5-2015

Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency Data to Predict Michigan Educational Assessment Program Outcomes

Lisa Lockman
Western Michigan University, ljlockman@yahoo.com

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Educational Assessment, Evaluation, and Research Commons, and the Education Policy Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/530

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
USING DYNAMIC INDICATORS OF BASIC EARLY LITERACY SKILLS (DIBELS)
ORAL READING FLUENCY DATA TO PREDICT MICHIGAN EDUCATIONAL
ASSESSMENT PROGRAM OUTCOMES

by

Lisa Lockman

A dissertation submitted to the Graduate College
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
Educational Leadership, Research, and Technology
Western Michigan University
May 2015

Doctoral Committee:

Louann Bierlein Palmer, Ed.D., Chair
Jessaca Spybrook, Ph.D.
Michelle Johnston, Ph.D.
Schools are under legislative pressure to ensure students make adequate progress each year, and therefore, must identify students in need of intervention at early grade levels. Educators use diagnostic tools to measure student progress, attempting to determine which children may be most at risk for future reading failure. The Dynamic Indicators of Early Literacy Skills (DIBELS) is widely used for early identification of students who have risk factors. This study incorporated a quantitative, correlational, ex post facto design to examine the relationship between students’ third grade scores on the spring DIBELS Oral Reading Fluency (ORF) assessment and their fifth grade Michigan Educational Assessment Program (MEAP) reading results. By examining the relationship between the DIBELS ORF assessment and the MEAP reading assessment, researchers can gain insight into the system many schools currently use to identify at risk students.

The study analyzed data from 2,220 students enrolled in seven rural school districts participating in a Regional Data Initiative grant program. The research questions examined relationships, including the effect of gender and socioeconomic status, between DIBELS ORF scores in the spring of third grade and MEAP reading assessment scores in
their third, fourth and fifth years. Data was examined using descriptive statistics, Linear Regression, Chi Square analysis and Hierarchical Linear Modeling.

Findings revealed lower DIBELS ORF at third grade were predictive of lower MEAP reading scores. After controlling for the above mentioned effects, each increase in DIBELS ORF score at third grade was associated with about a .60 increase in normed MEAP scores. This means that students scoring lower on their third grade DIBELS ORF assessment are most at risk for scoring lower on their third grade state assessment.

DIBLES ORF was also a significant predictor of the slope of student MEAP reading scores between third and fifth grade. These findings support the use of DIBELS ORF as a screener to help educators predict a student’s future reading proficiency scores, and therefore can help inform decision when selecting the students in need of interventions.
ACKNOWLEDGMENTS

This has been such a wonderful, long journey. I certainly would not have made it this far without the support of my family. Heartfelt thanks go to both Kevin and Ally, who helped make time for me to work and patiently reminded me that I could do this. I love you and appreciate your support more than you know.

When I started my course work at Western, I was privileged to be assigned a wonderful advisor and chairperson, Dr. Bierlein Palmer, who not only provided support and guidance, but also prodded me to get to work when I was behind (always in the nicest way). I did not take the easiest path to finishing my dissertation and truly appreciate you being a support system over the entire course of my program at WMU.

I also appreciate Dr. Bierlein Palmer, Dr. Spybrook, and Dr. Johnston for taking the time to really help me understand HLM and the statistical models I selected. You all supported me as I worked my way down this path.

Finally, I want to thank my parents, Daryl and Judy Morr, for instilling in me a lifelong love of learning. You were both wonderful role models and fabulous educators. I know I wouldn’t be where I am today without your love and support.

Lisa Lockman
# TABLE OF CONTENTS

ACKNOWLEDGMENTS ........................................................................................................... ii

LIST OF TABLES ................................................................................................................ vii

LIST OF FIGURES .............................................................................................................. viii

CHAPTER ............................................................................................................................ 1

I. OVERVIEW OF THE RESEARCH ................................................................................. 1

   Problem Statement and Research Questions ...................................................... 3

   Conceptual Framework ..................................................................................... 6

   Methodology Overview ..................................................................................... 10

   Significance of the Study ................................................................................ 11

   Chapter I Summary ........................................................................................ 11

II. LITERATURE REVIEW ............................................................................................ 13

   Introduction .......................................................................................................... 13

   Accountability .................................................................................................... 14

       Educational Standards Movement .............................................................. 14

       No Child Left Behind .................................................................................... 16

       Individuals with Disabilities Act ................................................................. 18
Table of Contents - continued

CHAPTER

Early Literacy Policy ............................................................... 19
Policy in Michigan .................................................................. 20
Assessment ................................................................................. 22
Assessment Standards ................................................................ 22
Reasons for Testing ................................................................... 23
Assessment Types ...................................................................... 24
Reading Instruction and Assessment .......................................... 25
Reading Instruction ................................................................... 25
Assessment of Reading Skills .................................................... 27
MEAP Overview ......................................................................... 30
Response to Intervention ............................................................ 33

Studies Comparing DIBELS and Criterion-Referenced Assessments ......................................................... 34
Chapter II Summary ................................................................ 36

III. METHODOLOGY ...................................................................... 38

Research Questions ................................................................... 38
Research Design ......................................................................... 39
Population and Sample ............................................................... 41
Table of Contents - continued

CHAPTER

Instrumentation ................................................................. 42

DIBELS ............................................................................. 42

MEAP ............................................................................... 44

Data Analysis ........................................................................ 45

Operationalization of Variables of Study ......................... 46

Test of Hypotheses According to Research Question ......... 49

Statistical Models ............................................................ 51

Power Analysis ............................................................... 52

Limitations and Delimitations ......................................... 52

Chapter III Summary ...................................................... 53

IV. RESULTS ........................................................................ 54

Population and Demographic Findings ......................... 55

Operationalization of Inferential Analysis Variables .......... 60

Linear Regressions ........................................................ 64

Correlational Analyses .................................................... 64

Multilevel Models .......................................................... 65

Chapter IV Summary ....................................................... 69
Table of Contents - continued

CHAPTER

V. DISCUSSION

Key Findings

Conclusions as Related to Research Question 1

Conclusions as Related to Research Question 2

Conclusions as Related to Research Question 3

Conclusions as Related to Research Question 4

Implications for School Leaders

Recommendations for Future Study

Conclusions

APPENDICES

A. Approval Letter from the Human Subjects Intuitional Review Board
## LIST OF TABLES

1. Data Analysis Cross-Walk Table ................................................................. 48
2. Demographic and Assessment Data ............................................................ 57
3. SES by District and by School ....................................................................... 59
4. Gender by District and by School .................................................................. 60
5. Summary of Linear Regressions Using DIBELS ORF 3rd Grade Predicting Normed MEAP Scores (3rd – 5th) ............................................................... 64
6. 5th Grade MEAP Score Categories by DIBELS ORF 3rd Grade Score Categories .... 65
7. Hierarchical Linear Model Predicting Normed MEAP Reading Scores in 3rd-5th Grade from Gender, SES, School, DIBELS, and Interactions ................................. 66
8. Growth Curve Predicting Normed MEAP Scores (3rd-5th Grade) From DIBELS and Covariates .................................................................................. 68
9. Key Findings/Results .................................................................................... 75
LIST OF FIGURES

1. The Relationship of DIBELS ORF and the MEAP ........................................... 7
2. Timeline of Educational Reform ..................................................................... 21
3. Mean Normed MEAP Scores in Each Grade by Each DIBELS ORF Score ....... 74
CHAPTER I

OVERVIEW OF THE RESEARCH

How can we ensure that students will be successful academically? Researchers, educators and legislators have been searching for the right formula for student success for years. Starting in 1983 with the publication of *A Nation at Risk*, American education ushered in an age of standards and accountability ("A nation at," 1983). Over the course of the last twenty years, the United States has created national educational standards and individual states have created their own standards and benchmarks for educational performance (Linn, 2000). In 1993, Michigan started developing a State Curriculum Framework, aligning core content areas with the national education standards. Gone are the days of closing the proverbial classroom door and teaching anything a teacher wants to cover.

Educators have long worked to ensure that their students would reach the highest achievement levels possible. Yet, national and state legislation have added pressure for increased student results as measured by standardized assessments. Heading into the new millennium, teachers and school leaders throughout the world were being held accountable for student achievement (Borkowski & Sneed, 2006; Fullan & Watson, 2000; Margolis, 2006; Reutzel & Michell, 2005; Saiger, 2005). This accountability movement was one of the driving forces behind the No Child Left Behind Act of 2001 (NCLB). NCLB requires all states to improve learning outcomes for students by, among other things, testing student progress each year, issuing school report cards based on the achievement of students, and indicating the Adequate Yearly Progress (AYP) state of each building (Linn, Baker, & Betebenner, 2002).
Under NCLB, AYP is also based on subgroup performance (Owens & Sunderman, 2006), specifically, low-income students, minority students, students with limited English proficiency, and students with disabilities (Wiener & Hall, 2004). Michigan uses the Michigan Educational Assessment Program (MEAP) to determine district AYP and measure subgroup and individual student growth. The MEAP is a criterion-referenced test based on the Michigan Curriculum Frameworks Grade Level Content Expectations (GLCE). Criterion-referenced assessments measure performance with respect to a specific criterion or standard (Reynolds & Fletcher-Janzen, 2007).

NCLB was designed to improve student achievement not only through increased accountability, but also with additional demands for early identification and intervention systems. The Act places an increased emphasis on reading and the implementation of instructional practices that are evidence based and scientifically validated, and was designed to ensure all students meet high academic standards by requiring states to create annual assessments in grades three through eight in reading and math (Allington, 2005). These assessments have become very high stakes for schools and students (Bober, 2004; Carsten, 2004). High-stake assessments often determine how the school organizes its plan of intervention and how students are afforded opportunities to learn. Specifically, these assessments are most often used to determine future placements for students (Chappuis, Chappuis, & Stiggins, 2009).

NCLB is not the only federal legislation currently driving school change. The Individuals with Disabilities Act (IDEA) stipulates that schools monitor student progress and collect data through frequent progress checks. The 2004 reauthorization of IDEA is aligned with the NCLB, specifically through its provisions for assessment and
accountability (Webber, 2006). Students with disabilities must be included in district-wide achievement measures and testing cycles, and districts must provide the data needed to assist with student placement and the design of student interventions. Schools struggle to find assessment systems that are reliable, valid, and sensitive enough for this purpose (Altwerger, Arya, Jin, & Jordan, 2004).

**Problem Statement and Research Questions**

Current educational reform initiatives have focused on increases in student achievement as demonstrated by statewide achievement tests (Chappuis et al, 2009). Schools are under legislative pressure to ensure students can make a year of progress each year, and schools must, therefore, identify students in need of intervention at early grade levels. Educators are using diagnostic tools to measure student progress, attempting to determine which children may be at risk for future reading failure. Schools have a limited number of assessment tools available to identify students who need additional support, and they must be assured that the tools used adequately identify students in need of interventions (Siu-Runyan, 2009).

Across the country, including in Michigan, the Dynamic Indicators of Early Literacy Skills (DIBELS) is widely used for early identification of students who are at risk for reading failure (Li & Zhang, 2008; Manzo, 2005). According to the DIBELS website, the assessments are given in more than 15,000 schools in both the United States and around the world (University of Oregon, 2010). Developed by Good and Kaminski (2002) at the University of Oregon, DIBELS was designed as a diagnostic instrument to provide information to teachers for instructional planning by measuring the acquisition of early literacy skills (Goodman, 2006). DIBELS uses a norm-referenced skills assessment
approach to assessing students’ reading ability. Norm-referenced assessments are those which a student’s achievement is ranked in accordance with all students who have taken the assessment (Colburn, 2009).

In sharp contrast the MEAP, is a criterion-referenced assessment aligned with the Michigan GLCEs. The MEAP test is developed by the Michigan Bureau of Assessment and Accountability (BAA) whose main mission is to fairly and accurately measure student achievement against the state’s content standards (Michigan Assessment Integrity Guidelines, 2009). The term “criterion-referenced measure” was first used by Glaser (1963), where he defined it as a measure that “depends on an absolute standard of quality” (p. 514). Criterion-referenced, competency-based assessments are a derivative of the mastery model of instruction, in which performance is interpreted in terms of defined criteria. Criterion-referenced testing describes a targeted behavior, defines the particular skill level, and then measures whether the behavior meets the standard of quality (Dyson, 2008; Horne, 1984; Reynolds & Fletcher-Janzen, 2007). For many decades, educators have endorsed the assessment of learning centered on mastering skills and domains rather than normative comparisons (Bloom, 1968; Guskey, 2010).

In Michigan, as across the country, the DIBELS assessment is one of the most commonly used early literacy assessments to determine student placement in intervention groups (Manzo, 2005). DIBELS is a performance-based interim assessment designed for monitoring reading progress, and it was not designed or validated specifically as a tool for predicting student performance on a criterion-referenced summative assessment, but as a tool to monitor early literacy skills (University of Oregon, 2010). DIBELS is not specifically aligned with the Michigan GLCEs or the national education standards. The
DIBELS ORF assessment is used widely for thousands of students in the State of Michigan and elsewhere, without adequate research on whether the students, flagged through its use, are in fact performing below grade level according to the state’s curriculum framework.

Therefore, an important overall issue to examine is the relationship between one interim assessment tool (i.e., DIBELS ORF assessment) commonly used in early grades to determine if students need interventions, and the state summative assessment tool, the MEAP. Specifically, my study addresses the following research questions:

1. To what extent is there a relationship between students’ DIBELS Oral Reading Fluency at the beginning of third grade and their MEAP Reading scores in third, fourth and fifth grades?

2. To what extent is there a relationship between students’ MEAP Reading proficiency statuses at fifth grade, and their DIBELS’ level (i.e., most at risk, some risk, low risk) at the beginning of third grade?

3. In reference to students’ third grade mean MEAP Reading scale score:
   
   a. to what extent are there difference between student labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessment); and
   
   b. how do the intervening variables of gender and socioeconomic status effect these difference?

4. In reference to student’ MEAP Reading score growth trajectories in grades three through five:
a. to what extent are there differences between students labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessments); and

b. how do the intervening variables of gender and socioeconomic status effect these differences?

**Conceptual Framework**

National and state education policies require students to read at grade level by the end of third grade (The Education Digest, 2010). Further, schools are required to assess students using scientifically valid assessments. This has led schools in Michigan and elsewhere, to adopt assessments, such as DIBELS, to identify which students should receive early interventions. As you can see in Figure 1, all students take the state mandated assessments, whether they have received interventions or not. These state level summative assessments are used to determine if schools meet AYP. It is important to look at whether a norm-reference test like DIBELS can accurately predict student results on criterion-referenced assessments like the MEAP.
Figure 1. The relationship of DIBELS ORF and the MEAP.
Schools across the country have adopted an early literacy approach with a goal of all students being proficient readers by the end of third grade. Reading skills have been targeted because reading impacts student achievement in multiple content areas (Chant, 2009). Many schools have adopted scientifically-validated diagnostic, norm-referenced skills assessments, such as DIBELS, as an interim or benchmark-level assessment in early grades to determine student progress towards this goal. These assessments are also often used to sort students into intervention groups. When students reach third grade, all students are tested using the state level assessment to determine reading proficiency. In Michigan, this is the MEAP assessment, a criterion-referenced assessment that is aligned with the Michigan Curriculum Framework. But, is DIBELS a good prediction of student success on the MEAP? Educators need to know that the right students are identified to receive interventions.

