 are strong, .50-.70 are moderate, and ≤.50 are weak.

The majority of these early validity studies discussed by Marston (1989) and included in Appendix Table A1 indicate that R-CBM passages are moderate to strong measures of reading. Marston indicated that of the early validity studies reviewed, criterion-related validity between R-CBM and other measures of reading, including the Woodcock Reading Mastery Test, Stanford Achievement Test, Peabody Individual Achievement Test, Science Research Associates, and California Achievement Test, which ranged from .53 to .88. Additionally, Good and Jefferson

(1998) cited concurrent, criterion-related validity between R-CBM passages to range from .62 to .73, the strongest correlations of which were at the third and fourth grades.

Shinn and Shinn (2002) reported validity coefficents ranging from .26 to .91. Specifically, the studies not included in the summary described above by Marston (1989) that compared fluency passages with standardized assessments, such as the Comprehensive Test of Basic Skills, California Achievement Test, Gates-MacGintie Reading Test, Metropolitan Achievement Test, and Kaufman Test of Educational Achievement-Brief, that ranged from .41 to .86, which is summarized in Appendix Table A2. Although the additional studies indicated validity coefficients somewhat lower than earlier studies, the majority of studies were within an acceptable range.

Wayman et al. (2007) summarized several more recent studies of the concurrent and predictive validity of CBM, as well as its reliability. Only the studies not already reviewed and summarized by Marston (1989) and Shinn and Shinn (2002) are included in Appendix Table A3. Wayman et al. reported criterion-related validity coefficients ranging from .42 to .91 between R-CBM and standardized tests such as the Woodcock Reading Mastery Test, Woodcock-Johnson, Third Edition, Iowa Test of Basic Skills, Stanford Diagnostic Reading Test, Minnesota Comprehensive Assessment, Diagnostic Reading Scales, Michigan Educational Assessment Program, Washington Assessment of Student Learning, and the Stanford Achievement Test. Test-retest reliability coefficients between R-CBM and other reading tests (i.e., Woodcock Reading Mastery Test Word Identification, Michigan Educational Assessment Program, and Stanford Achievement Test) ranged from .87 to .99. Alternate-form reliability with the Minnesota Comprehensive Assessment ranged from .83 to .87.

Wayman et al. (2007) also summarized concurrent and predict validity between R-CBM and the Stanford Achievement Test and Minnesota Comprehensive Assessment test, which is presented in Appendix Table A4. Concurrent validity coefficients ranged from .49 to 85, while predictive validity coefficients ranged from .62 to .72. These values indicate that R-CBM typically assesses and predicts the overall reading skills also assessed by states in their standardized reading assessments.

Additionally, the reliability studies reviewed by Shinn and Shinn (2002), Marston (1989), and Wayman et al. (2007) indicate that test-retest reliability coefficients ranged from .82 to .99, and parallel form reliability coefficients ranged from .90 to .98. All of these reviews indicate that the R-CBM is both a reliable and valid measure.

Tennessee Comprehensive Assessment Program. The Tennessee Comprehensive Assessment Program (TCAP) Achievement Test, taken by students in third through eighth grades each spring, is a timed, multiple choice, criterion-referenced, standardized assessment that measures skills in Reading and Language Arts, Mathematics, Science and Social Studies, based upon content standards and state performance indicators for each grade level (Tennessee Department of Education, 2007). Each subject area test is given in two parts, each of which is given in approximately one-hour segments. Students with IEPs may be permitted to take them in longer segments or in smaller group settings. Scaled scores were used to determine TCAP proficiency and provide normally distributed data.

The state of Tennessee utilizes a database of TCAP data through the Tennessee Value-Added Assessment System (TVAAS, 2006), a SAS Enterprise Intelligence Platform (SAS Institute, 2007). TVAAS is a statistical process that provides measures of the influence that previous statewide achievement test scores have on future test scores (Sanders & Horn, 1994). It uses scale scores from the nationally norm-referenced portions of the TCAP, called the Comprehensive Test of Basic Skills (CTBS, fourth edition). In 1998, Tennessee administered the TerraNova which was double-tested with the CTBS-4 in a sample of 1,500 students per grade to transition the CTBS-4 scores into TerraNova scaled scores. Finally, in 2004 the state transitioned to a criterion-referenced test from the norm-referenced tests that were being used. Both the TerraNova and the criterion-referenced test were administered to all children in 2004 which allowed the scales of the two tests to be combined. The scale that continues to be used is criterion-referenced (TVAAS, 2006). For the purposes of this study, the TCAP scores used were from the criterion-referenced scale used currently; there have been no significant changes to the TCAP since 2003.

Reliability and validity of TCAP. The TCAP tests are currently developed and analyzed by the Educational Testing Service (ETS). No reliability and validity for the TCAP were publicly available from the Tennessee Department of Education.

Procedures

Benchmark data and TCAP test data information were archival and were collected prior to this study. Benchmark data were entered into the AIMSweb system and a rate of improvement (ROI) was calculated when two or more scores for a student were compiled. AIMSweb's ROI is calculated in terms of the number of words read correctly per minute (WRC) gained per week and is estimated by calculating the difference between the first and last scores divided by the number of weeks in between (L. S. Fuchs et al., 1993; L. S. Fuchs & Shinn, 1989). Once the data were requested for this study from the school system, the school system collated the two sets of data (AIMSweb and TCAP) and assigned unique identification numbers for each case in order to protect the students' anonymity.

Fidelity of the CBM benchmarks. Prior to the first set of benchmarks each year, the school coordinator conducted training for new testers and provided a review of CBM procedures for those who had participated in benchmarks previously, including standardized directions and scoring, for each individual who would be administering R-CBM probes. The initial training lasted approximately 1.25 hours and involved practice in scoring using a scripted sample. At the time of the CBM administration, the school coordinator provided a script of all the directions that were needed during that session to ensure standardization of administration. At later benchmark testing periods, the school coordinator conducted training for any evaluator who had not yet received training and made reminder announcements before the first child was tested.

Procedure of student administration. Each evaluator was given a copy of the student's scoring sheets with the child's name at the top. Following a brief introduction, the child was provided the standardized instructions verbally and a copy of the student's version of the R-CBM, free of distracting marks, from which the child read. The administrator began the timer and cued the child to begin reading. The administrator marked on the child's scoring pages the errors in the moment and, when the timer beeped, cued the child to stop. The administrator then recorded the number of words read correctly over reading errors on that page. Two further R-CBMs were administered in the same manner; however, the standardized instructions were typically abbreviated at that point. The administrator or the test coordinator determined the median score for that administration by taking the median top (words read correctly) and the median bottom (words read incorrectly), and that set of scores was transferred to a separate page that was used to enter scores manually into AIMSweb.

Some administrations, especially during the first few years, were conducted outside the child's classroom by one or two testers until the entire class had been administered the test. In later years, depending on the school, administrations were conducted in the gymnasium or auditorium by several testers sitting at student desks with an extra chair across the desk with several feet in between, and students were prompted prior to entering the space to maintain a low reading volume during the administration.

TCAP administration and fidelity. The TCAP administration procedures that are required by the Tennessee State Department of Education were used. Students in third through eighth grades are evaluated over four consecutive school days within a two-week window permitted by the Tennessee State Department of Education, typically during the month of April each year. TCAP tests were timed and utilized standardized instructions and procedures for time and student-teacher interactions. The TCAP tests were also kept in a locked area by the school principal and/or building-level testing coordinator when TCAP testing was not in session.

Data Analysis

All data analyses were conducted using SPSS for Windows. The analyses were conducted on a secure password-protected computer.

Rates of improvement. Two methods of ROI were used in this study for comparison. One ROI calculated in this study was based on the winter and spring benchmark scores for each year using a simple linear slope. Winter and spring benchmark scores were chosen because Wiley and Deno (2005) found that fall scores are much lower than winter and spring scores, therefore indicating a possible false rate of improvement. A second ROI that was used was one similar to that used by AIMSweb, which was generated using a simple linear slope between the fall and spring scores (L. S. Fuchs et al., 1993; L. S. Fuchs & Shinn, 1989). These ROIs were used for comparison to determine if there is a difference among various methods of determining ROI and their degree of predictability with TCAP scores.

Prediction of TCAP with CBM. Although other researchers have chosen HLM to answer similar questions, multiple regression analysis was chosen as a desirable statistical procedure due to its ease of use and practicality for school-based practitioners. Multiple regression analysis examines the relationship between one dependent variable and several independent variables and develops an equation that is the best combination of independent variables for predicting the dependent variable. All assumptions of multiple regression analysis (linearity, independence of residuals, homogeneity of variance of residuals, normality of residuals, and no multicollinearity) were tested as part of the analyses.

Three stepwise multiple regression analyses were used in Study 1. Stepwise multiple regression indicates that the independent variables were entered in steps. The first analysis separately predicted the third-grade Reading/Language Arts TCAP (TCAP R/LA) score using third-grade spring R-CBM scores, third-grade winter-to-spring ROI, and the third-grade fall-to-spring (AIMSweb) ROI, entered stepwise into the regression analysis based on magnitude of zero-order and partial correlations. The second analysis separately predicted third-grade TCAP R/LA scores using the second-grade spring R-CBM scores, second-grade winter-to-spring ROI, and the second-grade fall-to-spring ROI, again entered stepwise into the regression analysis based on magnitude of zero-order and partial correlations. The third analysis predicted third-grade TCAP R/LA scores using the first-grade spring R-CBM scores, first-grade winter-to-spring ROI, and first-grade fall-to-spring ROI, entered stepwise into the regression analysis based on magnitude of zero-order and partial correlations.

