Relationships Among Early Literacy Curriculum Based Measurement and Reading State Criterion Tests Over Time

Lynn Utchell

Recommended Citation

Follow this and additional works at: https://dsc.duq.edu/etd
RELATIONSHIPS AMONG EARLY LITERACY CURRICULUM BASED MEASUREMENT AND READING STATE CRITERION TESTS OVER TIME

A Dissertation
Submitted to the School of Education

Duquesne University

In partial fulfillment of the requirements for the degree of Doctor of Philosophy

By
Lynn Ann Utchell, M.S.Ed.

August 2011
DUQUESNE UNIVERSITY
School of Education
Department of Counseling, Psychology and Special Education

Dissertation

Submitted in partial fulfillment of the requirements
for the degree
Doctor of Philosophy (Ph.D.)

School Psychology Doctoral Program

Presented by:

Lynn Ann Utchell
B.S. Psychology, Allegheny College, 2000
M.S.Ed. School Psychology, Duquesne University, 2001

June 2, 2011

RELATIONSHIPS AMONG EARLY LITERACY CURRICULUM BASED
MEASUREMENT AND READING STATE CRITERION TESTS OVER TIME

Approved by:

_______________________________________________, Chair
Ara J. Schmitt, Ph.D.
Assistant Professor
Department of Counseling, Psychology and Special Education
Duquesne University

_______________________________________________, Member
Kara E. McGoey, Ph.D.
Associate Professor
Department of Counseling, Psychology and Special Education
Duquesne University

_______________________________________________, Member
Elizabeth McCallum, Ph.D.
Assistant Professor
Department of Counseling, Psychology and Special Education
Duquesne University
ABSTRACT

RELATIONSHIPS AMONG EARLY LITERACY CURRICULUM BASED MEASUREMENT AND READING STATE CRITERION TESTS OVER TIME

By

Lynn Ann Utchell

August 2011

Dissertation supervised by Ara J. Schmitt, Ph.D.

Despite significant regulations regarding the reading proficiency of students in the United States, more than half of students graduating high school are continually reading below proficiency (National Center for Education Statistics, 2010). In 2000 the National Reading Panel (NRP) identified phonemic awareness, phonics, fluency, vocabulary and comprehension as the five main areas of reading development. As a result of the NRP’s findings and the low rates of literacy, No Child Left Behind (2002) mandated that all students are proficient readers by 2014 and that proficiency is measured on a yearly basis by state mandated high-stakes assessments, beginning in Grade 3. This study presents
findings of statistical analysis examining the relationship between early literacy benchmark data in kindergarten and Grade 1 and statewide high-stakes achievement tests taken two to seven years later. One hundred thirty kindergarten students, ages five to six in 2002, were followed through 2010 and were included in this study sample. Results indicated that the DIBELS demonstrates moderate predictability when estimating future PSSA performance, based upon kindergarten and Grade 1 winter benchmark testing. Oral Reading Fluency at the winter Grade 1 benchmark moderately predicted PSSA Reading domain performance in Grades 3 and 5. Letter Naming Fluency at the winter kindergarten benchmark moderately predicted PSSA Reading domain performance in Grade 7. Diagnostic accuracy of the DIBELS, using ROC analysis, was more acceptable than the diagnostic accuracy of the DIBELS using the recommended cut scores when screening for the performance on the PSSA Reading domain. These findings imply that the assessment of early literacy skills, as early as the winter of kindergarten, predicts future reading performance up to seven years later. Implications of these findings and recommendations for future research were discussed.
DEDICATION

This document is dedicated to my late mother, Marilyn, who always encouraged me to pursue my dreams and goals despite the obstacles that may appear. This document is also dedicated to my father, Robert, who has seen me through all stages of my academic career, calmed me when times were rough and always believed in me. Finally, this document is dedicated to my husband, Brian, who made significant sacrifices throughout this long process. While providing me with love and understanding, he helped me to realize my true potential. Without the support of all three of these individuals, this journey would not have been possible.
I would like to thank the members of my dissertation committee for their encouragement, support and expertise throughout this long process. To my chair, Dr. Ara Schmitt, thank you for your guidance, patience and sense of humor. The organization and structure that you provided throughout this process was invaluable. You have been the motivator that helped me to persevere and achieve my timelines. I would like to thank Dr. Kara McGoey for her willingness to be a part of my committee, without physically meeting until the day of my proposal. Your insightful comments and background knowledge of my study has been appreciated. I would also like to thank Dr. Elizabeth McCallum for her feedback during this journey. Your commitment to my completion was extremely helpful. Although Audrey Czwalga was not a member of my committee, I would like to thank her for her endless hours of support, encouragement and registration “corrections” in regards to the completion of my dissertation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>iv</td>
</tr>
<tr>
<td>Dedication</td>
<td>vi</td>
</tr>
<tr>
<td>Acknowledgement</td>
<td>vii</td>
</tr>
<tr>
<td>List of Tables</td>
<td>xii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xiii</td>
</tr>
<tr>
<td>Chapter I: Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Early Literacy Skills</td>
<td>2</td>
</tr>
<tr>
<td>Phonemic Awareness</td>
<td>2</td>
</tr>
<tr>
<td>Alphabetic Principle</td>
<td>3</td>
</tr>
<tr>
<td>Fluency</td>
<td>4</td>
</tr>
<tr>
<td>Curriculum-Based Measurement</td>
<td>5</td>
</tr>
<tr>
<td>High-Stakes Reading Assessment</td>
<td>6</td>
</tr>
<tr>
<td>Critical Analysis of Relevant Literature</td>
<td>7</td>
</tr>
<tr>
<td>Problem Statement</td>
<td>9</td>
</tr>
<tr>
<td>Research Questions and Hypotheses</td>
<td>10</td>
</tr>
</tbody>
</table>
Chapter II: Literature Review

Background and History

Big Five

Phonemic Awareness

Alphabetic Principle

Fluency

Vocabulary

Comprehension

Data-Based Decisions

Curriculum-Based Measurement

Assessment of reading skills

Phonemic awareness measurement

Alphabetic principle measurement

Fluency measurement

High-Stakes Assessment

Curriculum-Based Measurement and Future Reading Success

Curriculum-Based Measures and Statewide Assessments
Longitudinal research of CBM and high-stakes testing .................. 41

Early Literacy Skills and High-Stakes Testing: Current Study .................. 44

Chapter III: Methods ...................................................................................... 46

Participants ........................................................................................................ 46

Measures ........................................................................................................... 47

Dynamic Indicators of Basic Early Literacy Skills, 6th Edition ............ 47

Letter Naming Fluency .................................................................................... 48

Initial Sound Fluency ........................................................................................ 49

Phoneme Segmentation Fluency .................................................................... 50

Nonsense Word Fluency ................................................................................... 51

Oral Reading Fluency ...................................................................................... 52

Pennsylvania System of School Assessment (PSSA) ...................... 53

Research Design .............................................................................................. 56

Procedure .......................................................................................................... 56

Data Analysis ..................................................................................................... 58

Participant Demographics and Outliers ........................................ 58

Research Question One and Analysis .................................................... 58
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>DIBELS Benchmark Testing</td>
<td>57</td>
</tr>
<tr>
<td>Table 2</td>
<td>Descriptive Statistics</td>
<td>62</td>
</tr>
<tr>
<td>Table 3</td>
<td>Pearson Correlations Among Curriculum Based Measurement Probes and State Criterion Test Reading Performance</td>
<td>63</td>
</tr>
<tr>
<td>Table 4</td>
<td>Model Summary for Research Question 2a</td>
<td>67</td>
</tr>
<tr>
<td>Table 5</td>
<td>Coefficients for Final Model for Research Question 2a</td>
<td>67</td>
</tr>
<tr>
<td>Table 6</td>
<td>Model Summary for Research Question 2b</td>
<td>69</td>
</tr>
<tr>
<td>Table 7</td>
<td>Coefficients for Final Model for Research Question 2b</td>
<td>69</td>
</tr>
<tr>
<td>Table 8</td>
<td>Model Summary for Research Question 2c</td>
<td>71</td>
</tr>
<tr>
<td>Table 9</td>
<td>Coefficients for Final Model for Research Question 2c</td>
<td>71</td>
</tr>
<tr>
<td>Table 10</td>
<td>Partial Coordinates of the Curve: Grade 1 Winter Oral Reading Fluency and Grade 3 PSSA</td>
<td>74</td>
</tr>
<tr>
<td>Table 11</td>
<td>Partial Coordinates of the Curve: Grade 1 Winter Oral Reading Fluency and Grade 5 PSSA</td>
<td>76</td>
</tr>
<tr>
<td>Table 12</td>
<td>Partial Coordinates of the Curve: Kindergarten Winter Letter Naming Fluency and Grade 7 PSSA</td>
<td>78</td>
</tr>
<tr>
<td>Table 13</td>
<td>Reading Diagnostic Accuracy Results Including CBM and PSSA</td>
<td>80</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Page

Figure 1. PSSA 3 Reading Performance Distribution…………………………………….64

Figure 2. PSSA 5 Reading Performance Distribution…………………………………65

Figure 3. PSSA 7 Reading Performance Distribution…………………………………65

Figure 4. Receiver operator characteristic curve predicting Grade 3 PSSA Reading performance…………………………………………………………………….73

Figure 5. Receiver operator characteristic curve predicting Grade 5 PSSA Reading performance…………………………………………………………………….75

Figure 6. Receiver operator characteristic curve predicting Grade 7 PSSA Reading performance…………………………………………………………………….77
CHAPTER I
INTRODUCTION

As recent as 2009, the National Center for Education Statistics (NCES) reported that 62% of the United States’ twelfth grade students were not reading at the proficient level (National Center for Education Statistics, 2010). After 12 years of formal schooling, more than half of the nation’s students are graduating without basic reading skills, despite the vast amount of dedicated research in the area of reading. In 2000, the National Reading Panel (NRP) identified five main areas of successful reading development: phonemic awareness, phonics (alphabetic principle), fluency, vocabulary and comprehension. These areas are now the primary focus of reading research, assessment and instruction among researchers and educators (Chard & Kame’enui, 2000; Fletcher, Lyon, Fuchs, & Barnes, 2007; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001; Torgesen, Wagner, & Rashotte, 1994; Walker & Shinn, 2002).

In conjunction with the findings of the NRP, federal legislation has passed several acts, most recently No Child Left Behind (NCLB; 2002) and the Individuals with Disabilities Education Improvement Act (IDEIA; 2004), mandating that all students, regardless of disability, are proficient readers by 2014. Proficiency is measured on a yearly basis by state mandated high-stakes assessments starting in Grade 3. Concerns arise as the assessment of proficiency in Grade 3 assumes that the foundational skills of reading have been met, or at minimum appropriately introduced to students. Unfortunately with the substantial number of students reading below proficiency there is grave uncertainty regarding the gap in the acquisition of the fundamental reading skills. The continued prevalence of poor readers has resulted in a greater need for the accurate
assessment of early literacy skills in order to target areas of need and ultimately provide students with reading instruction and intervention based upon specific needs.

**Early Literacy Skills**

Reading research centers on the “Big Five” component reading skills identified by the NRP. These skills are investigated and assessed at various points in a student’s schooling due to known reading skills benchmarks. The early literacy skills of phonemic awareness, alphabetic principle and fluency, are the first of the early literacy skills assessed in progressive order on commercially-developed curriculum-based measurements (CBM), as well as teacher-developed CBM, to measure student progress and guide instruction. Acquisition theories of reading development propose that learning to read begins with constrained skills including basic components of sound and progresses toward unconstrained skills, such as the meaning of text (Ehri, 2001; Paris, 2005).

**Phonemic Awareness**

Phonemic awareness involves the ability to hear and use the smallest units of spoken language, phonemes (Blachman, 1991; Holland, McIntosh, & Huffman, 2004; National Reading Panel, 2000). Early, accurate assessment of phonemic awareness skills of students allows for the guided development of future reading skills and appropriate interventions. The ability to discriminate and manipulate sounds has been linked with reading performance as early as three years of age, making early assessment critical (Fox & Roth, 1975; Wagner & Torgesen, 1987). Continued research has also found that phonemic awareness is a unique predictor of future reading skills such as word identification, word analysis, and passage comprehension (Bradley & Bryant, 1985;
While phonemic awareness involves the manipulation of the sounds of the words, alphabetic principle involves alphabetic orthography to identify sounds of letters and words. The relationship between phonological awareness and alphabetic principle indicates that efficient phonological processing is essential to orthographic learning (Cunningham, Perry, Stanovich, & Share, 2002). Orthographic learning assessed in second grade students through homophonic pseudoword choice, spelling production and homophone naming showed that the number of target homophones correctly decoded in text is a considerable predictor of orthographic learning. Research also indicates that although closely related, phonological and orthographic skills are distinct components in reading acquisition and are taught and assessed through different measures (Cunningham, Perry, & Stanovich, 2001). Reading acquisition theories further indicate that letter naming skills and phonemic awareness prepare a child to understand the alphabetic principle in print (Adams, 1990; Torgesen & Mathes, 2000).

**Alphabetic Principle**

As mentioned, alphabetic principle focuses on printed words and the associated sounds. This skill is one of the earliest skills taught and assessed that impacts future reading success in regards to fluency and comprehension (Coyne & Harn, 2006). Alphabetic principle is closely related to phonemic awareness, in that mastering phonemic awareness is a necessary skill in the development of the alphabetic principle (Share & Stanovich, 1995). Understanding the sounds of spoken words allows for the development of understanding the sounds of the written word. Mastery of phonemic
awareness skills, however, is not enough for the mastery of the alphabetic principle. Further, phonemic awareness and alphabetic principle skills are not automatic and therefore require instruction as part of the process of learning to read. Inadequate letter-sound association and phonological decoding are frequently identified as the primary indicators of students with specific learning disabilities in reading (Fletcher et al., 2007; Share & Stanovich, 1995; Stanovich, 1986). Therefore, early identification of skills deficit, through assessments, is important in the continuation of reading development. Without the foundational skills of phonemic awareness and the alphabetic principle, fluently identifying printed words presents as a significant challenge.

**Fluency**

 Individuals are able to identify words with automaticity, through the recognition of letter sound correspondence in speech and in print. Fluency is described as the ability to read connected text with speed, accuracy, and proper expression (National Reading Panel, 2000). Reading fluency is more than the recognition of words and involves automatic decoding skills of the student. When fluent reading skills are present, students are able to spend more time on the meanings of words and comprehension and less time on decoding. Similar to phonemic awareness and alphabetic principle skills, fluency is necessary for the development of proficient reading comprehension skills, but it is not sufficient (Catts & Hogan, 2003). Simple measures of reading fluency have been cited as predictors of reading impairments and future reading success (Hintze, Ryan, & Stoner, 2003; Hintze & Silberglitt, 2005; Shapiro, Solari, & Petscher, 2008). Understanding that phonemic awareness and the alphabetic principle provide the necessary skills for reading
fluency provides support for the early assessment and intervention of these three important foundational skills.

Due to the young ages of students, the time factor for teachers and the cost for districts, it is advantageous for educators to use a quick, yet accurate tool to identify potential reading concerns. Numerous measures have been utilized over the years to monitor academic progress and screen for learning disabilities. Using measures of fluency as well as phonemic awareness and alphabetic principle to predict future reading proficiency is now an important part of directing reading instruction and focusing on the attainment of 100% reading proficiency by 2014.

Curriculum-Based Measurement

The most widely used assessment of early literacy skills is curriculum-based measurement (CBM) (Jenkins & O’Connor, 2002; National Reading Panel, 2000; Shinn, 1998). CBM was developed based upon the need for adequate (i.e., reliable and valid) and practical (i.e., simple, efficient, and inexpensive) tools to measure student performance (Deno, 1985). CBM is part of the problem-solving or data-based decision making approach to special education, and is now a guide for Response to Instruction and Intervention (RtII). CBM allows for the evaluation of student growth over a short period of time, therefore providing staff the opportunity to adapt instruction based upon student performance. Due to the ease of assessment, CBM is utilized throughout schools to quickly and frequently address performance on academic skills, particularly reading skills. Since CBM are continually used to guide reading instruction, the reliability and validity toward predicting future reading performance is quite intriguing to researchers and educators.
Recent research supports the idea that CBM can be utilized as one source or predictor of student success or failure on statewide measures (Goffreda, Diperna, & Pedersen, 2009; Keller-Margulis, Shapiro, & Hintze, 2008; Shapiro, Keller, Lutz, Santoro, & Hintze, 2006). The increase in screening and monitoring through data-based decision making has created a greater need to identify deficiencies in skills as early as possible to allow time for growth. Additionally, through the screening process, the prevention of deficits, versus the remediation of skills, is the ultimate goal. Screening for future reading success seems simple enough, but the determination of when the earliest future reading skills can be predicted from early literacy skills is under question in the research. Assessing too early may not be representative of true ability, and assessing too late does not provide time for the needed instruction prior to high-stakes assessments.