Sub-group reporting is relevant to my study of early interventions and assessments because, by measuring whether differences exist in SES and gender, we have a comprehensive look at the relationship between DIBELS and MEAP. As population trends change, we need to evaluate the relationship between SES and student achievement. As American schools attempt to make Adequate Yearly Progress and increase subgroup achievement levels, it is important to examine each of the sub-groups and determine how sub-groups within the larger population are functioning. Because the population of this study is located in a rural area with predominantly caucasian students, I did not include race in my analysis.

A meta-analytic review of research on SES, suggests that SES is probably the most widely used contextual variable in education research (Sirin, 2005). Regarding
SES, researchers engage in ongoing disputes about its conceptual meaning and empirical measurement (Bornstein & Bradly, 2003). In general, however, SES describes an individual's or a family's ranking on a hierarchy according to access to or control over some combination of valued commodities such as wealth, power, and social status (Mueller & Parcel, 1981). While researchers disagree on the conceptual meaning of SES, there seems to be a general agreement on Duncan, Featherman, and Duncan's (1972) definition of the tripartite nature of SES that incorporates parental income, parental education, and parental occupation as the three main indicators of SES (Meuller & Parcel, 1981). A fourth indicator, home resources, is not as commonly used as are the other three main indicators.

SES stands out as an important issue to examine for two very important reasons: (1) over the last 25 years, inequities in income and wealth have been growing, and (2) people have a predisposition to segregate themselves into different geographic areas, often based on their financial status (Hickrod, 2006). Cavanagh (2007) discusses the effects of the tendency for families with low SES to cluster together:

[lower] exam results are not surprising, given research showing that the U.S. system tends to provide underprivileged students with less demanding curricula, poorer-quality teachers, and fewer educational resources than their peers in wealthier U.S. communities. (p. 13)

Numerous studies have attempted to measure the relationship between SES and student achievement (e.g., Johnson, McGue, & Iacono, 2007; Sirin, 2005), with results from modest to strong.
It is also important to look at gender differences in student reading levels. Brozo (2010) points out that boys’ reading achievement is significantly lower than girls’ in nearly every state (p. 7). In fact by fourth grade, the average boy is two years behind the average girl in both reading and writing (Zambo & Brozo, 2008). Zambo and Brozo (2008) point out that this is not just the case in the United States, but internationally boys score lower than girls in reading. As schools across the country put interventions in place to increase student reading achievement, it is important to continue monitoring the achievement differences of boys and girls.

**Methodology Overview**

Based on the analysis needed, a quantitative approach is used in my study. Because the study variables that already exist and thus are not manipulatable for this study, my study is categorized as predictive non-experimental (Johnson, 2001). Creswell (2008) describes prediction research designs as those in which a correlation uses one or more predictive variables as a criterion for one or more outcome variables (p. 359). My study uses an ex post facto design to explore the relationship between third grade students’ Oral Reading Fluency (ORF) scores on the end of year (spring) DIBELS assessment and their fifth grade MEAP reading results, and an examination of any relationships found within grades 3, 4, and 5. Although the results cannot be used as proof for cause and effect relationships (Mertens, 2005), Johnson (2001) posited that non-experimental research is important to educators because further information is needed on these non-manipulable independent variables commonly used in the field of education.

The data for this study consists of a sample of 2,220 student records as drawn from a Northern Michigan regional service agency. The data set contains nominal data
(gender and socio-economic status), and interval data (DIBELS ORF, and MEAP Reading scores).

**Significance of the Study**

Schools, teachers, and administrators need to know if DIBELS testing adequately identifies students in need of intervention, allowing remediation programs to be put into place. By examining the relationship between the DIBELS ORF assessment and the MEAP reading assessment we gain insight into the system many schools currently use to identify at risk students. This study also provides data that can impact legislative decisions regarding student assessment and accountability.

**Chapter I Summary**

Identifying which instructional methods for assessing reading skills provide the greatest student success is important, as reading skills tend to impact student achievement in multiple content areas (Chant, 2009). The No Child Left Behind Act of 2001 puts an increased focus on reading, mandating that every student read at or above grade level by the end of third grade.

Schools and teachers are under increasing legislative pressure to ensure students make a year of progress each year. In January of 2010, Michigan passed new legislation that requires teacher evaluation to be in part based on student academic growth (Michigan School Code, 2010). Legislated accountability for schools increases the necessity of early identification of students who are falling behind. Schools must identify students in need of interventions at increasingly earlier grade levels. Yet schools have a limited number of assessment tools available to identify students who need additional support and we know little about the ones commonly used (e.g., DIBELS).
Chapter II provides a review of the literature regarding school accountability requirements, assessment types, and use of these assessments. Chapter III specifies the methods used in this study. Chapter IV contains the analysis of the data and Chapter V explains the conclusions as well as current and future implications.
CHAPTER II
LITERATURE REVIEW

All schools for miles and miles around
Must take a special test
To see who’s learning such and such –
To see which school’s the best.
If our small school does not do well,
Then it will be torn down,
And you will have to go to school
In dreary Flobbertown.

_Hooray for Diffendoofer Day!_
Dr. Seuss, Jack Prelutsky, and Lane Smith (1998)

Introduction

Schools across the country have come under intense pressure to conform to the public ideal of increasing student achievement. As these expectations have accelerated, accountably has become a political tool pushing the role of assessment and the use of high-stakes testing, causing them to increase proportionally (Bober, 2004; Carsten, 2004). Testing has become an instrument of policymakers, both for determining student progress and for driving policy decisions in the educational arena, particularly in the area of early literacy (Reutzel & Mitchell, 2005). This focus on accountability and public reporting of scores has schools scrambling to find assessments that produce data showing student progress toward expectations and educational progress (Ivernizzi, Landrum, Howell, & Warley, 2005). Major legislation such as NCLB, IDEA, and Race to the Top has prompted school districts to adopt tests such as DIBELS to monitor students reading progress prior to the high-stakes assessment such as the MEAP (Chappuis et al., 2009).
This review of the literature describes research related to accountability and national and state initiatives regarding reading. It also provides an overview of assessment types and includes an examination of DIBELS ORF assessment.

**Accountability**

As early as the 1950s, assessment has played a significant role in education, first as a means of identifying gifted and talented students and then as an indicator of college readiness (Linn, 2000). In the 1960s, assessments were used to appraise the effectiveness of federal programs like Title I. Minimum competency testing, the practice of requiring students to pass a test to receive a high school diploma, was popular in the 1970s and 1980s. A total of 34 states instituted some sort of basic skills test as a graduation requirement (Linn, 2000).

The current age of school reform has adopted *accountability* as a mantra for expanding testing in schools (Cook, 2003; Linn et al., 2002). Accountability reform programs that increase the real or perceived stakes of assessment results for teachers and administrators (Linn, 2000). Saiger (2005) goes as far as stating that “at the heart of the term ‘accountability’ is that failure should compel consequences” (p. 2). As far back as 1983, A Nation at Risk “seeks to generate reform of our educational system in fundamental ways” (p. 6). In their report to the nation, The National Commission on Excellence in Education went on to call for high standards and better monitoring of students’ academic skills ("A nation at," 1983).

**Educational Standards Movement**

Many of today’s educational reforms are linked to the creation of educational standards. In the spring of 1994, President Clinton signed the Goals 2000: Educate
Goals 2000 was a voluntary program that provided incentives for states to adopt standards in core subject areas. In order to receive additional educational funding, states would need to develop an improvement plan that provided for the following: (a) the adoption of content standards, (b) the development of valid and reliable assessments in grades 4, 8, and 12, (c) assure that all students have an opportunity to achieve these new standards, (d) institute provisions to assist students who have dropped out of school, (e) coordinate with other education initiatives, such as the Elementary and Secondary Education Act (ESEA) of 1965, and (f) develop strategies for involving parents and the community in planning for these new standards and assessments (Lewis, 1994).

At the same time that Congress was passing the Goals 2000 act, they were also redefining Title I services through the Improving America’s Schools Act (IASA) of 1994 which was a reauthorization of ESEA (Linn, 2000). Through IASA, we see the first standards based reporting requirements, specifically:

States are expected to adopt or develop student assessments in the same academic areas as their content and performance standards and use them to measure the yearly progress of Title I students, schools and districts. (Wong & Meyer, 1998, p. 8)

This legislation greatly expanded Title I funding, allowing more schools to access federal support for low-income students.

Policy makers and the public see mandated testing programs as an integral component of state accountability systems that monitor students’ ability to meet the standards in Michigan and across the country (Chappuis et al., 2009). Linn (2000) may
have said it best when he stated, “Testing and assessment have been both the focus of controversy and the darling of policymakers” (p. 4). Linn (2000) goes on to explain why testing and assessment appeal to policymakers: (a) compared to other educational changes, such as additional teacher training, tests are relatively inexpensive, (b) assessments can be mandated, (c) implementation of student assessment can happen quickly, and (d) the outcome is visible and can be reported easily in a public forum such as the newspaper.

No Child Left Behind

NCLB (2002), legislation that mandated students in grades three through eight be given annual tests in reading and that reading assessments for students in grades kindergarten through third grade be increased, served to propel the intensification of student assessment and monitoring. NCLB provided visible results by issuing school report cards based on student achievement and indicating the AYP status of each school on state websites (Borkowski & Sneed, 2006; Linn et al., 2002). One of the stipulations of NCLB, and a condition of receiving Title I funding, is that every state must implement an accountability system that aimed to get 100% of students’ proficient on the state adopted assessment by 2013-2014 (Cronin, Dahlin, Xiang, & McCahon, 2009). Another key component of the newest round of accountably mandates embeds huge consequences for schools that fail to reach the goal, including loss of local control (Saiger, 2005).

NCLB was enacted in the hopes of increasing student achievement in reading with one of the major goals of the legislation being that all children will be able to read at grade-level by the end of third grade (Allington, 2005; Linn et al., 2002). According to a Kids Count special report, a child’s reading ability at the end of third grade is a critical
marker of his or her future academic success ("Early warning: Why," 2010, p. 27). Children who fail to make this benchmark are at greater risk of dropping out of school and in 2009, 85% of third graders from low-income families failed to score proficient on the National Assessment of Educational Progress ("Early warning: Why," 2010).

NCLB legislation included some positives for public education. First and foremost, it was based on the principle that all children can learn and have the right to a high quality education. It ensured the implementation of rigorous standards for education and compelled schools to report students’ progress toward these standards (Borkowski & Sneed, 2006). NCLB also mandated that school and district test results would be disaggregated by race, ethnicity, socioeconomic status, disability, and English language learner status (Owens & Sunderman, 2006). Borkowski and Sneed (2006) observe that:

By mandating both annual assessments and the reporting of results by school, district, and particular subgroups, NCLB seeks to use these tendencies to inform parents and to promote school improvement. Moreover, there is some evidence that strong, clear state accountability systems may contribute to improved student performance. (p. 505)

By focusing on improving academic performance by all students, NCLB helped move schools and districts toward closing the achievement gap.

Questions of whether literacy programs were measuring up is not new. As long as standardized testing has been a part of school, with the advent of the first Scholastic Aptitude Test in 1923, public officials have harbored questions about student achievement (Bracey, 2000). By 1950, standardized testing had found a permanent place in school culture and was only escalated by reports like 1983’s Nation at Risk (Kohn, 2000). Both
Kohn (2000) and Bracey (2000) have raised concerns over using high-stakes testing to determine the quality of children’s education.

As we moved into the early 1990’s, a new round of high-stakes testing emerged throughout the country (Abrams & Madaus, 2003, p. 31). All across the nation, high-stakes tests are being pushed by politicians as the silver bullet to holding schools and teachers accountable. “Pressure to apply scientific methodology to the day-to-day work of teaching children to read and write is perhaps even stronger today than it was during the days of Sputnik and the Cold War” (Invernizzi et al., 2005, p. 610). This continued pressure for teachers to show the quantifiably results of their instruction is highlighted by requirements to use student growth measures as a large component of the teachers evaluation (Megan, 2012).

**Individuals with Disabilities Act**

Special education students have also been addressed through legislative acts. IDEA introduced progress monitoring of students through the collection of data and the use of frequent progress checks. IDEA supports the accountability and assessment measure introduced in NCLB by reinforcing that all students must be assessed with a standards-based measurement (Webber, 2006). IDEA also introduces early intervention services by allowing up to 15% of federal special education money to be used for services for children who have not been formally found to be special education eligible (Webber, 2006). IDEA was the catalyst for Response to Intervention (RtI) programs that introduced screening tools for evaluating student growth in areas such as early literacy (Bradley, Danielson, & Doolittle, 2005; Case, Speece, & Molloy, 2003; VanDerHeyden,
Snyder, Broussard, & Ramsdell, 2007). These assessments are then used to determine instructional programming for individual students (Chappuis et al., 2009).

**Early Literacy Policy**

Policy players are actively looking for ways to keep the United States competitive in a world market and ensure that students are ready for 21st Century jobs. Public concern over student achievement has helped to move forward policy focused on teacher accountability. Early literacy has been the focus of much of this policy reform, as reading skills have a direct impact on all content areas. According to The Educational Trust’s *Ed-Watch Online* (2005), “perhaps the most important task of elementary schools is to teach students to read well. Strong reading skills are the key to later success both in school and in life” (p. 2). The goal of all students reading on grade level by third grade has been set at the national level through both policy (*Goals 2000*, 1994; National Reading Panel, 2000; National Reading Summit, 1998) and federal legislative initiatives (Reading Excellence Act, 1999; NCLB, 2002). On January 8, 2002, when President Bush signed NCLB into law, every state was “presented with an unprecedented challenge: to implement a tightly prescribed accountability model with a goal of all students achieving grade-level proficiency in reading or language arts and mathematics within 12 years” (Erpenback, Forte-Fast, & Potts, 2003, p. 10).

Reading First is one of the biggest initiatives authorized by NCLB. Reading First grants were federal dollars that were distributed by the State Departments of Education to school districts who submitted proposals for how they would use scientifically based reading research to improve reading instruction ("A closer look," 2004). Reading First framed the four pillars of an effective reading program as follows: (a) valid and reliable
assessments, (b) instructional programs and aligned materials, (c) aligned professional
development, and (c) dynamic instructional leadership. It is the first pillar, valid and
reliable assessments that forced educators to reconsider how they were measuring reading

The accountability requirements in NCLB were designed to close the achievement
gap between high and low achieving students as well as between minority and
nonminority students (Gawlik, 2012). A decade after these legislative mandates, student
test scores in reading are not increasing at a rate that satisfies the public and a gap still
remains between subgroups. This has resulted in an increased emphasis on the use of
testing and diagnostic tools (Cooper, 2011; Weiss-Tisman, 2012).

Policy in Michigan

Some policy players in Michigan have tried to influence the political culture in
favor of policy matching their group and/or individual values, including a push to
increase school accountability. Michigan made its first steps into accountability in 1990
with Public Act 25 (Gawlik, 2012). Public Act 25 was the basis for the states first
accountability system by: (a) adopting a State Core Curriculum and proposing learning
outcomes by grade-level, (b) requiring schools to engage in a school improvement
process that included parents and teachers with a plan to monitor progress toward
goals, (c) creating a school accreditation system that evaluated curricula, staffing and
facilities, and (d) mandating annual reports for all schools ("Strengthening accountability
in," 2000). The Act was amended in 1995 to include the MEAP as a part of the
calculation for Michigan school accreditation, although the MEAP pre-dates any
accountability legislation (Burns, 2005).
Another shift took place in 2000 when Michigan aligned with Improving America’s Schools Act (IASA) and reported AYP status for all schools (Gawlik, 2012). The latest education reform in Michigan came in 2010, when the state passed legislation to align state policy with the Race to the Top initiative. This reform requires teacher evaluation to be in part based on student academic growth (Michigan School Code, 2010).