Study 2

The purpose of this study was to cross-validate and check the utility of the prediction models gained from Study 1 upon a separate data set in order to determine if the results of this study could be generalized to other school systems, which also answers research question 9. Because there was such a large sample for comparison in the first study, the sample was randomly divided in half and the equation generated on the first half was cross-validated on the second half. Participants for this study were randomly selected using SPSS and were not included in the previous study.

Study 2 included the remaining 1216 students who took the third-grade statewide achievement test in 2009. Demographic information that was retained included age at the time of the 2009 statewide achievement test, gender, ethnicity, socioeconomic status as measured by participation in the free and reduced lunch program, and special education and Section 504 certification. All other identifying information was removed from the data.

Data Analysis

The prediction models that were generated in Study 1 were run on SPSS using the assigned data set to determine if the regression coefficients obtained in the first set of analyses were stable. Stability was determined by using the equations from Study 1 that were found to have the greatest predictability. The equations were then used on the half of the sample that was not used in Study 1. Data analysis examined the accuracy of determining percentage of positive predictions—those predicted to pass and did pass—and percentage of negatives—those predicted to fail and who did fail. Finally, error in the predicted scores and the obtained scores was calculated by analyzing the relationship between the predicted TCAP scores and the obtained scores.

Study 3

The purpose of this study was to answer the research questions that applied to those children who completed the eighth-grade statewide achievement test in spring 2009. These research questions included the correlation models pertaining to the eighth-grade statewide achievement test using the third-, fourth-, and fifth-grade reading fluency measures and the subsequent rates of improvement.

Participants

Students in a moderate-sized Southeastern city from third through fifth grades were given CBM probes in reading fluency three times annually (September, January, and April) for six consecutive school years between the years of 2003 and 2009. Students from each grade level (1-5) were administered the R-CBMs during the first year in six elementary schools during the 2003-2004 school year. There were 304 who participated in the third-grade benchmark testing during the 2003-2004 school year. The use of CBM expanded each year, and during the 2008-2009 school year, CBM data were collected in 50 elementary schools. Students having IEPs were administered the R-CBM probes also, with the exception of students with severe cognitive or adaptive delays who behaviorally could not attempt the work.

The participants of this study were children who took the eighth-grade statewide achievement test in 2009 and who had complete R-CBM data from third, fourth, and fifth grades. Of the 304 students who began participating in the third-grade R-CBM benchmarks in 2003 and took the 2009 eighth-grade statewide achievement test, only 246 had complete data for all of the R-CBM benchmarks from third, fourth, and fifth grades. Therefore, due to the low number of children anticipated to be in this sample, a cross-validation was not recommended.

Demographic information that was retained included age at the time of the 2009 statewide achievement test, gender, ethnicity, socioeconomic status as measured by participation in the free and reduced lunch program, and special education and Section 504 certification. All other identifying information was removed from the data.

Measures

Curriculum-based measurements collected during the years of 2003 and 2006 were used, with the benchmark scores and the rates of improvement described in the previous study. The Reading/Language Arts subtest of the eighth-grade TCAP described in the previous study was also used.

Procedures

Benchmark data and TCAP test data information were archival and were collected prior to this study. The fidelity and administration were described in Study 1.

Data Analysis

All data analyses were conducted using SPSS. The rates of improvement that were used were described in Study 1, including the winter-to-spring benchmark scores' linear slope ROI and the fall-to-spring benchmark scores' linear slope ROI, similar to that generated by AIMSweb.

Three prediction models were estimated using sequential multiple regression. The first model separately predicted the eighth-grade Reading/Language Arts TCAP scores using the third-grade spring R-CBM scores, third-grade winter-to-spring ROI, and third-grade fall-to-spring ROI, entered stepwise into the regression analysis based on magnitude of zero-order and partial correlations. The second model predicted eighth-grade TCAP scores using the fourth-grade spring R-CBM scores, fourth-grade winter-to-spring ROI, and fourth-grade fall-to-spring ROI, again entered stepwise into the regression analysis based on magnitude of zero-order and partial correlations. The third model predicted eighth-grade TCAP scores using the fifth-grade spring R-CBM scores, fifth-grade winter-to-spring ROI, and fifth-grade fall-to-spring ROI, entered stepwise into the regression analysis based on magnitude of zero-order and partial correlations.

CHAPTER 4

RESULTS

Study 1

The purpose of Study 1 was to examine the oral reading fluency scores of students who took the third grade TCAP in 2009. The relationship between distal and proximal reading fluency scores with the third-grade TCAP scaled scores was examined. This study also tested the validity of two methods of calculating rate of improvement at each grade level and their effect on predicting TCAP scores.

Descriptive Analysis of Third-Grade Sample

There were 2,430 third-grade students who had complete data sets. An analysis of the descriptive statistics of all third-grade TCAP Reading/Language Arts (R/LA) scaled scores; oral reading fluency (ORF) scores from first, second, and third grades; and reading rates of improvement (ROI) was conducted. Descriptive statistics for these scores are presented in Table 1. The mean third-grade TCAP R/LA score suggested that more than half of the students scored above the proficient cut score of 495. When compared with the district's norms, the mean third-grade fall ORF score was above the 50th percentile of 87 words read correctly per minute (WRCM). The high standard deviation can be explained by the number of outliers at both the high and low end, as there were cases close to zero and also cases over 200 WRCM. The mean third-grade winter ORF score was above the district 50th percentile of 106 WRCM. The mean third-grade spring ORF score was above the district 50th percentile of 119 WRCM. These scores indicate that growth typically occurred throughout the third-grade year. The mean third-grade fall-to-spring ROI indicated an average rate of less than one word per week for 36 weeks, which is approaching the district 50th percentile of 0.9 WRCM increase per week. The mean first-grade winter-to-spring ROI also indicated an average rate of less than one word per week for 18 weeks.

Table 1

Descriptive Statistics for the Third-Graders' Scores at Third Grade (N = 2,430)

Third-Grade Scores	Minimum	Maximum	*M*	*SD*
TCAP R/LA Scaled Score	315.00	640.00	498.17	27.67
Fall ORF	0.00	247.00	93.67	40.05
Winter ORF	0.00	258.00	111.70	40.50
Spring ORF	2.00	276.00	125.14	42.20
Fall-to-Spring ROI	-.86	3.86	.87	.46
Winter-to-Spring ROI	-5.61	6.56	.75	.77

As presented in Table 2, the second-grade descriptive statistics indicated the mean second-grade fall ORF score was above the district 50th percentile of 66 WRCM. The mean second-grade winter ORF score was above the district 50th percentile of 87 WRCM. The mean second-grade spring ORF score was above the 50th percentile of 99 WRCM. The mean second-grade fall-to-spring ROI indicated an average rate of nearly one word per week for 36 weeks, which is above the district 50th percentile ROI norm of 0.9 WRCM increase per week. The mean first-grade winter-to-spring ROI also indicated an average rate of less than word per week for 18 weeks.

Table 2

Descriptive Statistics for the Third-Graders' Scores at Second Grade (N = 2,430)

Second-Grade Scores	Minimum	Maximum	*M*	*SD*
Fall ORF	0.00	218.00	72.67	36.50
Winter ORF	0.00	221.00	92.83	36.61
Spring ORF	2.00	221.00	106.30	37.99
Fall-to-Spring ROI	-1.47	3.56	.93	.43
Winter-to-Spring ROI	-5.00	6.44	.75	.69

The mean first-grade fall ORF score was well above the district 50th percentile of 12 WRCM. The mean first-grade winter ORF score was above the district 50th percentile of 34 WRCM. The mean first-grade spring ORF score was above the district 50th percentile of 58 WRCM. The mean first-grade fall-to-spring ROI indicated an average rate of over one word per week for 36 weeks, which is above the district 50th percentile norm of 1.3 WRCM increase per week. The mean first-grade winter-to-spring ROI also indicated an average rate of over one word per week for 18 weeks. Table 3 provides the descriptive statistics of all first-grade data.

Table 3

Descriptive Statistics for the Third-Graders' Scores at First Grade (N = 2,430)

First-Grade Scores	Minimum	Maximum	*M*	*SD*
Fall ORF	0.00	217.00	23.10	28.63
Winter ORF	0.00	230.00	47.51	35.34
Spring ORF	0.00	250.00	71.01	38.40
Fall-to-Spring ROI	-1.56	3.89	1.33	.66
Winter-to-Spring ROI	-5.61	6.56	.75	.77

A frequency analysis was also conducted on the entire third-grade sample. Of the 2,430 students who had complete data samples, 48.6% were female (n =1,181) and 51.4% were male (n = 1,249). Ethnicity percentages were similar to those of the school district, as reported on its website: 0.2% were American Indian/Alaskan Native (n = 6); 2.5% were Asian/Pacific

Islander (*n* = 61); 8.2% were Black (*n* = 199); 2.8% were Hispanic (*n* = 69); and 86.2% were White (*n* = 2,095).

The data set was split in order to both generate a multiple regression equation and to cross-validate the multiple regression equation, which is discussed in Study 2.

Descriptive Analysis of Third-Grade Sample After Split-Half Procedure

There were 1,214 third-grade students who were included in the first half from which the model of best fit was generated. An analysis of the descriptive statistics of the third-grade TCAP R/LA scaled scores; oral reading fluency scores from first, second, and third grades; and reading rates of improvement was conducted on this sample set. Descriptive statistics for these scores are presented in Table 4. The mean third-grade TCAP R/LA score suggested that more than half of the children scored above the proficient cut score of 495. The mean third-grade fall ORF score was above the district's 50th percentile of 87 WRCM. The mean third-grade winter ORF score was above the district's 50th percentile of 106 WRCM. The mean third-grade spring ORF score was above the 50th percentile of 119 WRCM. These scores indicated that growth typically occurred throughout the third-grade year. The mean third-grade fall-to-spring ROI indicated an average rate of less than one word per week for 36 weeks, which was approaching the district 50th percentile norm of 0.9 WRCM increase per week. The mean third-grade winter-to-spring ROI also indicated an average rate of less than one word per week for 18 weeks.

