2011

The Predictive Value of the Maze and Oral Reading Fluency Measures for Students with and without EBD

Jason W. Kurtz

THE PREDICTIVE VALUE OF THE MAZE AND ORAL READING FLUENCY MEASURES FOR STUDENTS WITH AND WITHOUT EBD

DISSERTATION

Presented in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy in

Leadership and Education in

the Adrian Dominican School of Education of

Barry University

by

Jason W. Kurtz, B.A., M.Ed.

*****

BARRY UNIVERSITY

2011

Area of Specialization: Exceptional Student Education

THE PREDICTIVE VALUE OF THE MAZE AND ORAL READING FLUENCY MEASURES FOR STUDENTS WITH AND WITHOUT EBD

DISSERTATION

by

Jason W. Kurtz

2011

APPROVED BY:

_____
Catherine Roberts, Ph.D.
Chairperson, Dissertation Committee

_____
Judy Harris-Looby, Ph.D.
Member, Dissertation Committee

_____
Lauran Sandals, Ph.D.
Member, Dissertation Committee

_____
Terry Piper, Ph.D.
Dean, Adrian Dominican School of Education

ABSTRACT

THE PREDICTIVE VALUE OF THE MAZE AND ORAL READING FLUENCY MEASURES FOR STUDENTS WITH AND WITHOUT EBD

Jason W. Kurtz

Barry University, 2011

Dissertation Chairperson: Dr. Catherine Roberts

Purpose

The purpose of this study was to evaluate and compare the effectiveness of curriculum based measures using the oral reading fluency measure (ORF) of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) and the maze measures of the Florida Assessments for Instruction in Reading (FAIR) to predict the success of third and fourth grade students both with and without emotional and behavioral disorders (EBD) on the Florida Comprehensive Assessment Test (FCAT). Knowledge of valid and early predictors of student performance on high stakes assessments is critical in the identification and instructional planning of students who require early intervention in order to support their academic success.

Method

Archival assessment data across a two year period were collected from a convenience sample of students with EBD (n = 55) and a comparison group of students without EBD (n = 55) matched across ethnicity, socio-economic status, gender and age in order to control for differential selection. All participants (N =110) were enrolled in several suburban public

elementary schools in southwest Florida. *t*-values were used to examine mean differences between both student groups. Additionally, correlational and multiple regression analyses were used to examine the relationship between the ORF and maze measures with the developmental scale score in reading on the FCAT and with specific demographic variables to determine which assessment was a better predictor of FCAT success for students with and without EBD.

Major Findings

The results of the study indicated that a significant difference in scores existed for all assessments for students with EBD when compared to their peers without EBD (with *t* values ranging from -3.110, $p < .05$ to -5.565, $p < .001$). Correlations of ORF and maze measures with the FCAT were significant for both populations, suggesting a high level of predictability of student performance for both assessments (*r* values ranging from .447 to .695, $p < .001$). The results of multiple linear regression analyses demonstrated both ORF and maze measures can be used with some accuracy to predict success on the FCAT for students with and without EBD while the demographic variables of ethnicity, socio-economic status, and gender did not significantly contribute to the model. Overall, the ORF measures appeared to be a stronger and more stable predictor of success for students with EBD.

ACKNOWLEDGEMENTS

"Who reads much and walks much sees much and knows much." -Miguel de Cervantes

This experience has been as much about the journey as the learning. While certainly the knowledge gained about research, statistics, and writing is important; what I have learned about myself is much more significant. However, none of it would have been possible without the support and love of my family, friends, and colleagues; for the dissertation journey is not experienced by the student alone, but by all those around him that provide support, love, and expertise. And while I am sure to miss someone, please know that I am indebted to all of you.

First and foremost, I must thank my loving wife, Jessica Hope. Her love and support for everything I do is invaluable, without which this journey would have been difficult if not impossible. I am especially grateful for the many days she spent at the park with our children so that I had a quiet house in which to work; the many times our personal lives had to be put on hold so I could "work on my dissertation;" and the countless times she encouraged me with a listening ear or gentle reminder to stay the course.

I would also like to thank my parents, Gary and Kathleen Kurtz who taught me that life is about seeing and doing; and what you make of it is only limited by the restrictions you place on yourself. For the endless support, guidance, and sacrifices that they have provided throughout my educational journey, I thank them. This dissertation and what it represents is as much a testimony to them as it is to me.

Thank you to my committee members- Dr. Catherine Roberts, Chair, Dr. Judy Harris-Looby, and Dr. Lauran Sandals- for their guidance, direction, support, and expertise throughout the writing of this dissertation. Without them this would simply be an idea on paper.

Lastly, thank you to all of my family, friends, and colleagues who have supported, encouraged, listened, and inspired me along the way. You are as much a part of this journey as I am.

DEDICATION

I dedicate this dissertation to my children, Carter William, Owen Whitaker, and Gwyneth Hope… may your journeys in life be only limited by your dreams and your lives as blessed as mine.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

Statement of the Problem

In 2001, the No Child Left Behind Act (NCLB) was passed as the United States Congress amended the Elementary and Secondary Education Act (ESEA) of 1965. This monumental legislation mandated that all students meet proficiency on challenging state academic achievement standards and state academic assessments in reading by the 2013-2014 school year (U.S. Department of Education, 2002b; Wright, Wright, & Heath, 2004). This was important because, for the first time, schools would be required to report progress towards these goals by subgroup. This included students with disabilities (U.S. Department of Education, 2002a). States and districts now had reason to not only ensure the participation in assessments of students with disabilities, but also to dramatically improve the educational results for these very same students.

Then in 2004 on December 3, President George W. Bush signed the reauthorized Individuals with Disabilities Education Act (IDEA). This piece of legislation, now called the Individuals with Disabilities Education Improvement Act (IDEIA) required that the states include the progress of children with disabilities in its evaluation of adequate yearly progress (AYP) (U.S. Department of Education, 2007). Additionally, (though not specifically referred to as Response to Intervention [RtI]) IDEIA mandates that schools use a tiered system of interventions to meet the needs of all students. Many understand RtI to be a problem-solving process in which assessment and interventions are integrated within a multi-level system in an effort to meet the needs of all students (Fuchs & Fuchs, 2006; 2009a; 2009b). The emphasis on grade level proficiency for all students and on a tiered approach of RtI for

meeting the needs of students highlights the need for reliable and accurate progress monitoring. Educators must now access information that can be used to predict how students with and without disabilities are achieving when compared with grade-level standards as they work to accelerate student progress (Quenemoen, Thurlow, Moen, Thompson, & Morse, 2004). Fuchs and Fuchs (1986; 1999) have long championed the importance of having an assessment process that incorporates ongoing data collection and its subsequent use in evaluating the effectiveness of instruction. Additionally, the President's Commission on Excellence in Special Education (U.S. Department of Education, 2002b) has emphasized the need for districts to implement a process of continued progress monitoring. As the use of progress monitoring has increased, so too has the search for an efficient and accurate means of assessing the performance of all students.

The assessment of the reading performance of students is essential in the elementary school setting. However, the assessment of the reading performance of students with emotional/behavioral disorders (EBD) is of particular importance, as most consistently perform below their peers in all academic areas (Epstein & Cullinan, 1983; Lane, Little, Redding-Rhodes, Phillips, & Welsh, 2007; Nelson, Benner, Lane, & Smith, 2004). In fact, in a six-state sample, Greenbaum and Dedrick (1996) found that 85% of students with EBD were below grade level in reading after seven years in school. This in itself is a very disturbing statistic. Even more disturbing is that students who experience academic failure, especially in reading, have been found to engage in higher rates of delinquency, violence, and substance abuse (Fleming, Harachi, Cortes, Abbott, & Catalano, 2004).