*Figure 2. Timeline of educational reform.*
Assessment

Throughout shifts in policy and changes in accountability requirements, schools, teachers, and administrators have struggled with how to assess students’ skills. The continued focus on school accountability and growing emphasis on using research-based practices for literacy assessment may cause disconnects between research and classroom practice (Invernizzi et al., 2005). Educators need to evaluate the purpose of testing as well as looking at assessment standards to determine the correct assessment tool.

Assessment Standards

Teachers and administrators may have a difficult time selecting assessments that meet the legislated requirements. To assist in this selection process the American Education Research Association, the American Psychological Association, and the National Council on Measurement in Education have recommended eight standards for evaluating educational assessment and assessment practices (Invernizzi et al., 2005). These assessment standards are:

1. Validity or the extent to which an assessment measures what it is designed to measure.
2. Reliability is an indicator that the assessment consistently measures what the test is designed to measure.
3. Test development suggests that test designers disclose the theoretical framework and disclose the process for test development procedures such as item analysis and field testing.
4. Fairness in testing refers to the responsibility of test makers to ensure that the assessment is unbiased and fair for all test takers.

5. Scales, norms, and score comparability must be fully described. Test developers must provide an overview of the way assessment results will be displayed as well as provide a full description of how to interpret the information and scores provided after the assessment.

6. Standardized administration, scoring, and reporting helps to ensure the accuracy and integrity of the assessment outcomes.

7. Testing individuals of diverse linguistic backgrounds ensures that students from all cultural backgrounds will get valid assessment results.

8. Responsibilities of policy decision makers include understanding that test results are used for multiple purposes and that schools may find themselves giving multiple assessments to meet the policy mandates currently in place. It is of paramount importance that assessments we are giving students were designed to measure the information we are trying to collect.

These assessment standards lay a framework for understanding the place of assessment in schools today.

**Reasons for Testing**

Students are given assessments for a variety of reasons. As early as 1973, the American Educational Testing Service had recognized five major categories of assessment: (a) placement which assists in placing children into a group within a class, or into a specific school, class, or treatment group, (b) prediction which is a measure of a student’s future performance, (c) assessment which measures a student’s competency at
one point in time, (d) diagnosis which is used to determine if a student is meeting the standard, as well as giving some insight into why the student is at a specific level, and (e) evaluation which focuses on using assessment results to determine things such as curriculum alignment or effectiveness of individual programs or teachers (Katz, 1973, p. 6). DIBELS is an assessment measure used by schools for prediction, placement, and diagnosis, while the MEAP is used primarily as an evaluation assessment.

**Assessment Types**

Understanding different assessment types and how to select an assessment that gives you the student information you seek, is critical to being able to understand assessment results. Yet, policymakers have been unable to come to consensus on the goal of mandated assessment programs (Goodman, 2006; Hinchey, 2010; Johnston & Costello, 2005). Do we want (a) to assess which content standards students understand, or (b) to know where our students rank locally, statewide and nationally (Sloane & Kelly, 2003)? Criterion-referenced assessments determine what concepts or skills students know while norm-referenced assessments rank students with respect to the achievement in broad areas of knowledge (Notar, Herrin, & Restauri, 2008).

Criterion-referenced assessments, such as the test you take to get you driver’s license or the MEAP, measure how well a person has learned a specific body of knowledge (Notar et al., 2008). These tests help teachers determine what students know or what students are able to do. These tests can be helpful in measuring the change in learning over time (Ornstein, 1993). Many researchers have offered definitions of criterion-referenced assessment, but perhaps Gray (1978) offered the most comprehensive definition, stating “Criterion-referenced tests are those designed to
produce measurements directly interpretable in terms of specified performance standards where the standards form a continuum of knowledge that is dependent on the prerequisite relations among the various levels of the continuum” (p. 227).

Norm-referenced tests, such as DIBELS or college entrance exams, are used to discriminate between high and low achievers by comparing test takers to each other (Notar et al., 2008). Many popular achievement tests are norm-referenced, such as the Iowa Test of Basic Skills and the National Assessment of Educational Progress (Ornstein, 1993).

**Reading Instruction and Assessment**

Mandates to ensure all children can read at grade-level by third grade have added a new sense of urgency to how we assess a student’s ability to read. Teaching reading requires expertise, is complex, and should be guided by a scientific knowledge base (Kame’enui & Simmons, 2008). How we go about assessing our students’ reading skills is no less complicated. To decipher the proper methods of assessment, we must first look at how children are taught to read.

**Reading Instruction**

In 1997, the National Reading Panel (NICHD, 2000) was convened by the United States Congress. Along with the United States Secretary of Education, this group analyzed more than 100,000 studies in search of the key factors to developing successful readers. Prior to this meta-analysis, schools had depended on state-wide criterion-referenced assessments to determine students’ achievement with very little regard for best practice reading instruction (Kiser, 2011).
The report by the National Reading Panel (NICHD, 2000) included the following five findings as significant in influencing reading instruction: (a) children’s parents are important in providing early literacy experiences, (b) children at risk for reading failure need to be identified early and provided with appropriate interventions, (c) phonemic awareness, phonics, and good literature are all important to the reading process and need to be integrated for instruction to be most effective for all students, (d) scientifically based information should be used to inform our policy and practice regarding reading instruction, and (e) the teachers’ instructional knowledge plays a big part in students’ reading instruction. Because of this, teachers need professional development as well as access to current research on reading instruction.

This initial work of the National Reading Panel (NICHD, 2000) clearly shows that successful reading instruction must address five critical areas: (a) phonemic awareness, (b) phonics, (c) fluency, (d) vocabulary, and (e) comprehension. It is these five components that were addressed in NCLB and the Reading First initiative as essential components of a reading program ("A closer look," 2004). This seminal report encouraged teachers to blend some of the approaches used for instruction in the past. Education professionals need to meld the best of past practice from the basal reading series that had dominated reading instruction in the 1930s, 1940s, and 1950s where 90% of children were taught to read through sight words and vocabulary building (Sadoski, 2004), to Kenneth Goodman’s whole reading method, built on the premise that skilled readers could derive meaning from context (Smith & Goodman, 2008) and the push for phonemic awareness introduced in the 1990s (Manning & Kato, 2006) into one comprehensive reading program.
As schools implemented these new reading programs, the need for assessments and monitoring tools became evident. High quality instruction relies on teachers having the best possible information about their students reading ability and progress toward individual goals.

**Assessment of Reading Skills**

The Reading First emphasis on valid and reliable assessments for all four pillars of reading pushed schools to look beyond traditional assessment methods of silently reading a passage and then answering questions based on what was read (Sadoski, 2004). Educators across the country turned to curriculum-based measurements (CBM) as the answer (Hintze, Christ, & Methe, 2006). CBMs are a set of measurement activities that use “direct observation and recording of a student’s performance in the local curriculum as a basis for gathering information to make instructional decisions” (Deno, 1987, p. 41). Hintze, Christ, and Methe (2006) describe how CBMs rely on measuring very specific sub-skills to determine mastery, an approach which breaks down curriculum outcomes into a set of smaller skills for which assessment material is developed.

Researchers from the psychology Department of the University of Oregon developed DIBELS based upon the procedures used for CBM (University of Oregon, 2003). It was designed to measure the core reading components as set forth by the National Reading Panel: alphabetic principle, phonological awareness, vocabulary, comprehension, and accuracy and fluency with connected text (Good & Kaminski, 2002). Seay (2005) describes DIBELS as a “quick, low-cost way for teachers and schools to monitor student progress and to identify students who need intensive reading interventions” (p. 22). According to the test developers, DIBELS is comprised of a set of
specific indicators that are applied to quantify student performance in reading (Good & Kaminski, 2002). The literacy measures used in the DIBELS assessments are generic, meaning that they do not reflect any specific school’s curriculum (University of Oregon, 2003).

The DIBELS assessment is comprised of two components. One set is used by the classroom teacher for progress monitoring. Progress monitoring was originally developed by clinical psychologists to help track changes in treatment progress and allow the practitioner to modify the treatment or, in the case of a teacher, the instruction (Overington & Ionita, 2012). Progress monitoring should take place weekly and should only take one to five minutes per student (McLane, 2006). If students do not show growth during progress monitoring, teachers may choose to use an intervention, such as changing teaching techniques, adding instructional time, offering additional support, offering additional tutoring, or change grouping arrangements (McLane, 2006). The second set is the benchmark tests that are to be given three times per year (Good & Kaminski, 2002). The administration and scoring of the benchmark assessments has been standardized, so that all students experience the testing and scoring in the same way (University of Oregon, 2010).

With the passage of NCLB and Reading First legislation, DIBELS has been increasingly used in schools across the country (Li & Zhang, 2008). Indeed, DIBELS was the only early literacy assessment to be recommended by the Reading Assessment Committee, and DIBELS became a required assessment for Reading First schools (Afflerback, 2007; Goodman, 2006).
DIBELS measures reading comprehension through the ORF and Retell Fluency (RF) subtests (University of Oregon, 2010). The Literacy Dictionary (Harris, Hodges & Association, 1995) defines fluency as “freedom from word-identification problems that might hinder comprehension” and a fluent reader as “any person who reads smoothly, without hesitation, and with comprehension” (p. 85). The ORF test also assesses a student’s accuracy and fluency with text. According to Good and Kaminski (2002), the ORF test has students read a standardized passage out loud for one minute. Hesitations greater than three seconds, omissions and mispronunciations are scored as errors. The student’s ORF is determined by the number of words read correctly during the testing period (Good & Kaminski, 2002). After completing the ORF test, students are asked to retell the passage. Students must be able to retell 25% of the information to be proficient on the RF test.

The authors of the DIBELS assessments have described the ORF measure as “a standardized, individually administered test of accuracy and fluency with connected text” (Good et al., 2001, p. 683). DIBELS offers “a standardized set of passages and administration procedures designed to (a) identify children who may need additional instructional support and (b) monitor progress toward instructional goals” (Good, Gruba, & Kaminski, 2001, p. 683). The ORF selections are adjusted for the reading-level of the students being tested.

The development of the DIBELS assessment is based on several underlying instructional theories of how students learn to read. These include both philosophical and psychological ideas about education. Some researchers feel that DIBELS reflects an essentialist philosophy of learning (Li & Zhang, 2008). Essentialists believe that there is
a core set of information that must be transmitted to students. An underlying assumption is that this information can be measured. In the essentialist classroom, teachers are the givers of knowledge while students receive information. This knowledge transfer culminates in some form of standardized assessment or evaluation, such as DIBELS (Ornstein & Hunkins, 2004).

From a psychological perspective, DIBELS reflects a behaviorist perspective. Behaviorists are concerned with observable or measurable aspects of human behavior and explain student learning in terms of observable responses (Li & Zhang, 2008). Like the DIBELS assessment, behaviorists attempt to break learning into the smallest observable elements. DIBELS also displays principles of LeBerge’s and Samuels’ automaticity theory (Seay, 2005). Processes are considered to be automatic when they demonstrate four traits: speed, effortlessness, autonomy, and lack of conscious awareness (Kuhn, Schwanenflugel, & Meisinger, 2010). From a reading instructional standpoint, students who struggle to sound out words divert all of their attention to decoding and are unable to concentrate on comprehension.

DIBELS reflects some of the values of the reading readiness developmental theory. Li and Zhang (2008) explain that “advocates of reading readiness theory argue that learning to read begins only after a set of prerequisite perceptual and pre-reading skills have been mastered” (p. 45). Advocates of reading readiness believe that all children master the same skills in a set sequence while learning to read.

**MEAP Overview**

The MEAP, a criterion-referenced assessment, had been the standard measurement of students’ proficiency in relation to the Michigan Curriculum Framework.
(Michigan Assessment Integrity Guidelines, 2009). The MEAP assessment was initiated in 1969 on the recommendation of the Michigan State Board of Education, with the support of then Governor, William G. Milliken. Funded by the Michigan legislature through Public Act 307, the test was first administered during the 1969-70 school year. At its inception, the MEAP was used to give teachers information about what students knew and were able to do as compared to the standards set by the State Board of Education (State of Michigan, 2004). Originally, the purpose of the MEAP was to allow teachers and schools to evaluate the strengths and weaknesses of adopted curriculum and then make modification to the instruction as needed (Michigan Board of Education, 2004). Over time and with the amendment of Michigan Public Act 25 in 1995, this purpose had morphed into an accountability measure (Gawlik, 2012).

The MEAP test had been aligned with the Michigan Essential Goals and Objectives for Reading and Math and the GLCS approved by the Michigan State Board of Education (McGlinchey & Hixson, 2004). The Michigan State Board of Education approved the standards and cut scores for the MEAP. These proficiency levels were set by teams of Michigan educators. Students that reached or exceeded a set score that represented the achievement of the minimum standard were considered proficient (Burns, 2005).

In 2015, the MEAP testing program was replaced with an updated testing program called Michigan Student Test of Education Progress (M-STEP). The MEAP reading assessment was a group-administered, untimed test with standard directions and procedures. Students typically had two test periods to complete a reading section, each lasting approximately 50 minutes. During the test, students read a story selection and
information text. Each reading passage was followed by 20 multiple-choice questions (McGlinchey & Hixson, 2004). These questions targeted three specific ways for students to use comprehension skills to construct meaning: intersentence, text, and beyond text (Michigan State Board of Education, 1999). Intersentence questions assessed a student’s ability to comprehend information from two or three sentences, while text questions evaluated understanding of a larger section of the passage. Beyond text questions measured the readers ability at integrate personal experience with what was read.

The development of criterion-referenced assessments such as MEAP, is based on the Mastery Learning model. This model is an evolution of the earlier programmed learning approaches and structural approaches employed by turn of the century educators (Horne, 1984). At the heart of Mastery Learning theory is the belief that learning should be measured against a set of predetermined criteria (Dyson, 2008; Horne, 1984; Reynolds & Fletcher-Janzen, 2007). Torshen (1977) describes the Mastery Learning as being comprised of six components including “objectives, pre-assessment, instruction, diagnostic assessment, prescription, and post-assessment” (p. 40). The post-assessment component of Mastery Learning has evolved into Mastery Testing.