Table 4

Descriptive Statistics for the First Half of the Third-Graders' Scores at Third Grade (n = 1,214)

Third-Grade Scores	Minimum	Maximum	*M*	*SD*
TCAP R/LA Scaled Score	315.00	640.00	497.85	28.11
Fall ORF	1.00	247.00	92.97	40.95
Winter ORF	0.00	258.00	111.23	41.56
Spring ORF	2.00	276.00	124.24	43.49
Fall-to-Spring ROI	-.86	3.86	.87	.47
Winter-to-Spring ROI	-2.56	6.56	.72	.78

The descriptive statistics for the second-grade data included in Study 1 are listed in Table 5. The mean second-grade fall ORF score was above the district 50th percentile of 66 WRCM. The mean second-grade winter ORF score was above the district 50th percentile of 91 WRCM. The mean second-grade spring ORF score was above the district 50th percentile of 99 WRCM. The mean second-grade fall-to-spring ROI indicated an average rate of nearly one word per week for 36 weeks, which was above the district 50th percentile for ROI of 0.9 WRCM increase per week. The mean first-grade winter-to-spring ROI also indicated an average rate of less than one word per week for 18 weeks.

Table 5

Descriptive Statistics for the First Half of the Third Graders' Scores at Second Grade (n = 1,214)

Second-Grade Scores	Minimum	Maximum	*M*	*SD*
Fall ORF	0.00	218.00	71.71	37.01
Winter ORF	0.00	221.00	91.87	37.14
Spring ORF	2.00	221.00	105.41	38.57
Fall-to-Spring ROI	-1.47	2.78	.94	.44
Winter-to-Spring ROI	-3.06	6.11	.75	.69

The mean first-grade fall ORF score was well above the district 50th percentile of 12 WRCM. The mean first-grade winter ORF score was above the district 50th percentile of 34

WRCM. The mean first-grade spring ORF score was above the district 50th percentile of 58 WRCM. The mean first-grade fall-to-spring ROI indicated an average rate of over one word per week for 36 weeks, which is above the district 50th percentile norm of 1.3 WRCM increase per week. The mean first-grade winter-to-spring ROI also indicated an average rate of over one word per week for 18 weeks. The descriptive statistics for the first-grade data included in Study 1 are presented in Table 6.

Table 6

Descriptive Statistics for the First Half of the Third Graders' Scores at First Grade (n = 1,214)

First-Grade Scores	Minimum	Maximum	*M*	*SD*
Fall ORF	0.00	217.00	23.21	29.55
Winter ORF	0.00	230.00	46.82	35.96
Spring ORF	0.00	250.00	70.31	39.48
Fall-to-Spring ROI	-1.56	3.72	1.31	.66
Winter-to-Spring ROI	-3.28	5.17	1.31	.81

A frequency analysis was also conducted on the first half of the third-grade sample. Of the 1,214 children included in the first half, 47% were female (*n* = 570) and 5% were male (*n* = 644). Ethnicity percentages were similar to those of the school district, as reported on its website: 0.2% were American Indian/Alaskan Native (*n* = 3); 2.4% were Asian/Pacific Islander (*n* = 29); 7.5% were Black (*n* = 91); 2.5% were Hispanic (*n* = 30); and 87.4% were White (*n* = 1,061).

Multiple Regression Analyses

Three stepwise multiple regression analyses were utilized to determine if the ORF scores and ROIs at each grade level were predictive of third-grade TCAP R/LA scores. The criterion for all three analyses was the third-grade TCAP R/LA score.

A stepwise multiple regression analysis was conducted with third-grade TCAP R/LA scores as the criterion and third-grade spring ORF, third-grade fall-to-spring ROI, and third-grade winter-to-spring ROI as the predictors. Variables were entered based on the largest contribution of partial and zero-order correlations. The assumptions about residuals in multiple regression analysis were checked in each analysis using the histogram of standardized residuals, the normal probability plot, and the plot of residuals. The linearity of the predictors to the criterion in each analysis of TCAP R/LA scaled scores was checked using the plot of residuals. While there were large outliers, the vast majority of residuals fall within ± 2 standard values on both axes. The assumption of normality was checked using the histogram of standardized residuals and normal probability plot. The histogram was leptokurtic, but it was symmetrical and demonstrated a reasonably normal shape. Inspection of the normal probability plot verified that the assumption of normality was met. The assumption of independence was checked by observing the random display on the plot of residuals. Homogeneity of variance was checked by verifying that there was no shape in the cluster of standard values on the plot of residuals. The tolerance values of each of the predictors in the model were also checked and verified to be above .2, indicating that the assumption of no multicollinearity was reasonably met. The assumption of no multicollinearity was met, as indicated by the high tolerance values of the predictor variables, which ranged from .63 to .91.

Third-grade ORF was found to be a significant predictor of TCAP R/LA scores, $b^* = .74$, $t(1211) = 34.25$, $p < .001$. Additionally, third-grade fall-to-spring ROI was a significant predictor of TCAP R/LA scores, $b^* = -.18$, $t(1211) = -7.85$, $p < .001$. Third-grade winter-to-spring ROI was not found to be a significant predictor of TCAP R/LA. The first step in the regression indicated that the third-grade spring ORF score was the best predictor, explaining

47% of the third-grade TCAP R/LA score, *F*(1, 1212) = 1073.99, $p < .001$, *Mse* = 419.16. A second step added third-grade ROI for fall-to-spring to the equation. The fall-to-spring ROI explained an additional 3% of variability of third-grade TCAP R/LA scores, *F*(1, 1211) = 61.56, $p < .001$, *Mse* = 399.21. Thus, these two third-grade variables were able to explain 50% of third-grade TCAP R/LA scores. Table 7 summarizes the third-grade models generated by the multiple regression analyses.

Table 7

Multiple Regression Analysis of Third-Grade TCAP Reading/Language Arts with Third-Grade Oral Reading Fluency

Variable	R	R^2	Adjusted R^2	Std. Error of Estimate
Step 1	.69	.47	.47	20.47
Spring ORF				
Step 2	.70	.50	.50	19.98
Spring ORF				
Fall-to-Spring ROI				

A stepwise multiple regression analysis was conducted with third-grade TCAP R/LA scores as the criterion and second-grade spring ORF, second-grade fall-to-spring ROI, and second-grade winter-to-spring ROI as the predictors. Variables were entered based on the largest contribution of partial and zero-order correlations. The assumptions of linearity, independence of residuals, homogeneity of variance of residuals, normality of residuals, and no multicollinearity were checked. The linearity of the predictors to the criterion in each analysis of TCAP R/LA scaled scores was checked, and while there were large outliers, the vast majority of residuals fell within ± 2 standard values on both axes. The assumption of normality was checked using the histogram of standardized residuals, which was leptokurtic, but it was symmetrical and demonstrated a reasonably normal shape. The assumption of no multicollinearity was verified by the high tolerance values of the predictors in the model, which ranged from .69 to .93.

Second-grade ORF was found to be a significant predictor of TCAP R/LA scores, $b^* = .74$, $t(1210) = 34.21$, $p < .001$. Additionally, second-grade fall-to-spring ROI was a significant predictor of TCAP R/LA scores, $b^* = -.12$, $t(1210) = -4.65$, $p < .001$. Second-grade winter-to-spring ROI was also found to be significant, $b^* = -.06$, $t(1210) = -2.29$, $p = .022$. The first step of the regression indicated that second-grade spring ORF score was the best predictor, explaining 48% of the third-grade TCAP R/LA score, $F(1, 1212) = 1095.17$, $p < .001$, $Mse = 415.31$. The second step added second-grade fall-to-spring ROI to the equation. The fall-to-spring ROI explained an additional 2% of variability of third-grade TCAP R/LA scores, $F(1, 1211) = 45.32$, $p < .001$, $Mse = 400.66$. The third step added second-grade winter-to-spring ROI to the equation. The winter-to-spring ROI explained 0.2% of the variability in third-grade TCAP R/LA scores, $F(1, 1210) = 5.24$, $p = .022$, $Mse = 399.26$. Thus, second-grade variables were able to explain 50% of third-grade TCAP R/LA scores. Table 8 summarizes the second-grade models generated by the multiple regression analyses.

Table 8

Multiple Regression Analysis of Third-Grade TCAP Reading/Language Arts with Second-Grade Oral Reading Fluency

Variable	*R*	R^2	Adjusted R^2	Std. Error of Estimate
Step 1	.69	.48	.47	20.38
Spring ORF				
Step 2	.70	.49	.49	20.02
Spring ORF				
Fall-to-Spring ROI				
Step 3	.70	.50	.50	19.98
Spring ORF				
Fall-to-Spring ROI				
Winter-to-Spring ROI				

A stepwise multiple regression analysis was conducted with third-grade TCAP R/LA scores as the criterion and first-grade spring ORF, first-grade fall-to-spring ROI, and first-grade winter-to-spring ROI as the predictors. Variables were entered based on the largest contribution of partial and zero-order correlations. The assumptions of multiple regression were checked in the same manner as the previous two multiple regressions. The assumptions of linearity, independence of residuals, homogeneity of variance of residuals, normality of residuals, and no multicollinearity were checked. The linearity of the predictors to the criterion in each analysis of TCAP R/LA scaled scores was checked, and while there were large outliers, the vast majority of residuals fell within ± 2 standard values on both axes. The assumption of normality was checked using the histogram of standardized residuals, which was leptokurtic, but it was symmetrical and demonstrated a reasonably normal shape. The assumption of no multicollinearity was verified by the high tolerance values of the predictors in the model, which ranged from .56 to .83.