**High-Stakes Reading Assessments**

Increased accountability through high-stakes assessments has evoked great interest in the relationship between CBM and state mandated testing. NCLB requires that all students and schools are held to standards that are set by individual states. Further, NCLB requires states to report student performance relative to proficiency standards on a yearly basis. These scores are criterion-based scores that are referenced on a developmental scale, focusing on proficiency standards rather than norm-referenced descriptions of student performance (Braden & Tayrose, 2008). Reporting of proficiency begins in the third grade and continues through most of a student’s educational career. The focus on proficient reading performance by the end of the third grade, however, is a significant undertaking by states, districts, and schools. Teachers provide instruction centered on specific skills necessary for proficient readers and continually assess these
skills through curriculum-based measures until the time of high-stakes testing. When goals of proficiency are not met, schools and districts stand to lose state and federal funding for programs, as well as prestige. Although schools are in jeopardy of losing financially, students are losing the most by not attaining foundational reading skills.

**Critical Analysis of Relevant Literature**

Curriculum-based measures of reading are imperative to understanding, predicting, and influencing reading achievement. Research has examined the linear relationship between phonemic awareness, alphabetic principle, fluency, and comprehension and the impact on reading proficiency. Much of the research investigating the predictability of future reading success focuses on reading fluency as the primary predictor (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Fuchs, Fuchs, & Maxwell, 1998; Harn, Stoolmiller, & Chard, 2008; Stage and Jacobsen, 2001). Due to the increased attention of state and district accountability for student performance, recent research has focused on the predictability of reading CBM in relation to high-stakes state-mandated testing.

Most recently, Goffreda, Diperna, and Pedersen (2009) and Keller-Margulis, Shapiro, and Hintze (2008) examined the relationship between reading CBM data and large scale achievement test data. Keller-Margulis et al. (2008) examined this relationship in several cohorts of students with results indicating moderate relationships between CBM performance and achievement performance one and two years later. Although this data was longitudinal in the sense that it reviewed data on particular groups of students over a one and two-year period, the study did not examine the long term
correlations of basic early literacy skills and continually measured state reading achievement performance.

Oral reading fluency (ORF) is the most prevalent CBM used in predicting future reading performance, as it has been shown to be an important predictor of reading comprehension (Adams, 1990; Fuchs et al., 2001; National Reading Panel, 2000; Shinn, 1998). Additionally, longitudinal data on ORF CBM is readily accessible in the research, as it is generally measured starting in Grade 1. Oral reading fluency is thought to have a strong relationship with future reading performance, as fluency is continually used when reading for comprehension. Overall, most research has focused on the relationship between ORF and future reading proficiency, neglecting the predictability of the early literacy skills of phonemic awareness and the alphabetic principle.

Recent additional research that focused on three of the early literacy skills (i.e., phonemic awareness, alphabetic principle and fluency) continued to find that ORF was the only significant predictor of future reading proficiency on high-stakes assessments (Goffreda et al., 2009). Performance in the first grade was examined, as it is a particularly critical developmental period in reading acquisition (Juel, 1988). Student performance was not examined over a continuous longitudinal timeframe, but rather just two years separated early literacy performance and high-stakes achievement assessment performance. Research findings proposed that only assessments utilizing ORF be used as predictive measures, as it would save time and resources. If research indicates the need for measures of early literacy to guide instruction, then why are researchers and districts ignoring data that is taken at the beginning of an individual’s schooling? Studies that examine early literacy in relation to future reading performance exist, but research is less
due to the overwhelming research and support to use fluency measures. Although numerous studies have investigated the relationship between fluency and future reading performance, an examination of the early literacy skills and continued future reading performance is lacking in the literature. Research is strong in the area of early literacy skills and in the area of high-stakes assessments, although the relationship between the two is not fully examined. Understanding how curriculum based measures of early literacy skills can explain future reading performance in later years will provide schools with the much needed information to continue to drive instruction and performance toward proficiency.

**Problem Statement**

Although significant emphasis has been placed on early intervention for reading problems, the rates of reading failure continue to be unacceptably high. This is also problematic as federal mandates, such as NCLB, require all students to meet proficiency on state-level reading assessments by 2014. As a nation we are far from this goal, with only 38% of twelfth grade students reaching proficiency in 2009 (National Center for Education Statistics, 2010).

Not presently known is the extent to which early literacy skills predict future reading performance up to seven years later. Examining longitudinal reading state achievement performance in light of kindergarten and first grade early literacy skills may produce data to help identify those early readers at greatest risk for long term reading failure and design interventions for those students. Further, understanding the implication of early literacy skills on later performance on high-stakes reading
assessments will help to inform instruction in the classroom and prepare more proficient readers.

**Research Questions and Hypotheses**

This study explored the predictive power of early literacy skills assessments designed to predict performance on state-level assessments taken no less than two years later and up to seven years later. Specifically, the following research questions were investigated:

1. What are the relationships among performance on Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good and Kaminski, 2002) early literacy probes (Initial Sound Fluency, ISF; Letter Naming Fluency, LNF; Phoneme Segmentation Fluency, PSF; and Nonsense Word Fluency, NWF) administered at the winter benchmark in kindergarten, ORF administered at the winter benchmark in Grade 1, and PSSA Reading performance in Grades 3, 5, and 7?

   Hypothesis 1: All variables will be moderately correlated with each other.

2. To what degree do ISF, LNF, PSF and NWF probes administered in kindergarten and ORF administered in Grade 1 predict PSSA Reading performance in Grades 3, 5, and 7, respectively?

   Hypothesis 2: Performance on Grade 1 ORF will best predict Grade 3 PSSA Reading performance.

   Hypothesis 3: Grade 1 ORF performance will best predict Grade 5 PSSA Reading performance.
Hypothesis 4: Grade 1 ORF performance will best predict Grade 7 PSSA Reading performance.

3. Using Receiver Operator Characteristic (ROC) curve analysis, what is the diagnostic accuracy (i.e., positive predictive power, negative predictive power, sensitivity, and specificity) of the early reading CBM measures that were found to be most predictive of Grades 3, 5, and 7 PSSA Reading performance, respectively?

Hypothesis 5: With respect to CBM measures found to be most predictive of Grades 3, 5, and 7 PSSA Reading performance, cut scores determined by ROC analyses will satisfactorily identify success or failure on Grades 3, 5, and 7 PSSA Reading performances more than DIBELS-provided cutoff scores.
CHAPTER II
LITERATURE REVIEW

Background and History

Statistics released by the National Assessment of Educational Progress (NAEP) have historically revealed that an alarming number of students in the United States are not able to read grade level material proficiently. For example, in 1992, 71% of the nation’s fourth and eighth grade students were not able to read at proficient levels (National Assessment of Educational Progress, 2009). Currently, 67% of fourth graders and 68% of eighth graders still read at the basic level or below (National Assessment of Educational Progress, 2009). As such, federal legislation has been enacted to stimulate the development of students’ reading skills.

Although the late 1980s are identified as the start of the standards-based educational reform movement, educational acts were initiated long before that time. The Elementary and Secondary Education Act (ESEA) passed in 1965 and was the largest federal government law effecting public education at the time. The ESEA was developed to increase the performance of education in kindergarten through twelfth grade. Despite devoting a large amount of the nation’s budget on the ESEA, it did not hold states accountable for the improvement of student achievement. Beginning in 1969, the NAEP began providing periodic data to the nation in the area of reading achievement at the national, state, and district level. Legislation had passed and academic achievement performance was made public, but reading proficiency was still subpar. A few short years later, in 1975, Public Law 94-142, the Education for All Handicapped Children Act, was passed and provided for the education of all students regardless of their
disability. Regardless of holding states and districts responsible for the education of all students, there still was no mandate for accountability for student performance. In 1983 the National Commission on Excellence in Education released “A Nation at Risk: The Imperative for Educational Reform.” This movement, led by George Bush and Bill Clinton, focused on educational outcomes and supported the development of national educational goals, starting in 1988. The United States was finally on the way to monitoring and holding educators responsible for academic performance of students. Through continued federal efforts, new amendments were added to Public Law 94-142, and in 1995 it was renamed the Individuals with Disabilities Education Act (IDEA). Accountability for students with disabilities in the schools was now on the forefront; however, academic success was still not regulated by law.

In continuation of the educational reform, the No Child Left Behind Act (NCLB) of 2001 was passed as a reauthorization of ESEA and sought to hold stakeholders accountable for student performance. NCLB is currently the largest federal funding program for education in the United States history and outlined the responsibility for increased student outcomes on assessments to be in the hands of states, districts, and schools. Accountability and assessment are the foundation of NCLB, and the creation and implementation of statewide standards is a requirement of NCLB. States are now held responsible for identifying “challenging” reading and mathematics standards and measuring student performance annually in grades three through eight, as well as once during high school. The mandate of NCLB is to have all students proficient in reading and mathematics by the spring of 2014.
NCLB’s mandate of proficiency of all students means that even students with disabilities are expected to achieve proficiency in the areas of reading and mathematics. IDEA attempted to align special education law with the accountability mandates of NCLB and was recently reauthorized in 2004, in which it was renamed the Individuals with Disabilities Education Improvement Act (IDEIA). The greatest change in regards to accountability in IDEIA is that students with disabilities are now included in the state accountability systems and are required to participate in the assessments. Furthermore, individuals with disabilities are held to the same standards as those individuals who do not have a disability. Researchers have indicated that the process of linking assessment and instruction practices to statewide content standards shows promise for students with disabilities. Focus is now on meeting the goal by improving access to the general education curriculum for students with disabilities.

As previously noted, NCLB mandates that all students must be at the proficient level on state level assessments in reading and mathematics by the 2013-2014 school year. Considering recent NAEP data, much improvement must take place for that goal to be achieved. According to federal legislation, like NCLB, identifying individuals who have not developed or mastered adequate reading skills is the responsibility of the educators and school districts. Such legislation also requires districts to be accountable for the learning of students under their watch. Due to the continued mandates, schools are in danger, now more than ever, of losing their funding used to educate the students who have fallen behind. With the introduction of NCLB, schools are now held more responsible for keeping track of students’ academic progress; making sure that no student fails to succeed. Public schools and districts are also required to make adequate yearly
progress (AYP). Meeting AYP means that all students, including those with disabilities and those who are economically disadvantaged, will meet the identified targets of proficiency each year. Additionally, with the reauthorization of IDEA, there is a greater focus on meeting the goal of improving access to the general curriculum for students with disabilities, while continuing to meet AYP (Parrish & Stodden, 2009). As of 2005, NCLB requires all students in grades three through eight to participate in a statewide assessment in the areas of reading and mathematics. This assessment requirement is used to establish AYP, provide feedback to parents and teachers regarding the performance of individual students, and help inform who is in need of additional academic intervention as early as possible.

Providing early intervention is essential to improving the reading performance of students throughout future grades. Recent studies indicate that up to 40% of students that enter kindergarten are one or more years behind their peers in critical language and early literacy skills (Fielding, Kerr & Rosier, 2007). This poses instructional challenges, since without early reading intervention, addressing the literacy needs of students in the regular education classroom may become futile (Chard & Kame’enui, 2000). Current research supports the concept of early intervention in grades kindergarten through third grade as the period of crucial reading development (Coyne, Kame’enui, & Simmons, 2001; Torgesen, 2000). Further, literacy skills acquired in kindergarten and first grade are thought to serve as the foundation for the development of subsequent reading skills (Coyne et al., 2001; Coyne & Harn, 2006). Knowing what skills are stepping stones to future reading skills will help educators guide their instruction. As with many skills, reading difficulties have been noted to be easier to prevent than to remediate, therefore
the stress on the attainment of early reading skills should continue (Juel, 1988; Stanovich, 1986; Torgesen, 2002). Researchers have also concluded that an element of effective early intervention is the use of tools that are reliable and valid for problem identification and intervention progress monitoring (Fletcher et al., 2007).

In recognition that American students are not reading to standard, and that early and targeted reading intervention is necessary, in 1997 Congress asked the National Institute of Child Health and Human Development and the Secretary of the United States Department of Education to form a committee. The committee, known as The National Reading Panel (NRP), was formed to examine the status of research-based knowledge of the process of learning to read and the effective methods of teaching someone to read. The NRP (2000) reviewed more than 100,000 studies and identified the five main areas that constitute effective reading and distinguish proficient readers from less proficient readers: phonemic awareness, phonics (i.e., alphabetic principle), fluency, vocabulary and comprehension.

**Big Five**

Current studies regarding the assessment and intervention of reading now focus on the NRP’s “Big Five” component reading skills. As mentioned, these include phonemic awareness, alphabetic principle, fluency, vocabulary, and comprehension. Each of these component reading skills may be targets of assessment and intervention for students throughout different time periods of learning. These five skills are acquired through various stages of reading development and follow a linear path of acquisition.
**Phonemic Awareness**

One of the earliest reading skills assessed in students is phonemic awareness. Researchers and educators frequently confuse phonemic awareness with phonological awareness, which refers to a broader category of awareness of the larger units of words including syllables and rhyming words. Phonemic awareness is the ability to hear and manipulate the phonemes in spoken words (Blachman, 1991; NRP, 2000). Phonological awareness is thought of as the “understanding of the rules about how words are divided into their component sounds and then how these sounds are subsequently blended together” (Holland et al., 2004). These units, or phonological concepts, are identified by the graphemes (letters) and phonemes (sounds) that are found in each word. Phonemes are the smallest units of the spoken language that combine to form syllables and words. In the English language there are approximately 41 identified phonemes (NRP, 2000).

Phonemic awareness does not involve letters or words in print, but is an auditory process and involves the sounds of letters. Although phonemic awareness involves only sounds, the components of the words (i.e., phonemes and graphemes) are ultimately represented by alphabetic orthography. Instruction and assessment of phonemic awareness skills are conducted orally and not through printed text.

Researchers in favor of the theory that phonemic awareness is critical to the acquisition of learning to read argue that individuals are then able to view the alphabet orthography in a sensible way, since they are accustomed to the segmenting and blending of sounds. Therefore learning to read new words and segmenting the words into parts is a sensible way to perform the task of reading. Understanding the sounds segmented in words allows for the visual segmentation of words to occur. Individuals must, however,
receive instruction of the spoken language system to understand the phonemic parts of
words and then to grasp the written language system.

In an attempt to further examine and explain early reading skills, Cunningham, Perry, and Stanovich (2001) found that phonological and orthographic skills are distinct components in reading acquisition. However, the decoding of words by individual sounds influences an individual’s reading and spelling skills. Cunningham, Perry, Stanovich, and Share (2002) investigated this area further and found that efficient phonological processing is a necessary, but not sufficient, condition for orthographic learning. Participants in the second grade were administered both connected text with post-test assessments of orthographic learning and measures designed to assess a variety of cognitive skills linked to early reading acquisition (Cunningham et al., 2002). Research showed that orthographic learning and decoding of sounds were different skills with separate levels of correlation regarding future reading skills. Recognition of the distinction of these skills, as well as the relationship between them is beneficial in understanding early literacy skills. Early instruction and assessment in skills of phonemic segmentation may prove to be the initial factor in future reading success. The ability to discriminate and manipulate sounds at the phonemic level is closely linked with reading ability at an early age, as evidenced through a review of longitudinal correlational studies (Wagner & Torgesen, 1987). In fact, some researchers have found that even children three years of age were able to segment at least some words into beginning and remaining sounds (Fox & Routh, 1975).

Phonemic awareness, in comparison to other phonological concepts, was noted as the only unique predictor of reading growth from kindergarten through fourth grade.
(Torgesen et al., 1994, 1997a; Wagner et al., 1994). Moderate correlations were noted between phoneme segmentation and future reading skills of word identification, word analysis and passage comprehension for students assessed in grades two and again in grade four (Torgesen et al., 1994). Similar correlations were also found when analyzing relationships of the same reading skills in students assessed in the third grade and again in the fifth grade. Previous longitudinal studies again reported moderate correlations between phonological awareness of four and five year olds and performance on reading achievement tests three years later (Bradley & Bryant, 1985). These results imply that phonemic awareness of prereaders is a factor in the success of early reading skills.