The philosophy behind Mastery Testing is that “a level of competency can be determined at or above which an individual can be said to have mastered the domain or unit of learning. Below this level all learners are deemed to be non-masters” (Horne, 1984, p. 162). Mastery Testing has been linked to the behavioral objective movement (Horne, 1984; Torshen, 1977).
Response to Intervention

Hailed as a solution for decreasing the achievement gap and supporting at risk students, RtI is a model to identify and remediate students at risk for failure (Bradley et al., 2005; Case et al., 2003; VanDerHeyden, Snyder et al., 2007). The National Association of State Directors of Special Education (NASDSE, 2006), define RtI as:

The practice of providing high-quality instruction and interventions matched to student’s need, monitoring progress frequently to make decisions about changes in instruction or goals, and applying child response data to important educational decisions. (p. 5)

Often RtI frameworks make reference to hierarchical or a three tiered model of intervention, with the tiers as follows: (I) the general education program with at least three screenings each year, the use of data to plan instruction and professional development for teachers, (II) targeted interventions within a fixed period of time, with additional progress monitoring, and (III) intensified individual services customized to the student, such as smaller groups or longer intervention blocks (Bradley et al., 2005; Wanzek & Cavanaugh, 2012). In this type of framework, all students are assessed initially to evaluate the classroom literacy environment and the core instruction program (VanDerHeyden et al., 2007). During Tier I intervention, if the performance of most students is at proficiency levels, the core program is considered to be acceptable. Students who are identified as high risk of developing reading difficulties are put into Tier II intervention groups such as small group instruction or they may receive additional assessment as a diagnostic tool for selecting precise instructional targets.
The core features of RtI include: (a) high quality, research-based classroom instruction, (b) universal screening, (c) continuous progress monitoring, (d) research-based secondary interventions, (e) progress monitoring during interventions, and (f) fidelity measures (Bradley et al., 2005). Wanzek and Cavanaugh (2012) describe effective RtI implementation as:

Effective RtI implementation requires application of instruction and intervention based on the scientific evidence along with the use of data to make key decisions regarding not only students’ instructional needs but also the intensity of interventions in terms of time, frequency, duration, and instructional group size. (p. 193)

Quality general education programing with reliable screening tools are essential to a successful RtI model.

Studies Comparing DIBELS and Criterion-Referenced Assessments

Current research on the DIBELS assessment has resulted in mixed outcomes concerning its ability to predict performance on criterion referenced assessments. DIBELS was the quality, early reading assessment tool recommended by the Reading First Panel, who found it to be a good measure of students’ reading abilities (Kiser, 2011). Several studies have been conducted by the DIBELS development team and have found it to be a valid and reliable indicator of a student’s future reading success (Good, Simmons, & Kame’enui, 2001; Kaminski, 2004; Kaminski & Good, 1998).

Other researchers have also found positive connections. For example, Shaw and Shaw (2002) set out to determine the usefulness of DIBELS ORF assessment as a predictive measure on the Grade 3 reading portion of the Colorado State Assessment
Program (CSAP). CSAP is a standards-based reading comprehension assessment administered annually to all students in Colorado. Although the study only included 52 students from one school, Shaw and Shaw (2002) found a high correlation among student reading levels on ORF and CSAP. In another study, Cook (2003) looked at the correlation of DIBELS and the Stanford Achievement Test, Ninth Edition (SAT 9). In a sample of 79 students from five first-grade classrooms, DIBELS was found to provide a valid inventory of early reading skills in comparison to the reading standards measured. Finally, in 2005, Seay examined the relationship between DIBELS ORF and the Stanford Achievement Test, Tenth Edition (SAT 10) Reading Comprehension sub-test given to third grade students. This study reviewed testing records from 2,881 students and found that DIBESL ORF did moderately explain the variance in SAT 10 Reading Comprehension scores. Seay (2005) found that students who did not reach benchmark level in one grade continued to score below benchmark in future years.

Some researchers, however, have voiced concerns over using DIBELS as the primary early literacy assessment. In a technical report on Michigan’s Reading First schools, Shillings, Carlisle, Scott, and Zeng (2004) found that the “predictive or criterion validity of Oral Reading Fluency with respect to standardized measures of reading comprehension is convincing” (p. 13). Yet, the same report goes on to indicate that one third of students in the low-risk category on the fall ORF measure were reading below grade-level on the spring Iowa Test of Basic Skills.

Often researchers who found concerns were like Hetch and Greenfield (2000) who compared the predictive validity and accuracy of DIBELS with first-grade-teachers ratings in predicting students’ future reading ability and the DIBELS assessments.
Teachers ratings and third grade DIBELS assessments reviewed from 170 students in the National Head Start/Public School Early Childhood Transition Demonstration Project found that DIBELS was no better than teachers at predicting which student would not reach proficiency levels by third grade. Goodman (2006) pointed out that DIBELS places an overemphasis on segmented reading skills at the expense of reading as a skilled process. Afflerback (2007) and Deeney (2010) both expressed concerns that DIBELS has a narrow definition of reading fluency, focusing on reading speed and virtually ignoring the students’ ability to comprehend what they have read, sending the wrong messages to students and teachers. DIBELS perpetuates the “teach to the test” attitude pervasive in the education community (Afflerbach, 2007; Goodman, 2006; Li & Zhang, 2008). Most of these concerns rest on trepidation regarding DIBELS’ ability to adequately assess vocabulary or comprehension, two benchmarks of fluent readers (Gordinier & Foster, 2004).

**Chapter II Summary**

There is no doubt that the role of high-stakes assessments has increased in schools over the past several decades (Bober, 2004; Carsten, 2004). This literature review summarized the national and state legislative and policy issues that have intensified the accountability movement, different assessment types, and the reliability of the DIBELS assessment. Researcher outcomes are mixed regarding the predictive value of the DIBELS assessment with no available research on DIBELS ability to predict students’ future success on the MEAP. These results indicate a need for further research to examine the relationship between the DIBELS assessment and other assessments of
reading performance as well as the ability of DIBELS to predict reading achievement on other high-stakes assessments such as the MEAP.
CHAPTER III

METHODOLOGY

While there has been research done on the DIBELS Oral Reading Fluency (ORF) (e.g., Shelton, Altweger, & Jordan, 2009) and MEAP assessment (e.g., Burns, 1998; Cunningham, 2004) respectively, prior to this proposed study, little research could be found on the relationship between the two assessments. Therefore, the purpose of this study is to explore the relationship between an interim assessment tool (i.e., DIBELS Oral Reading Fluency (ORF) assessments) commonly used in early grades to determine if students need interventions, and how well those students are reading at the end of elementary school (as measured by their fifth grade MEAP reading scores), as well as the trends within student reading scores from grades three through five. Chapter III provides a detailed description of the research methodology used for this study. In this chapter, the research design and its appropriateness are presented. Information on the study population, data collection procedures and rationale, analytical methods, and limitations are also discussed in this chapter.

Research Questions

1. To what extent is there a relationship between students’ DIBELS Oral Reading Fluency at the beginning of third grade and their MEAP Reading scores in third, fourth and fifth grades?

2. To what extent is there a relationship between students’ MEAP Reading proficiency statuses at fifth grade, and their DIBELS’ level (i.e., most at risk, some risk, low risk) at the beginning of third grade?

3. In reference to students’ third grade mean MEAP Reading scale score:
a. to what extent are there difference between student labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessment); and
b. how do the intervening variables of gender and socioeconomic status effect these difference?

4. In reference to student’ MEAP Reading score growth trajectories in grades three through five:
   a. to what extent are there differences between students labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessments); and
   b. how do the intervening variables of gender and socioeconomic status effect these differences?

**Research Design**

This study incorporated a quantitative, correlational, ex post facto design to examine the relationship between students’ third grade Oral Reading Fluency scores on their end of year (spring) DIBELS assessment and their fifth grade MEAP reading results. The ex post facto design is appropriate since the study examined student test scores and sub-group differences after the independent variables have occurred. Ex post facto is done retrospectively, that is, existing conditions are explored for possible relationships. Kerlinger and Lee (1999) described ex post facto design as beginning with the observed outcome, the dependent variable (e.g., the MEAP fifth grade assessment), and then investigating the independent variable (e.g., the DIBELS third grade end of year assessment) for possible relationship to the dependent variable.
A quantitative correlational research design is considered appropriate for this study, since investigation of relationships between variables, including their strength and direction of association, is the motive of this study. According to Creswell (2008), correlational designs are “procedures in quantitative research in which investigators measure the degree of association or relationship between two or more variables using statistical procedures” (p. 52). The quantitative method was selected to utilize an explanatory correlational design.

In correlational research, the two primary correlation designs are: explanatory and prediction (Creswell, 2008). Explanatory correlational research design is defined as “the extent to which two variables (or more) co-vary, that is, where changes in one variable are reflected in changes in the other” (Creswell, 2008, p. 327). “The objective of prediction design is to anticipate outcomes by using certain variables as predictors” (Creswell, 2008, p. 328). However, the intent of this study is not to make predictions about outcomes. In the case of this study, the purpose is to show the extent of the relationship between the variables of third grade DIBELS ORF and fifth grade MEAP assessment scores; therefore, an explanatory design is appropriate.

Quantitative research addresses questions about relationships between measured variables for the purpose of explaining, predicting, and controlling events (Leedy & Ormrod, 2005). The quantitative approach is appropriate because it reduces potential biases by focusing on direct responses without interpretation. Quantitative research involves the use of specific and narrow questions targeted toward measuring and explaining variable relationships (Cooper & Schindler, 2005; Creswell, 2008).
Qualitative research design was not selected for the proposed study. Qualitative research design is not appropriate for this current study because this process analyzes words or text from participants and inquiries are conducted in a more subjective and biased manner (Creswell, 2008).

A variety of methods are available to collect information to examine relationships between third grade DIBELS ORF assessment scores and fifth grade MEAP assessment scores. A retrospective study method was chosen for this study. The dataset used for this study was collected by the Regional Data Initiative (RDI) program. The dataset includes information collected for 2008 through 2013 with respect to interim and summative assessment data for students.

The RDI database provided more detailed information than could be collected by survey sampling or with focus groups due to temporal and cost considerations. Also, use of the RDI database allowed for more objective data collection than could be done if collecting more subjective participant answers on surveys or with focus groups.

**Population and Sample**

This study analyzed data from 2,220 students enrolled in seven school districts participating in a Regional Data Initiative (RDI) grant program. Michigan used $11.6 million of federal Title I Part D from the American Recovery and Reinvestment Act (ARRA) of 2009 funds to start up eight RDI programs. The goal of the RDI program is to combine state and local data to better inform instructional decisions and other educational practices at the district, building, and classroom levels. All school districts participating in this grant project agreed to allow data to be used in the State’s Research Collaborative, which allows researchers at all levels to use this data for research purposes.
Students included in the study were administered the DIBELS Oral Reading Fluency assessment at the beginning of third grade, as well as the MEAP reading assessment in third, fourth, and fifth grades. Students come from one intermediate school district (seven school districts) in one midwestern state. Student data for this study includes scores for DIBELS and MEAP, as well as, gender, SES, district and school attended. Such student data was accessed through Data Director, the RDI data warehouse software.

**Instrumentation**

**DIBELS**

The Dynamic Indicators of Basic Early Literacy Skills, 6th edition (DIBELS; University of Oregon, 2003) is an instructional assessment created by the University of Oregon and, since the early 1980s, has been used to collect data on reading fluency. The data collected for this study included the DIBELS Oral Fluency Monitoring Assessment (DIBELS ORF). The DIBELS ORF includes reading passages that reflect the curriculum for the grade in which the students are being assessed. Each grade level has benchmarked passages that are administered to students individually during the fall, winter, and spring. At each assessment period, three separate passages are administered,
with each passage yielding an oral reading fluency score. These scores consist of the number of words read correctly in one minute. The student’s final score, a median of the scores for the three passages, is recorded as the score for the benchmark period. The median scores are then used to classify a student’s oral reading fluency for the benchmark period. The end of year (spring) benchmark classifications are (a) Most At Risk (median score of 0 – 79), (b) Some Risk (median score of 80-109), and (c) Low Risk (median score of 110 and above). The DIBELS third grade benchmark classifications have been recorded for each student.

Cronbach’s alpha coefficients of reliability for the 6th edition of the DIBELS ORF for third grade have been measured with values between .92 and .98 (Dynamic Measurement Group, 2008). In a study by Baker et al, (2008), the DIBELS ORF returned a concurrent validity correlation of .67 and predictive validity correlations of .65 and .68 when compared to the Stanford Assessment Test – 10 (SAT-10). Barger (2003) reported concurrent validity of the DIBELS ORF with the North Carolina End-of-Grade Test (NC End of Grade) of .73. Concurrent validity of the DIBELS ORF with The Florida Comprehensive Assessment Test – Sunshine State Standards (FCAT-SSS) included correlations of \( r = .70 \) (Buck and Torgeson, 2003), and \( r = .71 \) (Roehring, Pettshcer, Nettles, Hudson, & Torgeson, 2008). Other concurrent validity measures included the DIBELS ORF with (a) The Iowa Tests of Basic Skills (ITBS) with correlations between .56 and .68 (Schilling et al., 2007), (b) The Pennsylvania System of School Assessment (PSSA) with a correlation of .68 (Shapiro, Solari, & Petscher, 2008), and (c) the Colorado State Assessment Program (CSAP) with a correlation of .80.
Predictive validity correlation coefficients ranging from .65 to .71 have been reported between the DIBELS ORF and SAT-10 (Baker et al., 2008; Roehring at al., 2008). Predictive validity coefficients ranging from .70 to .73 have been reported for the DIBELS ORF with the CSAP (Shaw & Shaw, 2002; Wood, 2006).

MEAP

The Michigan Educational Assessment Program (MEAP) was a testing program based on the Michigan Curriculum Framework and GLCE, as approved by the State Board of Education. The program assessed reading and math in third through eighth grade. This study used the MEAP English and Language Arts -Reading assessment scores (MEAP-ELA-Reading) for third through fifth grade for the students of included in the study. The MEAP reading assessment was a group-administered untimed test with standard directions and procedures. Test administration took place during two 50 minute testing sessions. Student performance was scored on twenty multiple choice questions on each of two reading passages, one story selection, and one informational selection (McGlinchey, 2004). The questions for each passage were classified as one of three different types of constructing meaning items: Inter-sentences, Text, or Beyond Text (Michigan State Board of Education, 1999). This study used the scaled scores of the MEAP ELA – Reading assessments for the students from third through fifth grade to determine student progress over time. Scale scores are statistical conversions of raw scores that adjust for slight differences in underlying ability levels at each score point and permit comparison of assessment results across different test administrations. The range of scale scores changed for each year of assessment, however, all scores were scaled with a standard deviation of 25 for comparison between assessment times (MEAP Technical
Report 2010-2011). The scale score is stable because it allows for students’ scores to be reported on the same scale regardless of which year or form of a particular assessment was completed. Students were also classified according to proficiency level in each grade as (a) Advanced, (b) Proficient, (c) Partially Proficient, and (d) Not Proficient. Descriptive statistics (frequency and percentages) are presented for the MEAP-ELA-Reading proficiency levels for the students over the three years of study (third grade through fifth grade).

According to the MEAP Technical Report 2011-2012, criterion validity had not been established for the MEAP assessment. However, median reliability for the MEAP ELA Reading assessment using Cronbach’s coefficient was determined as .85 for third grade and .86 for fourth, and fifth grades respectively.

**Data Analysis**

The data for this sample were analyzed through a variety of methods. Attendance, gender, SES, DIBELS benchmark ranking, and MEAP proficiency ranking were examined with frequency tables and shown in a tabular form. Measures of central tendency including means, standard deviations, and score ranges are presented for the scale scores of the MEAP-ELA-Reading assessment. The analyses were conducted using IBM SPSS statistical software version 22. The research questions were tested using the statistical platform R and IBM SPSS version 22. The R package `lmerTest` was used to test the hierarchical linear model (HLM) for research question one. SPSS was used for the cross tabulation and correlation used to test research question two. The R package `lavaan` was used to test research question three and four. Table 1 presents a cross-walk
table to readily identify the analysis procedures for each research question. All analyses are based on 0.05 level of significance ($\alpha=0.05$).