First-grade ORF was found to be a significant predictor of TCAP R/LA scores, $b^* = .66$, $t(1212) = 30.28$, $p < .001$. Neither of the first-grade ROI variables was significant. The first step of the regression indicated that first-grade spring ORF score was the greatest predictor, explaining 43% of the third-grade TCAP R/LA score, $F(1, 1212) = 916.55$, $p < .001$, *Mse* = 450.16. Thus, first-grade variables were able to explain 43% of third-grade TCAP R/LA scores. Table 9 summarizes the first-grade model generated by the multiple regression analyses.

Table 9

Multiple Regression Analysis of Third-Grade TCAP Reading/Language Arts with First-Grade Oral Reading Fluency

Variable	R	R^2	Adjusted R^2	Std. Error of Estimate
Spring ORF	.66	.43	.43	21.22

Study 2

The purpose of Study 2 was to validate the regression equation generated by Study 1. In Study 1, it was found that second-grade ORF scores, fall-to-spring ROI, and winter-to-spring ROI were significant predictors of TCAP R/LA scores and were no better or worse at predicting TCAP R/LA scores than the third-grade ORF or ROIs. Therefore, second-grade scores provide a more useful measure of predicting TCAP R/LA scores because they allow school officials to make programming decisions. Early identification of reading difficulties is important because research has demonstrated that the earlier that a child learns to read effectively, the better student that child becomes (Speece & Ritchey, 2005).

Descriptive Statistics

The following table (Table 10) presents the descriptive statistics for the second half of the sample of those who took the 2009 third-grade Reading/Language Arts TCAP. An analysis of the descriptive statistics for the second half of the third-grade sample was conducted (n = 1,216). The mean third-grade TCAP R/LA score suggested that more than half of the children scored above the proficient cut score. The mean third-grade fall ORF score was above the district 50th percentile norm of 87 WRCM. The mean third-grade winter ORF score was above the district 50th percentile of 106 WRCM. The mean third-grade spring ORF score was above the district 50th percentile of 119 WRCM. These scores indicated that growth typically occurred throughout the third-grade year. The mean third-grade fall-to-spring ROI indicated an average rate of less than one word per week for 36 weeks, which is approaching the district 50th percentile norm of 0.9 WRCM increase per week. The mean first-grade winter-to-spring ROI also indicated an average rate of less than one word per week for 18 weeks.

Table 10

Descriptive Statistics for the Second Half of the Third Graders' Scores at Third Grade (n = 1,216)

Third-Grade Scores	Minimum	Maximum	*M*	*SD*
TCAP Reading Scaled Score	315.00	640.00	498.49	27.24
Fall ORF	0.00	234.00	94.37	39.15
Winter ORF	5.00	254.00	112.17	39.43
Spring ORF	10.00	263.00	126.03	40.86
Fall-to-Spring ROI	-.50	3.31	.88	.44
Winter-to-Spring ROI	-5.61	6.22	.77	.77

The mean second-grade fall ORF score was above the district 50th percentile norm of 66 WRCM. Table 11 summarizes these descriptive statistics. The mean second-grade winter ORF score was above the district 50th percentile of 87 WRCM. The mean second-grade spring ORF score was above the district 50th percentile of 99 WRCM. The mean second-grade fall-to-spring ROI indicated an average rate of nearly one word per week for 36 weeks, which is above the district 50th percentile norm of 0.9 WRCM increase per week. The mean second-grade winter-to-spring ROI also indicated an average rate of less than one word per week for 18 weeks.

Table 11

Descriptive Statistics for the Second Half of Third-Graders' Scores at Second Grade (n = 1,216)

Second-Grade Scores	Minimum	Maximum	*M*	*SD*
Fall ORF	0.00	196.00	73.58	35.97
Winter ORF	2.00	204.00	93.78	36.06
Spring ORF	3.00	221.00	107.18	37.40
Fall-to-Spring ROI	-1.44	3.56	.93	.43
Winter-to-Spring ROI	-5.00	6.44	.74	.70

The mean first-grade fall ORF score was above the 50th percentile norm of 12 WRCM, as described in Table 12. The mean first-grade winter ORF score was above the 50th percentile of 34 WRCM. The mean first-grade spring ORF score was above the 50th percentile of 58 WRCM. The mean first-grade fall-to-spring ROI indicated an average rate of over one word per week for 36 weeks, which is above the district 50th percentile norm of 1.3 WRCM increase per week. The mean first-grade winter-to-spring ROI also indicated an average rate of over one word per week for 18 weeks.

Table 12

Descriptive Statistics for the Second-Half of the Third-Graders' Scores at First Grade (n = 1,216)

First-Grade Scores	Minimum	Maximum	*M*	*SD*
Fall ORF	0.00	202.00	22.98	27.69
Winter ORF	0.00	188.00	48.21	34.72
Spring ORF	3.00	201.00	71.71	37.30
Fall-to-Spring ROI	-.61	3.89	1.35	.65
Winter-to-Spring ROI	-1.22	4.44	1.31	.76

Prediction Model

The equation generated was Predicted TCAP Reading/Language Arts Score = 449.75 (constant) + (.54*second-grade spring ORF standardized coefficient beta weight score) + (-7.44*second-grade fall-to-spring ROI standardized coefficient beta weight score) +

(-2.29*second-grade winter-to-spring ROI standardized coefficient beta weight score). The equation was tested two ways: through calculation of the standard error of prediction and through checking accuracy of predictions. These were used to compare the concordance of students predicted to pass the TCAP R/LA with their actual scores. Upon examination of the correlation between the actual TCAP R/LA scores with the predicted TCAP R/LA scores, it was observed that there is a strong positive correlation, $r(1214) = .68$, $p < .001$.

When using the prediction equation, there is a better than 95% concordance that those students the prediction equation indicates will pass the TCAP R/LA will pass. As summarized in Table 13, there were 50 students in the second half of the sample who were predicted to score proficiently on the TCAP R/LA but actually were not proficient. This resulted in 4.1% of the sample that resulted in a false positive prediction. Additionally, there were five students who were predicted to be not proficient but scored as proficient, indicating that 0.4% of the total sample (n = 1,213) resulted in a false negative prediction. The prediction model has a high false positive, as all of those who were actually not proficient were predicted to be proficient, but a very low false negative prediction. There were only three cases in which a student was predicted to be not proficient and whose actual score was not proficient.

Table 13

Difference Between the Predicted and Actual TCAP Reading/Language Arts Scores of the Third-Grade Sample

TCAP Actual Rating	TCAP Predicted Rating			
	Not Proficient		Proficient	
	n	%	*n*	%
Not Proficient	3	0.2	50	4.1
Proficient	5	0.4	1157	95.2
Total	8	0.7	1207	99.3

In order to establish a cut score for second-grade spring ORF, a fluency score of 90 WRCM has been indicated to be the fluency goal by the end of second grade by Good et al. (2002), which is at the normative 40th percentile; scores below the 40th percentile are considered at risk. When using the prediction equation with the second-grade spring ORF scores to predict the students who will pass the TCAP R/LA, there is a 70.9% chance that those whose fluency score is at or above 90 WRCM would pass the TCAP R/LA, as described in Table 14. However, this also indicates that there were 29.1% false negatives, indicating that nearly 30 out of 100 students who were predicted to not pass the TCAP in third-grade would actually pass. Additionally, there were 13.2% false positives, indicating that those whose second-grade ORF was above 90 WRCM were predicted to pass the TCAP, but actually did not. In sum, it appears that the regression equation method of predicting TCAP proficiency was better than the use of a cut score to determine those who are expected to pass the TCAP.

Table 14

Difference Between the Predicted and Actual Second-Grade Spring Oral Reading Fluency and TCAP Reading/Language Arts Scores

	Spring R-CBM Score			
TCAP Rating	Below Benchmark		Above Benchmark	
	n	%	*n*	%
Not Proficient	46	86.8	7	13.2
Proficient	338	29.1	825	70.9
Total	385	27.8	832	68.4

Study 3

The purpose of Study 3 was to address the potential relationship between R-CBM oral reading fluency (ORF) scores collected as many as five years prior to the eighth-grade statewide achievement test in spring 2009. A model of relationship pertaining to the eighth-grade statewide achievement test using the third-, fourth-, and fifth-grade reading fluency measures and the subsequent rates of improvement (ROI) was generated. This study, as in Study 1, examined two forms of generating ROI based upon the fall and winter ORF scores compared with the spring ORF score.

Descriptive Analysis

There were 246 eighth-grade students with complete data sets. An analysis of the entire eighth-grade TCAP R/LA scaled scores, ORF scores from third-, fourth-, and fifth- grade students, and subsequent ROIs was conducted. Descriptive statistics for the eighth-graders' scores in third grade are presented in Table 15. The mean eighth-grade TCAP R/LA score suggested that most children scored above the proficient cut score of 495. The mean third-grade fall ORF score was above the district 40th percentile. The mean third-grade winter ORF score was above the district 40th percentile. The mean third-grade spring ORF score was above the district 40th percentile. The mean third-grade ROI between the fall and spring scores was above the district 10th percentile. The mean third-grade ROI between the winter and spring scores was very similar to the fall-to-spring rate.

Table 15

Descriptive Statistics for the Eighth Graders' Scores at Third-Grade (n = 246)

Third-Grade Scores	Minimum	Maximum	*M*	*SD*
Eighth-Grade TCAP R/LA Score	340.00	632.00	559.65	36.95
Fall ORF	2.00	176.00	83.61	36.48
Winter ORF	2.00	181.00	98.77	37.10
Spring ORF	7.00	196.00	113.20	40.39
Fall-to-Spring ROI	-2.08	2.28	.82	.44
Winter-to-Spring ROI	-4.61	2.39	.80	.70

As presented in Table 16, the mean fourth-grade fall ORF score was above the district 40th percentile. The mean fourth-grade winter ORF score was above the district 40th percentile.