Phonemic awareness and letter identification are identified as two of the strongest predictors of learning to read in students (Lyon, 1999; NRP, 2000). Poor letter sound associations and the failure to acquire these basic preliteracy skills are potential indicators of future reading-based disabilities (Jenkins & O’Connor, 2002; Share & Stanovich, 1995). Researchers have proclaimed that unless detected early and addressed through direct instruction, the deficits in phonemic awareness will negatively impact performance in elementary grades, as well as performance in adolescence and adulthood (Fletcher et al., 2007; Wolf & Bowers, 1999). Ultimately, phonemic awareness skills are the building blocks or the foundation for more advanced reading skills (Adams, 1990; Ehri, 1998; Shaywitz, 1996). The basic skills of letter naming skills and phonemic awareness prepare a student to understand the alphabetic principle in print (Adams, 1990; Torgesen & Mathes, 2000).
Alphabetic Principle

The ability to match sounds to letters and use this knowledge in reading and spelling is referred to as the alphabetic principle. More specifically, it is the mapping of print to speech and the establishment of a clear link between a letter and a sound (Baker, Kame’enui, Simmons, & Stahl, 1994). In contrast to phonemic awareness, which does not rely on the printed word, the alphabetic principle focuses on the printed words and the associated sounds. This specific skill is one of the earliest taught to students and is therefore used to remediate instruction that struggling readers may have missed early on. Once an individual has mastered the alphabetic principle, they are thought to be able to achieve, at minimum, average reading fluency and comprehension (Coyne & Harn, 2006).

When beginning to read, an individual looks at the smallest units of sound, phonemes, and then pairs those sounds to letters. Share and Stanovich (1995) reported that phonemic awareness skills provide necessary, but not sufficient, support for the development of good orthographic reading ability (i.e., alphabetic principle). Once the mastery of individual letters and sounds is accomplished, readers move on to identify letters and sounds in groups. These skills are not inherent or automatic, but rather learning these individual skills are part of the process of learning to read and require exposure and practice.

Reading acquisition theories support the notion that reading is an “unnatural act,” therefore asserting that individuals do not have the automaticity of reading skills prior to identifying phonemes and graphemes. Through repetition individuals are said to acquire the necessary skills for phonemic awareness and decoding. Gough and Hillinger (1980)
describe this repetitive learning process of reading acquisition as a two-stage theory comprised of paired associative learning and alphabetic coding. In the paired associative stage there is no generalization to unfamiliar words, although it does rely on visual cues and memory of visual cues. Ehri (1992) referred to this as the prealphabetic stage. In this first stage individuals are acquiring words spontaneously, as the spoken words are associated with the visual cues. In the second stage generalization begins to take place, but it requires continued intervention and external cues for the generalization to take place with unfamiliar words. Generalization of early reading skills is a learned process which leads to automaticity with fluency and comprehension. This unnatural process begins with skills that are very controlled and limited and then branches out into skills that are unrestricted.

The alphabetic principle is viewed by Paris (2005) as the most constrained variable in learning to read. There are a set number of letters and a defined number of sounds related to each of the 26 letters of the English language. The constrained skills have a finite amount of information necessary to learn, and unless students are able to master these skills it is difficult to begin learning or mastering another skill. Vocabulary and comprehension, the least constrained skills according to Paris, are related to a range of academic skills throughout life. Unconstrained skills continue to develop over time and are cited to create more of a gap between individuals than the finite, constrained skills. Correlations between mastered skills and other measures of reading are viewed by some researchers as indicators of future reading and not the cause of future reading difficulty. Although constrained skills are necessary for many reading skills, they are not sufficient, independent causes of reading development. According to Paris, there is no
evidence to warrant instructional priority of constrained skills over unconstrained skills. Numerous researchers would disagree and contend that that the priority should be on the instruction of phonemic awareness and alphabetic principle, as the first skills to introduce and assess due to the connection with later reading skills (Ehri, 1991; Juel, 1991; Lyon, 1995; Shaywitz, 2003).

The formation of complete connections between written letters in words and phonemes detected in pronunciation is described in literature as the full alphabetic phase (Ehri, 1992). Within this phase, individuals begin to decode words they have never seen before by use of graphemes and phoneme distinction. Overall, phonemic awareness and alphabetic principle are the stepping stones for continued reading, as they provide the foundation for students to identify printed words. Poor letter-sound association and phonological decoding are often the primary characteristic of students with reading-based learning disabilities and are therefore introduced and monitored prior to tasks requiring reading fluency (Share & Stanovich, 1995; Stanovich, 1986).

**Fluency**

The combination of adequate phonemic awareness skills and the alphabetic principle allows for the development of fluency. Through the recognition of letter sound correspondence in speech and print, individuals are able to identify words with automaticity. Fluency is described as the ability to read connected text with speed, accuracy, and proper expression (NRP, 2000). This is more than the recognition of words and involves automatic decoding skills. Individuals with automatic decoding have a seamless time orally reading the text and spend more time focusing on the meaning of the words.
Those who experience difficulty with fluency struggle with accessing word recognition skills at a fluid pace. Fluency, as measured by reading from a list and a passage, has been found to be predictive of reading impairments (Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003). Often these deficits are secondary to problems with higher-order processing, attention, and executive functioning. As fluency is measured by rate or speed, automaticity obviously plays a role in this reading skill. Difficulties with fluency, and subsequently comprehension, in the absence of phonological difficulties have been cited in research by Wolf and Bowers (1999) and Lovett, Steinbach, and Frijters (2000). These findings support the research that phonemic awareness and alphabetic principle are necessary, but not sufficient, skills for adequate reading fluency. The idea that naming speed, or fluency, may contribute to reading failure independent of phonological difficulties is conceptualized by Wolf and Bowers (1999) and referred to as the double deficit. The deficits identified were separated into three subtypes: (a) deficits with phonological processing and rapid naming; (b) deficits only in phonological processing; and (c) deficits in rapid automatized naming only. Students with deficits in phonological processing and rapid naming, compared to those with only a single deficit (e.g., phonological processing or rapid naming), display more severe reading difficulties (Wolf & Bowers, 1999). Fluency itself, however, is dissociable from phonemic awareness, word recognition, and comprehension, but they are all highly correlated.

As mentioned, although highly correlated, fluency is distinct from word recognition, and therefore is a standalone reading disability. With the authorization of IDEIA 2004, a specific learning disability in reading fluency skills was recognized as separate from basic reading skills and reading comprehension. Studies indicate, however,
that individuals who are not fluent oral readers also display poor reading comprehension (Fuchs et al., 2001). Difficulty with accessing words, subsequently vocabulary, can then lead to challenges with reading comprehension.

**Vocabulary**

Vocabulary is described as the ability to understand and use words (NRP, 2000). To further understand this idea, it should be explained that the relationship between vocabulary, phonemic awareness, and alphabetic principle is reciprocal. Individuals that do not have the words in their spoken vocabulary will not understand the unknown “real” written word any more than a pseudoword. Therefore, oral vocabulary is stated to be the bridge between the oral and written forms of words and reading vocabulary is then noted to be the link to reading comprehension (NRP, 2000).

Individuals enter school with a vocabulary based upon what they hear at home and in the community. At school entry, researchers report a wide range of exposure based upon socioeconomic status (Hart & Risley, 1995). This discrepancy between individuals who enter school with a limited vocabulary and those who enter with a rich vocabulary remains constant throughout education. Although students will continue to add words to their vocabulary, the gap will grow as all students continue to expand their vocabulary (Baker, Simmons, & Kame’enui, 1997). This does not mean that individuals with limited vocabulary will not reach the average vocabulary count, but rather they will not acquire the same overall number of words in their vocabulary as students who begin school with a rich vocabulary.

Researchers and educators frequently discuss several types of vocabulary, as well as the disparity between each. The most prominently discussed forms of vocabulary are
receptive, expressive, oral, reading, and sight word. Receptive vocabulary refers to the vocabulary that an individual understands and that is presented orally in conversation or through the written words in text. Expressive or productive vocabulary is the vocabulary individuals produce when speaking or writing. An individual’s receptive vocabulary is noted to be greater than their expressive vocabulary, as individuals are able to understand and recognize a larger number of words than generally used in written or spoken communication. Vocabulary is further broken into oral vocabulary and reading vocabulary. Oral vocabulary is the understanding of words presented orally, while reading vocabulary is the recognition of words used in print. Reading vocabulary is further segmented into sight word vocabulary and non-sight word vocabulary, where the sight word vocabulary is made up of the words that do not require decoding or processing skills to recognize the words. Basic sight word recognition relies on the development of memorized spelling patterns and automaticity. However, the acquisition of the spelling patterns depend upon, another early literacy skill, an individual’s phonemic decoding skills (Ehri, 1998).

Instruction regarding vocabulary is not as straightforward as the instruction for phonemic awareness and alphabetic principle. The NRP identified five separate methods in which vocabulary instruction takes place. The instruction methods are explicit instruction, indirect instruction, multimedia instruction, capacity methods, and association methods. The various methods are not recommended by the NRP for use in isolation, but rather in conjunction with another method. Students receiving explicit instruction are provided with the definitions of words to be learned, while implicit or indirect methods propose that exposure to the words through reading is beneficial. The
combination of these two methods allows for additional exposure to vocabulary outside of the targeted vocabulary words. Multimedia instruction permits individuals to utilize other forms of exposure to text through mediums such as the computer or American Sign Language. The capacity method of instruction regarding vocabulary focuses on the meaning of the word rather than the orthographic or oral representation of the words. Lastly, the association method relies on the individual drawing connections between the known and the unknown vocabulary. Overall, repetition and variety of each method are crucial to reading success, as mastered vocabulary is the gateway to reading comprehension. Regardless of the lack of research on the vocabulary of the primary learner, early literacy skills are thought to have a direct impact on vocabulary, subsequently impacting comprehension. Gains in vocabulary support an individual’s reading comprehension, as they have greater word meaning to access.

**Comprehension**

Comprehension is commonly referred to as the core of reading, in which there is interaction between the text and the reader (Durkin, 1993). Individuals who obtain meaning from the text are able to comprehend what they have read. Reading comprehension is described as the result of an individual’s language comprehension ability and word identification skills (Torgesen, 2002).

Similar to vocabulary, limited studies have been conducted on students younger than the third grade in regards to comprehension. Research has noted that factors such as an individual’s phonemic awareness skills, alphabetic understanding, fluency, vocabulary, prior knowledge, and interest in the material can influence comprehension (Wagner & Torgesen, 1987). Additionally, difficulty with reading comprehension can
occur without difficulties in phonemic awareness and fluency; however, proficient reading comprehension occurs when accurate and fluent decoding are present (Fletcher et al., 2007).

Assessing reading comprehension is complex, particularly if an individual has poor decoding or fluency skills. When other deficits are present it is challenging to discern if the problem is poor decoding and poor comprehension or poor comprehension as a result of poor decoding (Fletcher et al., 2007). Proficient readers display automaticity of phonological awareness, decoding and vocabulary and are able to focus on the meaning of the text (Kuhn & Stahl, 2000). Individuals who are not reading at a proficient rate are unable to efficiently discern the meaning of the text; however, that could be attributed to poor phonemic awareness, limited accuracy and speed, poor fluency, limited vocabulary, or lack of comprehension.

The vast amount of research regarding early literacy skills explains the disconnect in regards to future reading. Provided with the knowledge of how to assess and instruct students in the area of early literacy, educators can spend time remediating the skills of older learners and hopefully decreasing the prevalence of deficient young readers. Due to the continued prevalence of reading difficulties, federal and state laws have continued to evolve and hold schools and states more responsible for academic failure and success. This focus on educational improvement is a continued journey spanning several decades and numerous governmental leaderships. Current educational reform and legislators are focused on informing instruction through data. The promotion of problem solving teams and the Response to Instruction and Intervention (RtII) initiative supports the notion of
early assessment and early identification of skill deficits in hopes of preventing future reading difficulties (Coyne et al., 2001; Coyne & Harn, 2006; O’Connor, 2000).

**Data-Based Decisions**

Review of the research indicates the need for the prevention of poor reading skills in contrast to the remediation of these skills. Further, the prevention of reading deficits needs to start when students begin school in kindergarten. NCLB and IDEIA call for high quality instruction, assessment of learning rates and informed educational decisions. A tiered model of prevention and intervention is a way to address curricular issues and improve instruction in order to meet the needs of students (Coyne et al., 2001; O’Connor, 2000). With the reauthorization of educational law, Response to Instruction and Intervention (RtII) is a systematic manner to review assessment data and to assign specific resources to enhance the learning of all students (Jimerson, Burns, & VanDerHeyden, 2007). Through this approach, schools are able to manage the interventions for students in the early grades before learning problems become severe and students fall drastically behind. The three-tier approach to problem solving, specific to reading difficulties, includes screening students for reading failure, identifying specific reading skill deficits, monitoring student progress, informing instructional practice, and assisting in making eligibility decisions.

The primary tier of intervention involves a universal screening for all students using developmentally-appropriate measures. Benchmark data is collected through universal screenings on all students throughout the school year to determine if students are on target to meet the outlined skills. Universal screening of students for difficulties in reading was one of the issues in the NRP’s report (NRP, 2000).
The secondary tier of intervention involves small group instruction and is provided to students who are unresponsive to the primary tier of intervention, as identified by the universal screening measure. In the secondary tier of intervention, students are progress monitored for a specified period of time to determine the effectiveness of the small group instruction. Small group instruction is provided within the general education classroom for the students that are identified as “at-risk” for not meeting benchmarks (Jimerson et al., 2007).

Tertiary intervention is reserved for students that were unresponsive to the primary and secondary interventions. Students receiving tertiary intervention are generally provided with pull-out instruction with multiple opportunities to practice skills, receive feedback, and have their skills monitored. Students with direct instruction via the tertiary level of intervention have shown improvement in the critical early literacy skills, with some students achieving grade-level performance (Kamps et al., 2008).

**Curriculum-Based Measurement**

Curriculum-based measurement is the method of examining student progress, as well as observing the effects of instruction on the student progress (Deno, 1985). The development of CBM stemmed from the need for technical adequacy and practicality regarding the “tools” used to assess the growth of student performance. CBM were designed to be short measures that would be administered on a frequent and repeated basis. Due to the developmental nature of learning, student instruction and assessment requires regular records of skills and an assessment of these skills and deficits.

While CBM is a quick and inexpensive manner to assess students’ skills, it is also important to make sure that these measures are reliable and valid measures of monitoring
student growth. In 1989, an extensive review of the literature regarding CBM was conducted. Marston (1989) revealed strong reliability in regards to test-retest reliability and alternate-form reliability for reading CBM. Validity was noted to be slightly lower, ranging from moderate to strong using criterion-related validity coefficients. A more recent review of CBM reading literature continues to support the use of CBM as a reliable and valid measure of student performance with correlations consistently in the moderate to strong range (Wayman, Wallace, Wiley, Ticha, & Espin, 2007). Further research supports the idea that CBM can be utilized as one source or predictor of student success or failure on statewide measures (Shapiro et al., 2006). A more thorough and direct examination of curriculum-based measures of reading indicate that there might be a seasonal effect in reading growth, which is larger in the lower primary grades than in the upper primary grades (Christ, Silberglitt, Yeo, & Cormier, 2010). The seasonal effect indicates that larger growth is seen in the fall rather than in the spring. This finding would support the use of winter benchmark scores or the growth from fall to winter benchmarks as the predictors for future reading performance. Further investigation of this finding will be examined in the current study.

**Assessment of reading skills.** Reading skills can be measured in a multitude of ways, including standardized testing, informal classroom assessments, and benchmark testing. Historically, reading skills in students have been assessed through classroom tests and standardized group achievement tests. Only when significant difficulties arose, and learning disabilities were suspected, were students assessed through standardized individual achievement tests (Shapiro & Derr, 1987).
In the past, standardized tests such as the Wide Range Achievement Test-Revised (Jastak & Wilkinson, 1984) and the Peabody Individual Achievement Test-Revised (Markwardt, 1989) were analyzed regarding their correlation to reading curricula. Through hypothetical analysis, researchers found such standardized tests scores may not represent what is being taught, but rather what is being tested (Armbruster, Stevens, & Rosenshine, 1977; Shapiro & Derr, 1987). Such research supports the use of curriculum-based assessment and CBM to monitor student progress and to drive instruction. With CBM, assessments are performed at benchmark intervals for students in an attempt to target those students in need of more support to reach the identified goal. The benchmarks are set to follow the reading development of a typically-developing child and how they are thought to process through the stages of learning to read. Measures such as the DIBELS (Good and Kaminski, 2002) and AIMSweb (Edformation, 2005) begin assessing students at the start of the kindergarten year through elementary school. Updated versions of DIBELS are now able to determine the level of reading proficiency in preschool students. This change will give educators the ability to predict the reading of individuals at a much younger age. Early literacy skills performance is the current focus of this study, specifically on the basic early literacy skills of phonemic awareness, alphabetic principle, and fluency, and their connection to future reading success.