**Operationalization of Variables of Study**

**SES.** A student was classified as having low socio-economic status (SES) if they had participated in the free or reduced price lunch program in third grade. The variable is dichotomous and was coded as 1 = Low SES, 0 = Not low SES. Not low SES was the reference category for correlational and HLM analysis. SES was included as a Level 2 variable in the HLM and also included in the growth curve analysis.

**Gender.** Gender is dichotomous and was coded as 1 = female, 0 = male. The reference category for the HLM analysis was male. Gender was included as a Level 2 variable in the HLM and also included in the growth curve analysis.

**DIBELS ORF.** The DIBELS ORF variable consists of the end-of-year (spring) benchmark classifications of the DIBELS ORF assessment and coded as (a) Most At Risk (score of 0 – 79) = 1, (b) Some Risk (score of 80-109) = 2, and (c) Low Risk (score of 110 and above) = 3. Higher rankings on the DIBELS ORF variable are associated with lower risk. The DIBELS benchmark classifications were analyzed for each student in the third grade. The DIBELS benchmark classification in third grade was included as a nominal Level 2 variable in the HLM model and represented with indicator variables for (a) some risk, and (b) low risk. The classification of most “at risk” was the reference category. The indicator variables were coded as 0 = no, 1 = yes.

**School and District Site.** The school and school district site variable are nominal and were recorded for each student’s third grade year. The School Site variable was
included as a level 2 variable in the HLM. School District was only used in the preliminary analysis to describe the sample.

**Grade Level.** Student grade-level (3, 4, and 5) were included as a within-groups time measure for the MEAP-ELA-Reading assessment outcome in the HLM.

**MEAP-ELA-Reading.** Scaled scores of the MEAP ELA reading assessments for the students from third to fifth grade were used in this study. The range of scale scores change for each year of assessment, however, all scores were scaled with a standard deviation of 25 for comparison between assessment times. (MEAP Technical Report 2010-2011). The scaled score is continuous, and was included as a dependent variable in the HLM and growth curve analysis.

Student proficiency levels of the MEAP ELA reading assessment were also collected. The proficiency levels are ordinal and were coded as (a) Advanced = 4, (b) Proficient = 3, (c) Partially Proficient = 2, and (d) Not Proficient = 1.

Table 1 summarizes my research questions, key variables and the statistical analysis for each.
Table 1

*Data Analysis Cross-Walk Table*

<table>
<thead>
<tr>
<th>Research Question / Hypotheses</th>
<th>Variables</th>
<th>Statistical Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To what extent is there a relationship between students’ DIBELS Oral Reading Fluency at the beginning of third grade and their MEAP Reading scores in third, fourth and fifth grades?</td>
<td>DV: MEAP Reading Scale Score (Grade 3)</td>
<td>Linear Regression</td>
</tr>
<tr>
<td></td>
<td>MEAP Reading Scale Score (Grade 4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MEAP Reading Scale Score (Grade 5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>IV: DIBELS ORF Score (Grade 3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender (control IV)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SES (control IV)</td>
<td></td>
</tr>
<tr>
<td>2. To what extent is there a relationship between students’ MEAP Reading proficiency statuses in fifth grade, and their DIBELS’ label (i.e., most at risk, some risk, low risk) at the beginning of third grade?</td>
<td>DV: MEAP Reading proficiency level (Grade 5)</td>
<td>Pearson’s Chi Square</td>
</tr>
<tr>
<td></td>
<td>IV: DIBLE-ORF label (Grade 3)</td>
<td></td>
</tr>
<tr>
<td>3. In reference to students’ third grade mean MEAP Reading scale scores: to what extent are there differences between students labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessment); and how do the intervening variables of gender and socioeconomic status effect these differences?</td>
<td>DV: MEAP Reading proficiency level (Grade 3)</td>
<td>Hierarchical Linear Model</td>
</tr>
<tr>
<td></td>
<td>Level I IV’s: DIBLE-ORF label (Grade 3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Between-groups IV)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender (control IV)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SES (control IV)</td>
<td></td>
</tr>
<tr>
<td>4. In reference to students’ MEAP Reading score growth trajectories in grades three through five: to what extent are there differences between students labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessment); and how do the intervening variables of gender and socioeconomic status effect these differences?</td>
<td>DV: MEAP-ELA-Reading Scale Score</td>
<td>Hierarchical Linear Model</td>
</tr>
<tr>
<td></td>
<td>Level II IV’s: DIBLE-ORF label (Grade 3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(Between-groups IV)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gender (control IV)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>SES (control IV)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* DV = Dependent Variable; IV = Independent Variable.
Test of Hypotheses According to Research Question

The data for this sample was examined through a variety of methods matching each specific research question.

Research question one asks the following: To what extent is there a relationship between each students’ DIBELS Oral Reading Fluency and each students’ MEAP Reading scores in third, fourth and fifth grade? In order to answer the research questions posed, I used a linear regression model. First, a scatter plot was used to determine a correlation between variables. Before employing a regression analysis, the basic assumptions were reviewed and met: the observations are independent, MEAP scores have a normal distribution, the variance of error is consistent across observations (homoscedasticity), and the scatter plot indicates the possibility of a linear relationship. A regression test with a significance level of 0.05 was then applied to test for a relationship between the categorical variables.

Research question two asks the following: To what extent is there a relationship between each students’ MEAP Reading proficiency status at fifth grade and each students’ DIBELS label (most at risk, some risk, low risk) at the beginning of third grade? Pearson’s chi-square test with a significance level of 0.05 was applied to test for independence between the categorical variables. The basic assumptions were reviewed. One of the assumptions of chi-square is that the expected frequency of each cell is greater than 5, or that no more than 20% of cells have an expected count of 5 or less. The data must also meet the other assumptions: the observations must be independent, only appear in one cell, and must be measured as frequencies. A Spearman’s correlation was also conducted between the DIBELS ORF and MEAP scores. This correlation adds to the
understanding of the relationship between the two variables because it treats the
categories as ordinal instead of simply categorical.

Research Questions three and four were addressed with a HLM growth curve
model. Research question three asks the following: In reference to students’ third grade
mean MEAP Reading scale scores, to what extent are there differences: (a) between
students labeled most at risk, some risk and low risk based on their third grade DIBELS
assessment; and (b) how do various intervening variables (i.e., gender, socioeconomic
status) effect these differences? One section of the growth curve analysis, namely the
regressions onto the intercept of the model, was examined to address research question
three. The intercept is set at the starting time point for the dependent variable which is
third grade normed MEAP scores for the current data. The DIBELS ORF categories,
gender, SES, and school were all dummy coded and entered as predictors of the intercept.

Research question four asks the following: In reference to students’ MEAP
Reading score growth trajectories in grades three through five, to what extent are their
differences: (a) between students labeled most at risk, some risk and low risk based on
their third grade DIBELS assessment; and (b) how do various intervening variables (i.e.,
gender, socioeconomic status) effect these differences? The same HLM growth curve
analysis was used to investigate both research question three and four, thus adequate
model fit was already established before interpreting the statistical analysis used to
answer question four. The regressions onto the slope of the normed MEAP scores were
used to evaluate this question. Based on the preliminary analysis which showed that the
average rate of change in normed MEAP scores from third to fifth grade was fairly
consistent, a linear growth rate was specified as part of the model. The dummy codes for
the DIBELS ORF categories and covariates were all used to predict this linear change in MEAP scores. The regression coefficients were interpreted as the rate of change in normed MEAP scores for each predictor. In addition, the covariance of the intercept with the slope was interpreted as the relationship between MEAP scores in third grade and the rate of change in MEAP scores. For instance, did students who started with lower MEAP scores increase or decrease more rapidly than students with higher initial MEAP scores. The overall normed means for the MEAP scores were plotted for the three DIBELS ORF scores and are shown in a figure in order to better visualize the rate of change in the dependent variable by the primary independent variable.

**Statistical Models**

HLM is a powerful tool used in modeling cross-level effects and partition variance and covariance elements of fitted models (Raudenbush & Bryk, 2002). In this way, HLM analyzes growth over time. HLM also allows us to look at the effect of the school on student scores and a two-level model was implemented for this study. Level I represents the repeated measures across grades, Level II represents the student and includes variables SES, Gender, DIBELS benchmark classification for third grade using the following model:

**Level 1**

MEAP score for individual i at time t.

\[ Y_{it} = \pi_{0ij} + \pi_{1ij} \text{ (Grade Level)} + e_{it} \]

Where:

\[ Y_{ij} = \text{MEAP score for person i in school j at grade t} \]

\[ \text{Grade level} = \text{grade at time t for person i in district j} \]

\[ \pi_{0ij} = \text{MEAP score at grade 3 for person i in district j} \]
\[ \pi_{1ij} = \text{the linear change in MEAP score for person } i \text{ for each grade} \]

\[ e_{ij} = \text{random error} \]

**Level 2:**

\[ \pi_{0i} = \beta_{00} + \beta_{01} (\text{DIBELS@3}) + \beta_{02} (\text{DIBELS low}) + \beta_{03} (\text{Male}) + \beta_{04} (\text{not low SES}) + \sum_{i=6}^{13} \beta_{0i} (\text{school}) + R_{oi} \]

\[ \pi_{1ij} = \beta_{10j} + \beta_{11} (\text{DIBEL some risk}) + \beta_{12} (\text{DIBELS low}) + \beta_{13} (\text{Male}) + \beta_{14} (\text{not low SES}) + \sum_{i=6}^{13} \beta_{1i} (\text{school}) + R_{oi} \]

\[ \beta_{00} = \text{mean at third grade} \]

**Power Analysis**

Power for the primary analysis was estimated to detect differences in growth between low and middle, and between middle and high scoring students on the DIBELS-ORF assessment. Power was estimated using the Optimal Design software, for longitudinal and multilevel research (Version 1.55; Raudenbush, Spybrook, Liu, & Congdon, 2005).

**Limitations and Delimitations**

Students are primarily from rural school districts with the majority of participants falling into the economically disadvantaged range. This means that generalizations of the results of this study are delimited to similar populations. The study also had a fairly small “N” in cell counts of the HLM due to the levels of nesting, which is a limitation. Because students were nested in schools, many students who did move between school did not have enough data points to be included in the study and, a large number of these who were excluded, fell in the economically disadvantaged category. Another limitation
is that the databases used did not have data on whether they had interventions, what type of intervention was provided, or the duration of interventions that students may have received once classified as most at risk as identified through the DIBELS ORF assessment.

**Chapter III Summary**

This study used a quantitative design utilizing ex post facto data from student’s third grade Oral Reading Fluency scores on the end-of-year DIBELS assessment and their fifth grade MEAP reading results, with third, fourth, and fifth grade trend data included. This data was accessed using the Regional RDI data warehouse system. Chapter IV presents the results of the statistical analyses detailed in this Chapter.
CHAPTER IV
RESULTS

In Chapter 4, the results of the research are presented in a descriptive textual format as well as with tables. The results chapter is separated into four sections (a) descriptive findings as related to the selected sample of students for each of the schools and districts included in the study, (b) presentation of variables used for analysis, (c) investigation of assumptions as they relate to the inferential analysis, and (d) inferential analysis results. The preliminary analysis section describes the study sample and shows the bivariate relationships among variables for the covariates, DIBELS ORF scores, and normed MEAP scores.

The four research questions are addressed in the primary analysis section. Alpha was set at .05 and IBM SPSS version 22 was used for all preliminary analyses and to answer research question two. The statistical platform R was used for the primary analyses. In particular, the lmerTest package was used for research question one and the lavaan package was used for research questions three and four.

The purpose of this study was to explore the relationship between an interim assessment tool (i.e., DIBELS Oral Reading Fluency (ORF) assessment) commonly used in early grades to determine if students need interventions, and how well those students are reading at the end of the elementary school (as measured by their fifth grade MEAP reading scores), as well as the trends within student reading scores from grades three through five. Specifically, the study addressed the following research questions:
1. To what extent is there a relationship between students’ DIBELS Oral Reading Fluency at the beginning of third grade and their MEAP Reading scores in third, fourth and fifth grades?

2. To what extent is there a relationship between student’s MEAP Reading proficiency statuses in fifth grade, and their DIBELS’ label (i.e., most at risk, some risk, low risk) at the beginning of third grade?

3. In reference to students’ third grade mean MEAP Reading scale scores:
   a. to what extent are there differences between students labeled most at risk, some risk, and low risk (as based on their third grade DIBELS assessment); and
   b. how do the intervening variables of gender and socioeconomic status effect these differences?

4. In reference to students’ MEAP Reading score growth trajectories in grades three through five:
   a. to what extent are there differences between students labeled most at risk, some risk, and low risk (as based on their third grade DIBELS assessment); and
   b. how do the intervening variables of gender and socioeconomic status effect these differences?

**Population and Demographic Findings**

The study included students enrolled in 12 schools across seven school districts participating in a Regional Data Initiative (RDI) grant program. The students were administered the DIBELS Oral Reading Fluency assessment (DIBELS ORF) at the
beginning of third grade, as well as the MEAP reading assessment (MEAP ELA Reading) in third, fourth, and fifth grades. A total of \(N=2,220\) student records were included for study.

Table 2 shows the frequencies and percentages for the third grade DIBELS ORF variables as well as MEAP frequencies and percentages for third, fourth and fifth grades. The school districts and schools themselves have been de-identified for privacy purposes but the schools within each district are shown underneath that district. The largest district represented was coded as District B (32.3\%) with five schools in it. The largest school represented was school E (13.4\%). The children were evenly divided between males (41.5\%) and females (40.7\%), with the gender of the remaining children not identified. For SES, of the majority of children who had an SES level identified (27.1\% were missing SES), 66.7\% were identified as not being considered low SES. For the DIBELS ORF score category in third grade, the largest category was most at risk, representing 44.0\% of the children. The MEAP score categories were of interest at each grade levels. The largest category at each grade level was proficient (third grade was 42.7\%, fourth grade at 44.6\%, and fifth grade at 40.1\%). Nearly one-third of children were missing a MEAP score in each grade.
Table 2  

Demographic and Assessment Data (n=2,200)  

<table>
<thead>
<tr>
<th>District &amp; School</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>District 1</td>
<td>115</td>
<td>4.5</td>
</tr>
<tr>
<td>School A</td>
<td>115</td>
<td>4.5</td>
</tr>
<tr>
<td>District 2</td>
<td>822</td>
<td>32.3</td>
</tr>
<tr>
<td>School B</td>
<td>172</td>
<td>6.8</td>
</tr>
<tr>
<td>School C</td>
<td>188</td>
<td>7.4</td>
</tr>
<tr>
<td>School D</td>
<td>195</td>
<td>7.7</td>
</tr>
<tr>
<td>School F</td>
<td>163</td>
<td>6.4</td>
</tr>
<tr>
<td>School K</td>
<td>104</td>
<td>4.1</td>
</tr>
<tr>
<td>District 3</td>
<td>342</td>
<td>13.4</td>
</tr>
<tr>
<td>School E</td>
<td>342</td>
<td>13.4</td>
</tr>
<tr>
<td>District 4</td>
<td>244</td>
<td>9.6</td>
</tr>
<tr>
<td>School G</td>
<td>53</td>
<td>2.1</td>
</tr>
<tr>
<td>School L</td>
<td>239</td>
<td>9.4</td>
</tr>
<tr>
<td>District 5</td>
<td>130</td>
<td>5.1</td>
</tr>
<tr>
<td>School H</td>
<td>244</td>
<td>9.6</td>
</tr>
<tr>
<td>District 6</td>
<td>275</td>
<td>10.8</td>
</tr>
<tr>
<td>School I</td>
<td>130</td>
<td>5.1</td>
</tr>
<tr>
<td>District 7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School J</td>
<td>275</td>
<td>10.8</td>
</tr>
<tr>
<td>Missing</td>
<td>323</td>
<td>12.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>1036</td>
<td>40.7</td>
</tr>
<tr>
<td>Male</td>
<td>1055</td>
<td>41.5</td>
</tr>
<tr>
<td>Missing</td>
<td>452</td>
<td>17.8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SES</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low SES</td>
<td>159</td>
<td>6.3</td>
</tr>
<tr>
<td>Not Low SES</td>
<td>1695</td>
<td>66.7</td>
</tr>
<tr>
<td>Missing</td>
<td>689</td>
<td>27.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DIBELS ORF Score Category</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Most At Risk</td>
<td>1120</td>
<td>44.0</td>
</tr>
<tr>
<td>Risk Level</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>--------------------</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>Some Risk</td>
<td>692</td>
<td>27.2</td>
</tr>
<tr>
<td>Low Risk</td>
<td>366</td>
<td>14.4</td>
</tr>
<tr>
<td>Missing</td>
<td>365</td>
<td>14.4</td>
</tr>
</tbody>
</table>