The mean fourth-grade spring ORF spring score was above the district 50th percentile. The mean fourth-grade fall-to-spring ROI was above the district 10th percentile. The mean fourth-grade winter-to-spring ROI was lower than the fall-to-spring rate.

Table 16

Descriptive Statistics for the Eighth Graders' Scores at Fourth Grade (n = 246)

Fourth-Grade Scores	Minimum	Maximum	*M*	*SD*
Fall ORF	11.00	199.00	107.90	37.05
Winter ORF	13.00	208.00	122.94	38.26
Spring ORF	13.00	227.00	135.15	39.12
Fall-to-Spring ROI	-.56	2.89	.76	.41
Winter-to-Spring ROI	-1.44	7.17	.68	.79

The mean fifth-grade fall ORF score was above the district 50th percentile. The mean fifth-grade winter ORF score was above the district 50th percentile. The mean fifth-grade spring ORF score was above the district 40th percentile. The mean fifth-grade fall-to-spring ROI was below the district 10th percentile. The mean fifth-grade winter-to-spring ROI was higher than the fall-to-spring rate. Table 17 provides the descriptive statistics for the eighth graders' scores at fifth grade.

Table 17

Descriptive Statistics of the Eighth Graders' Scores at Fifth Grade (n = 246)

Fifth-Grade Scores	Minimum	Maximum	*M*	*SD*
Fall ORF	13.00	216.00	123.67	37.48
Winter ORF	21.00	226.00	139.02	39.43
Spring ORF	26.00	237.00	147.79	40.48
Fall-to-Spring ROI	-1.50	3.33	.49	.66
Winter-to-Spring ROI	-.42	2.81	.67	.42

A frequency analysis was also conducted on the eighth-grade sample. Of the 246 subjects in the sample, there were 120 girls (48.8%) and 126 boys (51.2%). Additionally, 91.5% were Caucasian, 1.2% were Hispanic, 6.5% were Black, and .8% were Asian/Pacific Islander.

Multiple Regression Analysis

Stepwise multiple regression analyses were utilized. The assumptions about residuals in regression analysis were addressed in the same manner with which they were addressed in Study 1. Three stepwise multiple regression analyses were utilized to determine if the ORF scores at each grade level were predictive of eighth-grade TCAP R/LA scores. The criterion for all three analyses was the eighth-grade TCAP R/LA score.

A stepwise multiple regression analysis was conducted with eighth-grade spring TCAP R/LA scores as the criterion and third-grade spring ORF, third-grade fall-to-spring ROI, and third-grade winter-to-spring ROI as the predictors. Variables were entered based on the largest contribution of partial and zero-order correlations. The assumptions about the residuals in multiple regression analysis were checked, as described in Study 1. The assumptions of multiple regression were checked in the same manner as the previous two multiple regressions were checked. The assumptions of linearity, independence of residuals, homogeneity of variance of residuals, normality of residuals, and no multicollinearity were checked. The linearity of the predictors to the criterion in each analysis of TCAP R/LA scaled scores was checked, and while there were large outliers, the vast majority of residuals fell within ± 2 standard values on both axes. The assumption of normality was checked using the histogram of standardized residuals, which was leptokurtic, but it was symmetrical and demonstrated a reasonably normal shape. The assumption of no multicollinearity was met, as indicated by the high tolerance values of the predictor variables, which ranged from .82 to .84.

Third-grade ORF was found to be a significant predictor of eighth-grade TCAP R/LA scores, $b^* = .51$, $t(244) = 9.22$, $p < .001$. The first step of the regression indicated that third-grade spring ORF score was the greatest predictor, explaining 26% of the eighth-grade TCAP

R/LA score, *F*(1, 244) = *85.02*, $p < .001$, *Mse* = 1016.71. The addition of either ROI value was not significant. Table 18 summarizes the third-grade model generated by the multiple regression analyses.

Table 18

Multiple Regression Analysis of Eighth-Grade TCAP Reading/Language Arts with Third-Grade Oral Reading Fluency

Variable	R	R^2	Adjusted R^2	Std. Error of Estimate
Third-Grade Spring ORF	.51	.26	.26	31.89

A stepwise multiple regression analysis was conducted with eighth-grade spring TCAP R/LA scores as the criterion and fourth-grade spring ORF, fourth-grade fall-to-spring ROI, and fourth-grade winter-to-spring ROI as the predictors, the results of which are summarized in Table 19. Variables were entered based on the largest contribution of partial and zero-order correlations. The assumptions of multiple regression analysis were checked, as described in Study 1. The assumptions of linearity, independence of residuals, homogeneity of variance of residuals, normality of residuals, and no multicollinearity were checked. The linearity of the predictors to the criterion in each analysis of TCAP R/LA scaled scores was checked, and while there were large outliers, the vast majority of residuals fell within ± 2 standard values on both axes. The assumption of normality was checked using the histogram of standardized residuals, which was leptokurtic, but it was symmetrical and demonstrated a reasonably normal shape. The assumption of no multicollinearity was met, as indicated by the high tolerance values of the predictor variables, which ranged from .89 to .94.

Fourth-grade ORF was found to be a significant predictor of eighth-grade TCAP R/LA scores, $b^* = .53$, $t(244) = 9.77$, $p < .001$. The first step of the regression indicated that fourth-grade spring ORF was the greatest predictor, explaining 28% of the eighth-grade TCAP R/LA score, $F(1, 244) = 95.50$, $p < .001$, *Mse* = 985.32. The addition of the fourth-grade ROI variables was not significant.

Table 19

Multiple Regression Analysis of Eighth-Grade TCAP Reading/Language Arts with Fourth-Grade Oral Reading Fluency

Variable	R	R^2	Adjusted R^2	Std. Error of Estimate
Fourth-Grade Spring ORF	.53	.28	.28	31.39

A stepwise multiple regression analysis was conducted with eighth-grade spring TCAP R/LA scores as the criterion and fifth-grade spring ORF, fifth-grade fall-to-spring ROI, and fifth-grade winter-to-spring ROI as the predictors, the results of which are summarized in Table 20. Variables were entered based on the largest contribution of partial and zero-order correlations. The assumptions of multiple regression analysis were checked, as described in Study 1. The assumption of no multicollinearity was met, as indicated by the high tolerance values of the predictor variables, which ranged from .86 to .95.

Fifth-grade ORF was found to be a significant predictor of eighth-grade TCAP R/LA scores, $b^* = .56$, $t(244) = 10.67$, $p < .001$. The first step of the regression indicated that fifth-grade spring ORF was the greatest predictor, explaining 32% of the eighth-grade TCAP R/LA score, $F(1, 244) = 113.79$, $p < .001$, *Mse* = 934.98. The addition of ROI variables was not significant.

Table 20

Multiple Regression Analysis of Eighth-Grade TCAP Reading/Language Arts with Fifth-Grade Oral Reading Fluency

Variable	R	R^2	Adjusted R^2	Std. Error of Estimate
Fifth-Grade Spring ORF	.56	.32	.32	30.58

CHAPTER 5

DISCUSSION

The purpose of this research was to examine the predictive value of oral reading fluency upon statewide achievement test scores, primarily in reading. Researchers have increasingly examined aspects of this question since the 2001 NCLB legislation. Researchers have established that there is fairly strong predictive validity of curriculum-based measures. However, few research articles have examined how a student's rate of improvement in oral reading fluency may impact the predictive value upon statewide achievement scores. School administrators, teachers, and school psychologists need to have more convenient means of identifying students who are at risk for reading problems as well as effective means of preventing further academic problems. Reading fluency has been identified as a simple method of monitoring the progress of students who receive reading interventions; these fluency assessments can also be used to identify and monitor at-risk students. Although reading fluency is easy to assess, predicting success on statewide achievement tests is not as easy. Three studies were conducted to examine the long-term predictive value of oral reading fluency upon statewide achievement test scores in reading at third and eighth grades.

In Study 1, the results indicate that of the effects of each grade-level variable (spring oral reading fluency score, fall-to-spring reading rate of improvement, and winter-to-spring reading rate of improvement), the ones that demonstrated the most significant effects on statewide achievement test scores in reading in third grade were the second-grade spring oral reading fluency scores, with a close second being the third-grade spring oral reading fluency scores. Reading rates of improvement at second and third grade did contribute a significant amount of variance to the prediction models generated, primarily the fall-to-spring rates of improvement, though these contributed only 2-3% of the variance. The first-grade reading rates of improvement did not significantly contribute to the prediction of statewide achievement test scores in reading. These results are consistent with the findings of Stage and Jacobsen (2001), whose research indicated that a single score was a better predictor of achievement test scores than the rate of improvement.

Additionally, the results of Study 1 indicate that while the winter-to-spring reading rate of improvement in second and third grades did significantly contribute to the prediction of the statewide achievement test score in third grade, it contributed much less variance than the spring oral reading fluency scores in first, second, and third grades, which ranged from 47-50%. This indicates that a single score collected at the end of the grade level has much more predictive accuracy in helping schools make decisions about student programming for the next school year than does the numerical growth value attributed to the student's reading improvements made in the course of a school year. However, because the contribution of reading rate of improvement was significant, its addition may indicate an increase in accuracy in predicting future scores, even as early as the end of second grade.

Study 2 examined the value of the prediction model generated in Study 1. The prediction model had a tendency to generate a relatively high number of false positives of TCAP proficiency. Additionally, the prediction model resulted in very few false negatives, resulting in only two students who were predicted to be not proficient, yet who scored as proficient. The prediction model also was used to determine the strength of using oral reading fluency scores in the spring of second grade, one year before the student takes the third-grade TCAP, which would be extremely useful for school administrators in making programming decisions and providing reading interventions that can improve scores needed for AYP. Based on whether or not a student scored above the at-risk range of 90 words correct per minute at second-grade level (Good et al., 2002), there were approximately 13% who were not proficient on the TCAP (false positive); however, of the students who scored below the 40th percentile, 29% actually scored as proficient on the TCAP (false negative). Schools can use the quick measure of oral reading fluency as a means of identifying students who typically might not be identified for interventions and support through a response-to-intervention model, and create a program for these students with plenty of time to improve the outcome of the third-grade TCAP.