*Phonemic awareness measurement.* Several forms of phonemic segmentation tools are utilized to assess phonemic awareness, such as phoneme isolation, phoneme identity, phoneme categorization, phoneme blending, phoneme segmentation, and phoneme deletion. Each type of assessment allows the student to practice skills associated with phonemic awareness. Initial sound fluency is one well-known
individualized assessment used to assess a child’s recognition and production of initial sounds in orally-presented words during the kindergarten year (Kaminski & Good, 1996; 1998). Another well-known form of measurement of phonemic awareness is PSF. Both ISF and PSF are presented orally to students to assess for the basic skill of phonemic awareness. Additionally, these are found on the DIBELS in kindergarten (ISF and PSF) and grade 1 (PSF).

**Alphabetic principle measurement.** The ability to identify letter-sound associations is developed earlier than the ability to apply letter-sound associations to words read (Ritchey, 2008). Therefore, while ISF and PSF are utilized to measure phonemic awareness, NWF is used assess letter-sound associations in words in text. Nonsense word fluency is designed to determine if a student has an effective strategy for reading any novel word. It was not designed to determine which letter sounds or words students can read. For this reason, pseudowords are used to ensure students are not reading or recalling memorized words. This skill is typically measured from the middle of kindergarten to the middle of first grade. The 2008 study by Harn, Stoolmiller, and Chard found that students who displayed unitization of sounds not only scored higher on NWF but read significantly more fluently both in the winter (i.e., 11 points higher on ORF) and in the spring (i.e., 29 points higher on ORF) of first grade than students without this strategy. Even those students who used a partial blending strategy scored higher than students who utilized the sound-by-sound or recoding approach to reading. Harn and colleagues (2008) further indicated that students using the unitization approach would be in the consolidated alphabetic phase proposed by Ehri, indicating they are able to recognize words in their entirety versus individual sound segments. An additional
measure utilized to assess letter-sound knowledge is LSF. On the basis of the total score, correct sounds per minute, LSF and NWF appear to demonstrate similar predictive relationships with reading outcomes (Ritchey, 2008). Together or separate, measurements of phonemic awareness and alphabetic principle can be predictive of performance on measures of fluency, vocabulary, and comprehension.

**Fluency measurement.** The Test of Reading Fluency (TORF; Children’s Educational Services, Inc., 1987) has been the basis of many classroom and individualized checks of fluency in the past. Reading fluency is measured by taking into consideration the individual’s rate and accuracy of reading. Oral reading fluency (ORF), as a curriculum-based measure, is currently the most common method used to assess fluency. Additionally, numerous formal standardized assessments contain measures of ORF as a subtest that are utilized in formal evaluations or in progress monitoring of students with known reading difficulties.

As mentioned above, ORF is the most widely researched and recognized CBM of reading. Although it is the most widespread CBM, it is also the most vulnerable to criticism. Ardoin and Christ (2009) report that the variability of passage difficulty on measures of ORF takes away from accurately representing growth. Concerns with the readability of passages and the level at which “cut scores” are set continues to be a focus of many researchers. Wayman, Wallace, Wiley, Ticha, and Espin (2007) also report concerns regarding the influence of passage characteristics and the validity of measures in kindergarten and Grade 1, as well as in Grades 6-12. Although cautions and concerns arise, research shows that ORF is an important predictor of future reading success (Fuchs et al., 2001; Shinn, 1998). Further, regardless of the concerns raised, the flexibility and
durability of CBM supports the use of CBM for progress monitoring and screening purposes in the area of reading. Screening for reading difficulties is extremely important as educators continue to work toward the goal of proficiency in elementary school.

**High-Stakes Assessments**

NCLB calls for increased accountability and high-stakes testing, challenging schools to ensure that all students become proficient readers by the end of the third grade (NCLB, 2002). Struggling readers have become a priority of the state and federal government, but how are they looking to measure the reading skills of students? With the use of CBM to guide instruction in the schools, districts are able to track student progress prior to the administration of any high-stakes testing and adjust instruction as necessary.

The accountability of school districts is now measured based on standards and criterion-based assessments due to the standards-based educational reform movement that started in the 1980s. At that time, states were responsible for the development of academic standards outlining what knowledge students should have and what they could do with the knowledge, and then creating assessments that were aligned to the standards to measure the knowledge. Accountability for implementing these measures throughout the states, as well as supports to improve the quality of teachers, were all part of the new reform (NCLB, 2002).

As stated above, the proficiency of young learners begins in the third grade in the Commonwealth of Pennsylvania. In the spring of Grade 3 students are currently assessed in both reading and mathematics for the first time as part of the Pennsylvania System of School Assessment (PSSA; Pennsylvania Department of Education, 2002). Advantages and disadvantages are noted with the first high-stakes measurement period. The concern
with waiting until the third grade for the first formal assessment is that unless other assessments have been administered prior to third grade, students with skill deficits will be gravely behind their peers. The issue with assessing reading at an earlier time frame than third grade is that the foundational reading skills will not be mastered prior to the third grade. In an attempt to address this major concern, progress monitoring tools, curriculum based assessments and universal screening tools are now widely used among schools (Jimerson et al., 2007).

**Curriculum-Based Measurement and Future Reading Success**

Curriculum-based measurement in the area of reading is one of most widely researched areas in education. Studies continually indicate the need for early identification of students at risk for developing reading problems. Reading performance has been measured by nationally-normed achievement tests such as the Iowa Test of Basic Skills (ITBS; Hoover, Dunbar, & Frisbe, 2005) and the Stanford Achievement Test-Tenth Edition (SAT-10; Harcourt Brace, 2003), which have been administered as local assessments to measure student performance in early elementary grades. As group achievement tests have been widely used, student performance on reading CBM (R-CBM) and those achievement tests have been available for analysis of correlations throughout the years. Overall, concurrent and construct validity of ORF CBM with norm-referenced standardized reading tests has been established (Deno, 1985; Fuchs & Deno, 1992; Hintze, Owen, Shapiro, & Daly, 2000). The high levels of validity have prompted schools to use CBMs more frequently than the lengthy and more expensive nationally-normed reading assessments.
Changes in accountability at the federal and state level, specifically the passage of NCLB, have prompted school districts to administer these more specific and valid local assessments in an attempt to predict student progress on the statewide achievement tests. Curriculum-based measurements have been utilized in the past to measure student progress and rate of growth on specific skill areas in regards to special education referrals or the need for Title I services. Due to the ramifications of high-stakes testing, districts are now looking for assessments that will help to measure student progress and identify areas of curricular gaps or student deficits as early as possible. The assessments that are being used at the state level are based upon state established curriculum standards and are purported to be aligned with the NRP’s reading elements (Braden & Tayrose, 2008). Therefore, indentifying early measures that align with future performance would be monumental in education.

**Curriculum-Based Measures and Statewide Assessments**

Good, Simmons, and Kame’enui (2001) examined the decision making utility of the DIBELS and a first grade ORF CBM with respect to reading outcomes. The strength of relationship between CBM and high-stakes testing was further investigated. Four separate cohorts were followed from kindergarten through third grade. Correlations between earlier and later literacy skills indicated weak to strong correlations. Additionally, data indicated that students who attained previous benchmark goals were likely to attain subsequent literacy benchmark goals 90% of the time or greater. The only exception to the data involved the spring PSF, where the prediction of performance was inconsistent and uncertain. The relationship between ORF and the Oregon Statewide assessment indicated a moderate correlation. Support for a correlation between CBM and
high-stakes assessment was again indicated, with additional corroboration that early literacy skills are building blocks for one another. A need for follow up in the higher elementary grades was recommended to investigate the relationship between early literacy skills and future reading achievement performance.

Late elementary reading performance measured through ORF and high-stakes state assessment was investigated by Stage and Jacobson (2001). These researchers investigated the rate of growth in a small sample of fourth grade students \((n = 173)\). Performance on an ORF measure in September, January, and May was compared to students’ performance on the Washington Assessment of Student Learning (WASL; 1998). Additionally, the slope of student ORF performance was measured against performance on the WASL. A growth curve analysis was conducted to determine individual slope and to test for change in student slope. Results indicated that level of performance at benchmark testing was a better predictor of WASL performance than slope of ORF. Correlation between ORF performance and WASL was moderate. Further examination found the use of cut scores as a stronger predictor of failure and pass rates on the WASL than slope. The September ORF predicted failure of the WASL at 41%, while it predicted passing of the WASL at 90%. Further, use of the September ORF cut rate scores increased negative and positive predictive power by 30% over the base rates. According to this study, student rate of growth at the intermediate elementary level was not significant in determining performance on a state standardized assessment. While this study examined performance in later elementary grades, it did not examine the early literacy performance of students in grades other than fourth grade. This study limits
results to CBM conducted within the same year as later elementary grade high-stakes assessments.

As mentioned previously, ORF is the most commonly utilized CBM for reading performance. Correlations and predictability between ORF and state-mandated assessments at the elementary level have been noted throughout the nation. Buck and Torgesen (2003) investigated the predictability of DIBELS ORF on performance of third graders on high-stakes testing in Florida. Results indicated a moderate relationship ($r = .70, p < .001$) between third grade spring benchmarks of ORF and the reading comprehension portion of the Florida Comprehensive Assessment Test-Sunshine State Standards (FCAT-SSS; Florida Department of Education, 2001) and a moderate relationship ($r = .74, p < .001$) between ORF and the reading comprehension portion of Florida Comprehensive Assessment Test-Norm Referenced Test (FCAT-NFT; Florida Department of Education, 2001). Further, students who were identified as low risk or at risk on the DIBELS ORF spring benchmark were correctly classified at performing adequately or inadequately 91% and 81% of the time, respectively. Predictions for students identified as some risk show little accuracy for adequate or inadequate performance on high-stakes testing, as students are equally likely to perform at either level. This finding was not specific to this study, as previous research indicated similar results (Good et al., 2001). Continued conclusions regarding lack of predictability in the middle category supports the need for increasing early interventions in reading to move these students from some risk to low risk. A similar investigation was conducted regarding DIBELS ORF and the North Carolina end-of-grade reading test (Barger, 2003). Although a small study, correlations between ORF spring scores and an end of grade test
were again moderate and unclear in predicting performance of students who attain scores in the some risk category. Again, both studies provided limited information regarding the relationship among early literacy skills and future high-stakes reading assessments administered in different years.

Continued examination of the relationship between CBM and state mandated high-stakes testing was found when McGlinchey and Hixon (2004) replicated the work of Stage and Jacobson (2001). McGlinchey and Hixon (2004) examined the relationship between CBM oral reading probes and the Michigan Educational Assessment Program (MEAP; Michigan State Board of Education, 1999) in students in the fourth grade over an eight year period. Cut scores on the CBM were set at 100 wcpm, as previous research identified this as an appropriate cut score (Fuchs & Deno, 1982; Stage & Jacobson, 2001). Correlations between CBM reading rate and MEAP remained fairly consistent and moderate over the eight year period. Data also indicated that 74% of students who met proficiency on the MEAP read 100 wcpm or greater. This moderately strong relationship between CBM and MEAP reading provided a stronger correlation than Stage and Jacobsen (2001) found between the CBM and WASL. While this study was not longitudinal, it did investigate the correlation of the same assessments over an extended period of time and it provided support of the correlations between CBM and high-stakes reading achievement tests. Unfortunately this study is again limited to examining performance of CBM and high-stakes reading assessments administered within the same year.

Shapiro, Keller, Lutz, Santoro, and Hintze (2006) continued the research on CBM and high-stakes assessments when they examined the relationship between R-CBM and
the Pennsylvania System of School Assessment in addition to published norm-referenced achievement tests as part of their study. Students in the third, fourth and fifth grades from two different districts involved in curriculum-based norming projects were the participants in this study. Correlations between the data in the same school year in R-CBM (ORF) and the reading PSSA revealed a moderate relationship, with the strongest predictor of PSSA scores at the winter benchmark. Analysis showed that diagnostic accuracy maintained hit rates between .68 and .86. Further, positive predictive power ranged from .83 to .94 and negative predictive power spanned from .43 to .68. Statistics of this level indicate that the R-CBM are an acceptable tool for screening purposes related to statewide achievement tests. Results from this study again provided evidence that early literacy CBM from Grade 3 is correlated with high-stakes assessments, specifically the PSSA, however it does not provide evidence of correlations past the third grade. Further, as was the case in research by Stage and Jacobson (2001), results indicate that CBM were correlated with high-stakes assessments conducted in the same year.

Baker and colleagues (2008) examined the relationship between ORF and high-stakes testing in Grades 1, 2, and 3. Four separate cohorts were followed to examine the link between ORF and lower elementary grade level (first and second) high-stakes reading tests. The major focus of the study was to determine what contribution slope makes to predicting performance on an outcome, after controlling for initial level of performance. Correlations were noted to be moderate to strong. ORF slope added to the accuracy of predicting performance on high-stakes testing in year 2 of assessments, indicating that increasing slope on ORF is likely to lead to better performance on
comprehensive measures of reading. Although this study followed a longitudinal path, the investigation focused only on skills in the primary elementary grades.

The accuracy of an ORF measure was further assessed in predicting performance on third grade reading comprehension tests. Roehrig, Petscher, Nettles, Hudson, and Torgesen (2008) investigated the validity of the predictive performance of the DIBELS ORF in relation to the Florida Comprehensive Assessments Test (FCAT) and the SAT-10. Participants in the study were third grade students enrolled in Reading First elementary schools in Florida (N = 35, 207). Results indicated that DIBELS ORF (ORF) predicted reading comprehension on the FCAT-SSS equally as well as on the SAT-10. The strongest correlations were found between the third administration of the ORF (February/March), the FCAT-SSS, and the SAT-10. Although moderate to strong correlations were found, researchers examined the cut-scores identified through DIBELS to determine the possibility of more accurate predictions. ROC curve analysis indicated the recalibrated cut scores improved the predictive power of ORF scores resulting in more students appropriately identified as at risk or not at risk. No significant difference was found between several identified demographic groups. Nonbiased prediction of comprehension outcomes across groups was consistent with previous research (Hintze, Callahan, Matthews, Williams, & Tobin, 2002; Hixon & McGlinchey, 2004). This study is one of the more recent in regards to correlations between R-CBM and high-stakes assessments, but again it lacks the cohesiveness of following students over a significant period of time and into the intermediate grade levels.

**Longitudinal research of CBM and high-stakes testing.** A review of the literature indicates that there have been limited studies that investigated the longitudinal
relationship between R-CBM and statewide achievement tests. The available literature has primarily focused on the early grades and the correlation between early literacy skills and statewide assessments in the elementary grades. This emphasis on early identification is important, however little research has investigated the impact of CBM of early literacy skills and the statewide assessments in late elementary and middle school.

One of the first longitudinal studies regarding CBM-R as a predictor of state-mandated high-stakes testing was performed by Hintze and Silberglitt (2005). Data was collected on five separate cohorts from the winter of the first grade through the spring of third grade. The criterion and predictive validity of R-CBM in relation to the Minnesota Comprehensive Assessment (MCA; Minnesota Department of Education, 2003) was examined. Moderate correlations ($r = .49 - .69, p < .01$) were found with all R-CBM benchmarks and the MCA, indicating that R-CBM appears to be an efficient method for predicting passing performance on high-stakes testing as far back as the first grade. Diagnostic accuracy was analyzed through discriminant analysis, logistic regression, and ROC curves. Data analysis also considered the difference in using the R-CBM as a predictor of performance on the MCA versus using successive R-CBM benchmarks to predict MCA performance. Results indicated that using successive R-CBM benchmarks as the criterion results in greater specificity and negative predictive power and lower sensitivity and positive predictive power. Ultimately, viewing the data through whichever of the three mentioned analysis showed consistent diagnostic accuracy for R-CBM as a predictor of MCA.