Other bodies of research indicate that antisocial behaviors, like aggression, and attention problems are inversely related to academic achievement in school (Barriga, Doran, Newell, Morrison, Barbetti, & Robbins, 2002). In the United States there are close to 500,000 students with EBD being served in federally funded programs and this number has continued to increase over the last 15 years (U.S. Department of Education, 2006). In the state of Florida, this equates to well over 31,000 students. Of these, only 45% successfully complete their education (Florida Department of Education, 2007). These numbers highlight the importance of finding a curriculum based measure (CBM) to effectively monitor the progress of students with EBD as well as the ensuing challenge for school districts to increase the levels of academic success experienced by students with EBD.

For the past several years, the Dynamic Indicators of Basic Early Literacy Skills (DIBELS) has been used in the State of Florida to monitor the progress and growth of students with and without disabilities in reading. In third grade, the DIBELS measures use oral reading fluency (ORF) to assess three of the five big ideas of early literacy: alphabetic principle**,** fluency with connected text**,** and comprehension (Kaminski, Cummings, Powell-Smith, & Good, 2008). The literature suggests that DIBELS is a good predictor of success on standardized assessments in reading for all students (Buck & Torgeson, 2003; Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008; Schilling, Carlisle, Scott, & Zeng, 2007). In 2008 with the nearing of the end of the Reading First grant program, the Florida Department of Education began a pilot project of the Florida Assessments for Instruction in Reading (FAIR) (Florida Center for Reading Research, 2009). During the 2008-2009 academic year, a preliminary version of the FAIR was used

by approximately 150 teachers and 2,000 students in three Florida school districts: Leon, Alachua, and Manatee counties. Changes suggested by teachers and administrators were then incorporated into the final version implemented statewide in 2009-2010 (Carlson, Römhild, McCormick, Chin, Geiseinger, Shaw, & Foley, 2010). The twenty pilot schools in three Florida counties (chosen on their representative demographics of Florida and also on the availability of local coordinators and testers) administered the tests during three different assessment windows (Florida Center for Reading Research, 2008b). Based on the project's findings, the Florida Department of Education suggested that the assessments were an improvement over DIBELS, as they included the pre-K population and evaluated both vocabulary and comprehension. Additionally, the FAIR could be used to assess standards in grades 3-12 and predict student end-of-year performance on standardized measures (Florida Center for Reading Research, 2008a). These diagnostic inventories can be used to guide instruction and to accurately monitor the progress of students in the area of reading. In August of 2009, the new Florida Assessments for Instruction in Reading was made available to K- 12 public schools free of charge.

The FAIR does not include an oral reading fluency measure as a part of the assessment battery. Research over the past several years suggests that ORF or words correct per minute (WCPM) is a valid way for educators to monitor the progress of their students' reading success for all students (Baker & Smith, 2001; Coyne, Kame'enui, & Simmons, 2004; Fuchs & Fuchs, 1992; Fuchs, 1999; Hagans, 2008) as well as for predicting the success on standardized assessments like the FCAT (Roehrig et al., 2008; Shapiro, Keller, Lutz, Santoro,

& Hintze, 2006; Wood, 2006). However, the FAIR differs from the DIBELS in that it uses a maze task.

A maze task is created by leaving the first sentence of the passage intact and deleting every nth word in subsequent sentences. In place of deleted words, three alternate words are offered in a multiple choice format. The alternatives consist of two incorrect and one correct word. Generally these passages are read silently rather than aloud as in the ORF. However, maze tasks are scored in much the same manner as the ORF in that only correct word choices are counted (Wiley & Deno, 2005). Like the ORF, the literature suggests that maze tasks can be an effective way to monitor and measure the reading comprehension of all students (Deno, Maruyama, Espin, & Cohen, 1990; Guthrie, 1973; Guthrie, Seifert, Burnham, & Caplan, 1974).

There is much in the literature regarding the efficacy of using curriculum based assessments like the DIBELS and FAIR with the general population of students. As a result, there is widespread use of curriculum based assessments across the field of education. However, very little research has looked specifically at its use with students with EBD; and notwithstanding, many important instructional decisions with far-reaching ramifications are made everyday for this population based on these assessment outcomes. This becomes the basis for the purpose of this study.

Purpose of the Study

This study investigated the effectiveness of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS, Good & Kaminski, 2002) at predicting success for students with EBD and a matched (by ethnicity, SES, gender and age) group of students without EBD on

the Florida Comprehensive Assessment Test (FCAT). The DIBELS uses timed oral reading passages to measure fluency. Correlational research procedures were used to compare the oral reading fluency (ORF) score of the DIBELS with the developmental scale score in reading on the FCAT to determine if DIBELS is a predictor of FCAT success for third and fourth grade students with and without EBD. Secondly, this study investigated the effectiveness of the Florida Assessments for Instruction in Reading (FAIR) at predicting success for students with EBD with and without EBD on the Florida Comprehensive Assessment Test. The FAIR utilizes a maze task to measure a student's reading growth. Correlational research procedures were used to compare the maze percentile rank score of the FAIR with the developmental scale score in reading on the FCAT to determine if the FAIR is a predictor of FCAT success for students with and without EBD. Finally, this study utilized multiple regression analyses to determine which tool is the overall better predictor of student success for these populations on the FCAT.

This study examined the effectiveness of the two types of assessments commonly used for identifying students in need of academic support. The use of these assessments will ultimately determine the effectiveness of curriculum-based measurements for monitoring the progress of students with and without EBD and at predicting their success on the FCAT. The determination of this information is extremely beneficial for teachers of students with and without EBD.

Research Questions

The purpose of this study was to extend the research on the use of the ORF as measured by the DIBELS and maze measures as measured by the FAIR with students with

and without EBD. Assessments were administered over a two-year period in the fall (AP1) and then again in the winter (AP2). Of particular interest is which curriculum based measure is a better assessment for predicting success on the FCAT. The following research questions were investigated:

1. How do the mean scores for students with EBD compare to the mean scores for students without EBD on the ORF, maze, 3rd grade FCAT, and 4th grade FCAT assessments?
2. What is the relationship between ORF scores on the DIBELS and total reading scores on the FCAT for students with and without EBD?
    a. What is the relationship between AP1 ORF scores on the DIBELS and total reading scores on the 3rd grade FCAT for students with and without EBD?
    b. What is the relationship between AP2 ORF scores on the DIBELS and total reading scores on the 3rd grade FCAT for students with and without EBD?
    c. What is the relationship between AP1 ORF scores on the DIBELS and total reading scores on the 4th grade FCAT for students with and without EBD?
    d. What is the relationship between AP2 ORF scores on the DIBELS and total reading scores on the 4th grade FCAT for students with and without EBD?

3. What is the relationship between AP1 and AP2 ORF scores on DIBELS for 3rd grade students with and without EBD?
4. What is the relationship between Maze Percentile Rank scores on the FAIR and the total reading scores on the 4th grade FCAT for students with and without EBD?
    a. What is the relationship between AP1 Maze Percentile Rank scores on the FAIR and the total reading scores on the 4th grade FCAT for students with and without EBD?
    b. What is the relationship between AP2 Maze Percentile Rank scores on the FAIR and the total reading scores on the 4th grade FCAT for students with and without EBD?
5. What is the relationship between AP1 and AP2 Maze Percentile Rank scores on the FAIR for 4th grade students with and without EBD?
6. What is the relationship between ORF scores on the DIBELS and Maze Percentile Rank scores on the FAIR for students with and without EBD?
    a. What is the relationship between AP1 ORF scores on the DIBELS and AP1 Maze Percentile Rank scores on the FAIR for students with and without EBD?
    b. What is the relationship between AP2 ORF scores on the DIBELS and AP2 Maze Percentile Rank scores on the FAIR for students with and without EBD?

7. What is the relationship between total reading scores on the 3rd grade FCAT and total reading scores on the 4th grade FCAT for students with and without EBD?

Significance of the Study

The benefits of being able to predict with accuracy the reading performance of students with and without EBD is crucial for teachers. Identifying those students who are at-risk of failing early on goes a long way towards increasing students' reading success. This is especially important for students with EBD as teachers work to close the achievement gap between these students and their peers. Thus far, there is very little research that has looked at the use of CBM with students with EBD. In fact, Rones and Hoagwood (2000) found in a comprehensive review of the literature that there are "few studies specifically focused on this group of children. This is a major shortcoming in the knowledge base," (p. 238). Third grade is a particularly important year as the ability to read by the end of third grade is vital to children's success in school, their life-long earning potential, and their ability to contribute to both our state and national economy (The Annie E. Casey Foundation, 2009). In the state of Florida there is a mandatory retention requirement for all third graders (regardless of disability) who are not proficient on the FCAT.