**MEAP Score Category (Grade 3)**

<table>
<thead>
<tr>
<th>Score Level</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 4 - Not Proficient</td>
<td>145</td>
<td>5.7</td>
</tr>
<tr>
<td>Level 3 - Partially Proficient</td>
<td>404</td>
<td>15.9</td>
</tr>
<tr>
<td>Level 2 - Proficient</td>
<td>1087</td>
<td>42.7</td>
</tr>
<tr>
<td>Level 1 - Advanced</td>
<td>198</td>
<td>7.8</td>
</tr>
<tr>
<td>Missing</td>
<td>709</td>
<td>27.9</td>
</tr>
</tbody>
</table>

**MEAP Score Category (Grade 4)**

<table>
<thead>
<tr>
<th>Score Level</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 4 - Not Proficient</td>
<td>106</td>
<td>4.2</td>
</tr>
<tr>
<td>Level 3 - Partially Proficient</td>
<td>404</td>
<td>15.9</td>
</tr>
<tr>
<td>Level 2 - Proficient</td>
<td>1135</td>
<td>44.6</td>
</tr>
<tr>
<td>Level 1 - Advanced</td>
<td>146</td>
<td>5.7</td>
</tr>
<tr>
<td>Missing</td>
<td>752</td>
<td>29.6</td>
</tr>
</tbody>
</table>

**MEAP Score Category (Grade 5)**

<table>
<thead>
<tr>
<th>Score Level</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 4 - Not Proficient</td>
<td>184</td>
<td>7.2</td>
</tr>
<tr>
<td>Level 3 - Partially Proficient</td>
<td>352</td>
<td>13.8</td>
</tr>
<tr>
<td>Level 2 - Proficient</td>
<td>1021</td>
<td>40.1</td>
</tr>
<tr>
<td>Level 1 - Advanced</td>
<td>168</td>
<td>6.6</td>
</tr>
<tr>
<td>Missing</td>
<td>818</td>
<td>32.2</td>
</tr>
</tbody>
</table>

*Note.* Frequencies and percentages are shown for 3rd grade data except where specified (MEAP score categories).

Table 3 presents the frequencies and percentages for each district and school by SES level. The results suggest that the highest proportion of low SES students were in School A (in District 1). However, these results should be interpreted with caution due to the high percentage of missing SES in the other schools, which range from 14.5% to 35.8% of students in each school. Table 4 shows the proportion of males and females in
each district and school. The proportions of reported gender were fairly evenly
distributed for each school.

Table 3

SES by District and by School

<table>
<thead>
<tr>
<th>District</th>
<th>Low SES</th>
<th>Not Low SES</th>
<th>Missing SES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>District 1</td>
<td>17</td>
<td>14.8</td>
<td>96</td>
</tr>
<tr>
<td>School A</td>
<td>17</td>
<td>14.8</td>
<td>96</td>
</tr>
<tr>
<td>District 2</td>
<td>55</td>
<td>6.7</td>
<td>611</td>
</tr>
<tr>
<td>School B</td>
<td>15</td>
<td>8.7</td>
<td>132</td>
</tr>
<tr>
<td>School C</td>
<td>11</td>
<td>5.9</td>
<td>135</td>
</tr>
<tr>
<td>School D</td>
<td>15</td>
<td>7.7</td>
<td>145</td>
</tr>
<tr>
<td>School F</td>
<td>11</td>
<td>6.7</td>
<td>122</td>
</tr>
<tr>
<td>School K</td>
<td>3</td>
<td>2.9</td>
<td>77</td>
</tr>
<tr>
<td>District 3</td>
<td>28</td>
<td>8.2</td>
<td>246</td>
</tr>
<tr>
<td>School E</td>
<td>28</td>
<td>8.2</td>
<td>246</td>
</tr>
<tr>
<td>District 4</td>
<td>15</td>
<td>5.1</td>
<td>205</td>
</tr>
<tr>
<td>School G</td>
<td>3</td>
<td>5.7</td>
<td>31</td>
</tr>
<tr>
<td>School L</td>
<td>12</td>
<td>5.0</td>
<td>174</td>
</tr>
<tr>
<td>District 5</td>
<td>17</td>
<td>7.0</td>
<td>174</td>
</tr>
<tr>
<td>School H</td>
<td>17</td>
<td>7.0</td>
<td>174</td>
</tr>
<tr>
<td>District 6</td>
<td>5</td>
<td>3.8</td>
<td>85</td>
</tr>
<tr>
<td>School I</td>
<td>5</td>
<td>3.8</td>
<td>85</td>
</tr>
<tr>
<td>District 7</td>
<td>13</td>
<td>4.7</td>
<td>210</td>
</tr>
<tr>
<td>School J</td>
<td>13</td>
<td>4.7</td>
<td>210</td>
</tr>
<tr>
<td>Missing</td>
<td>9</td>
<td>2.8</td>
<td>68</td>
</tr>
</tbody>
</table>
Table 4

**Gender by District and by School**

<table>
<thead>
<tr>
<th>District</th>
<th>Female n</th>
<th>Female %</th>
<th>Male n</th>
<th>Male %</th>
<th>Missing n</th>
<th>Missing %</th>
</tr>
</thead>
<tbody>
<tr>
<td>District 1</td>
<td>58</td>
<td>50.4</td>
<td>57</td>
<td>49.6</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>School A</td>
<td>58</td>
<td>50.4</td>
<td>57</td>
<td>49.6</td>
<td>0</td>
<td>0.0</td>
</tr>
<tr>
<td>District 2</td>
<td>352</td>
<td>42.8</td>
<td>320</td>
<td>38.9</td>
<td>150</td>
<td>18.2</td>
</tr>
<tr>
<td>School B</td>
<td>76</td>
<td>44.2</td>
<td>71</td>
<td>41.3</td>
<td>25</td>
<td>14.5</td>
</tr>
<tr>
<td>School C</td>
<td>77</td>
<td>41.0</td>
<td>70</td>
<td>37.2</td>
<td>41</td>
<td>21.8</td>
</tr>
<tr>
<td>School D</td>
<td>86</td>
<td>44.1</td>
<td>77</td>
<td>39.5</td>
<td>32</td>
<td>16.4</td>
</tr>
<tr>
<td>School F</td>
<td>73</td>
<td>44.8</td>
<td>61</td>
<td>37.4</td>
<td>29</td>
<td>17.8</td>
</tr>
<tr>
<td>School K</td>
<td>40</td>
<td>38.5</td>
<td>41</td>
<td>39.4</td>
<td>23</td>
<td>22.1</td>
</tr>
<tr>
<td>District 3</td>
<td>129</td>
<td>37.7</td>
<td>158</td>
<td>46.2</td>
<td>55</td>
<td>16.1</td>
</tr>
<tr>
<td>School E</td>
<td>129</td>
<td>37.7</td>
<td>158</td>
<td>46.2</td>
<td>55</td>
<td>16.1</td>
</tr>
<tr>
<td>District 4</td>
<td>109</td>
<td>37.3</td>
<td>117</td>
<td>40.1</td>
<td>66</td>
<td>22.6</td>
</tr>
<tr>
<td>School G</td>
<td>17</td>
<td>32.1</td>
<td>19</td>
<td>35.8</td>
<td>17</td>
<td>32.1</td>
</tr>
<tr>
<td>School L</td>
<td>92</td>
<td>38.5</td>
<td>98</td>
<td>41.0</td>
<td>49</td>
<td>20.5</td>
</tr>
<tr>
<td>District 5</td>
<td>105</td>
<td>43.0</td>
<td>90</td>
<td>36.9</td>
<td>49</td>
<td>20.1</td>
</tr>
<tr>
<td>School H</td>
<td>105</td>
<td>43.0</td>
<td>90</td>
<td>36.9</td>
<td>49</td>
<td>20.1</td>
</tr>
<tr>
<td>District 6</td>
<td>47</td>
<td>36.2</td>
<td>48</td>
<td>36.9</td>
<td>35</td>
<td>26.9</td>
</tr>
<tr>
<td>School I</td>
<td>47</td>
<td>36.2</td>
<td>48</td>
<td>36.9</td>
<td>35</td>
<td>26.9</td>
</tr>
<tr>
<td>District 7</td>
<td>106</td>
<td>38.5</td>
<td>122</td>
<td>44.4</td>
<td>47</td>
<td>17.1</td>
</tr>
<tr>
<td>School J</td>
<td>106</td>
<td>38.5</td>
<td>122</td>
<td>44.4</td>
<td>47</td>
<td>17.1</td>
</tr>
<tr>
<td>Missing</td>
<td>130</td>
<td>40.2</td>
<td>143</td>
<td>44.3</td>
<td>50</td>
<td>15.5</td>
</tr>
</tbody>
</table>

**Operationalization of Inferential Analysis Variables**

The variables collected for the study were operationalized according to the criteria as summarized in this section.
Scaled scores of the *MEAP ELA Reading* assessments for the students from third to fifth grade were used in this study. The range of scale scores changed for each year of assessment, however the scores for fourth and fifth grade were normalized to the third grade level scale of scores, with a score range from 0 – 420. The scaled score was continuous and was included as the dependent variable in the HLM models.

A student was classified as having low *socio-economic status (SES)* if they participated in the free or reduced price lunch program in third grade. The variable was dichotomous and was coded such that students classified as “low SES” were used as the comparison group. “Not low” SES was therefore the reference category for the correlational, HLM and growth curve analyses. SES was included as a Level 2 variable in the HLM.

*Gender* was dichotomous and males were coded as the reference group for the correlation and HLM. Results from the HLM were for female students in reference to the male students. Gender was included as a Level 2 variable in the growth curve model.

The *DIBELS ORF* variable consisted of the end of year (spring) benchmark classifications of the DIBELS ORF assessment, and coded as (a) Most At Risk (score of 0 – 79) = 1, (b) Some Risk (score of 80-109) = 2, and (c) Low Risk (score of 110 and above) = 3. Higher ranking on the DIBELS ORF variable was associated with lower risk. The DIBELS benchmark classifications were collected for each student for grades three, four, and five. However, only the third grade DIBELS ORF ranking for each student was included in the HLM as a Level 2 variable. The DIBELS ORF was an ordinal variable. However, it was re-coded into two dummy categories of “some risk” and “at risk” for use in the HLM with the reference category of “low risk.”
The school site variable was nominal and was recorded for each student’s third grade year. Each school site was given a numeric code from 1 to 12 and the schools were grouped inside a variable representing District level. Although the school districts were given numeric codes, they were analyzed as nominal level variables. The schools included in District 1 were the reference category for the HLM, and all inferences for school level variables were made in reference to the schools in District with information collected from the students over the three grade levels (third, fourth, and fifth). Many designs using ANOVA or OLS regression require assumptions of (a) no missing data, (b) equal numbers of cases and time intervals, (c) independence of errors, and in the case of ANOVA (b) sphericity. The advantages of the multilevel modeling over repeated measures ANOVA and OLS regression are that the four assumptions are not required. However, the assumption of normality for the dependent variable is required, as well as the assumption of absence of multicollinearity between the independent variables.

Multilevel modeling can be used with missing data, when missing data is treated as missing at random (Tabachnick & Fidell, 2007). For growth models such as the one modeled in this study, missing data and student mobility can be incorporated directly into the analyses using maximum likelihood (ML) estimation, which reduces parameter bias that would result from eliminating those cases (Peugh & Enders, 2004). Using the long format in SPSS, where each student’s data was recorded into three rows of data (one row for each grade level), data was preserved for analysis because this format allowed for students to have incomplete data on the dependent variable of MEAP ELA Reading, such that the students record was included in the model if a student had at least one MEAP ELA Reading score (Heck, 2012).
Multilevel modeling accommodates studies in which the number of cases nested within groups are not equal, and studies for which independence of errors is not required or assumed. In fact, multilevel models often violate the assumption of independence of errors due to the influence of the nesting (Tabachnick & Fidell, 2007). Sphericity (uncorrelated errors over the three grade levels of the study) is not a concern for multilevel modeling, because the grade level predictor was entered into the equation as a single time-related predictor in the unconditional (intercepts only) model, and was thus evaluated as a single df test, as a longitudinal growth curve with a linear relationship between time and the dependent variable of MEAP ELA Reading (Tabachnick & Fidell, 2007).

Univariate normality for the dependent variable of MEAP ELA Reading was investigated using SPSS. A visual check of histograms and Normal Q-Q plots for the students’ MEAP ELA Reading scores indicated a distribution close to normal, with a few scores of zero contributing to a slight skew to the left. A score of zero is possible, so the records were retained for analysis and the assumption of univariate normality was assumed as met. Multivariate normality was investigated via a SPSS REGRESSION with the dependent variable of MEAP ELA Reading regressed onto grade level, to return a value for Mahalanobis distance. “Mahalanobis distance is the distance of a particular case from the centroid of the remaining cases, where the centroid is the point created by the means of all the variables” (Tabachnick & Fidell, 2007). None of the records had a Mahalanobis distance above the critical $\chi^2$ of 10.83 for 1 df at $\alpha = .001$. Therefore, the assumption of multivariate normality was met.
Multicollinearity between the independent variables was investigated with a series of bi-variate Spearman’s rank order correlation analyses. Multicollinearity may be assumed if a correlation coefficient between two variables is .90 or greater (Pallant, 2005). No violations were noted, and the assumption of absence of multicollinearity was met.

**Linear Regressions**

Research question one was addressed with three simple linear regression models predicting normed MEAP reading scores in third, fourth and fifth grades from the continuous score of DIBELS ORF 3rd grade (see Table 5).

Table 5

<table>
<thead>
<tr>
<th></th>
<th>Normed MEAP 3rd Grade</th>
<th>Normed MEAP 4th Grade</th>
<th>Normed MEAP 5th Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>DIBELS ORF 3rd Grade</td>
<td>.553 ***</td>
<td>.555 ***</td>
<td>.506 ***</td>
</tr>
<tr>
<td>$F$</td>
<td>771.76 ***</td>
<td>723.80 ***</td>
<td>511.63 ***</td>
</tr>
<tr>
<td>$(df)$</td>
<td>(1,1750)</td>
<td>(1,1622)</td>
<td>(1, 1484)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.306</td>
<td>.309</td>
<td>.256</td>
</tr>
</tbody>
</table>

*Note.* Coefficients are displayed as standardized $\beta$ weights, ***$p < .001$.