Keller-Margulis and others (2008) continued to investigate the longitudinal relationships between CBM data and large scale achievement tests (i.e., PSSA and Terra
Nova) across one and two years through rate of growth and diagnostic accuracy. Results indicated moderate to strong correlations both one and two years later between CBM scores and subsequent statewide achievements tests (i.e., PSSA). Stronger correlations were found between PSSA performance and the winter and spring CBM assessment scores, a finding noted in previous research (Shapiro et al., 2006). Further, insignificant relationships between the slope of CBM and statewide achievement test data one and two years later in reading was noted with the exception of a moderate correlation between rate of growth on the first grade CBM and performance two years later on the PSSA ($r = .60$). In regards to diagnostic accuracy, success on the statewide achievement test in third grade was related to expected outcomes on the first grade spring R-CBM for 85% of the students in the cohort. Diagnostic accuracy was strong for third grade spring R-CBM and fourth grade reading PSSA (.80) and for fourth grade spring R-CBM and fifth grade reading PSSA (.76). In general, negative predictive power was greater than the level of positive predictive power across the samples. Thus, these results indicated that students who performed poorly on the CBM, and were predicted to perform poorly on the PSSA, may in fact have performed well on the PSSA. This research was one of the first to compare CBM rate of improvement across a school year to outcomes on statewide achievement tests in the same year and following year. Results of this study offered strong evidence for the long term diagnostic accuracy for CBM scores and performance on statewide achievement tests. Overall, rate of growth in first grade reading was significantly and moderately related to the third grade PSSA with the relation decreasing to nonsignificance by third grade. Further research regarding efforts to improve the
predictability and diagnostic decision making of student performance at the primary level is suggested.

Due to the limitations of high-stakes testing and the mandates for accountability, it is important to monitor student progress and have a reasonable idea of whether students’ measured performance is indicative of future performance on statewide assessments. Studies have indicated a connection between curriculum based assessments, nationally-normed achievement, and state criterion assessments tests. Research now needs to focus on how early success on high-stakes statewide assessments can be predicted, and if performance over time is indicative of success in late elementary and middle school.

**Early Literacy Skills and High-Stakes Testing: Current Study**

Individuals begin to acquire reading skills prior to entering kindergarten. Upon entrance into kindergarten educators are able to collect a baseline of students’ skills in which they can use to drive instruction and monitor growth of skills. Previous studies have focused on ORF as the primary predictor of performance on high-stakes reading assessments. Research has noted a concern when correlations between mastered skills and other measures of reading have been used as scientific evidence without regard for the lack of significant correlations among the same skills in the same individuals a year later (Paris, 2005). Vast numbers of correlational studies exist between the early literacy skill of fluency and reading achievement, with little regard to the phonemic awareness and the alphabetic principle. Overall, current research is missing information regarding performance on all early literacy skills and performance on high-stakes assessments after the initial testing in the third grade.
While Keller-Margulis and colleagues (2008) examined the CBM in only the first through third grades, they did compare CBM performance to high-stakes testing in the higher elementary grades. Overall, studies have not centered on a longitudinal data set that follows a single cohort of students to determine the predictive power of CBM starting in kindergarten and comparing this data to high-stakes testing in upper elementary and middle school grades. Goffreda and others (2009) recently examined the predictability of first grade performance on LNF, NWF, PSF and ORF on the DIBELS with Grade 3 performance on the PSSA. Results indicated that only ORF was a significant predictor of PSSA performance. Further investigation of the mentioned DIBELS measures was suggested to determine predictability in different years. Overall, long term predictability of CBM, other than ORF, has not been widely investigated. Another area in need of additional research is the inclusion of students with disabilities in the study of CBM predictability. Recent studies have eliminated students that currently have an identified disability, other than a speech and language impairment and gifted, and an active Individualized Education Plan (Shapiro et al., 2006). As NCLB indicated that all students should be proficient, it follows that all students should be examined through research.

The current study examined a cohort of students from kindergarten through the seventh grade, to examine the relationship among the early literacy skills and performance on the PSSA in Grades 3, 5, and 7. This study has strong implications for change in emphasis on R-CBM and classroom instruction at the primary grade level. Understanding the impact of early literacy skills assessed in kindergarten on performance on high-stakes assessments in the seventh grade is paramount.
CHAPTER III

METHOD

Participants

Participants were students ages five and six enrolled in kindergarten during the 2002-2003 academic year at a rural/suburban school district in southwestern Pennsylvania. The district is small in terms of enrollment (N = 2532) and consisted of four elementary schools (at the time of data collection), one junior high school and one senior high school. Elementary schools in the district house kindergarten through Grade 4, the junior high spans Grades 5 through 8, and the senior high spans Grades 9 through 12. The district historically reports 25% of the students as economically disadvantaged (based on free and reduced lunch), 15% of the population as receiving special education (excluding gifted), and less than 1% of students as having limited English proficiency. Racial/ethnic background of students within the district historically includes 95% Caucasian, 2% Black (Non-Hispanic), 1% Hispanic and less than 1% Asian/Pacific Islander students. Across the district in 2009-2010, the percentage of students who scored Proficient or Advanced on a state achievement test in reading were as follows: Grade 3 (84.6%), Grade 5 (69%), and Grade 7 (74.5%) (Data Recognition Corporation, 2009).

Data used in this study were extracted from a pre-existing database maintained by the school district. The de-identified data provided by the school district included descriptive information such as participant age, gender, race, and socioeconomic status; eligibility for special education; scores on the Dynamic Indicators of Basic Early Literacy Skills (DIBELS; Good & Kaminski, 2002) subtests (K-Grade 3); and scores on the
reading PSSA in Grades 3, 5, and 7. This study examined data for students enrolled in kindergarten at the start of the 2002-2003 school year who remained enrolled in the district through the 2009-2010 school year (Grade 7). Only participants with kindergarten and Grade 1 CBM data and Grades 3, 5, and 7 PSSA data were included in this study. The cohort of students used in this study began with 185 kindergarten students who received kindergarten fall benchmark testing. Data was available in the spring of 2010 for a total of 130 participants who remained as part of the initial data collection 7 years prior. Overall, 70% of the original participates remained over the course of eight years. All identifying information was removed from the data prior to release of the data for this study.

**Measures**

**Dynamic Indicators of Basic Early Literacy Skills, 6th Edition (DIBELS)**

DIBELS is one of the most widely used assessments for measuring early literacy skills. DIBELS is an assessment comprised of a group of five subtests designed to assess the three early literacy skills of phonemic awareness, alphabetic principle, and ORF. These skills are assessed at the start of kindergarten through the spring of the sixth grade. Phonemic awareness is assessed through ISF and PSF tasks. Nonsense word fluency is the measure of alphabetic principle skills. Lastly, ORF assesses reading fluency. These individually administered measures are standardized, and student performance is reported in terms of the degree of risk for failure to master these three early literacy skills compared to grademates. Kindergarten LNF, ISF, PSF, NWF and Grade 1 ORF data of the cohort of participants were used in this study.
**Letter Naming Fluency.** Letter Naming Fluency is based on research by Marston and Magnusson (1988) and is designed for use with students starting in the fall of the kindergarten year through the fall of the first grade year. Students have one minute to correctly identify the letter name of randomly presented uppercase and lowercase letters on one page. A student’s score is the number of correctly identified letters in one minute. In the current study, cut scores used by the district followed those indicated in the DIBELS administration manual. At risk performance was indicated by students who identified less than 2 letters (fall kindergarten), 15 letters (winter kindergarten), 29 letters (spring kindergarten) or 25 letters (fall first grade) in one minute.

Alternate-forms reliability on LNF ranges from .80 to .94 (Good et al., 2004; Hintze et al. 2003). Concurrent criterion-related validity is reported to be moderate with the Woodcock Johnson Psycho Educational Battery-Revised Readiness Cluster ($r = .57 -.71$; Woodcock, 1987), the Comprehensive Test of Phonological Processing (CTOPP) Rapid Naming, Phonological Awareness and Phonological Memory Composites ($r = .52 -.58$; Wagner, Torgesen, & Rashotte, 1999), the Test of Word Reading Efficiency (TOWRE) Sight Word Efficiency subtest ($r = .62$; Torgesen, Wagner, & Rashotte, 1997b), and the TOWRE Phonetic Decoding Efficiency subtest ($r = .47$; Good et al., 2004; Hagan-Burke, Burke, & Crowder, 2006; Hintze et al., 2003). Further, predictive criterion-related validity was indicated in several studies. Rouse and Fantuzzo (2006) reported moderate correlations ($r = .48 -.63$) with the Terra Nova Reading, Vocabulary, and Language composites (CTB/McGraw-Hill, 1997). While weak to moderate correlations were reported with the SAT-10 Comprehension ($r = .30$) and Vocabulary ($r$
= .22) as well as the Iowa Test of Basic Skills-Reading Total Index ($r = .57$; Chard et al., 2008; Schilling, Carlisle, Scott, & Zeng, 2007).

Although LNF is not identified as a formal measure of phonemic awareness, alphabetic principle, or fluency, as it involves letter names rather than sounds, there are direct relationships between LNF and NWF, sight word efficiency, and phoneme decoding efficiency. For example, the strongest path coefficient found is between LNF and NWF (.47), followed by sight word efficiency (.36) and phoneme decoding efficiency (.26), providing support of the reading acquisition theory including letter naming (Burke, Hagan-Burke, Kwok, & Parker, 2009).

**Initial Sound Fluency.** Phonemic awareness is initially assessed on the DIBELS through the ISF subtest. Initial Sound Fluency may be administered to students from the start of preschool through the winter months of kindergarten. Regarding this subtest, students are presented with four separate pages, one at a time, that have four distinct pictures displayed. When the pages are presented, the pictures are named by the examiner, and the student is then asked to identify the picture that begins with a particular sound stated by the examiner. The student is also asked to produce the beginning sound of a word presented orally and that matches one of the four pictures. The time allotment to complete the ISF task is three minutes. The timer is initially set for 180 seconds and then counts down after each question has been asked by the examiner. The total score is then reported in terms of correct responses per minute (i.e., number of correct responses multiplied by 60 and then divided by the number of seconds necessary to complete the task). The DIBELS administrative manual characterizes at risk performance as less than four initial sounds in one minute in the fall of kindergarten. Students who identified less
than 10 initial sounds in 1 minute are identified as deficient in ISF in the winter of kindergarten.

Initial Sound Fluency alternate-form reliability is reported between .61 and .95, indicating moderate to very strong reliability (Good et al., 2004; Hintze et al., 2003). Weak to moderate predictive validity ($r = .37 - .44$) is found between ISF and the Woodcock Johnson Total Reading Cluster Standard Score (Good et al., 2004). Discriminate validity of ISF with CTOPP Rapid Naming Composite (Wagner et al., 1999) is reportedly low ($r = 20$), while moderate concurrent validity is displayed between the CTOPP Phonological Awareness Composite ($r = .60$) and the CTOPP Phonological Memory Composite ($r = .46$; Hintze et al., 2003). These studies indicate strong relationships between the Ellison, Blending Words, Sound Matching, and Nonword Repetition subtests of the CTOPP. Overall, moderate concurrent criterion validity was found between ISF and the Phonological Awareness Composite of the CTOPP in kindergarten students during the winter benchmarks (Good et al., 2004; Hintze et al., 2003).

Initial Sound Fluency in kindergarten has been found to be predictive of reading performance in the first and second grades (Burke et al., 2009). Specifically, ISF was found to moderately correlate ($r = .45$) with sight word performance on the TOWRE in the first grade, and moderately correlate ($r = .46$) with the Passage Comprehension subtest on the Woodcock Reading Mastery Test-Revised in the second grade.

**Phoneme Segmentation Fluency.** The PSF subtest is the second measure of phonemic awareness on the DIBELS. Students may be assessed using PSF starting in the middle of the kindergarten year through the end of the first grade year. Phoneme
Segmentation Fluency measures a student’s ability to segment three and four-phoneme words into their individual phoneme, the smallest unit of sound within words. Words are presented orally to the child and they are then required to verbally produce the corresponding phonemes of the words. Students have one minute to correctly identify the phonemes. Scores are calculated by the number of correct phonemes provided. In the current study, cut scores for the district were determined by the DIBELS administration manual indicating at risk performance for students who identified less than 7 phonemes correct per minute (winter kindergarten) and identified students with a skill deficit when less than 10 phonemes correct per minute in the spring of kindergarten through the spring of Grade 1.

Good and colleagues (2004) reported strong to very strong alternate forms reliability (r = .74 - .90). Strong alternate forms reliability was also reported for PSF, at .79 and .88 (Kaminski & Good, 1996). Concurrent, criterion validity was noted as moderate (r = .54) for PSF reported at the kindergarten spring benchmark and the Woodcock Johnson Psycho-Educational Battery Readiness Cluster score (Good et al., 2004).

Nonsense Word Fluency. The only measure of alphabetic principle on the DIBELS is the NWF subtest. Nonsense Word Fluency is assessed from the middle of kindergarten through the beginning of the second grade in the DIBELS system. Students are presented with one page of randomly ordered consonant-vowel-consonant (CVC) CVC and vowel-consonant (VC) “nonsense words” (e.g., vug, lut, ov) and are asked to verbally produce the word as a whole or to produce the individual letter sounds of each word. One minute is allotted for the student to produce as many correct sounds as
possible. Scores are calculated by the number of correct letter sounds in isolation or correct phonologically recoded words. Students identifying less than the predetermined number of letter sounds are identified as at-risk using the following criteria: (a) less than 5 in winter kindergarten; (b) 15 in spring kindergarten; and (c) 13 in fall Grade 1. Students identifying less than 30 letter sounds in the winter and spring of Grade 1 are identified as having a deficit in this skill.

Very strong test-retest reliability ($r = .94$) and moderate to strong alternate-forms reliability ($r = .67 - .88$) have been reported for NWF (Good et al., 2004; Harn et al., 2008; Speece, Mills, Ritchey, & Hillman, 2003; Ritchey, 2008). The concurrent validity of the kindergarten NWF is .65 with the CTOPP and .91 with the Woodcock Johnson-Revised (WJ-R) Letter Word Identification Subtest (Speece et al., 2003). Concurrent validity at the first grade level is similar, with moderate to strong reported values of .68 and .75 in relation to the TOWRE and a range of moderate concurrent validity of .54 to .60 with the Iowa Test of Basic Skills (ITBS; Burke & Hagan-Burke, 2007; Schilling et al., 2007). Predictive concurrent validity has been reported at similar moderate rates between .51 and .67 with the TOWRE, WJ-R and ITBS (Burke et al., 2009; Schilling et al., 2007; Speece et al., 2003).

**Oral Reading Fluency.** Dynamic Indicators of Basic Early Literary Skills ORF passages and procedures are based on the work of Deno (1985) and Shinn (1989). Students may be assessed using this measure starting in the winter of first grade through the spring of the sixth grade. Students are provided three separate passages and asked to read each passage aloud for one minute. Omissions, substitutions, and hesitations of more than three seconds are marked as an error. The ORF rate is the number of correct
words read per minute. The median score of the three passages is recorded as the overall ORF score. At risk performance, as stated in the DIBELS administration manual, is noted when students read less than 8 words correct per minute in the winter of the first grade and less than 20 words correct per minute in the spring of the first grade.

Test-retest reliability for R-CBM is strong to very strong and has ranged from .92 to .97 with alternate forms reliability ranging from .89 to .94 (Tindal, Marston, & Deno, 1983). Reliability measures reported in the DIBELS Technical Adequacy Information Report indicates reliability rates of ORF ranging from .89 to .99 (Dynamic Measurement Group, 2008). Concurrent criterion-related validity in the first through third grades ranged from .31 to .97 with the Test of Reading Fluency-TORF, ITBS, and TOWRE (Burke et al., 2009; Schilling et al., 2007; Warrington, 2003). Predictive criterion-related validity indicates moderate to very strong correlation with a smaller range from .58 to .94 with the SAT-10, ITBS, and 4Sight (Baker et al., 2008; Schilling et al., 2007; Shapiro et al., 2008; Success for All Foundation, 2007).

**Pennsylvania System of School Assessment (PSSA)**

The PSSA is the criterion-referenced measure designed for educational accountability purposes in Pennsylvania (Pennsylvania Department of Education, 2002). Student performance is assessed in the areas of reading, writing, mathematics, and science at specific grade levels mandated by the state. Reading assessment is required of all Pennsylvania students in the third through eighth and eleventh grades. The PSSA is published in several different formats and alternate forms per grade level. Scoring of the PSSA is conducted by the Data Recognition Company (DRC), a professional testing service in Minnesota. Performance is classified into four categories: (a) Advanced; (b)
Proficient (c) Basic, and (d) Below Basic. Scores in the proficient and advanced ranges are considered “passing” in the Commonwealth of Pennsylvania. The Pennsylvania Department of Education approved the following specific criteria for Advanced, Proficient, Basic, and Below Basic levels of performance (Pennsylvania Department of Education, 2010, General Performance Descriptors, para. 1-4):

1. The *Advanced* descriptor reflects superior academic performance. Advanced work indicates an in-depth understanding and exemplary display of the skills included in the Pennsylvania Academic Content Standards.