Additionally, students with EBD are required to have an individualized education plan (IEP). Under IDEIA (2004), this plan should address the needs of the individual student. It is crucial that a teacher understand the present levels of the student in order to best meet his or her needs. Giving teachers the correct tools to adequately assess and monitor progress across the curriculum improves the quality of the IEP and helps ensure that the student

continues to show growth. This is equally important for students receiving supports through the Response to Intervention model.

Response to Intervention (RtI) is a problem-solving process for addressing the needs of students. The National Research Center on Learning Disabilities defines RtI as "an assessment and intervention process for systematically monitoring student progress and making decisions about the need for instructional modifications or increasingly intensified services using progress monitoring data" (Johnson, Mellard, Fuchs, & McKnight, 2006, p. i.2). As a result of this mandate many students now have an individualized plan designed to address a student's needs. While the RtI problem-solving process is not specifically required for students with disabilities, the lines between special education and general education are blurring. NCLB and IDEIA both mandate a shared responsibility and accountability of general education and special education teachers in an effort to improve the educational outcomes of students with disabilities (Elliot, 2003). The state of Florida is using the RtI model as a means of not only meeting student needs but also as a way of identifying and labeling students with specific learning disabilities. As a result, many school districts rely heavily on curriculum-based measurements in order to monitor the progress of students with and without disabilities.

These monitoring tools assist educators in meeting the needs of all students and therefore increasing the likelihood that these same students meet grade level standards on state-wide assessments like the FCAT. This ensures that there is a focused attempt to provide meaningful and specific interventions that clearly address a student's individual needs and

that every student is provided appropriate instruction using proven research-based strategies in order to meet grade level standards.

By increasing an educator's ability to predict the reading success of students with and without EBD, effectively monitoring their progress across the curriculum, and ensuring appropriately matched reading interventions, the use of the DIBELS and FAIR can disrupt the current cycle of educational failure that students with and without EBD experience. With the current mandates of NCLB, the reauthorization of IDEA, and the implementation of RtI, the full impact of this study will only be effectively evaluated over time.

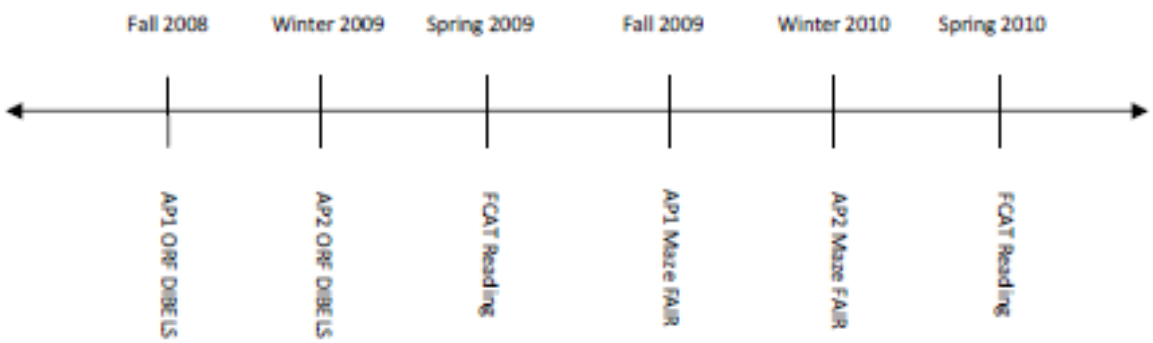
Overview of Methodology

The purpose of this study was to assess the effectiveness of maze and oral reading fluency measures at predicting success for third and fourth grade students with and without emotional and behavioral disorders on the Florida Comprehensive Assessment Test. This study focused on the accuracy of maze and oral reading fluency measures for the progress monitoring and predicting success of third and fourth grade students with and without EBD on a standardized reading assessment.

A group of fourth grade students with and without EBD who were enrolled in suburban public elementary schools in southwest Florida participated in the study. A retrospective look at the scores of two different types of assessments for this group of students was examined over a two-year time span (See Figure 1.). The first was the oral reading fluency measure of the Dynamic Indicators of Basic Early Literacy Skills. The oral reading fluency measure of the DIBELS was administered at two different assessment periods in 2008-2009 while students were enrolled in the third grade. The first and second

scores were used as a predictor for success on the FCAT administered in March of 2009. The third score was not included as it was administered at the end of the year, after the administration of the FCAT. The following year, the same group of students (now fourth graders) participated in two administrations of the maze measures of the Florida Assessments for Instruction in Reading during the 2009-2010 school year. These scores were used to determine success on the FCAT administered in March of 2010.

Figure 1. Timeline for Assessment Administrations



Note: Timeline of the assessment administrations which shows the dates of different administrations of ORF, maze, and FCAT measures.

Theoretical Framework

John Carroll (1963) proposed a model that school learning is a function of time. His model suggested that the learning rate for students was actually a function of the opportunity and perseverance to learn (or time spent) and a student's aptitude, ability to

understand instruction, and the quality of instruction (time needed). Defining the variable of time needed is important for the purposes of this study. First, as previously stated, one of the variables defined by Carroll was a student's aptitude or ability to learn. He suggested that there is no definitive way to measure aptitude. Therefore, he noted that aptitude must be estimated by measures of past performance (Carroll, 1963). Additionally, understanding the amount of time needed allows a teacher to manipulate the classroom environment to increase the time allocated for learning and improve the amount of time a student spends academically engaged.

Assessment in education is one way of measuring past performance and is therefore important when attempting to quantify aptitude. As Carroll suggests, aptitude determines the amount of time needed, which can be useful to educators when defining the present levels of their students in order to effectively guide instructional decision making.

Improving time on task or academic engaged time is one way to change the way in which instruction is provided and is paramount for student success in learning. This variable can be manipulated by a teacher to improve learning. However, the teacher must understand how much time is necessary in order to plan effectively.

Carroll's Model of School Learning served as the framework for this study. CBM assessments were used to determine the performance levels of students with and without EBD. This is important information for teachers as understanding how a student is progressing can influence an educator's instructional decisions for a student. This can

assist in effectively establishing the amount of time that a student needs to be successful within the curriculum.

Definition of Terms

Adequate Yearly Progress (AYP) is a requirement of NCLB and must be defined by individual states as a means of ensuring that all students are proficient in reading, math and science. Data must be reported in disaggregate by sub-groups to ensure that all students are making learning gains (Wright, Wright, & Heath, 2004).

Curriculum based measurement (CBM) is an approach to measuring the academic growth of individual students for the purpose of evaluating the effectiveness of the instruction being provided to individual students (Deno, 1985).

Dynamic Indicators of Basic Early Literacy Skills (DIBELS) are a set of procedures and measures for assessing the acquisition of early literacy skills in kindergarten through sixth grade. They are planned to be short (one minute) fluency measures utilized to monitor the development of early literacy and reading skills (Good & Kaminski, 2002).

An *Emotional/Behavioral Disorder* is defined in the State of Florida as a persistent (not responsive to evidence based interventions) and consistent emotional or behavioral response that negatively impacts the educational performance of students that cannot be attributed to age, culture, gender, or ethnicity (FLDOE, 2009b).

The *Florida Assessments for Instruction in Reading* (FAIR) is a relatively new reading assessment designed to identify students who will most likely be on or above grade level in reading by the end of the school year. It consists of two broad screening tasks as well

as targeted diagnostic formative assessments that can be used to gain information about the student in order to guide instruction (Florida Center for Reading Research, 2008c).

A *maze task* is created by leaving the first sentence of the passage intact and deleting every nth word in subsequent sentences. In place of deleted words, three alternate words are offered in a multiple choice format. The alternatives consist of two incorrect and one correct word (Wiley & Deno, 2005).