**Correlational Analyses**

In order to answer research question two, the relationship between DIBELS ORF and MEAP proficiency statuses fifth grade were tested using cross tabulations with Pearson’s chi square. The overall intercepts, or means, for the intercept and slope are provided in Table 6 along with the variances.
Table 6

5th Grade MEAP Score Categories by DIBELS ORF 3rd Grade Score Categories

<table>
<thead>
<tr>
<th>DIBELS ORF 3rd Grade Score Category</th>
<th>Most At Risk</th>
<th>Some Risk</th>
<th>Low Risk</th>
<th>( \chi^2 )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5th Grade MEAP Score Categories</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Level 4 – Not Proficient</td>
<td>126</td>
<td>17.2 a</td>
<td>22</td>
<td>4.5 b</td>
<td>3</td>
</tr>
<tr>
<td>Level 3 - Partially Proficient</td>
<td>226</td>
<td>30.8 a</td>
<td>66</td>
<td>13.4 b</td>
<td>16</td>
</tr>
<tr>
<td>Level 2 - Proficient</td>
<td>364</td>
<td>49.7 a</td>
<td>340</td>
<td>68.8 b</td>
<td>178</td>
</tr>
<tr>
<td>Level 1 - Advanced</td>
<td>17</td>
<td>2.3 a</td>
<td>66</td>
<td>13.4 b</td>
<td>62</td>
</tr>
</tbody>
</table>

Note. Proportions across rows with different superscripts differ significantly using Z-test of proportion comparisons, \( p < .05^* \). No superscripts are shown for proportions of non-significant \( \chi^2 \) values. Spearman’s Rho correlation coefficient computed with ordinal DIBELS ORF 3rd Grade and MEAP scores in 5th Grade revealed a significant positive correlation rho (1484) = .421, \( p < .001 \).

Multilevel Models

HLM was used in examining both research questions three and four. The HLM for this study had two levels. The first level measurement units included the repeated measure element of grade level, with three repeated measures of third, fourth, and fifth grade. The second level measurement units were the students (\( N = 2,220 \) students), who were nested into the third level units of school districts (\( N = 7 \) districts). Students were declared as random effects in the HLM in order to assess variability among the students.
within the schools. All predictors were included as fixed effects. The model was a main effects model. Table 7 presents the results of the 2-level HLM.

Table 7

*Hierarchical Linear Model Predicting Normed MEAP Reading Scores in 3rd–5th Grade From Gender, SES, School, DIBELS, and Interactions*

<table>
<thead>
<tr>
<th>Unstandardized Estimate</th>
<th>SE</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed Effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>.19</td>
<td>2.08</td>
<td>.09</td>
<td>1684</td>
</tr>
<tr>
<td>Not Low SES (vs. Low SES)</td>
<td>12.74</td>
<td>3.26</td>
<td>3.91</td>
<td>1826</td>
</tr>
<tr>
<td>School B (vs. School A)</td>
<td>5.10</td>
<td>1.89</td>
<td>2.69</td>
<td>2033</td>
</tr>
<tr>
<td>School C (vs. School A)</td>
<td>6.67</td>
<td>1.86</td>
<td>3.58</td>
<td>2136</td>
</tr>
<tr>
<td>School D (vs. School A)</td>
<td>2.58</td>
<td>1.83</td>
<td>1.41</td>
<td>2223</td>
</tr>
<tr>
<td>School E (vs. School A)</td>
<td>2.32</td>
<td>1.68</td>
<td>1.38</td>
<td>1864</td>
</tr>
<tr>
<td>School F (vs. School A)</td>
<td>.51</td>
<td>1.88</td>
<td>.27</td>
<td>2226</td>
</tr>
<tr>
<td>School G (vs. School A)</td>
<td>-1.89</td>
<td>3.02</td>
<td>-.62</td>
<td>1978</td>
</tr>
<tr>
<td>School H (vs. School A)</td>
<td>1.23</td>
<td>1.82</td>
<td>.68</td>
<td>1978</td>
</tr>
<tr>
<td>School I (vs. School A)</td>
<td>3.54</td>
<td>2.10</td>
<td>1.69</td>
<td>1916</td>
</tr>
<tr>
<td>School J (vs. School A)</td>
<td>-2.10</td>
<td>1.75</td>
<td>-1.20</td>
<td>2037</td>
</tr>
<tr>
<td>School K (vs. School A)</td>
<td>-.47</td>
<td>2.14</td>
<td>-.22</td>
<td>2703</td>
</tr>
<tr>
<td>School L (vs. School A)</td>
<td>-2.24</td>
<td>1.79</td>
<td>-1.25</td>
<td>1933</td>
</tr>
<tr>
<td>Grade</td>
<td>14.32</td>
<td>.80</td>
<td>17.87</td>
<td>2615</td>
</tr>
<tr>
<td>DIBELS 3rd Grade</td>
<td>.62</td>
<td>.06</td>
<td>10.67</td>
<td>3343</td>
</tr>
<tr>
<td>Male x DIBELS 3rd Grade (interaction)</td>
<td>-.01</td>
<td>.02</td>
<td>-.53</td>
<td>1685</td>
</tr>
<tr>
<td>Not Low SES x DIBELS 3rd Grade (interaction)</td>
<td>-.05</td>
<td>.05</td>
<td>-.97</td>
<td>1803</td>
</tr>
<tr>
<td>Grade x DIBELS 3rd Grade (interaction)</td>
<td>-.05</td>
<td>.01</td>
<td>-5.73</td>
<td>2623</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>244.51</td>
<td>4.57</td>
<td>53.50</td>
<td>3760</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>IDs (N = 1,755)</td>
<td>159.9</td>
</tr>
<tr>
<td>Residual</td>
<td>167</td>
</tr>
</tbody>
</table>

*Note.* Chi Square testing on the -2 log likelihoods revealed that the model specified was significantly better fitting than the models without the interactions (p < .001)* and was significantly better fitting than the model without the covariates (p < .001). The MEAP scores given are based on school attended in 3rd grade.
This information is provided for the purpose of replicability but do not directly address the research questions. The intercepts, also referred to as means, of the intercept and slope are significant indicating that mean normed MEAP score in third grade was significantly higher than 0 ($Z = 124.10, p < .001$) and the mean slope was significantly higher than 0 ($Z = 12.36, p < .001$). In other words, overall, MEAP scores increased between third and fifth grade.

Table 8 looks at the growth curve predictions for the MEAP Reading scores in third, fourth and fifth grades while controlling for covariants of DIBELS ORF score categories, SES and gender. This table helps us look at the slope or growth rates for students scoring most at risk, some risk or low risk on the DIBELS ORF assessment. Overall, third grade students who are identified as most at risk on their DIBELS ORF assessment have the fastest rate of growth on their third, fourth and fifth grade MEAP Readings assessments.
Table 8

Growth Curve Predicting Normed MEAP Scores (3rd−5th Grade) From DIBELS and Covariates

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>Standardized Estimate</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intercept</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIBELS ORF: Some Risk (vs. High Risk)</td>
<td>17.37</td>
<td>1.06</td>
<td>.40</td>
<td>16.44</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>DIBELS ORF: Low Risk (vs. High Risk)</td>
<td>28.33</td>
<td>1.30</td>
<td>.54</td>
<td>21.86</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male (vs. Female)</td>
<td>-1.23</td>
<td>.91</td>
<td>-.03</td>
<td>-1.36</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Not Low SES (vs. Low SES)</td>
<td>16.30</td>
<td>1.75</td>
<td>.22</td>
<td>9.30</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>School B (vs. School A)</td>
<td>7.45</td>
<td>2.50</td>
<td>.10</td>
<td>2.99</td>
<td>.003</td>
</tr>
<tr>
<td>School C (vs. School A)</td>
<td>10.02</td>
<td>2.47</td>
<td>.14</td>
<td>4.06</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>School D (vs. School A)</td>
<td>5.27</td>
<td>2.42</td>
<td>.08</td>
<td>2.18</td>
<td>.030</td>
</tr>
<tr>
<td>School E (vs. School A)</td>
<td>3.34</td>
<td>2.21</td>
<td>.06</td>
<td>1.51</td>
<td>.130</td>
</tr>
<tr>
<td>School F (vs. School A)</td>
<td>-0.03</td>
<td>2.54</td>
<td>.00</td>
<td>-0.01</td>
<td>.990</td>
</tr>
<tr>
<td>School G (vs. School A)</td>
<td>-0.46</td>
<td>3.86</td>
<td>.00</td>
<td>-1.2</td>
<td>.905</td>
</tr>
<tr>
<td>School H (vs. School A)</td>
<td>2.85</td>
<td>2.34</td>
<td>.04</td>
<td>1.22</td>
<td>.223</td>
</tr>
<tr>
<td>School I (vs. School A)</td>
<td>3.20</td>
<td>2.76</td>
<td>.04</td>
<td>1.16</td>
<td>.248</td>
</tr>
<tr>
<td>School J (vs. School A)</td>
<td>-0.45</td>
<td>2.30</td>
<td>-.01</td>
<td>-0.2</td>
<td>.844</td>
</tr>
<tr>
<td>School K (vs. School A)</td>
<td>0.68</td>
<td>2.89</td>
<td>.01</td>
<td>0.24</td>
<td>.814</td>
</tr>
<tr>
<td>School L (vs. School A)</td>
<td>-3.33</td>
<td>2.36</td>
<td>-.05</td>
<td>-1.41</td>
<td>.158</td>
</tr>
<tr>
<td><strong>Slope</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIBELS ORF: Some Risk (vs. High Risk)</td>
<td>-2.57</td>
<td>.51</td>
<td>-.23</td>
<td>-5.09</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>DIBELS ORF: Low Risk (vs. High Risk)</td>
<td>-3.88</td>
<td>.63</td>
<td>-.29</td>
<td>-6.16</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male (vs. Female)</td>
<td>-.09</td>
<td>.44</td>
<td>-.01</td>
<td>-.21</td>
<td>.831</td>
</tr>
<tr>
<td>Not Low SES (vs. Low SES)</td>
<td>-3.23</td>
<td>.87</td>
<td>-.17</td>
<td>-3.71</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>School B (vs. School A)</td>
<td>-1.00</td>
<td>1.19</td>
<td>-.05</td>
<td>-.84</td>
<td>.402</td>
</tr>
<tr>
<td>School C (vs. School A)</td>
<td>-3.16</td>
<td>1.19</td>
<td>-.17</td>
<td>-2.66</td>
<td>.008</td>
</tr>
<tr>
<td>School D (vs. School A)</td>
<td>-1.93</td>
<td>1.17</td>
<td>-.11</td>
<td>-1.66</td>
<td>.098</td>
</tr>
<tr>
<td>School E (vs. School A)</td>
<td>-1.40</td>
<td>1.06</td>
<td>-.10</td>
<td>-1.33</td>
<td>.185</td>
</tr>
<tr>
<td>School F (vs. School A)</td>
<td>.58</td>
<td>1.21</td>
<td>.03</td>
<td>.48</td>
<td>.631</td>
</tr>
<tr>
<td>School G (vs. School A)</td>
<td>2.12</td>
<td>2.12</td>
<td>.06</td>
<td>1.00</td>
<td>.316</td>
</tr>
<tr>
<td>School H (vs. School A)</td>
<td>-.49</td>
<td>1.12</td>
<td>-.03</td>
<td>-.44</td>
<td>.659</td>
</tr>
<tr>
<td>School I (vs. School A)</td>
<td>-.74</td>
<td>1.33</td>
<td>-.03</td>
<td>-.55</td>
<td>.581</td>
</tr>
<tr>
<td>School J (vs. School A)</td>
<td>.64</td>
<td>1.11</td>
<td>.04</td>
<td>.58</td>
<td>.565</td>
</tr>
<tr>
<td>School K (vs. School A)</td>
<td>.07</td>
<td>1.39</td>
<td>.00</td>
<td>.05</td>
<td>.958</td>
</tr>
<tr>
<td>School L (vs. School A)</td>
<td>1.91</td>
<td>1.19</td>
<td>.11</td>
<td>1.61</td>
<td>.107</td>
</tr>
</tbody>
</table>
Table 8 (cont’d)

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>SE</th>
<th>Standardized Estimate</th>
<th>Z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Covariances</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept With Slope</td>
<td>-34.58</td>
<td>8.43</td>
<td>-.49</td>
<td>-4.10</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>307.07</td>
<td>2.47</td>
<td></td>
<td>124.10</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Slope</td>
<td>15.08</td>
<td>1.22</td>
<td></td>
<td>12.36</td>
<td>&lt; .001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Variances</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normed MEAP 3rd Grade</td>
<td>175.62</td>
<td>15.99</td>
<td>175.62</td>
<td>.300</td>
<td></td>
</tr>
<tr>
<td>Normed MEAP 4th Grade</td>
<td>167.14</td>
<td>8.04</td>
<td>167.14</td>
<td>.349</td>
<td></td>
</tr>
<tr>
<td>Normed MEAP 5th Grade</td>
<td>67.32</td>
<td>12.33</td>
<td>67.32</td>
<td>.202</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>232.83</td>
<td>16.81</td>
<td>.57</td>
<td>.568</td>
<td></td>
</tr>
<tr>
<td>Slope</td>
<td>21.60</td>
<td>6.86</td>
<td>.80</td>
<td>.801</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The intercepts for the normed MEAP scores for 3rd–5th Grade were set to 0. The MEAP scores given are based on school attended in 3rd grade.

**Chapter IV Summary**

This chapter presented the results of the statistical analysis detailed in Chapter III and laid out in the research questions. In summary, the preliminary analyses were conducted in order to better understand the sample. There were significant relationship among the demographic characteristics and between the independent, dependent, and dependent variables. These relationships were accounted for in the primary analyses by including the covariates in the models where appropriate. The interpretation, implications, and ideas for future research are discussed in the discussion Chapter V.
CHAPTER V
DISCUSSION

The purpose of this study was to investigate the relationship between how well students score on a formative assessment tool (DIBELS ORF assessment) commonly used in early grades to determine interventions assigned to students, and how well those students score on the MEAP test at the end of third, fourth and fifth grade. The format of this chapter includes the following: (a) key findings form the HLM, (b) implications of the study’s findings for school leaders, and (c) recommendations for future research.

Key Findings

A total of four overarching research questions guided this study and the results of the HLM were used to address the questions. Each of the four research questions are addressed individually. A linear regression was used to address research question one (see Table 5, Chapter IV). Research Question two was addressed via Spearman’s rank order correlation (see Table 6, Chapter IV). The HLM specifications from Table 8 (Chapter IV) were used to address Research Questions three and four.

Conclusions as Related to Research Question 1

Research question one examined the extent to which there is a relationship between students’ DIBELS Oral Reading Fluency at the beginning of third grade and their MEAP Reading scores in third, fourth and fifth grades. This was addressed with three linear regressions looking at normed MEAP reading scores in third, fourth, and fifth grades and DIBELS ORF (continuous version of the score) in third grade. The results revealed that the model predicting normed MEAP 3rd grade scores from DIBELS ORF 3rd grade scores ($\beta = .553, p < .001$) was significant, $F (1, 1750) = 771.76, p < .001$, and
accounted for 30.6% of the variance. DIBELS ORF 3rd grade scores ($\beta = .555, p < .001$) also significantly predicted normed MEAP 4th grade scores, $F(1, 1622) = 723.80, p < .001$, and accounted for 30.9% of the total variance. Finally, the model predicting normed MEAP scores in 5th grade from DIBELS ORF 3rd grade scores ($\beta = .506, p < .001$) was significant, $F(1, 1484) = 511.63, p < .001$, and accounted for 25.6% of the variance in the model.