The purpose of Study 3 was to examine the extended predictive effects that reading fluency can have on statewide achievement test scores. Using the eighth-grade statewide achievement test score as the object of three analyses, Study 3 indicated that end-of-year oral reading fluency scores at third, fourth, and fifth grades significantly predicted eighth-grade test scores, contributing between 26-32% of the variance. The predictive value did increase with proximity to eighth grade. However, until further research can be conducted regarding this upward extension, generalization is not warranted. For example, Rasinski et al. (2006) found that over half of their ninth-grade sample had a reading fluency score below the eighth-grade 25th percentile.

Several findings in the current study are of interest. First, the current findings add to the literature indicating that oral reading fluency is a valid predictor of high-stakes statewide tests (Barger, 2003; Crawford et al., 2001; Goffreda et al., 2009; Hosp et al., 2011; McGlinchey &

Hixson, 2004; Merino & Beckman, 2010; Roehrig et al., 2008; Shapiro et al., 2006; Shaw & Shaw, 2002; Silberglitt et al., 2006; Silberglitt & Hintze, 2005; Vander Meer et al., 2005; Wilson, 2005). This finding adds to the ability of schools to use existing, readily available data to assign resources to students who need them in enough time to make a difference to the students and school-wide reading goals (Baker et al., 2008; Chard et al., 2008; Vander Meer et al., 2005). Also, the results of Study 2 indicate that the regression formula developed from a correlation can be used to make future predictions, which further validates the utility of the findings. Because school psychologists are often responsible for conducting research, especially in smaller districts, their ability to make effective, research-based decisions at the local level will help them to make recommendations to teachers and school administrators about those students who require intervention and those who do not. Prevention of failure is paramount to meeting AYP goals.

When comparing two methods of predicting students who would be proficient on the statewide achievement test, the simpler method of using a cut score was not preferred; the regression equation proved to be a much better predictor, with fewer errors. This is unfortunate, as it may take more effort for school psychologists to generate such an equation to predict those who would be proficient. Additionally, neither of the prediction methods was very accurate in identifying those who were expected to perform poorly on the statewide achievement test. This may have been due to the low number of students in the sample whose scores were below proficient. Further investigation of factors that lead to poor achievement may be warranted in future research.

Similar to the proposed reasons for using oral reading fluency and reading rate of improvement in predicting statewide achievement test scores well in advance, the findings of

Study 3 indicate that oral reading fluency data can be used to predict eighth-grade statewide achievement tests. This is the first study to document the extended utility of the use of initial screeners in longitudinal regression analysis. However, additional research is needed to replicate the findings and to validate this upward extension of the abundant research present for elementary students.

The utility of conducting research on predicting failure in higher grades can help to address other aspects of meeting AYP goals. For example, a possible outcome for students who score poorly on middle-school statewide achievement tests is increased risk of dropping out of high school, which thus reduces a school's graduation rate. Therefore, being able to identify risk factors easily through annual screening tools, such as CBM, can help school systems to identify those who are at greater risk of failure earlier and provide preventative measures to prevent failure and improve graduation rates. The findings of Study 3 demonstrate that there is a significant correlation between statewide achievement tests in reading and oral reading fluency as much as five years prior to the administration of that test. This correlation was significant, such that if, for example, a series of interventions were to be implemented in sixth grade with students with poor reading fluency scores, there might be improvement in the students' eighth grade achievement test scores. Of course, it is probable that factors other than oral reading fluency (e.g., behavior, motivation, and demographic characteristics) impact poor performance on statewide achievement tests, as was described by Chard et al. (2008) and Roehrig et al. (2008). Examination of these factors would be warranted.

Another finding of interest was the lack of prediction from the rate of improvement values. This finding is of particular interest, given some support from previous research on the potential utility of rate of progress in prediction (Baker et al., 2008; Chard et al., 2008). One reason for this may have been the limited number of students who did not pass the statewide achievement test. Given the small percentage of non-proficient students, there were insufficient data to predict correctly those who were not proficient. Also, with students who have passed, research indicates less growth at the upper end of the reading spectrum (Silberglitt & Hintze, 2007). It is likely that if there were fewer students passing, then there may have been greater variability in the scores, which might have allowed for greater accuracy in creating a useful prediction model for early identification of students at risk for failure. However, some of the previous research had the same difficulty with few students falling in the below proficient range (Barger, 2003; Shaw & Shaw, 2002). Additionally, the question of comparing those who score below proficient on statewide achievement tests is difficult to answer because each state sets its own proficiency levels (Bandeira de Mello et al., 2009).

Overall, there continues to be evidence that practitioners should utilize reading fluency scores to make programming decisions for students at risk for failure. Additional research with a greater number of students, especially those in the higher grades, is needed before dismissing the possibility of early identification and prevention of academic problems in middle school. Additionally, more examination of predicting statewide achievement test scores could prove useful, especially if a school sets a goal higher than minimal proficiency.

Limitations

The benefit for school administrators in being able to predict with a high level of accuracy those who are at risk for scoring poorly on a high-stakes test is extremely useful in preventing the consequences of inadequate yearly progress. The results here suggest that an intervention for at-risk students could be developed and implemented for nearly an entire year prior to the administration of the TCAP. While there is the risk of false negatives, there is little argument that providing support to struggling readers will only help them improve. There were very few students who were identified as at risk for not passing and yet did pass. The problem lies in the number of students whom the model did not predict to perform below proficiency (false positives). Future research may need to more closely examine which oral reading fluency percentile cut scores provide better accuracy.

One of the purposes of this study was to perform these analyses using statistical modeling procedures that are more readily available to school district personnel. This may have been a limitation to developing an accurate model for predicting students who are at risk for non-proficiency on high-stakes tests. Other studies that have examined the relationship between similar variables have utilized hierarchical linear modeling (Silberglitt & Hintze, 2007), monotonic linear modeling (Christ, 2006), and binary logistic regression with ROC curves (Goffreda et al., 2009). However, stepwise multiple regression was utilized and, although the order of inclusion of the predictors was not determined by the researcher, as Hintze and Silberglitt (2005), Stage and Jacobsen (2001), and Wiley and Deno (2005) established, the single ORF score closest to the time of the TCAP administration was found to be the largest single predictor in each regression analysis.

Additionally, there is evidence from Study 3 indicating that reading rate in elementary school (third, fourth, and fifth grades) can significantly predict proficiency on statewide tests in eighth grade; however, future studies will need to confirm this prediction by testing a prediction equation for eighth-grade scores. A larger sample of longitudinal data would be needed in order to perform such a study. However, it is possible that rate of improvement may not be as useful in these grades due to the deceleration of reading rate over time and because the nature of reading changes as the student gets older (Rasinski et al., 2006).

Implications

School districts need a way to improve the scores on high-stakes testing for all children in order to meet the guidelines for AYP. The current study demonstrates that there are simple, easily available, cost-effective methods for collecting data that can be used to identify students who are at risk for failing to achieve proficiency on statewide tests. The predominance of students who did not achieve adequate levels were appropriately identified a year in advance of the statewide test in third grade.

The current studies also provided preliminary evidence to suggest that oral reading fluency may be strongly related to proficiency on statewide tests as distant as five years. Previous research has indicated that identification and intervention in earlier grades are strongly recommended and are greatly linked to future success in all academic subjects (Juel, 1988). Juel's (1988) conclusion continues to be supported by the findings of the current study, as oral reading fluency in third through fifth grades accounted for 25-30 % of the variance in statewide achievement test scores in eighth grade. However, a student's rate of improvement did not provide the sensitivity needed to contribute to such a model at any of these grades.

These studies did demonstrate the significant contributions that oral reading fluency makes upon performance on statewide achievement test scores in advance of the test administrations. This information will prove to be useful for school administrators and school psychologists who are looking for easily accessible means of identifying students who may not pass the statewide achievement tests. These studies provide strong evidence for the position of allocating educational resources toward improving reading fluency skills among all students as this may prove to be beneficial for improving long-term student achievement and success.

REFERENCES

Aaron, P. G. (1997). The impending demise of the discrepancy formula. *Review of Educational Research, 67,* 461-502. doi:10.3102/00346543067004461

Aaron, P. G., & Joshi, R. M. (1992). *Reading problems: Consultation and remediation*. New York, NY: Guilford Press.