Response to Intervention (RtI) is defined by the National Research Center on Learning Disabilities (NRCLD) as "an assessment and intervention process for systematically monitoring student progress and making decisions about the need for instructional modifications or increasingly intensified services using progress monitoring data" (Johnson et al., 2006, p. i.2).

CHAPTER II

Review of the Literature

One of the most important pieces of information for a teacher is developing an understanding of the time necessary for a student to meet benchmarks. In the area of reading, research suggests that given enough time students that are reading disabled can acquire the necessary skills to become proficient readers (Shaywitz, 2003; Stanovich, 2000). It is well documented that students with EBD are more at-risk for developing reading disabilities (Kauffman, Cullinan, & Epstein, 1987); therefore, it is imperative that teachers of students with EBD be able to ascertain how much instruction is necessary for these students to overcome their reading deficits.

In this chapter an overview of the research literature that pertains to assessment in reading for students with and without emotional and behavioral disorders was explored. Topics include model of learning, students with emotional/behavioral disabilities, laws and mandates, and assessments. The review of the literature provided the knowledge and understanding necessary to draw conclusions about the predictive value of reading assessments for students with emotional and behavioral disorders.

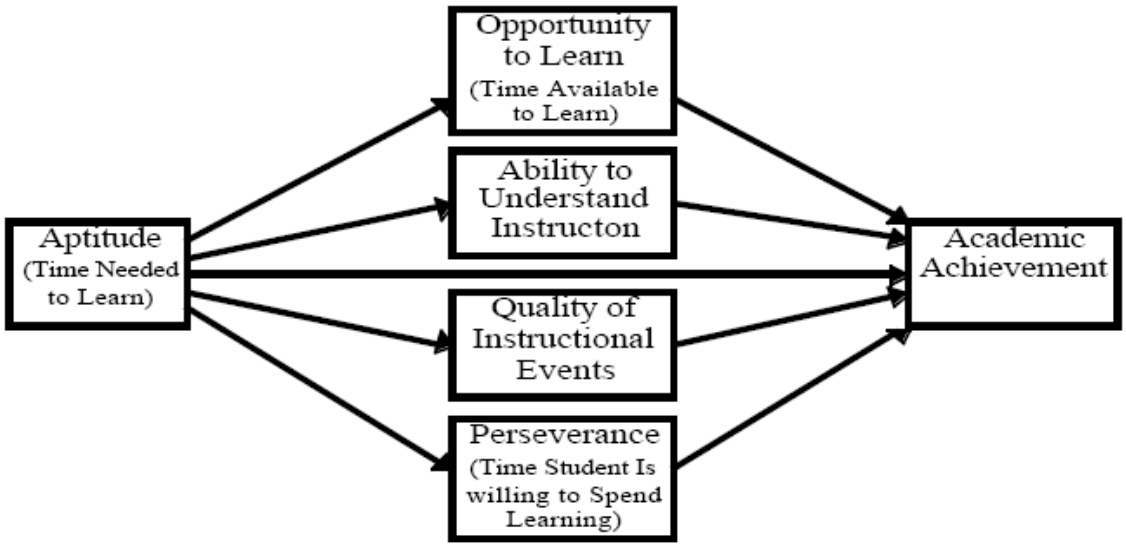
Theoretical Framework

The theoretical framework for this study is John Carroll's model of school learning (1963). His model has influenced educational researchers and developers for over three decades (Clark, 1987; Cooley & Lohnes, 1976). In fact, Benjamin Bloom used Carroll's model in the development of his concept of mastery learning (Carroll, 1989). In 1989, in a 25-year retrospective look at his model, Carroll stated that his model had

brought much more attention than he had initially expected, but felt that “the model could still be used to solve current problems in education” (p. 26).

In his model, Carroll accounted for five different classes of variables that influenced academic achievement. Three of these are in actuality expressions of time: aptitude, opportunity to learn, and perseverance. The last two are related to achievement: quality of instruction and ability to understand instruction. As seen in Figure 2, these variables work together to influence academic achievement. As defined by Carroll, academic achievement can be measured by standardized achievement tests such as the FCAT administered in this study.

Figure 2. John Carroll’s (1963, 1989) Model of School Learning



Note: Structure of relatedness of the five variables on academic achievement of Carroll’s model. From “A model to guide the integration of the world wide web as a cognitive tool in K-12 education,” by T. C. Reeves (1999).

As mentioned previously, Carroll's model of learning suggests that the learning rate for students is actually a function of the opportunity and perseverance to learn (or time spent) and a student's aptitude, ability to understand instruction, and the quality of instruction (time needed). Specifically, he proposed that:

$$Degree\ of\ Learning\ = f\left(\frac{time\ actually\ spent}{time\ needed}\right)$$

Defining the variable of time needed is important for the purposes of this study. Given that there is no definitive way to measure aptitude, Carroll interprets it as "the amount of time a student needs to learn a given task, unit of instruction, or curriculum to an acceptable criterion of mastery under optimal conditions of instruction and student motivation" (Carroll, 1989, p. 26). This suggests that most learners (with and without disabilities) are therefore capable of desirable levels of academic achievement if provided enough time.

The time components of aptitude can further be examined in terms of the opportunity to learn and to demonstrate perseverance. In a typical school schedule, the opportunity to learn may be hampered for many students with lower aptitude as they may be provided less time than they need to achieve a given set of objectives. Often, teachers may be unaware of the amount of time a student with disabilities may need to achieve the same benchmarks as his general education peers. Additionally, many educators may believe that the perseverance factor is simply a factor of student motivation and assume that all students have similar aptitudes. This equates to all students needing approximately the same amount of time to accomplish a certain learning task. Given

Carroll's definition, it is quite possible that a student may in fact be willing to spend the proper time on learning a given task; however, the necessary amount of time may not be available in school for this extra effort to be performed. In this case then, the perseverance factor will have little impact on academic achievement as measured through assessments.

Assessment in education is one way of measuring past performance and is therefore important when attempting to quantify aptitude. As Carroll suggests, aptitude determines the amount of time needed. "High aptitude is indicated when a student needs a relatively small amount of time to learn; low aptitude is indicated when a student needs much more than average time to learn" (Carroll, 1989, p. 26). Educators must therefore understand the present levels of their students in order to effectively provide instruction. This model provides the framework necessary to study CBM assessments used to determine the performance levels of students with and without EBD; additionally, this model traces the influence on an educator's instructional decisions for a student by effectively establishing the element of time as a student requirement for success within the curriculum. In other words, the intensity of instruction should change based on the current level of a student.

In the area of reading, time becomes an important variable for students who may be reading disabled and struggling academically as many students with EBD are. Stanovich (2000) suggests that these types of students are impaired in their phonological processing. This is especially important for students with EBD as language and communication are known to play a critical role in emotional development and academic

achievement (Wagner, Kutash, Duchnowski, Epstein, & Sumi, 2005). The good news is that the literature suggests that given direct and implicit instruction, students can overcome their reading disabilities as long as the proper amount of time is spent (Morris et al., 1998; Shankweiler et al., 1995; Stanovich, 1999; Torgeson, Morgan, & Davis, 1992).

Students with Emotional Behavioral Disabilities

Prevalence.

In 1999, it was estimated that there were 6-9 million children and adolescents living in the United States that suffer from emotional and behavioral disabilities. This large number reflects approximately 9-13% of all youth in the United States (U.S. Department of Health & Human Services, 1999). While not all of these students require specialized instruction or supports, public schools have identified and provide special education and related services to approximately 500,000 students with EBD; over the last 10-15 years, that number has grown by 20% (U.S. Department of Education [USDOE], 2002a; 2006). The good news is that in the last five years the opposite is true. From 2004 through 2007, there has been a 6% decrease in students with EBD as a percentage of the population of all students (USDOE, 2008). This is also true in the state of Florida where there is a similar trend.