**Conclusions as Related to Research Question 2**

Research question two examined the extent to which there is a relationship between students’ MEAP Reading proficiency statuses in fifth grade, and their DIBELS’ label (i.e., most at risk, some risk, low risk) at the beginning of third grade. Question two was addressed by conducting a crosstabulation with Pearson’s chi square test (see Table 12, Chapter 4). Results revealed a significant relationship between DIBELS ORF score category in third grade and the fifth grade MEAP score categories, ($\chi^2 (6) = 272.05, p < .001$, Cramer’s $V = .303$). Posthoc $Z$-test of proportion tests revealed that a significantly greater proportion of most at risk students according to the DIBELS ORF were not proficient on the MEAP in fifth grade (17.2%) compared to the proportions who were at some risk (4.5%) or low risk (1.2%) on the DIBELS ORF. In addition, a significantly greater proportion of those at some risk on the DIBELS ORF (4.5%) were not proficient on the MEAP compared to those at low risk on the DIBELS ORF (1.2%). A significantly greater proportion of most at risk students according to the DIBELS ORF were partially proficient on the MEAP in fifth grade (30.8%) compared to the proportions were at some risk (13.4%) or low risk (6.2%) on the DIBELS ORF. In addition, a significantly greater proportion of those at some risk on the DIBELS ORF (13.4%) were
partially proficient on the MEAP compared to those at low risk on the DIBELS ORF (6.2%). A significantly greater proportion of some risk (68.8%) and low risk (68.7%) students according to the DIBELS ORF were proficient on the MEAP in fifth grade compared to the proportions were at at-risk (49.7%) on the DIBELS ORF. A significantly greater proportion of low risk students according to the DIBELS ORF were advanced on the MEAP in fifth grade (23.9%) compared to the proportions were at some risk (13.4%) or at-risk (2.3%) on the DIBELS ORF. In addition, a significantly greater proportion of those at some risk on the DIBELS ORF (13.4%) were advanced on the MEAP compared to those at at-risk on the DIBELS ORF (2.3%).

**Conclusions as Related to Research Question 3**

Research questions three and four were addressed with a HLM growth curve model predicting normed MEAP scores in third through fifth grades from the DIBELS ORF and covariates. Research question three examined students’ third grade mean MEAP Reading scale scores, asking the (a) extent there are difference between student labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessment); and (b) how the intervening variables of gender and socioeconomic status effect these differences. This question was evaluated by examining the predictors of the intercept. Conceptually, this part of the growth curve model reveals how well the independent variables and covariates in the model predict the first time point, which are the third grade scores in the current model. As shown in Table 14 (Chapter 4), DIBELS ORF score categories, SES, and several of the schools were all significant predictors of the normed MEAP scores in third grade. Controlling for the covariates, students categorized as some risk according to the DIBELS ORF had significantly higher MEAP
scores compared to students categorized as high risk ($\beta = .40, p < .001$). In addition, students categorized as low risk had significantly higher MEAP scores compared to students categorized as high risk ($\beta = .54, p < .001$). For the covariates, students not categorized as low SES had significantly higher MEAP scores compared to students categorized as low SES ($\beta = .22, p < .001$). In addition, students in School B, School C, School D all had significantly higher third grade MEAP scores compared to students in School A ($\beta s = .08 - .14, ps < .05$).

**Conclusions as Related to Research Question 4**

The final research question examined students’ MEAP Reading score growth trajectories in grades three through five, asking the (a) extent there are differences between students labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessment); and (b) how the intervening variables of gender and socioeconomic status effect these differences. To address research question four, the predictors of the slope were examined. Conceptually, this part of the growth curve model reveals how well the independent variables and covariates in the model predict the change in the dependent variable, normed MEAP scores, over time. As shown in Table 14 (Chapter 4), the significant predictors of the slope were the DIBELS ORF score categories, SES, and School C compared to School A. Controlling for the covariates, MEAP scores of students at some risk according to the DIBELS ORF in third grade increased at a significantly slower rate over time compared to students at high risk ($\beta = -.23, p < .001$). In addition, MEAP scores of students at low risk in third grade increased at a significantly slower rate over time compared to students at high risk ($\beta = -.29, p < .001$). In other words, MEAP scores of student most at risk in third grade increased at a
significantly faster rate over time compared with students who were either at some risk or at low risk. The actual means are displayed in Figure 3.

Figure 3. Mean normed MEAP scores in each grade by each DIBELS ORF score.

Note that these are raw means which have not been adjusted for covariates and are shown to display the overall pattern of change over time by DIBELS ORF score category. For the covariates, MEAP scores of students who were not classified as low SES increased at a significantly slower rate over time compared to students with low SES ($\beta = -.17, p < .001$). Finally, MEAP scores of students at School C increased at a significantly slower rate over time compared to students at School A ($\beta = -.17, p = .008$). The results of the model also reveal a significant negative relationship between the slope and intercept ($\beta = -.49, p < .001$) indicating that students who had lower MEAP scores in third grade tended to see a faster increase in MEAP scores over time compared to students who started with higher MEAP scores in third grade.

Overall, this study had several key findings. Table 9 gives an overview of the results by research question.
### Key Findings / Results

<table>
<thead>
<tr>
<th>Research Question / Hypotheses</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. To what extent is there a relationship between students’ DIBELS Oral Reading Fluency at the beginning of third grade and their MEAP Reading scores in third, fourth and fifth grades?</td>
<td>DIBELS ORF scores accounted for:</td>
</tr>
<tr>
<td></td>
<td>• 30.6% of the variance in 3rd grade MEAP</td>
</tr>
<tr>
<td></td>
<td>• 30.9% of the variance in 4th grade MEAP</td>
</tr>
<tr>
<td></td>
<td>• 25.6% of the variance in 5th grade MEAP</td>
</tr>
<tr>
<td>2. To what extent is there a relationship between students’ MEAP Reading proficiency statuses in fifth grade, and their DIBELS’ label (i.e., most at risk, some risk, low risk) at the beginning of third grade?</td>
<td>Of students not proficient on the 5th grade MEAP:</td>
</tr>
<tr>
<td></td>
<td>• 17.2% were most at risk on 3rd grade DIBELS ORF</td>
</tr>
<tr>
<td></td>
<td>• 4.5% were some risk on 3rd grade DIBELS ORF</td>
</tr>
<tr>
<td></td>
<td>• 1.2% were low risk on 3rd grade DIBELS ORF</td>
</tr>
<tr>
<td></td>
<td>Of students partially proficient on the 5th grade MEAP:</td>
</tr>
<tr>
<td></td>
<td>• 30.8% were most at risk on 3rd grade DIBELS ORF</td>
</tr>
<tr>
<td></td>
<td>• 13.4% were some risk on 3rd grade DIBELS ORF</td>
</tr>
<tr>
<td></td>
<td>• 6.2% were low risk on 3rd grade DIBELS ORF</td>
</tr>
<tr>
<td>3. In reference to students’ third grade mean MEAP Reading scale scores: to what extent are there differences between students labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessment); and how do the intervening variables of gender and socioeconomic status effect these differences?</td>
<td>Controlling for covariates:</td>
</tr>
<tr>
<td></td>
<td>• students categorized as low risk on DIBELS ORF scored .54 higher on the MEAP Reading Assessment than students who were most at risk</td>
</tr>
<tr>
<td></td>
<td>• students categorized as some risk on DIBELS ORF scored .40 higher on the MEAP Reading Assessment than students who were most at risk</td>
</tr>
</tbody>
</table>
Table 9 (cont’d)

<table>
<thead>
<tr>
<th>Research Question/Hypotheses</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. In reference to students’ MEAP Reading score growth trajectories in grades three through five: to what extent are there differences between students labeled most at risk, some risk and low risk (as based on their third grade DIBELS assessment); and how do the intervening variables of gender and socioeconomic status effect these differences?</td>
<td>Students who had lower MEAP scores in 3rd grade tended to see a faster increase in MEAP scores over time compared to students who started with higher MEAP scores in 3rd grade.</td>
</tr>
<tr>
<td></td>
<td>Controlling for covariates:</td>
</tr>
<tr>
<td></td>
<td>• MEAP scores for students who score some risk on 3rd grade DIBELS ORF increased at a slower rate than the most at risk students (-.23)</td>
</tr>
<tr>
<td></td>
<td>• MEAP scores for students who scored low risk on 3rd grade DIBELS ORF increased at a slower rate than the most at risk students (-.29)</td>
</tr>
</tbody>
</table>

**Implications for School Leaders**

These results have implications for administrators and educators in positions to make decisions about how student performance data is collected and used to decide students who will receive reading interventions. Since the inception of NCLB (2002), schools have been under increasing pressure to ensure students are proficient readers. Schools continue to look for fast and efficient methods for identifying students at risk of performing poorly on end-of-year summative assessments. School leaders cannot afford to wait until a state level assessment, like the MEAP, is administered. It is necessary for schools to evaluate the instructional value of the assessments they are administering to students to be sure that they are reliable and accurate in measuring student skills.

A concern that some educators have expressed about using DIBELS ORF as the primary reading screener for students is that it is a measure of a student’s ability to decode words, while the MEAP Reading assessment measures a student’s ability to comprehend what they read. One of the purposes of completing this study was to analyze
the data and to share the findings with educators so that they did not overlook students in need of intervention because those students were good at the decoding process, but lacked the comprehension skills to be successful in upper grades. Yet, in this sample, 92.6% of students who were identified as low risk on the third grade DIBELS ORF assessment were proficient on the fifth grade MEAP Reading assessment. This is a key finding as it validates the ability of DIBELS ORF as a tool that can be used to successfully group students, and help educators predict which students are in fact at low risk for future reading problems.

Practitioners may also be interested in looking at the MEAP reading growth trajectories for MEAP reading scores in third, fourth and fifth grade based on the DIBELS ORF labels (most at risk, some risk and low risk). Students who had lower MEAP scores in third grade tended to see a faster increase in MEAP scores over time compared with students who started with high MEAP scores in third grade. In other words, MEAP scores of student most at risk in third grade increased at a significantly faster rate over time compared with students who were either at some risk or at low risk. The effect could be attributed to the interventions, or lack thereof, provided to students who scored some risk and low risk. While these increases are good, I have some concerns that the growth trajectory for students with low risk factors declines as they increase grade levels. This may indicate a problem with Tier I instruction in upper elementary, whereby educators need to enhance core instruction to ensure that all students reading skills increase at an acceptable rate.

My study revealed that an interim assessment, in this case DIBELS ORF, is able to help educators predict a student’s reading proficiency scores and therefore can help
inform decisions when educators are identifying students in need of interventions. A teacher can administer the DIBELS ORF assessment to a student in approximately five minutes. This method proves to be a quick and effective way to measure student’s reading ability and monitor their reading progress, leaving teachers with more time to implement appropriate interventions. The DIBELS ORF can provide schools with an assessment that is predictive of future performance, efficient, and informs instruction. It is important to note that this study used only the spring DIBELS ORF benchmark testing. While the DIBELS testing package offers both benchmark testing to be given three times a year, and progress monitoring to be administered by the classroom teacher as often as weekly, I feel schools need to look at a map of all assessments given and evaluate which are most effective at identifying the students in need of interventions. This study helps to affirm that the DIBELS ORF benchmark testing is an effective way to diagnose reading issues and to group students for interventions.

Findings of this study were consistent with studies conducted in other states (Cook, 2003; Good, Simmons, & Kame’enui, 2001; Seay, 2005; Shaw & Shaw, 2003), although all but one of these studies used a significantly small number of students and often did not track student scores beyond one year. School administrators can look at the results of this study and feel confident that DIBELS ORF is successful at identifying students in need of interventions. Overall, this study adds to the body of knowledge supporting the use of ORF assessments to identify struggling readers.
Recommendations for Future Study

This study clearly supports the use of interim assessments to identify student who are struggling with reading. This leads us to look further into how students develop as readers. It would be interesting to look at student scores within different intervention groups to identify the best methods for supporting increases in student’s reading skills. Of the students included in this study, only about half (52%) of student who scored in the most at risk range on the third grade DIBELS ORF were proficient on the MEAP Reading assessment at the end of fifth grade. It is important for teachers to understand how to help students reach optimal achievement level and what types of intervention would best support different types of learns. More information in needed to understand the interventions that are successful for students and which did not successfully increase students ability to read. This study also did not touch on the Tier I instructional method used in the classroom during whole group instruction and whether instructional methods influence how students reading progress. Ultimately, both of these pieces of information would be valuable in the classroom and help teacher make better decisions about effective instruction and intervention for students. Future research should explore the extent in which these practices impact student success in reading.

Conclusions

Since the early days of standards and accountability, schools and educators have looked for reliable ways to identify student reading progress. NCLB increased this pressure on schools by stipulating that schools collect student data and monitor learning with progress checks. Political proponents of NCLB set a goal of all children being proficient readers by the end of third grade. In order to meet these demands, educators
need to identify reliable assessment to track student progress and indicate appropriate interventions to students not progress at an appropriate pace.

My research has provided data to support the use of DIBELS ORF as a valid assessment to use in the early identification of student who are likely to struggle reaching proficiency on the end of year state assessment (i.e., MEAP). The results of my study showed that there is an indirect correlation, suggesting that at the third grade DIBELS risk levels increased, the MEAP proficiency status decreases. It is also important to note that while gender did not have an impact on predicted scores, SES did significantly impact scores, with students classified as “low SES” scoring substantially lower on end of year MEAP assessments. As educators continue to evaluate and monitor student progress, DIBELS ORF continues to be a valuable tool in their assessment arsenal.
REFERENCES


The Educational Trust (n. d.). *EdWatch Online.*


Reading Excellence Act, (1999). HR2614. Title VIII, ESEA.


University of Oregon. (2003). *Dynamic Indicators of Basic Early Literacy Skills (DIBELS)*. Retrieved from [http://dibels.uoregon.edu](http://dibels.uoregon.edu)

University of Oregon. *DIBELS data system*. Retrieved from [https://dibels.uoregon.edu/](https://dibels.uoregon.edu/)


Wong, K. K., & Meyer, S. The Mid-Atlantic regional Educational Laboratory, (1998). *Title I schoolwide programs: Federal legislative expectations*

Appendix A

Approval Letter from the Human Subjects Institutional Review Board
Date: February 13, 2014

To: Louann Bierlein Palmer, Principal Investigator
    Lisa Lockman, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: HSIRB Project Number 14-02-17

This letter will serve as confirmation that your research project titled "The Relationship of Dynamic Indicators of Basic Early Literacy Skills (DIBELS) Oral Reading Fluency and the Michigan Educational Assessment Program (MEAP)" has been approved under the exempt category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note: This research may only be conducted exactly in the form it was approved. You must seek specific board approval for any changes in this project (e.g., you must request a post approval change to enroll subjects beyond the number stated in your application under "Number of subjects you want to complete the study"). Failure to obtain approval for changes will result in a protocol deviation. In addition, if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

Reapproval of the project is required if it extends beyond the termination date stated below.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: February 13, 2015