Ardoin, S. P., Witt, J. C., Connell, J. E., & Koenig, J. L. (2005). Application of a three-tiered response to intervention model for instructional planning, decision making, and the identification of children in need of services. *Journal of Psychoeducational Assessment, 23,* 362-380. doi:10.1177/073428290502300405

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Beck, C. T. (2008). Reading fluency as a predictor of reading proficiency in low-performing, high-poverty schools. *School Psychology Review, 37,* 18-37. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=37&issue=1

Bandeira de Mello, V., Blankenship, C., & McLaughlin, D. H. (2009). *Mapping state proficiency standards onto NAEP scales: 2005-2007* (NCES 2010-456). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2010456

Barger, J. (2003). *Comparing the DIBELS oral reading fluency indicator and the North Carolina end of grade reading assessment.* Asheville, NC: North Carolina Teacher Academy. Retrieved from https://dibels.uoregon.edu/techreports/NC_Tech_Report.pdf

Center on Education Policy. (2008). *Instructional time in elementary schools: A closer look at changes for specific subjects.* Retrieved on from http://www.cep-dc.org/cfcontent_file.cfm?Attachment= McMurrer%5FReport%5FInstructionalTime%5F022008%2Epdf

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. doi:10.1177/0022219407313588

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,* 128-133. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=35&issue=1

Christ, T. J., Burns, M. K., & Ysseldyke, J. E. (2005). Conceptual confusion within response-to-intervention vernacular: Clarifying meaningful differences. *Communique, 34*(3), Bethesda, MD: National Association of School Psychologists. Retrieved from http://www.nasponline.org/publications/cq/cqindexlist.aspx

Crawford, L., Tindal, G. & Stieber, S. (2001). Using oral reading rate to predict student performance on statewide achievement tests. *Educational Assessment, 7,* 303-323. doi:10.1207/S15326977EA0704_04

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232. Retrieved from http://www.cec.sped.org/Content/NavigationMenu/Publications2/ExceptionalChildren/default.htm

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. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=30&issue=4

Edformation. (2005). AIMSweb benchmark probes. Retrieved from http://www.aimsweb.com/

Fletcher, J. M., Coulter, W. A., Reschly, D. J., & Vaughn, S. (2004). Alternative approaches to the definition and identification of learning disabilities: Some questions and answers. *Annals of Dyslexia, 54,* 304-331. doi:10.1007/s11881-004-0015-y

Fletcher, J. M., & Vaughn, S. (2009). Response to intervention: Preventing and remediating academic difficulties. *Child Development Perspectives, 3,* 30-37. doi:10.1111/j.1750-8606.2008.00072.x

Fuchs, D., & Fuchs, L. S. (2006) Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41,* 93-99. doi:10.1598/RRQ.41.1.4

Fuchs, D., Fuchs, L. S., & Compton, D. L. (2004). Identifying reading disabilities by responsiveness-to-instruction: Specifying measures and criteria. *Learning Disability Quarterly, 27,* 216-227. Retrieved from http://www.cldinternational.org/Publications/LDQ.asp

Fuchs, D., Mock, D., Morgan, P. L., & Young, C. L. (2004). Responsiveness to intervention: Definitions, evidence, and implications for the learning disabilities construct. *Learning Disabilities Research & Practice, 18,* 157-171. doi:10.1111/1540-5826.00072

Fuchs, L. S., & Fuchs, D. (1997). Use of curriculum-based measurement in identifying students with disabilities. *Focus on Exceptional Children, 30*, 1-15. doi:10.1002/pits.20113

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. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=22&issue=1

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading, 5,* 239-256. doi:10.1207/S1532799XSSR0503_3

Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disability Quarterly, 25,* 33–46. Retrieved from http://www.jstor.org/stable/1511189

Fuchs, L. S., & Shinn, M. R. (1989). Writing CBM IEP objectives. In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 131–152). New York, NY: Guildford Press. Retrieved from https://public.me.com/markshinn

Goffreda, C. T., Diperna, J. C., & Pedersen, J. A. (2009). Preventive screening for early readers: Predictive validity of the dynamic indicators of basic early literacy skills (DIBELS). *Psychology in the Schools, 46,* 539-552. doi:10.1002/pits.20396

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

Good, R. H., Kaminski, R. A., & Dill, S. (2002). DIBELS oral reading fluency. In R. H. Good & R. A. Kaminski (Eds.), *Dynamic indicators of basic early literacy skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Retrieved on from http://dibels.uoregon.edu/

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,* 257-288. doi:10.1207/S1532799XSSR0503_4

Hale, J., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., . . . Yalof, J. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: An expert white paper consensus. *Learning Disability Quarterly, 33,* 223-236. Retrieved from http://www.cldinternational.org/Publications/LDQ.asp

Hale, J. B., Kaufman, A., Naglieri, J. A., & Kavale, K. A. (2006). Implementation of IDEA: Integrating response to intervention and cognitive assessment methods. *Psychology in the Schools, 43,* 753-770. doi:10.1002/pits.20186

Hasbrouck, J. E., & Tindal, G. (1992). Curriculum-based oral reading fluency norms for students in grades 2 through 5. *Teaching Exceptional Children, 24,* 41-44. Retrieved from http://journals.cec.sped.org/tec/

Hintze, J. M., & Christ, T. J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review, 33*, 204-217. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=33&issue=2

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. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=34&issue=3

Hosp, J. L., Hosp, M. K., & Dole, J. K. (2011). Potential bias in predictive validity of universal screening measures across disaggregation subgroups. *School Psychology Review, 40,* 108-131. Retrieved from http://www.nasponline.org/publications/spr/spr401index.aspx

Individuals with Disabilities Education Improvement Act, Pub. L. No. 108-466. 118 Stat. 2647 (2004).

Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*, 719-729. doi:10.1037/0022-0663.95.4.719

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80,* 437-447. doi:10.1037/0022-0663.80.4.437

Kober, N. (2007). *Why we still need public schools: Public education for the common good.* Retrieved from http://www.cep-dc.org/cfcontent_file.cfm?Attachment= Kober_Report_WhyWeStillNeedPublicSchools_010107.pdf

Lee, J. (2006). *Tracking achievement gaps and assessing the impact of NCLB on the gaps: An in-depth look into national and state reading and math outcome trends.* Cambridge, MA: The Civil Rights Project at Harvard University. Retrieved from http://civilrightsproject.ucla.edu/research/k-12-education/integration-and-diversity/tracking-achievement-gaps-and-assessing-the-impact-of-nclb-on-the-gaps/lee-tracking-achievement-gaps-2006.pdf

Linn, R. L. (2005). Conflicting demands of No Child Left Behind and state systems: Mixed messages about school performance. *Education Policy Analysis Archives, 13*(33). Retrieved from http://epaa.asu.edu/epaa/v13n33/

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, NY: Guilford Press.

Marston, D. (2005). Tiers of intervention in responsiveness to intervention: Prevention outcomes and learning disabilities identification patterns. *Journal of Learning Disabilities*, *38,* 539-544. doi:10.1177/00222194050380061001

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33,* 193-203. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=33&issue=2

McMaster, K. L., & Wagner, D. (2007). Monitoring response to general education instruction. In S. R. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *The handbook of response to intervention: The science and practice of assessment and intervention* (pp. 223-233). New York, NY: Spring Science & Business Media. Retrieved from http://www.springerlink.com/content/978-0-387-49052-6

Mellard, D. (2004). *Understanding responsiveness to intervention in learning disabilities determination.* Retrieved from National Research Center on Learning Disabilities website: http://www.nrcld.org/about/publications/papers/mellard.html

Merino, K., & Beckman, T. O., (2010) Using reading curriculum-based measurements as predictors for the measure of academic progress (MAP) standardized test in Nebraska. *International Journal of Psychology: A Biopsychosocial Approach, 6,* 85-98. Retrieved from http://www.psyjournal.vdu.lt/uploads/ Main/2010_6%20Using%20Reading%20Curriculum-based%20Measurements.pdf

National Reading Panel. (2000). *Teaching children to read.* Washington, DC: U.S. Department of Health and Human Services. Retrieved from http://www.nichd.nih.gov/publications/nrp/upload/report.pdf

Nelson, J. M, & Machek, G. R. (2007). A survey of training, practice, and competence in reading assessment and intervention. *School Psychology Review, 36,* 311-327. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=36&issue=2

No Child Left Behind Act, Pub. L. No. 107-110. H.R. 1, 115 Stat. 1425 (2001).

President's Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and their families.* Washington, DC: U. S. Department of Education Office of Special Education and Rehabilitative Services,. Retrieved from http://www2.ed.gov/inits/commissionsboards/whspecialeducation/reports/ images/Pres_Rep.pdf

Rasinski, T. V., Padak, N. D., McKeon, C. A., Wilfong, L. G., Friedauer, J. A., & Heim, P. (2006). Is reading fluency a key for successful high school learning? *Journal of Adolescent and Adult Literacy, 49,* 22-27. doi:10.1598/JAAL.49.1.3

Reeves, S., Bishop, J., & Filce, H. G. (2010). Response to intervention and tier systems: Questions remain as educators make challenging decisions. *The Delta Kappa Gamma Bulletin, 76,* 31-35. Retrieved from http://www.dkg.org/

Reschly, D. J. (2002). Change dynamics in special education assessment: Historical and contemporary patterns. *Peabody Journal of Education, 77,* 117-136. doi:10.1207/S15327930PJE7702_9

Reschly, D. J., & Wilson, M. S. (1995). School psychology faculty and practitioners: 1986 to 1991 trends in demographic characteristics, roles, satisfaction, and system reform. *School Psychology Review, 24,* 62-80. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=24&issue=1

Restori, A. F., Katz, G. S., & Lee, H. B. (2009). A critique of the IQ/achievement discrepancy model for identifying specific learning disabilities. *Europe's Journal of Psychology, 4*, 128-145. Retrieved from http://www.ejop.org/images/11%202009/9.%20A%20Critique%20of%20the%20IQ%20%20Achievement%20Discrepancy%20Model%20.pdf

Reynolds, C. R., & Shaywitz, S. E. (2009). Response to intervention: Prevention and remediation, perhaps. Diagnosis, no. *Child Development Perspectives, 3,* 44-47. doi:10.1111/j.1750-8606.2008.00075.x

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS oral reading fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46,* 343-366. doi:10.1016/j.jsp.2007.06.006

Sanders, W. L., & Horn, S. P. (1994). The Tennessee value-added assessment system (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education, 8,* 299-311. doi:10.1007/BF00973726

SAS Institute. (2007). SAS© EVAAS© for K-12. Retrieved from http://www.sas.com/govedu/edu/k12/evaas/index.html

Schulte, A. C., Villwock, D. N., Whichard, S. M., & Stallings, C. F. (2001). High stakes testing and expected progress standards for students with learning disabilities: A five-year study of one district. *School Psychology Review, 30,* 487-506. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=30&issue=4