Florida has seen a decrease in the enrollment of all students, including those with EBD. In 2004-2005 there were almost 36,000 students with EBD. That number has dropped to 28,936 students with EBD who were identified and receiving services in 2009. This represents a 5.2% decrease in the number of students with EBD when compared to the total student population in Florida (FLDOE, 2009c).

Of the over 82,000 students in the Lee County School District (LCSD), there are 827 students with EBD currently receiving services and support. Of these, approximately 32% are in an elementary school (K-5) setting, with the remaining 70% in middle and high. Typical of students with EBD, a total of 83% of the students with EBD in grades K-5 receive services and support in self-contained and/or special center school classrooms (LCSD, 2009).

Reading instruction.

The behavioral and academic struggles of students with EBD are well documented in the literature (Gresham, Lane, McMillan, & Bocian, 1999; Kauffman, 2005; Ruhl & Berlinghoff, 1992; Sutherland & Wehby, 2001). While the behavioral deficits present a challenge in and of themselves, the academic needs can be especially difficult to remediate (Nelson et al., 2004). Historically, students with EBD are difficult to teach and their success in school lags behind that of their non-disabled peers (Jacobson, 1998; Kauffman, 2005; Landrum, Tankersley, & Kauffman, 2003; Scott & Shearer-Lingo, 2002), and they often experience reading difficulties as well (Barton-Arwood, Wehby, & Falk, 2005; Forness, Bennett, & Tose, 1983; Rock, Fessler, & Church, 1997).

Nearly 80% of all referrals for special education involve reading problems (Nelson & Machek, 2007). Therefore, the best thing to do is to attempt to be proactive and prevent reading difficulties from developing (Torgeson, 2002). Students with EBD typically are one to two years below grade level in reading comprehension (Kauffman et al., 1987), earn lower grades (Barton-Arwood et al., 2005), and have a higher drop-out rate (Buckley, 2009; Lassen, Steele, & Sailor, 2006; Nelson, Johnson, & Marchand-Martella, 1996) than that of

their general education peers. As mentioned previously, reading difficulties lead to higher rates of delinquency, violence, and substance abuse (Fleming et al., 2004).

Laws and Mandates

There have been many advances in the understanding of reading development in students with and without disabilities (Lyon, Shaywitz, & Chhabra, 2004 as cited in Brown-Chidsey & Steege, 2005; McCardle & Chhabra, 2004; Shaywitz, 2003; Stanovich, 2000). This research points to the importance of providing high–quality scientifically-based instructional practices for students learning how to read. As a result, recent education policies in the United States have also reflected this development in the literature. No Child Left Behind (NCLB) and the Individuals with Disabilities Education Improvement Act (IDEIA) both have incorporated requirements for the use of evidence-based practices.

No Child Left Behind Act (NCLB).

As mentioned previously, the NCLB Act was passed by the United States Congress in 2001. This legislation emphasized accountability, assessment, parent options, highly qualified teachers, and research based instruction for all students. The biggest impact was in the area of reading and NCLB mandated that all students meet proficiency on challenging state academic achievement standards and state academic assessments in reading by the 2013-2014 school year (USDOE, 2002b, Wright et al., 2004). This was important, because for the first time, schools would be required to report progress towards these goals by subgroup. This included students with disabilities (USDOE, 2002a). The aspect of NCLB that had the most impact on education was the establishment of the Reading First grant program.

The purpose of Reading First was to ensure that all children in America learn to read well by the end of third grade. Reading First grants were designed to prevent reading difficulties in grades kindergarten through grade three by assisting states and school districts to implement proven methods of scientifically based reading instruction in classrooms. The state of Florida received over 300 million dollars to help reach the goal of every child reading at or above grade level by 2012 (FLDOE, 2009). These funds were provided to establish scientifically based reading programs for students enrolled in kindergarten through grade three and to increase the professional development of all teachers so that they have the skills necessary to teach these programs effectively. Lastly, Reading First supported the use of screening and diagnostic tools and classroom-based instructional reading assessments to measure how well students are reading and to monitor their progress (USDOE, 2002b).

Inclusion.

Traditionally, students with disabilities were excluded from state proficiency assessments. NCLB (2001) changed this practice with the requirement that states include all children in their statewide testing programs. The inclusion of students with disabilities in these assessment programs has resulted in many changes in how all students are being instructed, especially in the area of reading. As mentioned previously, this has blurred the lines between general and special education and has resulted in school districts implementing similar instructional practices and using the same curricular materials to instruct both general and special education students, including those with EBD (C. Brunick, personal communication, November 11, 2010). Additionally, the ways in which student success is measured are also similar.

Adequate Yearly Progress.

NCLB also mandated that states define adequate yearly progress (AYP) of all students and report the achievement results of various subgroups, including students with disabilities (USDOE, 2007). AYP is a concept that applies to schools and districts rather than individual students (Wright et al., 2004). If achievement gaps exist, then schools are required to take specific action to close these gaps (Guskey, 2007). This too has resulted in significant change in the way students with disabilities are instructed and their progress monitored alongside their general education peers.

Individuals with Disabilities Education Improvement Act (IDEIA).

President George W. Bush signed the reauthorized Individuals with Disabilities Education Improvement Act (IDEIA) into law in 2004. This piece of legislation specifically included three elements of evidence-based practice: (1) use of scientifically-based reading instruction, (2) evaluation of student response to interventions, and (3) utilization of data for decision-making. (Brown-Chidsey & Seege, 2005).

The first element, the use of scientifically-based reading instruction, came directly out of the body of research about good reading instruction. Using empirically based interventions is considered best practice, because the use of these interventions increases the likelihood of a positive outcome for students. Secondly, many common educational practices have later been found to be ineffective (Brown-Chidsey & Seege, 2005). These practices result in poor academic growth for students. Lastly, the goal of any intervention is progress. The understanding of whether or not an intervention is working clearly has implications for students and educators alike.

The evaluation of a student's response to interventions is the second element mandated by IDEIA. Fuchs and Fuchs (1986; 1999) have supported the importance of having an assessment process that incorporates ongoing data collection and its subsequent use in evaluating the effectiveness of instruction. The President's Commission on Excellence in Special Education (USDOE, 2002a) also emphasized the need for districts to implement a process of continued progress monitoring. In the case of reading skills, curriculum based measures are well researched ways of documenting student skills (Good & Kaminski, 2002; Shin, Deno, & Espin, 2000; Wayman, Wallace, Wiley, Ticha, & Espin, 2007).

The last element required by IDEIA is the use of data for decision-making. Specifically, it specifies that decisions about a student's response must be data-based. This increases the likelihood that educators will make valid decisions regarding instruction for students, thereby improving student achievement and increasing the probability that all students will meet benchmarks. These three elements together form the basis of what has become known in the field of education as the response to intervention model.

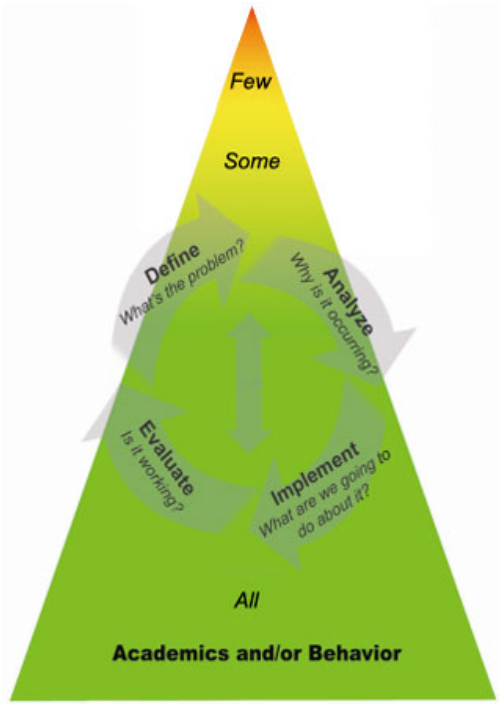
Response to Intervention (RtI).