2. The *Proficient* descriptor reflects satisfactory academic performance. Proficient work indicates a solid understanding and adequate display of the skills included in the Pennsylvania Academic Content Standards.

3. The *Basic* descriptor reflects marginal academic performance. Basic work indicates a partial understanding and limited display of the skills included in the Pennsylvania Academic Content Standards. This work is approaching satisfactory performance but has not yet reached it. There is a need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient Level.

4. The *Below Basic* descriptor reflects inadequate academic performance. Below Basic work indicates little understanding and minimal display of the skills included in the Pennsylvania Academic Content Standards. There is a major need for additional instructional opportunities and/or increased student academic commitment to achieve the Proficient Level.
Regarding the current study, the participants completed the Grade 3 PSSA during the 2005-2006 school year. Reading questions consisted of multiple choice and open-ended response. Five unique forms of the reading portion were developed and circulated for assessment. For the 2006 Grade 3 Reading domain, the internal consistency reliability of the multiple choice portion was calculated by Cronbach’s alpha and ranged from .91 to .92 with a standard error of measurement that ranged from 2.45 to 2.61 (CTB/McGraw-Hill, 2006). Open-ended questions were found to have a very high degree of inter-rater agreement with percentages ranging from 95.29 to 99.59. Validity was further calculated through Cohen’s kappa indicating good agreement with kappa of 0.40 – 0.65. Thus, appropriate, meaningful, and useful inferences may be made from the test results. Performance levels set for the third grade reading PSSA for the 2006 spring administration were as follows: (a) Below Basic, 1097 and below; (b) Basic, 1098-1234; (c) Proficient, 1235-1441; and (d) Advanced, 1442 and above (CTB/McGraw-Hill, 2006).

The participants for this study were administered the Grade 5 PSSA in the spring of 2008. Regarding the reading domain, reliability was noted to fall between .76 and .91 with a standard error of measurement of 3.00 (Data Recognition Corporation, 2009). Internal validity as calculated by Pearson’s correlation coefficient was reported for both Form A and Form B between .76 and .93. Performance levels indentified for the 2008 fifth grade reading PSSA were as follows: (a) Below Basic, 700-1136; (b) Basic, 1137-1274; (c) Proficient, 1275-1496; and (d) Advanced, 1497 and above.

The participants were administered the Grade 7 PSSA in the spring of 2010. Reliability and validity data regarding the reading domain remained similar to previous years’ administration of grade 7 PSSA reading domain. Reliability for the reading
domain was strong to very strong, reported between .82 and .90 for the 2010 administration. Additionally, standard error of measurement was reported between 2.1 and 3.0. Further, strong validity ($r = .83$) was also reported for the spring 2010 administration of the grade 7 reading domain and the Pennsylvania Assessment Anchors for Reading (Data Recognition Corporation, 2010). Performance levels, identified in 2007 for the seventh grade PSSA, remained consistent for the 2010 administration of the reading PSSA. The cuts scores were as follows: (a) Below Basic, 700-1130; (b) Basic, 1131-1278; (c) Proficient, 1279-1469; and (d) Advanced, 1470 and up.

**Research Design**

This study used a sample of convenience. Correlational techniques were used to examine the relationships among winter assessments of DIBELS ISF, LNF, PSF, NWF, ORF, and PSSA Reading. This investigation also examined the ability of the aforementioned DIBELS measures to predict PSSA Reading domain performance across time (i.e., Grades 3, 5, and 7). Descriptive statistics regarding mean and standard deviation were investigated.

**Procedure**

DIBELS data were collected prior to this study as part of universal screening efforts of a district. School psychologists, instructional support teachers, Title I Reading teachers, regular education teachers in kindergarten through third grade, special education teachers, and Title I Reading paraprofessionals were trained per the recommendations of the DIBELS Administration Manual to administer and score the DIBELS measures. DIBELS data were collected on three separate occasions (fall, winter, and spring) across
each school year. Specific benchmark measures and the time period in which DIBELS measures were administered within the district are detailed in Table 1.

Table 1

_DIBELS Benchmark Testing_

<table>
<thead>
<tr>
<th>Grade</th>
<th>Fall</th>
<th>Winter</th>
<th>Spring</th>
</tr>
</thead>
<tbody>
<tr>
<td>K</td>
<td>ISF</td>
<td>ISF</td>
<td>LNF</td>
</tr>
<tr>
<td></td>
<td>LNF</td>
<td>LNF</td>
<td>PSF</td>
</tr>
<tr>
<td></td>
<td>PSF</td>
<td>NWF</td>
<td>NWF</td>
</tr>
<tr>
<td>1</td>
<td>LNF</td>
<td>PSF</td>
<td>PSF</td>
</tr>
<tr>
<td></td>
<td>PSF</td>
<td>NWF</td>
<td>NWF</td>
</tr>
<tr>
<td></td>
<td>NWF</td>
<td>ORF</td>
<td>ORF</td>
</tr>
<tr>
<td>2</td>
<td>NWF</td>
<td>ORF</td>
<td>ORF</td>
</tr>
<tr>
<td>3</td>
<td>ORF</td>
<td>ORF</td>
<td>ORF</td>
</tr>
</tbody>
</table>

*Note.* LNF = Letter Naming Fluency; ISF = Initial Sound Fluency; PSF = Phoneme Segmentation Fluency; NWF = Nonsense Word Fluency; ORF = Oral Reading Fluency

The participants’ PSSA Reading domain performances were used in this study. The cohort of students for this study were administered the PSSA Reading domain in the spring (March or April) of Grades 3, 4, 5, 6, and 7. Data for the third, fifth, and seventh grades were utilized in this study. Standardized administration as detailed in the PSSA Administration Manual was followed for each year.

After Institutional Review Board approval, the Director of Curriculum and Instruction of the school district was contacted by formal letter to request DIBELS and PSSA data for the above mentioned grades and years. Data was requested in a Microsoft Excel file, with all identifying information removed prior to release for analysis.
Descriptive statistics such as sex, age, socioeconomic status, and disability status were also requested.

**Data Analysis**

**Participant Demographics and Outliers**

Analyses for this study were conducted using SPSS 19.0. The study investigated the relationship between CBM of early literacy and performance on criterion-referenced, statewide reading assessments over an eight year period. Frequency analyses, correlation analyses, and descriptive statistics were computed to describe the sample and reading performance across time. Mean age was reported for each assessed time period. Correlation analyses examined the relationship between the winter R-CBM data from the 2002-2003 and 2003-2004 school years and the PSSA reading data in the 2005-2006, 2007-2008, and 2009-2010 school years. Only participants with full longitudinal data were included in this study. Further, individuals with data points that were extreme in value (i.e., outliers), as determined through the use of Mahalanobis distance analyses, were eliminated.

**Research Question One and Analysis**

What are the relationships among performance on DIBELS early literacy probes (ISF, LNF, PSF, and NWF) administered at the winter benchmark in kindergarten, ORF administered at the winter benchmark in Grade 1, and PSSA Reading performance in Grades 3, 5, and 7? DIBELS scores were reported as raw scores and PSSA scores were coded according to scaled scores. Pearson correlation coefficients were computed as all data is continuous. It was hypothesized that performance on early measures of the DIBELS and performance on the PSSA Reading domain would be moderately correlated.
**Research Question Two and Analysis**

To what degree do ISF, LNF, PSF, and NWF probes administered in kindergarten and ORF administered in Grade 1 predict PSSA Reading performance in Grades 3, 5, and 7, respectively? Studies exist that focus on the winter CBM data and the predictability of performance on high-stakes measures of reading. Likewise, winter benchmark was thought to be the most representative of early literacy performance (Shapiro et al., 2006).

Three separate stepwise regressions, one at each grade level, were used to determine CBM measures that are most predictive of Grades 3, 5, and 7 PSSA Reading domain performances. The CBM raw scores of ISF, LNF, PSF, NWF and ORF were compared to Grades 3, 5, and 7 PSSA Reading domain scaled scores. It was hypothesized that performance on ORF at the winter Grade 1 benchmark would be most predictive of Grade 3 PSSA Reading domain performance. Second, it was hypothesized that performance on ORF at the winter Grade 1 benchmark would be most predictive of Grade 5 PSSA Reading domain performance. Lastly, performance on ORF at the winter Grade 1 benchmark was hypothesized to be the most predictive of Grade 7 PSSA Reading domain performance.

**Research Question Three and Analysis**

Using Receiver Operator Characteristic (ROC) curve analysis, what is the diagnostic accuracy (positive predictive power, negative predictive power, sensitivity, and specificity) of the early R-CBM measures that were found to be most predictive of Grades 3, 5, and 7 PSSA Reading performances, respectively?

Receiver operator characteristic curves were used in order to determine more accurate cut scores on the five R-CBM measures. A ROC curve is a statistical tool used
to assess the accuracy of predictions (Gönen, 2007). The ROC curves are a visual way to summarize the accuracy of predictions in regards to sensitivity and specificity. Recent studies have utilized ROC procedures with higher results in sensitivity and specificity, as well as positive and negative predictive power (Hintze & Silberglitt, 2005; Margulis-Keller et al., 2008; Shapiro et al., 2006). The ROC-determined cut point for the winter data point of the five subtests was used to examine the lasting relationship between the DIBELS and the PSSA. CBM winter cut scores were compared to performance on the third, fifth, and seventh grade reading PSSA to determine diagnostic accuracy. The dependent variables were dichotomous, indicating proficient or not proficient performance on the PSSA.

Diagnostic accuracy was analyzed through the descriptive statistics of sensitivity and specificity, as well as positive and negative predictive power. Sensitivity is the likelihood that the CBM score will identify, with reasonable accuracy, students who are not proficient on the PSSA. Specificity refers to the probability that a student who is proficient on the PSSA would have been predicted to be proficient on the PSSA based on low risk CBM scores. Negative predictive power is the probability that students identified as at risk on the CBM will also be not proficient on the PSSA, while positive predictive power is the probability that students identified as low risk on the CBM will be proficient on the PSSA. It is hypothesized that DIBELS cut scores for ISF, LNF, PSF, NWF, and ORF measures that are identified as most predictive of Grade 3, 5, and 7 PSSA Reading domain performance, as identified by ROC curves would in turn predict proficiency on the Grade 3, 5, and 7 PSSA Reading domain.
CHAPTER IV
RESULTS

This chapter presents findings of statistical analyses that were conducted to examine this study’s three research questions. The initial objective was to determine the relationship between early literacy skills and future reading performance. Pearson correlations are presented in a correlation matrix to display these relationships. Predictability of PSSA reading performance was investigated separately at Grades 3, 5, and 7 through multiple regression, using early literacy skills as predictors. Finally, diagnostic accuracy of early literacy measures in predicting future PSSA reading performance was calculated utilizing receiver operating characteristic curves (ROC).

Prior to running these analyses, descriptive statistics were obtained.

Descriptive Statistics

Data provided for this study involved 130 students who were administered winter kindergarten DIBELS probes (ISF, LNF, PSF, and NWF), winter DIBELS ORF in Grade 1, as well as the Spring PSSA in Grades 3, 5, and 7. The original cohort of students included 185 students, but through attrition, only 130 met criteria for this study and were provided in the data set. Mean age for students at the winter kindergarten DIBELS administration was 5 years, 7 months. Table 2 displays descriptive statistics including means and standard deviations for each CBM and PSSA administration.
Table 2

Descriptive Statistics

<table>
<thead>
<tr>
<th>Measure</th>
<th>M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISF-WK</td>
<td>24.95</td>
<td>14.26</td>
</tr>
<tr>
<td>LNF-WK</td>
<td>32.78</td>
<td>16.06</td>
</tr>
<tr>
<td>PSF-WK</td>
<td>16.79</td>
<td>12.17</td>
</tr>
<tr>
<td>NWF-WK</td>
<td>15.67</td>
<td>13.33</td>
</tr>
<tr>
<td>ORF-W1</td>
<td>36.39</td>
<td>29.05</td>
</tr>
<tr>
<td>PSSA3</td>
<td>1427.96</td>
<td>190.17</td>
</tr>
<tr>
<td>PSSA5</td>
<td>1380.25</td>
<td>177.59</td>
</tr>
<tr>
<td>PSSA7</td>
<td>1425.12</td>
<td>189.87</td>
</tr>
</tbody>
</table>

Note. ISF-WK = Initial Sound Fluency (winter kindergarten); LNF-WK = Letter Naming Fluency (winter kindergarten); PSF -WK = Phoneme Segmentation Fluency (winter kindergarten); NWF-WK = Nonsense Word Fluency (winter kindergarten); ORF -WK= Oral Reading Fluency (winter kindergarten); PSSA3 = Grade 3 Reading PSSA; PSSA5 = Grade 5 Reading PSSA; PSSA7 = Grade 7 Reading PSSA.

Research Question One Results

The first research question utilized a correlation matrix to examine the relationships among performance on DIBELS early literacy probes (ISF, LNF, PSF, NWF, and ORF), administered in the winter of kindergarten and Grade 1, with PSSA reading performance in Grades 3, 5, and 7. Pearson correlations for these measures are contained in Table 3. All early literacy probes and future reading performance indicators were found to be significantly and positively correlated at the $p < .01$ level. Winter kindergarten PSF and Grade 7 PSSA Reading performance displayed the lowest correlation with $r = .27$, while winter Grade 1 ORF and Grade 5 PSSA Reading performance displayed the greatest correlation at $r = .52$. The strongest correlation
among early literacy probes was found with LNF and NWF ($r = .77$), both administered at the winter kindergarten benchmark. The strongest correlation among PSSA Reading performance was found with Grades 5 and 7 ($r = .77$).

Table 3

*Pearson Correlations among CBM Probes and State Criterion Test Reading Performance*

<table>
<thead>
<tr>
<th>Measure</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ISF-WK</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. LNF-WK</td>
<td>.38*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. PSF-WK</td>
<td>.43*</td>
<td>.38*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. NWF-WK</td>
<td>.40*</td>
<td>.77*</td>
<td>.30*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. ORF-W1</td>
<td>.33*</td>
<td>.71*</td>
<td>.25*</td>
<td>.74*</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. PSSA3</td>
<td>.41*</td>
<td>.40*</td>
<td>.43*</td>
<td>.38*</td>
<td>.46*</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. PSSA5</td>
<td>.29*</td>
<td>.47*</td>
<td>.31*</td>
<td>.41*</td>
<td>.52*</td>
<td>.73*</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>8. PSSA7</td>
<td>.33*</td>
<td>.48*</td>
<td>.27*</td>
<td>.41*</td>
<td>.45*</td>
<td>.67*</td>
<td>.77*</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note.* ISF-WK = Initial Sound Fluency (winter kindergarten); LNF-WK = Letter Naming Fluency (winter kindergarten); PSF-WK = Phoneme Segmentation Fluency (winter kindergarten); NWF-WK = Nonsense Word Fluency (winter kindergarten); ORF-W1 = Oral Reading Fluency (winter Grade 1); PSSA3 = Grade 3 Reading PSSA; PSSA5 = Grade 5 Reading PSSA; PSSA7 = Grade 7 Reading PSSA. *$p < .01$

**Research Question Two Results**

The second research question utilized three stepwise regression procedures to determine which of the early literacy indicators were most predictive of Grade 3, 5, and 7 PSSA reading performance, respectively. As further described below, predictor variables
were the five early literacy indicators and the criterion variable was PSSA Reading performance at each grade.

Assumptions for multiple correlations were satisfied. First, dependent variables, Grade 3, 5, and 7 PSSA Reading Performance, were normally distributed for each combination of the independent variables (see Figures 1, 2, and 3).

*Figure 1*. PSSA 3 reading performance distribution.
Second, the population variances of the dependent variable are the same for all levels of the independent variables, as the sample of 130 students was constant through
analysis of the data. Lastly, the sample is representative of a random sample of the population and scores are independent of each other from one individual to the next.