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention*. New York, NY: Guilford Press.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests. *Journal of Psychoeducational Assessment, 24,* 19-35. doi:10.1177/0734282905285237

Shaw, R., & Shaw, D. (2002). *DIBELS oral reading fluency-based indicators of third-grade reading skills for Colorado State Assessment Program (CSAP).* Eugene, OR: University of Oregon. Retrieved from https://dibels.uoregon.edu/techreports/shaw_csap_technical_report.pdf

Shin, J., Espin, C. A., Deno, S. L., & McConnell, S. (2004). Use of hierarchical linear modeling and curriculum-based measurement for assessing academic growth and instructional factors for students with learning difficulties. *Asia Pacific Education Review, 5,* 136-148. doi:10.1007/BF03024951

Shinn, M. M., & Shinn, M. R. (2002). *AIMSweb training workbook: Administration and scoring of reading curriculum-based measurement for use in general outcome measurement* (pp. 35-42). Bloomington, MN: NCS Pearson. Retrieved from http://www.aimsweb.com/uploads/pdfs/scoring_rcbm.pdf

Silberglitt, B., Burns, M. K., Madyun, N. H., & Lail, K. E. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43,* 527-535. doi:10.1002/pits.20175

Silberglitt, B., & Hintze, J. M. (2005). Formative assessment using CBM-R cut scores to track progress toward success on state-mandated achievement tests: A comparison of methods. *Journal of Psychoeducational Assessment, 23,* 304-325. doi:10.1177/073428290502300402

Silberglitt, B., & Hintze, J. M. (2007). How much growth can we expect? A conditional analysis of R-CBM growth rates by level of performance. *Exceptional Children, 74,* 71-84. Retrieved from http://cec.metapress.com/content/l6514707h0565964/?p=8c1c109776304c3494d0473fbf4659dc&pi=3

Speece, D. L., & Ritchey, K. D. (2005). A longitudinal study of the development of oral reading fluency in young children at-risk for reading failure. *Journal of Learning Disabilities, 38,* 387-399. doi:10.1177/00222194050380050201

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,* 407-419. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=30&issue=3

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *22,* 360-406. doi:10.1598/RRQ.21.4.1

Stecker, P. M., Fuchs, L. S., & Fuchs, D. (2005). Using curriculum-based measurement to improve student achievement: Review of research. *Psychology in the Schools, 42,* 795-819. doi:10.1002/pits.20113

Tennessee Department of Education. (2007). *Assessment, evaluation, and research division.* Retrieved from http://www.state.tn.us/education/assessment/

Tennessee Value-Added Assessment System. (2006). Interpreting the 2006 Tennessee Trend Analyses. Retrieved from http://tvaas.sas.com/trend/help/Trend_Help_Main.htm

Thattai, D. (2001). A history of public education in the United States. *Journal of Literacy and Education in Developing Societies,1*(2). Retrieved from http://www.servintfree.net/~aidmn-ejournal/

U.S. Department of Education. (2000). *To assure the free appropriate public education of all children with disabilities: Twenty-second annual report to congress on the implementation of the Individuals with Disabilities Education Act.* Washington, DC: Office of Special Education Programs. Retrieved from http://www2.ed.gov/about/reports/annual/osep/2000/preface.pdf

U.S. Office of Education. (1977). Assistance to states for education of handicapped children: Procedures for evaluating specific learning disabilities. *Federal Register, 42*, 65082-65085.

Vander Meer, C. D., Lentz, F. E., & Stollar, S. (2005). *The relationship between oral reading fluency and Ohio proficiency testing in reading*. Eugene, OR: University of Oregon. Retrieved from https://dibels.uoregon.edu/techreports/ohio.pdf

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & 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 learning disability. *Journal of Educational Psychology, 88,* 601-638. doi:10.1037/0022-0663.88.4.601

Wanzek, J., & Vaughn, S. (2007). Research-based implications from extensive early reading interventions. *School Psychology Review, 36,* 541-561. Retrieved from http://www.nasponline.org/publications/spr/index.aspx?vol=36&issue=4

Wayman, M. M., Wallace, T., Wiley, H. I., Tichá, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education, 41,* 85-120. doi:10.1177/00224669070410020401

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26,* 207-214. doi:10.1177/07419325050260040301

Willis, J. O., & Dumont, R. (2006). And never the twain shall meet: Can response to intervention and cognitive assessment be reconciled? *Psychology in the Schools, 43,* 901-908. doi:10.1002/pits.20197

Wilson, J. (2005). The relationship of dynamic indicators of basic early literacy skills (DIBELS) oral reading fluency to performance on Arizona instrument to measure standards (AIMS).

Research Brief: Assessment and Evaluation Department, Tempe School District No. 3. Tempe, Arizona. Retrieved from https://dibels.uoregon.edu/techreports/arizona.pdf

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Eduational Assessment, 11*, 85-104. doi:10.1207/s15326977ea1102_1

Yell, M. L., Shriner, J. G., & Katsiyannis, A. (2006). Individuals with Disabilities Improvement Act of 2004 and IDEA regulations of 2006: Implications for educators, administrators, and teacher trainers. *Focus on Exceptional Children, 39*, 1-24. Retrieved from http://www.lovepublishing.com/catalog/focus_on_exceptional_children_31.html

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*. Advance online publication. doi:10.1037/a0023021.

APPENDIX: SUMMARY OF VALIDITY STUDIES FOR CURRICULUM-BASED MEASUREMENTS IN READING

Table A1

Summary of Validity Studies in Marston (1989) for Curriculum-Based Measures of Reading

Criterion Measure		Validity Coefficients
Woodcock Reading Mastery Test		
Word Identification (compared with passages)		.87
Word Comprehension (compared with passages)		.82
Passage Comprehension (compared with grade-level basal reader)		.85
Stanford Achievement Test		
Word Study (compared with grade-level word list)		.65
Word Study (compared with third-grade word list)		.74
Word Study (compared with grade-level basal reader)		.84
Vocabulary (compared with grade-level word list)		.53
Vocabulary (compared with third-grade word list)		.85
Vocabulary (compared with grade-level basal reader)		.59
Stanford Achievement Test		
Reading Comprehension (compared with grade-level word list)		.54
Reading Comprehension (compared with third-grade word list)		.72
Reading Comprehension (compared with grade-level basal reader)		.88
Inferential Comprehension (compared with word lists)	(range)	.74-.75
Inferential Comprehension (compared with passage)		.80
Literal Comprehension (compared with word lists)	(range)	.68-.71
Literal Comprehension (compared with passage)		.78
Peabody Individual Achievement Test		
Reading Comprehension (compared with word lists)	(range)	.76-.78
Reading Comprehension (compared with passages)		.76
Science Research Associates (compared with grade-level basal reader)		
Vocabulary	(two separate studies)	.80, .67
Comprehension	(two separate studies)	.80, .59
California Achievement Test		
Vocabulary (with grade-level basal reader)		.69
Comprehension (with grade-level basal reader)		.75

Table A2

Summary of Validity Studies in Shinn & Shinn (2002) for Curriculum-Based Measures of Reading

Criterion Measure	Validity Coefficients
Comprehensive Test of Basic Skills	
First-Grade reader	.62
Second-Grade reader	.79
Third-Grade reader	.72
California Achievement Test	
Second-Grade basal reader	.63
Third-Grade basal reader	.52
Fourth-Grade basal reader	.54
Fifth-grade basal reader	.51
Gates-MacGintie Reading Test	
Second-grade reading of upper-first-grade passage	.86
Third-grade reading of upper-first-grade passage	.82
Fourth-grade reading of upper-first-grade passage	.86
Fifth-grade reading of upper-first-grade passage	.68
Sixth-grade reading of upper-first-grade passage	.63
Metropolitan Achievement Test	
Second-grade reading of upper-first-grade passage	.84
Third-grade reading of upper-first-grade passage	.67
Fourth-grade reading of upper-first-grade passage	.82
Fifth-grade reading of upper-first-grade passage	.64
Sixth-grade reading of upper-first-grade passage	.58
Kaufman Test of Educational Achievement-Brief (fourth-grade)	.41

Table A3

Summary of Validity Studies in Wayman et al. (2007) for Curriculum-Based Measures of Reading

Criterion Measure	Validity	Reliability
Woodcock Reading Mastery Test (first through fourth grades)		
Word Attack	.71-.82	
Word Identification	.73-.91	(test-retest) .92-.97
Woodcock-Johnson, Third Edition (third grade)		
Letter-Word Identification	.62	
Reading Fluency	.74	
Passage Comprehension	.42	
Iowa Test of Basic Skills (third grade)		
Vocabulary	.35	
Reading Comprehension	.58	
Stanford Diagnostic Reading Test (second grade)	.51	
Minnesota Comprehensive Assessment (third grade)	.71	(alternate form) .83-.87
Minnesota Comprehensive Assessment (fifth grade)	.57	
Diagnostic Reading Scales (first through third grades)		
Decoding	.60-.65	
Comprehension	.78-.82	
Michigan Educational Assessment Program (fourth)	.49-81	(test-retest) .87-.95
Washington Assessment of Student Learning (fourth)	.51	
Stanford Achievement Test		(test-retest) .91-.99
Reading Comprehension (second grade)	.80	
Reading Comprehension (third grade)	.89	

Table A4

Summary of Concurrent and Predictive Validity Studies in Wayman et al. (2007) for Curriculum-Based Measures of Reading

Criterion Measure	Concurrent Validity	Predictive Validity
Stanford Achievement Test-9		
Total Reading	.63-.85	.62-.72
Minnesota Comprehensive Assessment (MCA)		
First-Grade MCA (two studies)	.49-.58, .57	
Second-Grade MCA (two studies)	.61-.68, .67	
Third-Grade MCA (two studies)		.69, .71