IDEIA (2004) mandates that schools use a tiered system of interventions to meet the needs of all students. RtI is a problem-solving process in which assessment and interventions are integrated within a multi-level system in an effort to meet the needs of all students (Fuchs & Fuchs, 2006; 2009a; 2009b). IDEIA (2004) does not define the number of tiers or what this system of interventions should look like. In fact, across the country RtI frameworks look very different. Berkeley, Bender, Gregg Peaster, and Saunders (2009) found that some schools had only two tiers of intensifying interventions while others had as many as seven.

The state of Florida has adopted a three-tier model and defines RtI as “a multi-tiered approach to providing high quality instruction and intervention matched to student needs, using learning rate over time and level of performance to inform instructional decisions. RtI involves the systematic use of assessment data to most efficiently allocate resources in order to improve learning for all students” (FLDOE, 2011, p. 1,). As such, the framework in Florida includes the following: (1) multiple tiers of evidence-based instruction, (2) a problem-solving method used to develop interventions, and (3) a data collection/assessment system to inform decisions. Recently, there has been a move to describe this model as a multi-tiered system of student supports (MTSSS). The conceptual framework for the state of Florida’s model is given in Figure 3 below.

The tiered approach of RtI for meeting the needs of students and the emphasis on grade level proficiency for all students in NCLB (2001) and IDEIA (2004) highlight the need for reliable and accurate progress monitoring. As a result, educators must now access information that can be used to predict how students are achieving when compared with the grade-level standards, in order to accelerate student progress (Quenemoen et al., 2004). As the use of progress monitoring increased, so too has the search for an efficient and accurate means of assessing student performance.

Figure 3. Florida's RtI Model.



Note: Model depicting increasingly more intense instruction and interventions. From Florida Department of Education (http://www.florida-rti.org/flMod/threeTierModel.htm).

Curriculum based monitoring.

A key component of RtI is the use of screening and progress-monitoring assessments. As mentioned above, this is required under IDEIA (2004) for measuring a student's responsiveness to instruction and/or intervention. Curriculum based measurements are designed to do just that. Additionally, these measures are capable of informing teachers' decisions about their teaching; what to teach, how to teach, and how long to teach. Therefore it makes the perfect tool for the monitoring of student progress. In order to do this, these assessment measures must be aligned with the curriculum, sensitive to the instruction,

repeatable, and criterion-referenced in order to determine mastery of the skills (Hosp, Hosp, & Howell, 2007). A more in-depth look at curriculum based monitoring will follow.

Assessments

Curriculum based measurement.

As educators, it is fairly easy to gather a large amount of assessment data. The questions that remain are: What data should be collected? What do we do with that data? Curriculum-based measurement (CBM) is one of the more successful models for gathering data and using it to make educational decisions in a problem solving model (Good & Kaminski, 1996). CBM is one way to measure the progress of a student in reading. CBM can be used to continuously monitor progress and assist educators in making formative instructional decisions (Deno, 1985). The benefit of using CBM is that it is an efficient means to monitor progress that is sensitive to student growth and can be administered frequently. Additionally, teachers can easily score these measures and chart the progress both numerically and graphically. CBM is designed as part of a problem-solving approach to address the academic difficulties of students (Deno, 1990), thereby focusing a teacher's attention on solving the problems a student faces rather than the characteristics of the student. The problem-solving model requires that the teacher constantly monitors and evaluates a student's progress so that modifications can be made to the curriculum and instruction. Therefore, a tool that can quickly evaluate a student's growth and response to the instruction is necessary. Usually, CBM in reading is a 1-minute reading probe that is administered regularly during the school year (Wiley & Deno, 2005).

Read aloud.

Research indicates that the general procedures of CBM are valid for indicating levels of student performance even when the passages are not taken from the students' curriculum directly (Espin & Deno, 1995; Fuchs & Deno, 1994; Wiley & Deno, 2005). This approach is referred to as general outcome measurement (GOM) by Fuchs and Deno (1994). Reading aloud (more recently known as oral reading fluency) has been used as a means of measuring a students' reading growth, and is an example of a GOM in which a student reads a passage for one minute and the number of words read correctly are counted to indicate the level of performance.

Despite the technical adequacy of the read-aloud measures, teachers and many researchers have continued to doubt whether simply measuring a student's rate of reading aloud from a text for one minute can be a true indictor of the student's reading proficiency. This is especially true in the area of reading comprehension (Mehrens & Clarizio, 1993; Yell, 1992). Yet researchers in their examination of the relationship between reading aloud and general reading proficiency, especially reading comprehension, have found this GOM to be a valid measure for overall reading performance and have established a relation between reading aloud and reading comprehension (Fuchs, Fuchs, & Maxwell, 1988; Hosp & Fuchs, 2005; Kranzler, Brownell, & Miller, 1998).

Examining students with mild disabilities in Grades 4 to 8, Fuchs, Fuchs, and Maxwell (1988) compared the validity of CBM read-aloud measures to that of other measures used to assess reading comprehension, including maze, story retell, and

question-answering measures. They found that reading aloud scores correlated more strongly with scores on the comprehension and word skills subtests of a standardized achievement test (r = .91 and r = .80) than did the scores from the other comprehension measures (rs = .76 to .82 for the reading comprehension and rs= .66 to .76 for the word skills subtests). This suggests that reading aloud is more than just decoding fluently and that a reader who is more fluent will also exhibit better reading comprehension skills.

Like many teachers, Kranzler, Brownell, and Miller (1998) hypothesized that the number of words read aloud from text in one minute might merely be a reflection of general speed of processing. They investigated how general cognitive ability, speed, and efficiency of elemental cognitive processing and reading aloud impacted the reading comprehension for students in grade 4. Using multiple regression analyses, Kranzler et al. (1998) found that there was a significant relationship between reading aloud and reading comprehension. They were unable to explain this relationship with general cognitive ability or speed or the efficiency of elemental cognitive processing. The results of their study suggest therefore, that reading aloud is not just a measure of general cognitive processing speed, but also of reading comprehension.

In 2005, Hosp and Fuchs found that the relationship between reading aloud and reading proficiency changed with age. Through the examination of the relationships between CBM reading aloud and the Decoding, Word Reading, and Comprehension subtests of the Woodcock Reading Mastery Test Revised (1987), Hosp and Fuchs found similar correlations between students in Grades 2 and 3 (ranging from .82 to .88). However, lower correlations were observed for students in grade 4 for the Decoding and

Word Reading subtests (*rs* = .72 and .73) than for the Reading Comprehension subtest (*r* = .82). They suggested that this was because by fourth grade students have normally acquired the skills necessary to read words quickly and may be more focused on comprehension rather than parts of unknown words. More importantly, they propose that the use of CBM at different grades may be tapping different skill sets.

Shinn, Good, Knutson, Tilly, and Collins (1992) examined the role of reading aloud as it related to decoding, fluency, and comprehension skills for students in Grades 3 and 5 using a confirmatory factor analysis. They found a single-factor model of "reading competence" that was validated for third-graders in which all reading skills make significant contributions. A two-factor model that included decoding and comprehension as two separate but highly related factors was validated for fifth graders, with reading aloud loading on the decoding factor. This seems to add to the findings of Hosp and Fuchs (2005) as discussed previously.

Another concern being raised by teachers is the existence of word callers. Word callers are students who read fluently but demonstrate low reading comprehensions skills. Hamilton and Shinn (2003) investigated a teacher's ability to recognize word callers. In their study, third-grade teachers were asked to name one to two students who were word callers and one to two similarly fluent peers. These similarly fluent peers were those students who demonstrated similar fluency rates but were judged to demonstrate higher levels of comprehension. The results confirmed that there were significant differences in the comprehension levels between the students in the word caller and similarly fluent groups; however, there were also differences in reading fluency. The word callers

demonstrated lower fluency rates than the similarly fluent readers group, thus questioning the ability of the teachers to accurately identify students as word callers.

More recently, researchers have focused their attention on the measurement of oral reading fluency. Fluency is taken to mean oral reading fluency in connected text. The National Reading Panel (2000) defined fluency as "the ability to read a text quickly, accurately, and with proper expression." Furthermore, fluency goes beyond accurate word recognition and is a causal determinant of higher order skills such as reading comprehension (NRP, 2000).