**Research Question 2a Results**

For the first stepwise regression, ISF, LNF, PSF, NWF, and ORF were entered as predictor variables and Grade 3 PSSA Reading performance was entered as the criterion variable. Results indicate that ORF was significantly and positively related to Grade 3 PSSA Reading performance, $R^2 = .213$, $R^2_{adj} = .213$, $F(1, 128) = 34.56, p < .001$. Twenty one percent of the variance of Grade 3 PSSA Reading performance can be accounted for by ORF alone. PSF then contributed an additional 10.5% to the variance in Grade 3 PSSA Reading performance, $\Delta R^2 = .105$. Together, ORF and PSF accounted for a total of 31.7% of the variance in Grade 3 PSSA Reading performance. Finally, ISF accounted for an additional 2.7% of variance, $\Delta R^2 = .027$. Overall, a total of 34.4% of the variance in Grade 3 PSSA Reading performance was accounted for by performance on ORF, PSF, and ISF. The other CBM did not further predict Grade 3 PSSA Reading performance beyond that accounted for by ORF, PSF, and ISF. Tables 4 and 5 further reference the multiple regression predicting Grade 3 PSSA Reading performance.
Table 4

*Model Summary for Research Question 2a*

<table>
<thead>
<tr>
<th>Step</th>
<th>$R$</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>$\Delta R^2$</th>
<th>$F_{chg}$</th>
<th>$df_1$</th>
<th>$df_2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ORF-W1</td>
<td>.461</td>
<td>.213</td>
<td>.206</td>
<td>.213</td>
<td>34.564</td>
<td>1</td>
<td>128</td>
<td>.000</td>
</tr>
<tr>
<td>2. PSF-WK</td>
<td>.563</td>
<td>.317</td>
<td>.307</td>
<td>.105</td>
<td>19.495</td>
<td>1</td>
<td>127</td>
<td>.000</td>
</tr>
<tr>
<td>3. ISF-WK</td>
<td>.587</td>
<td>.344</td>
<td>.329</td>
<td>.027</td>
<td>5.196</td>
<td>1</td>
<td>126</td>
<td>.024</td>
</tr>
</tbody>
</table>

*Note.* ORF – W1 = Oral Reading Fluency (winter Grade 1); PSF - WK = Phoneme Segmentation Fluency (winter kindergarten); ISF-WK = Initial Sound Fluency (winter kindergarten).

Table 5

*Coefficients for Final Model for Research Question 2a*

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
<th>Bivariate $r$</th>
<th>Partial $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF-W1</td>
<td>2.175</td>
<td>.332</td>
<td>4.310</td>
<td>.000</td>
<td>.461</td>
<td>.358</td>
</tr>
<tr>
<td>PSF-WK</td>
<td>4.136</td>
<td>.265</td>
<td>3.285</td>
<td>.001</td>
<td>.429</td>
<td>.281</td>
</tr>
<tr>
<td>ISK-WK</td>
<td>2.511</td>
<td>.188</td>
<td>2.279</td>
<td>.024</td>
<td>.412</td>
<td>.199</td>
</tr>
</tbody>
</table>

*Note.* ORF – W1 = Oral Reading Fluency (winter Grade 1); PSF - WK = Phoneme Segmentation Fluency (winter kindergarten); ISF-WK = Initial Sound Fluency (winter kindergarten).
Research Question 2b Results

Next, when the criterion variable was Grade 5 PSSA Reading performance, again the predictors were ISF, LNF, PSF, NWF, and ORF. Results indicate that ORF was significantly and positively related to Grade 5 PSSA Reading performance, $R^2 = .275$, $R^2_{adj} = .269$, $F(1, 128) = 48.511$, $p < .001$. Twenty seven and one half percent of the variance of the Grade 5 PSSA performance can be accounted for by ORF alone. PSF then contributed an additional 3.2% to the variance, $\Delta R^2 = .032$. Together, ORF and PSF combined for a total of 30.7% of the variance of Grade 5 PSSA Reading performance. Again, the other CBM did not further predict Grade 5 PSSA Reading performance beyond that accounted for by ORF and PSF. Tables 6 and 7 further reference the multiple regression predicting Grade 5 PSSA Reading performance.
Table 6

*Model Summary for Research Question 2b*

<table>
<thead>
<tr>
<th>Step</th>
<th>$R$</th>
<th>$R^2$</th>
<th>$R_{adj}^2$</th>
<th>$\Delta R^2$</th>
<th>$F_{chg}$</th>
<th>$df_1$</th>
<th>$df_2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. ORF-W1</td>
<td>.524</td>
<td>.275</td>
<td>.269</td>
<td>.275</td>
<td>48.511</td>
<td>1</td>
<td>128</td>
<td>.000</td>
</tr>
<tr>
<td>2. PSF-WK</td>
<td>.554</td>
<td>.307</td>
<td>.296</td>
<td>.032</td>
<td>5.936</td>
<td>1</td>
<td>127</td>
<td>.016</td>
</tr>
</tbody>
</table>

*Note.* ORF –W1= Oral Reading Fluency (winter Grade 1); PSF -WK= Phoneme Segmentation Fluency (winter kindergarten).

Table 7

*Coefficients for Final Model for Research Question 2b*

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
<th>Bivariate $r$</th>
<th>Partial $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF-W1</td>
<td>2.920</td>
<td>.478</td>
<td>6.258</td>
<td>.000</td>
<td>.524</td>
<td>.485</td>
</tr>
<tr>
<td>PSF-WK</td>
<td>4.136</td>
<td>.265</td>
<td>3.285</td>
<td>.001</td>
<td>.306</td>
<td>.211</td>
</tr>
</tbody>
</table>

*Note.* ORF –W1= Oral Reading Fluency (winter Grade 1); PSF -WK= Phoneme Segmentation Fluency (winter kindergarten).
Research Question 2c Results

Lastly, when Grade 7 PSSA Reading performance was the criterion variable, ISF, LNF, PSF, NWF, and ORF were again used as the predictors. Results indicate that LNF was significantly and positively related to Grade 7 PSSA Reading performance, $R^2 = .231$, $R^2_{adj} = .225$, $F(1, 128) = 38.36, p < .01$. Twenty three percent of the variance of the Grade 7 PSSA Reading performance can be accounted for by LNF alone. ISF then contributed an additional 2.7% to the variance, $\Delta R^2 = .027$. Overall, LNF and ISF combined contributed to a total of 25.8% of the variance of Grade 7 PSSA Reading performance. The other CBM (PSF, NWF, and ORF) did not further add to the predictive value of Grade 7 PSSA reading performance. Tables 8 and 9 further reference the multiple regression predicting Grade 5 PSSA Reading performance.
Table 8

*Model Summary for Research Question 2c*

<table>
<thead>
<tr>
<th>Step</th>
<th>$R$</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>$\Delta R^2$</th>
<th>$F_{ch}$</th>
<th>$df_1$</th>
<th>$df_2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. LNF-WK</td>
<td>.481</td>
<td>.231</td>
<td>.225</td>
<td>.231</td>
<td>38.457</td>
<td>1</td>
<td>128</td>
<td>.000</td>
</tr>
<tr>
<td>2. ISF-WK</td>
<td>.507</td>
<td>.258</td>
<td>.246</td>
<td>.027</td>
<td>4.537</td>
<td>1</td>
<td>127</td>
<td>.035</td>
</tr>
</tbody>
</table>

*Note.* LNF-WK= Letter Naming Fluency (winter kindergarten); ISF-WK= Initial Sound Fluency (winter kindergarten).

Table 9

*Coefficients for Final Model for Research Question 2c*

<table>
<thead>
<tr>
<th></th>
<th>$B$</th>
<th>$\beta$</th>
<th>$t$</th>
<th>$p$</th>
<th>Bivariate $r$</th>
<th>Partial $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>LNF-WK</td>
<td>4.901</td>
<td>.415</td>
<td>5.025</td>
<td>.000</td>
<td>.481</td>
<td>.407</td>
</tr>
<tr>
<td>ISK-WK</td>
<td>2.339</td>
<td>.176</td>
<td>2.130</td>
<td>.035</td>
<td>.332</td>
<td>.186</td>
</tr>
</tbody>
</table>

*Note.* LNF-WK= Letter Naming Fluency (winter kindergarten); ISF-WK= Initial Sound Fluency (winter kindergarten).
Research Question Three Results

This research question looked at the accuracy of early literacy skills, which were shown through multiple regression, to predict students who would not be successful on the PSSA. Winter Grade 1 ORF and Grade 3 PSSA Reading performances were examined through ROC curves (Area Under the Curve; $AUC = .84$), identifying a new cut score that maximized both sensitivity and specificity, producing diagnostic accuracy in the .80 range and a hit rate of 80.76% (see Figure 4). In comparison to diagnostic accuracy for DIBELS cut score of 20 wcpm, the ROC cut score of 18 indicated a sensitivity rate that was 7% greater. As ROC cut scores are noted to be the average of two of consecutive ordered observed test values, the ROC cut score of 17.50 was rounded up to the next whole number maximize the potential of identifying more students who would not be successful on the PSSA Reading domain. Table 10 shows the partial cut scores for ORF and Grade 3 PSSA Reading. Positive predictive power was significantly higher than the level of negative predictive power, demonstrating that the probability that proficient ORF scores accurately predicted proficient Grade 3 PSSA Reading scores was higher than the probability of nonproficient ORF scores accurately predicting nonproficient Grade 3 PSSA Reading scores.
Figure 4. Receiver operator characteristic curve predicting Grade 3 PSSA Reading performance. The x-axis displays the false positive rate, while the y-axis displays the true positive rate.
Table 10

*Partial Coordinates of the Curve: Grade 1 Winter ORF and Grade 3 PSSA*

<table>
<thead>
<tr>
<th>Cut Score</th>
<th>Sensitivity</th>
<th>1 – Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.00</td>
<td>1.00</td>
<td>.76</td>
</tr>
<tr>
<td>9.50</td>
<td>.97</td>
<td>.71</td>
</tr>
<tr>
<td>10.50</td>
<td>.95</td>
<td>.67</td>
</tr>
<tr>
<td>11.50</td>
<td>.95</td>
<td>.62</td>
</tr>
<tr>
<td>12.50</td>
<td>.93</td>
<td>.57</td>
</tr>
<tr>
<td>13.50</td>
<td>.90</td>
<td>.52</td>
</tr>
<tr>
<td>14.50</td>
<td>.88</td>
<td>.48</td>
</tr>
<tr>
<td>15.50</td>
<td>.86</td>
<td>.48</td>
</tr>
<tr>
<td>16.50</td>
<td>.84</td>
<td>.38</td>
</tr>
<tr>
<td>17.50</td>
<td>.82</td>
<td>.19</td>
</tr>
<tr>
<td>18.50</td>
<td>.78</td>
<td>.14</td>
</tr>
<tr>
<td>19.50</td>
<td>.75</td>
<td>.14</td>
</tr>
<tr>
<td>20.50</td>
<td>.72</td>
<td>.14</td>
</tr>
</tbody>
</table>

*Note.* Cutoff values are the averages of two consecutive ordered observed test values.

Next winter Grade 1 ORF and Grade 5 PSSA Reading performances were examined through ROC curves (*AUC* = .81), identifying a new cut score that maximized both sensitivity and specificity, producing diagnostic accuracy in the .70 and .80 ranges and a hit rate of 70.77% (see Figure 5). Again ROC cut scores showed a 7% greater sensitivity rate, in comparison to the DIBELS recommended cut score of 20 wcpm, which only indicated a 76% sensitivity rate. Table 11 shows the cut scores for ORF and
Grade 5 PSSA Reading. Positive predictive power was again significantly higher than the level of negative predictive power, demonstrating that the probability that proficient ORF scores accurately predicted proficient Grade 5 PSSA Reading scores was higher than the probability of nonproficient ORF scores accurately predicting nonproficient Grade 5 PSSA Reading scores.

*Figure 5.* Receiver operator characteristic curve predicting Grade 5 PSSA Reading performance. The x-axis displays the false positive rate, while the y-axis displays the true positive rate.
Table 11

Partial Coordinates of the Curve: Grade 1 Winter ORF and Grade 5 PSSA

<table>
<thead>
<tr>
<th>Cut Score</th>
<th>Sensitivity</th>
<th>1 – Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.00</td>
<td>1.00</td>
<td>.82</td>
</tr>
<tr>
<td>9.50</td>
<td>.99</td>
<td>.71</td>
</tr>
<tr>
<td>10.50</td>
<td>.97</td>
<td>.68</td>
</tr>
<tr>
<td>11.50</td>
<td>.96</td>
<td>.64</td>
</tr>
<tr>
<td>12.50</td>
<td>.95</td>
<td>.57</td>
</tr>
<tr>
<td>13.50</td>
<td>.94</td>
<td>.46</td>
</tr>
<tr>
<td>14.50</td>
<td>.92</td>
<td>.43</td>
</tr>
<tr>
<td>15.50</td>
<td>.91</td>
<td>.43</td>
</tr>
<tr>
<td>16.50</td>
<td>.87</td>
<td>.36</td>
</tr>
<tr>
<td>17.50</td>
<td>.83</td>
<td>.29</td>
</tr>
<tr>
<td>18.50</td>
<td>.78</td>
<td>.29</td>
</tr>
<tr>
<td>19.50</td>
<td>.76</td>
<td>.29</td>
</tr>
<tr>
<td>20.50</td>
<td>.72</td>
<td>.29</td>
</tr>
</tbody>
</table>

Note. Cutoff values are the averages of two consecutive ordered observed test values.

Finally, winter kindergarten LNF and Grade 7 PSSA reading performance were examined through ROC curves ($AUC = .77$), identifying a new cut score that maximized both sensitivity and specificity, producing diagnostic accuracy in the .60 and .70 ranges and a hit rate of 60.77% (see Figure 6). ROC cut scores again displayed greater sensitivity rates, this time by 6%, than the recommended DIBELS cut scores of 27 letters correct per minute. Table 12 shows the cut scores for LNF and Grade 7 PSSA Reading.
Positive predictive power was again higher than the level of negative predictive power, demonstrating that the probability that proficient LNF scores accurately predicted proficient Grade 7 PSSA Reading scores was higher than the probability of nonproficient LNF scores accurately predicting nonproficient Grade 7 PSSA Reading scores.

*Figure 6.* Receiver operator characteristic curve predicting Grade 7 PSSA Reading performance. The x-axis displays the false positive rate, while the y-axis displays the true positive rate.
Table 12

*Partial Coordinates of the Curve: Kindergarten Winter LNF and Grade 7 PSSA*

<table>
<thead>
<tr>
<th>Cut Score</th>
<th>Sensitivity</th>
<th>1 – Sensitivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>13.50</td>
<td>.89</td>
<td>.71</td>
</tr>
<tr>
<td>16.00</td>
<td>.88</td>
<td>.61</td>
</tr>
<tr>
<td>18.00</td>
<td>.88</td>
<td>.54</td>
</tr>
<tr>
<td>20.00</td>
<td>.85</td>
<td>.50</td>
</tr>
<tr>
<td>21.50</td>
<td>.84</td>
<td>.50</td>
</tr>
<tr>
<td>22.50</td>
<td>.84</td>
<td>.46</td>
</tr>
<tr>
<td>23.50</td>
<td>.83</td>
<td>.43</td>
</tr>
<tr>
<td>24.50</td>
<td>.83</td>
<td>.39</td>
</tr>
<tr>
<td>25.50</td>
<td>.76</td>
<td>.36</td>
</tr>
<tr>
<td>26.50</td>
<td>.70</td>
<td>.36</td>
</tr>
<tr>
<td>27.50</td>
<td>.66</td>
<td>.25</td>
</tr>
<tr>
<td>28.50</td>
<td>.64</td>
<td>.25</td>
</tr>
</tbody>
</table>

*Note.* Cutoff values are the averages of two consecutive ordered observed test values.

Diagnostic accuracy, as determined through ROC curves, includes variables such as specificity, sensitivity, negative predictive power, positive predictive power, false positives, false negatives, true positives, and true negatives. Specificity in this study refers to the probability that DIBELS CBM in kindergarten or Grade 1 will accurately identify or predict those students who are successful on the PSSA Reading in Grades 3, 5, and 7. Sensitivity then refers to the probability that DIBELS CBM in kindergarten or Grade 1 will accurately identify or predict those students who are not successful on the
PSSA Reading in Grades 3, 5, and 7. A false positive in the current study is the probability that the kindergarten or Grade 1 DIBELS CBM will fail to accurately identify or predict PSSA failure in Grades 3, 5 and 7. Therefore, a false positive occurs when the student is predicted to perform at the nonproficient level and the student performs at the proficient level on the PSSA Reading assessment. A false negative in this study is the probability that the kindergarten or Grade 1 DIBELS CBM will fail to accurately identify or predict PSSA success in Grades 3, 5, and 7. Thus, a false negative occurs when a student performs in the nonproficient range on the PSSA Reading, but they were predicted to be proficient based upon the CBM score. Lastly, negative predictive power and positive predictive power were determined in the current study. Negative predictive power refers to the probability that students identified as nonproficient on the DIBELS will be nonproficient on the PSSA Reading. Positive predictive power then refers to the probability that the students identified as proficient on the DIBELS will be proficient on the PSSA Reading. Table 13 displays statistics specific to this study.
Table 13

*Reading Diagnostic Accuracy Results Including CBM and PSSA*

<table>
<thead>
<tr>
<th>Best Predictor/Criterion</th>
<th>Cut score</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>PPP</th>
<th>NPP</th>
<th>Hit rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF/PSSA3</td>
<td>18</td>
<td>.82</td>
<td>.81</td>
<td>.97</td>
<td>.45</td>
<td>.81</td>
</tr>
<tr>
<td>ORF/PSSA5</td>
<td>18</td>
<td>.83</td>
<td>.71</td>
<td>.90</td>
<td>.40</td>
<td>.71</td>
</tr>
<tr>
<td>LNF/PSSA7</td>
<td>25</td>
<td>.83</td>
<td>.61</td>
<td>.85</td>
<td>.30</td>
<td>.61</td>
</tr>
</tbody>
</table>

*Note.* PSSA: Pennsylvania System of School Assessment; CBM: curriculum-based measurement; ORF: Oral Reading Fluency (winter Grade 1); LNF: Letter Naming Fluency (winter kindergarten); PPP: positive predictive power; NPP: negative predictive power.
CHAPTER V
DISCUSSION

The objective of this study was to examine the relationships among CBM, early literacy indicators, and performance on a state reading criterion assessment (i.e., PSSA) across time. CBM, particularly ORF, has been shown to be predictive of high-stakes testing up to two years later (Hintze & Silberglitt, 2005; Keller-Margulis et al., 2008; Shapiro et al., 2006). Previous literature, however, has not explained how well early literacy CBM predicts reading performance further into the future. For example, unknown was the extent to which early literacy CBM probes administered in kindergarten were predictive of middle school reading performance. This chapter will address the findings of the current study in light of the extant literature. Implications for school psychologists and recommendations for future research will be reviewed.

The first research question examined the relationships among early literacy CBM probes and subsequent performance on high-stakes state reading assessments. It was hypothesized that all variables would be moderately correlated with each other. The creation of a Pearson correlation matrix indicated that all early literacy variables were significantly and positively related to high-stakes reading assessments ranging from two to seven years later. Positive correlations among the R-CBM measures and statewide assessments were consistent with previous research (Baker et al., 2008; Buck & Torgeson, 2003; Good et al., 2001; McGlinchey & Hixon, 2004; Roehrig et al., 2008; Shapiro et al., 2006; Stage & Jacobson, 2001). The strongest of the correlations, a moderate correlation, was noted between statewide reading assessments in Grade 5 and the CBM of ORF in the winter of Grade 1. More specifically, the moderate correlation between ORF and Grade 5 PSSA Reading performance is consistent with previous
research indicating moderate to strong correlations between ORF and future reading performance on high-stakes Grade 4 state assessment (McGlinchey & Hixon, 2004; Stage & Jacobson, 2001). The weakest correlation was noted between Grade 7 PSSA Reading performance and the CBM of PSF in the winter of kindergarten. The strongest correlation among CBM was found with LNF and NWF. This correlation was also found in research by Burke and colleagues (2009); however, the current study found much stronger correlations among the CBM. Additionally, moderate to strong correlations were noted between Grade 3, 5, and 7 PSSA Reading performance. These correlations suggest relatively stable performance on PSSA Reading performance from Grade 3 through 7, a finding that is useful for teacher and school psychologists in regards to high-stakes reading achievement. Overall, these correlations indicate the presence of a moderate relationship between initial early literacy skills and future reading achievement up to seven years later.

The second question examined the extent to which early literacy CBM indicators would predict PSSA Reading performance in Grades 3, 5, and 7. Previous studies only examined ORF performance and reading achievement performance up to two years later. ORF was found to be predictive of Grades 3 and 5 PSSA Reading performance, when CBM and PSSA were administered in the same year (Shapiro et al., 2006). Further research on PSSA previously noted that ORF CBM administered in Grades 2 and 4 predicted PSSA Reading performance one year later in Grades 3 and 5, while ORF CBM administered in Grades 1 and 3 predicted PSSA Reading performance two years later in Grades 3 and 5 (Keller-Margulis et al., 2008).
Based upon previous research, it was hypothesized that performance on winter Grade 1 ORF would best predict performance on PSSA Reading performance in Grades 3, 5, and 7. The CBM identified to be most predictive of PSSA performance at Grades 3, 5, and 7 were all found to be of moderate strength and at the $p < .001$ significance level. Similar to previous studies, the present study also found that ORF CBM administered in the winter of Grade 1 was a strong predictor of PSSA Reading performance in Grades 3 and 5, therefore indicating a longer range of predictability up to four years. Results showed that ORF, followed by PSF and then ISF, were most predictive of Grade 3 reading achievement. In regards to Grade 5 reading achievement, ORF followed by PSF were found to be most predictive. Previous research on Grade 1 CBM indicated that ORF was the only significant predictor of reading performance in Grade 3, excluding PSF, LNF, and NWF (Goffreda et al., 2009). Phoneme Segmentation Fluency, LNF, and NWF have been shown to be predictive of ORF and therefore may not have shown the greatest correlation over time. However, in the current study, PSF, ISF, and LNF were found to significantly contribute to the variance of PSSA Reading performance in Grades 3, 5, and 7.

Unlike previous studies, this investigation found that LNF, followed by ISF, was the strongest predictor of Grade 7 PSSA Reading. This study contributed to the literature by establishing that early literacy skills, as measured by CBM, are predictive of reading achievement at least seven years into the future. Counter to this study’s hypothesis that ORF in Grade 1 would be most predictive of Grade 7 reading performance, LNF and ISF were found to be most predictive of Grade 7 PSSA Reading Domain performance. This unexpected finding may be related to task demands of reading criterion tests across
grades. As abstract and inferential comprehension are put at a premium by the PSSA Reading test in Grade 7, it is possible that the influence of reading fluency is decreased in favor of other cognitive skills. For example, McGrew and Wendling’s (2010) review of the empirical literature found that crystallized, language-related skills and background knowledge are significantly related to reading comprehension, and perhaps more so than many fluency indicators. Although crystallized language and background knowledge skills were not assessed in the current study, it appears that LNF being most predictive of long term reading achievement is an artifact of this phenomenon. Additionally, it is possible that the noted correlations are specific to this population and may not be found in future studies using other populations of students.

The third question examined the diagnostic accuracy of CBM that were found to be most predictive of future PSSA Reading performance. The original hypothesis stated that with respect to CBM measures found to be most predictive of grades 3, 5, and 7 PSSA Reading performances, cut scores determined by ROC analyses would satisfactorily identify success or failure on grades 3, 5, and 7 PSSA Reading performances more than DIBELS provided cutoff scores.

Results showed that ORF at Grade 1 predicted success or failure on the PSSA 3 Reading domain at a rate of 81%. Additionally, sensitivity, or the probability that ORF accurately predicted PSSA failure, was 82%. ORF at Grade 1 also predicted success or failure on the PSSA 5 Reading Domain at a rate of 71%, with sensitivity at 83%. Finally, LNF in kindergarten predicted success or failure on the PSSA 7 Reading Domain with a slightly lower rate of 61%, and a sensitivity rate of 83%. Findings using the ROC cut scores showed that the use of ROC scores in the current study were consistent with
previous research eliminating high rates of overestimating individuals as successful on the PSSA, when they actually fail (Hintze et al., 2003). The ROC DIBELS optimal scores for the current study also showed consistently higher levels of sensitivity, but not higher levels of specificity, when compared to the DIBELS recommended cut scores. Additionally, results using the optimal ROC scores were similar with previous research, whereas the optimal cutoff scores were actually lower than the DIBELS recommended cutoff scores (Goffreda et al., 2009; Hintze et al., 2003; Shaw & Shaw, 2002). This is in contrast to what Keller-Margulis and colleagues (2008) found with a sample of students in Pennsylvania, which suggested that a significant number of students scoring below the DIBELS suggested cutscores may be successful on subsequent reading assessments. Although in contrast of the studies conducted in Pennsylvania, results were more consistent with national research in the use of DIBELS scores as an accurate predictor of future performance. Additionally, the current study was able to accurately identify more students who were at risk for failure on the PSSA through use of the ROC analysis versus the use of the DIBELS recommended cut scores.

**Implications for Practice**

The results of this study have implications on the current practice of problem solving teams, including the use of RtII in schools. First, the findings of the current study confirm the significant positive correlations between early literacy skills, specifically correlations with ORF. Identifying this correlation in the current research endorses the use of early and frequent intervention on the early literacy skills of students. Literacy skills as early as the winter of kindergarten play a crucial role in the continued development of reading skills.
Second, the current findings indicate that CBM of initial early literacy skills can be used as an accurate predictor of future reading performance through Grade 7. While early intervention to increase early literacy skills performance is critical to immediate performance, these findings reiterate the use of CBM probes as a valuable screening tool regarding long term reading performance.

As noted in previous research, the use of DIBELS cut scores may not be appropriate in a specific location. Specific analyses should be conducted by districts to determine the most appropriate cut scores for their location. The use of more stringent cutoff scores when using the DIBELS assessments may indicate the use of lower or higher cut scores. Through the use of more appropriate cut scores, personnel will be able to utilize their time and resources more efficiently.

School psychologists will need to understand the relationships among each of the early literacy skills, as well as high-stakes state-level reading assessments. Their ability to administer, score, and interpret CBM data is fundamental to the education and success of students. While interpreting the data, school psychologists should guide districts toward the use of the most appropriate cut scores for their district.

More specific implications include the use of CBM in Pennsylvania. Findings from the current study may be more applicable for problem solving teams in Pennsylvania, as the PSSA was examined in this study. Teams in Pennsylvania now have more research to utilize when trying to establish which students are at risk of long term reading failure and which students should not be considered at risk for long term reading problems. Of course, the results of the current study can only be understood within the context of the curricula that were used, the problem solving procedures (or lack thereof)
implemented by district personnel, and the quality of the interventions attempted with individual and groups of students across the 7-year time span studies. As these factors do vary across districts, the strict results of this study may not apply to each school district in Pennsylvania.

Overall, the importance of utilizing a screening tool as early as possible in the educational career of students is extremely important with the tremendous focus on high-stakes assessments. This particular CBM could be utilized as an effective screening tool that predicts outcomes on the PSSA as early as kindergarten and Grade 1. This study contributed to the literature showing that early literacy probes were able to predict reading performance through Grade 7. As a result of this finding, districts are able to identify who is at risk for long term reading failure as early as kindergarten. The predictability of the future reading will help to facilitate the implementation of RtII models across the Commonwealth of Pennsylvania, as well as other states.

Limitations

The purpose of this study was to explore the longitudinal relationship between initial early literacy skills in kindergarten and Grade 1 and future reading achievement up to seven years later. Despite the significant results, there are limitations. The first limitation is that the data obtained from a preexisting database was derived primarily from suburban/rural, Caucasian students. A more heterogeneous geographical, ethnical, and racial sample may produce different results. Additionally, this study only represents findings from one school district in Pennsylvania and may not be generalizable to other districts within the Commonwealth of Pennsylvania or other states. Although this is a limitation, it should be noted that previous research has noted similar findings in
Pennsylvania school districts and other states based upon sufficient sample size with one school district. (Hintze & Silberglitt, 2005; Keller-Margulis et al., 2008).

Another limitation of the study was that only students with a complete data set were examined, due to the longitudinal nature of the study. As students throughout the state and the country are particularly transient due to current economic conditions, it may be beneficial to examine the performance of students who have moved between schools. Future studies should obtain data on students who have moved between schools to increase generalizability of results.

The use of a particular CBM, the DIBELS, and a particular high-stakes state-level assessment, the PSSA, is also a limitation to this study, as there are several CBM and state-level assessments currently designed to screen early literacy skills and measure academic performance. Although previous research in a number of states has identified correlations among CBM and high-stakes assessments, these findings have been confined to a 2-year prediction of performance. Research in other states needs to expand in order to examine the long term correlations between reading CBM and high-stakes reading assessments.

Lastly, due to the longitudinal nature of the study, the data set does not include the most current CBM for the DIBELS, making conclusions concerning current research a challenge. Future research should address more recent version of CBM assessments to provide for more practical findings.

**Recommendations for Future Research**

Future research should focus on a larger, more diverse population to provide for a greater representation of the Commonwealth’s demographics. Studies should include an
array of students with varying racial and ethnical backgrounds, socioeconomic status, and geographical locations (e.g., urban and rural). Including special populations of students, such as students receiving special education services and English language learners, would also provide a greater representation of students in other states.

The diverse population of students examined in future research should also examine students who are transient between districts or states. As mentioned in the limitations of this study, due to the economic hardships in the United States many families are frequently moving between districts and states. Eliminating this specific population while conducting research would also limit the ability of the district and staff in examining the reading development of these students. Understanding the reading trajectory of transient students is another area for future research, as these students often lack consistent reading instruction and are exposed to inconsistent teaching strategies due to their enrollment in multiple schools. As stated earlier, the results of this study can only be understood in the context of the reading curriculum and intervention procedures in place during this study. Future research should consider predicting reading performance across time within the context of varying curricula and intervention approaches.

While this study is specific to Pennsylvania, future research outside of Pennsylvania might examine other CBM in relation to their state specific high-stakes assessments for students. It is important for each state to carefully examine the relationships among widely used screening tools and performance on high-stakes reading assessments that are specific to their state. Future research in Pennsylvania could examine the relationship among various CBM published by other companies, such as AIMSweb and state criterion reading tests. Furthermore, unlike DIBELS, AIMSweb has
additional early literacy probes available for study, such as letter-sound fluency, that would provide additional measures to examine in relationship to future PSSA performance. Although reading assessments were the focus on the current study, researchers might investigate early CBM in the area of mathematics and writing to determine if there is a strong predictor of long term future academic performance. Previous research has examined mathematics performance up to two years after the administration of CBM, which continues to limit the predictability of CBM on future high-stakes assessments. On the whole, future research regarding screening assessments and future reading achievement performance is paramount to the success of education in the United States.

**Conclusions**

The current study examined the longitudinal relationship between early literacy skills and future reading achievement performance up to seven years later. Each DIBELS measure administered at the winter kindergarten benchmark and ORF administered at the winter first grade benchmark was significantly and positively correlated with Grades 3, 5, and 7 PSSA Reading performance. Similar to previous studies regarding early literacy skills, the strongest prediction was found with ORF and future high-stakes reading performance. Unlike previous research, LNF was found to predict high-stakes reading performance in middle school. Results of this study provide evidence of the long term diagnostic accuracy of initial early literacy CBM scores and performance on high-stakes statewide reading assessments.
References


11. Retrieved from
http://www.portal.state.pa.us/portal/server.pt/community/technical_analysis/7447.

http://www.portal.state.pa.us/portal/server.pt/community/technical_analysis/7447.


APPENDIX

INSTITUTIONAL REVIEW BOARD APPROVAL LETTER

DUQUESNE UNIVERSITY
INSTITUTIONAL REVIEW BOARD
424 RANGOS BUILDING • PITTSBURGH PA 15282-0202

February 4, 2011

Dr. Ara Schmitt
School of Education
Duquesne University
Pittsburgh PA 15282

Re: Relationships among early literacy CBM and reading state criterion tests over time
(Protocol # 11-09)

Dear Dr. Schmitt:

Thank you for submitting the research protocol from your student, Ms. Lynn Utchell, to the IRB.

Based on the review of Dr. Sarah Peterson, IRB Representative, and my own review, your study is approved as Exempt based on 45-CODE of Federal Regulations-46.101.b.4, regarding data without identifiers extracted from existing records.

This exempt approval pertains strictly to the research described in the protocol. If you and Ms. Utchell intend to make a change in the research, you must submit a formal amendment for review before proceeding. In addition, you should inform the IRB if any adverse events or procedural problems occur impacting subjects. In correspondence about the research, please refer to the protocol number shown after the title above.

Once the study is complete, provide our office with a short summary (one page) of your results for our records.

Thank you for contributing to Duquesne’s research endeavors.

Sincerely yours,

Paul Richer, Ph.D.

C: Ms. Lynn Utchell
Dr. Sarah Peterson
IRB Records