ORF measures are convenient as they can be directly observed through read-aloud tasks, are quick to give, and are easy to administer. They play a pivotal role in the assessment of reading skills (Hasbrouck & Tindal, 2006; Marcotte & Hintze, 2009) and have been found to be strong indicators of success on reading comprehension measures (Fuchs, Fuchs, & Maxwell, 1988; Hintze, Callahan, Matthews, Williams, & Tobin, 2002; Hintze, Owen, Shapiro, & Daly, 2000; Hintze, Shapiro, Conte, & Basile, 1997; Jenkins, Fuchs, van den Broek, Espin & Deno, 2003; Shinn et al., 1992). Additionally, multiple studies have demonstrated an association between ORF and overall reading proficiency, including reading comprehension (Markell & Deno, 1997).

LaBerge and Samuals first hypothesized in 1974 that automaticity of reading was directly connected to high levels of reading comprehension. Posner and Snyder (1975) suggested two context-based expectancy processes that facilitate word recognition. This has further been supported by multiple researchers (i.e., Adams, 1990; Jenkins et al., 2003; NRP, 2000). Additionally, Stanovich (1980) notes that good reading fluency is the

result of two processes. One process he described as from the bottom-up (print driven) and the other from the top-down (meaning driven). These two processes operate concurrently when a reader confronts a word in context. He believes that a skilled reader rarely needs to use conscious bottom-up processes to read words because there is a high level of automatic word recognition. Poor readers, on the other hand, must rely more on the context of the sentence to accurately read words due to inefficient and unreliable word automaticity skills, thus decreasing their comprehension levels (Stanovich, 2000). While there are some important differences among these models, all of them support the notion that efficient word recognition processes free up resources for comprehension.

As a result, oral reading fluency may be the most thoroughly studied CBM measure and therefore has the most support in the literature for its use (Baker et al., 2008). This makes oral reading fluency outcomes useful for guiding instruction as well as for predicting student performance on statewide high stakes assessments (Barger, 2003; Buck & Torgesen, 2003; Hintze & Silberglitt, 2005; Shaw & Shaw, 2002; Stage & Jacobsen, 2001).

ORF as an indicator of reading proficiency.

One of the first studies to investigate the validity of using ORF as an index of overall reading proficiency was Deno, Mirkin, and Chiang (1982). Students in grades one through five were given five different CBM measures (read from word lists, read underlined words in passages, read an intact passage, identify missing words in a passage, and state meaning of underlined words). ORF was found to have the strongest correlation (r= .71 to .91) with published measures of reading comprehension (i.e. Stanford

Diagnostic Reading Test, & Woodcock Reading Mastery Test). ORF correlated higher than maze and word meaning, which at the time were considered to be more direct measures of overall reading.

Then in 1988, Fuchs, Fuchs, and Maxwell investigated several CBM reading measures (including ORF, answering questions, recall, and cloze tests) with middle school students in special education. Again the authors found that ORF correlated higher than any of the other measures with two of the SAT-9 subtests. Interestingly enough, ORF showed a higher correlation with the reading comprehension criterion measure (.92) than it did with the decoding criterion measure (.81). This suggests that ORF is more strongly related to comprehension than decoding. This is a pattern that has been replicated in other studies as well (i.e., Shinn et al., 1992).

Several earlier studies were published establishing the validity of ORF as a measure of overall reading proficiency. One of the major conclusions of these studies is that correlations between ORF and published measures of reading proficiency, including reading comprehension, are consistently moderate to strong in value, generally ranging from .60 to .90 (see Marston, 1989, for a review of the research on ORF).

After the passage of NCLB in 2002 and the subsequent reauthorization of IDEIA in 2004, the use of ORF again was being investigated as more and more educators began to use ORF to assess reading performance in attempts to provide intensive and early intervention as required by these educational reforms. In fact, one study found that over 90% of the approximately 1,600 districts and 5,283 schools designated as Reading First schools, used oral reading fluency (ORF) to screen students for reading problems and

monitor their progress (Greenberg, Howe, Levi, & Roberts as cited in Baker et al., 2008). Both NCLB (2001) and IDEIA (2004) mandate that schools focus on intervening early and intensively to address reading problems. The literature provides a strong support for the use of ORF as a measure of reading proficiency in general and special education.

In 2006, Wood examined the relationship between oral reading fluency and performance on a statewide reading test (Colorado Student Assessment Program) in grades three through five. It is important to note that curriculum based measures of ORF were used. The author found significant correlations (.67 to .75) for all three grade levels. Additionally, it was determined that ORF was a better predictor of success on the statewide reading test than the previous year's performance on the same test. This would suggest that, regardless of previous performance, ongoing progress-monitoring is important for best meeting the needs of students.

Maze task.

Another type of GOM is the maze task. As mentioned previously, a maze task is created by leaving the first sentence of the passage intact and deleting every nth word in subsequent sentences. In place of deleted words, three alternate words are offered in a multiple choice format. The alternatives consist of two incorrect and one correct word. Generally these passages are read silently rather than aloud as in ORF. However, maze tasks are scored in much the same manner as ORF in that only correct word choices are counted (Wiley & Deno, 2005).

While the literature demonstrates that measuring the number of words read aloud in one minute is a technically sound assessment, the measure is limited because it must be

administered individually. Additionally, while it has been shown useful for younger students, there continues to be disagreements about its validity for older students (Shelton, Altwerger, & Jordan, 2009; Valencia, Smith, Reece, Li, Wixson, & Newman, 2010). With the improvements in technology and the demand to actively progress monitor more students, many have begun to consider the maze task. The maze can be administered in groups and appears to be more of a measure of reading comprehension measure than a reading fluency measure. Additionally, it can be administered using a computer and there seems to be a more general acceptance for its use with older students.

The maze task is not a new measure. An untimed version of the maze was studied in the 1970s by Guthrie as a measure of reading comprehension. Like ORF, the literature suggests that maze tasks can be an effective way to monitor and measure the reading comprehension of students (Deno et al., 1990; Guthrie et al., 1974). Guthrie found the maze measure to have good stability and a high correlation with standardized measures of reading proficiency for students with and without disabilities (Guthrie, 1973; Guthrie et al., 1974). However, the use of the maze as a timed measure within a CBM framework did not appear until the late 1980s and early 1990s.

The original maze task was an untimed passage which students read and made correct word choices and was called a multiple choice cloze activity (Cranney, 1972-1973; Guthrie, 1973; Kingston & Weaver, 1970). In the early 1990's, the maze structure changed and began being used as a timed activity equivalent to oral reading measures (Deno et al., 1990). Students read a passage to themselves and choose the correct choice in a maze within a time limit (Shin et al., 2000). Research has shown that when timed,

maze scores are likely to increase validity coefficients and are less negatively skewed (Parker, Hasbrouck, & Tindal, 1992).

Espin, Deno, Maruyama, and Cohen (as cited in Fuchs & Fuchs, 1992) investigated the technical adequacy of a maze measure used as part of a screening instrument called the Basic Academic Skills Samples (BASS) (Deno, Maruyama, Espin, & Cohen, 1989) administered in a group. Three one-minute maze selection tasks were included in the reading portion of the BASS. It was administered to more than 2,000 students in Grades 1 through 6. Correlations between the BASS maze and one-minute read-aloud passages for a random sample of students from Grades 3, 4, and 5 were .77, .86, and .86 respectively. A stable pattern of increase in maze scores from Grades 1 to 6 was revealed in the data, as well as from winter to spring within each grade.

In 1990, Fuchs & Fuchs (as cited in Wayman et al., 2007) obtained a correlation of .83 between scores on a maze task and a read-aloud measure, and a correlation of .77 between scores on a maze task and the reading comprehension subtest of the Stanford Achievement Test [SAT] (Gardner, Rudman, Karlsen, & Merwin, 1982). Then in 1992, Fuchs and Fuchs extended their research on the maze measure as they searched for a computer assisted CBM reading measure. They thought that such a CBM might be more acceptable to teachers than the typical read-aloud measures. Fuchs and Fuchs compared the technical adequacy and level of teacher acceptance for several alternative CBM measures. These included question answering, story recall, cloze, and maze selections. The maze measure was administered two times per week for 18 weeks in a row. It took approximately 2.5 minutes to complete and was administered using a computer. The

results of this study revealed that the maze measure is sensitive to a change in performance over time. The study also revealed a relatively small ratio of slope to standard error of estimate. This makes it easier to progress monitor using a graph, thereby easily displaying whether or not progress is being made. It also makes it extremely useful for teachers. Additionally, the teachers and students alike reported a high rate of satisfaction with the maze measure.

Jenkins and Jewell (1993) compared the technical adequacy of both a read-aloud measure and a maze task for Grades 2 to 6. Students participating in the study were given three one-minute maze tasks and three one-minute read-aloud passages. All of the passages used were written at the first or second grade reading level. Scores on the measures were correlated with scores on the Gates-MacGinitie Reading Tests (MacGinitie, Kamons, Kowalski, MacGinitie, & McKay, 1978) and the Metropolitan Achievement Tests [MAT] (Prescott, Balow, Hogan, & Farr, 1984). They found correlations ranging from .63 to .88 across the grade levels with the Gates and from .58 to .87 with the MAT. There tended to be a stronger correlation for the read-aloud tasks in the primary grades than for the maze task. When looking across all grade levels, correlations between the read-aloud tasks and the criterion measures dropped from second grade (.80) through grades four, five, and six (.60 to .70). However, the correlations for the maze remained consistent across the grade levels (.65 to .75). This suggests that the read-aloud may be better suited for the primary grades than the maze measure. This is supported by Ardoin et al. (2004), who also found that the addition of a

maze task did not increase the chances for third grade students' success on a standardized achievement test in reading.

CBM as a predictor of test performance.

The majority of studies investigating the use of CBM have been in the area of reading (Marston, 1989). In an extensive summary, Marston provides a clear explanation of CBM and addresses the issues of technical adequacy and some of the inherent problems of the more traditional systems that have been used to evaluate student performance. Overall, CBM in reading have been shown to generally have good construct validity (Fuchs & Fuchs, 1998; Hosp & Fuchs, 2005; Shinn et al., 1992). Additionally, CBM in reading has been shown to have high levels of correlation with other acceptable and published measures in reading (Deno et al., 1982; Fuchs et al., 1988). A review of the literature did not reveal studies specific to the use of CBM and students with EBD. However, the literature does suggest that students with and without mild/moderate disabilities demonstrate comparable performance on many reading-related tasks (Fletcher et al., 1994; Stanovich & Siegel, 1994). Therefore it is assumed that the previously mentioned studies are relevant to the purposes of this study.

DIBELS.

The Dynamic Indicators of Basic Early Literacy Skills (DIBELS) is an assessment tool developed by the University of Oregon. The DIBELS "are a set of procedures and measures for assessing the acquisition of early literacy skills from kindergarten through sixth grade. They are designed to be short (one minute) fluency measures used to regularly monitor the development of early literacy and early reading

skills" (University of Oregon Center on Teaching and Learning, 2009, p. 1). The most common measurement used is the Oral Reading Fluency (ORF). ORF is a measurement used to determine the number of words correctly read in one minute from a passage and can be used to help identify students who may require additional instructional support. An example of a third grade passage from DIBELS used to measure ORF can be found in Appendix A. Several studies have found a correlation between ORF and success on the Florida Comprehensive Assessment Test. In fact, Buck and Torgeson (2003) found a significant correlation between ORF scores and the reading portion of the FCAT (r=.70, p<.001) for third graders. This would suggest that ORF may be a good predictor of future performance on other standardized assessments.

Schilling, Carlisle, Scott, and Zeng (2007) examined the predictive validity of the DIBELS fluency measures for first through third graders in Michigan. They found that the DIBELS significantly predicted reading achievement on the Iowa Tests of Basic Skills at the end of the year. Hierarchical regression analyses found that ORF made the largest contribution of variance in reading total across the grade levels. From this, the authors concluded that, after second grade, ORF scores can accurately be used for predicting reading success. However, they suggest supplementing the DIBELS measures with measures of reading comprehension and vocabulary and increasing the frequency of progress monitoring for at-risk students.

In another study, the Florida Center for Reading Research evaluated the validity of DIBELS for predicting student performance on the reading comprehension measures of the FCAT and SAT-10 (Roehrig, Petscher, Nettles, Hudson, & Torgesen, 2008). The

study looked specifically at two samples of third grade students from Reading First schools and evaluated risk-level cut scores for success. The authors found that correlations of ORF with the FCAT and SAT-10 were high (r= .70-.71) and that "DIBELS is related equally well to a common measure of reading comprehension used across states as it is to a state-developed measure" (p. 360). It is important to note, however, that no students with EBD were identified as part of the sample.

Riedel (2007) found that ORF was a better predictor of comprehension than other assessments (like the retell fluency task included in DIBELS) that are specifically designed to measure comprehension. Additionally, it was found that ORF scores correctly classified 80% of the first and second grade students in the sample, indicating whether or not their comprehension status was satisfactory. This suggests that ORF can be used with a high level of effectiveness for identifying students that may be at-risk and require more supports or even specialized instruction.

FAIR.

The FAIR consists of several smaller measures that make up the entire assessment depending on the grade level. The focus for the purpose of this study was on the maze task. An example of a FAIR maze task is provided in Appendix B. The maze task is only one of the two tasks (maze and word analysis) that are part of the Targeted Diagnostic Inventory (TDI) of the FAIR. The maze and word analysis tasks are designed for progress monitoring of fundamental reading skills (Florida Department of Education, 2009). The purpose of the word analysis task is to assess a student's knowledge of the phonological, orthographic, or morphological information required to accurately identify

words in text. It is part of the targeted diagnostic inventory that is only given to students who score below a certain cut point on the other tasks. For this reason it will not be used in this study. The purpose of the maze is to measure whether or not a student can read a grade-level text with a basic level of comprehension by selecting one of three words that best completes the cloze items that are embedded within the passage. The FAIR derives three scores from the maze task; percentile rank, standard score, and an adjusted maze score. These scores can be used by the teacher to determine if the student is struggling with the fundamental reading skills of accuracy and fluency or basic text comprehension. It can also be used to monitor progress in text reading efficiency in struggling readers if they are receiving interventions focused on accuracy, fluency, and basic comprehension skills. As mentioned previously, it has been suggested that the maze measure appears to work better for older students, and in a recent comprehensive review of GOM in reading, Wayman et al. (2007) concluded that the maze test was indeed the best format for monitoring the growth of basic reading skills in middle and high school students.

As reported by the state of Florida Department of Education, scores on the maze tasks were adjusted for difficulty using equipercentile equating. This is a process whereby a raw score on a new form and a raw score on the anchor (or reference) form are equivalent for a group of test takers if they have the same percentile rank (FLDOE, 2009a). Student performance for fourth grade students is very similar for the two passages; passage 1 (X = 20.64, *S.D.* = 9.73); passage 2 (X= 21.94, *S.D.* = 9.57). The performance on the two maze passages were correlated with each other as a measure of parallel form reliability (mean correlation = 0.78). The state of Florida maintains that the

raw correlations underestimate the reliability of the maze task because they assume a test length of one passage, when in fact, the average of the two passages is used to score the task. After correcting for test length, the raw correlations reveal a more accurate estimate of parallel form reliability (reliability = 0.88).

To date, there is relatively little research on the FAIR as a predictor of test performance. The state of Florida has only recently developed the FAIR and has had little time to collect and measure its effectiveness. A pilot study was implemented in 2008-2009 by twenty pilot schools in four Florida counties using a preliminary version of the FAIR. It was used by approximately 150 teachers and 2,000 students and changes suggested by teachers and administrators were then incorporated into the final version implemented statewide in 2009-2010 (Carlson et al., 2010). The resulting data from the study was presented in the Florida Department of Education's (2009a) publication entitled *Florida Assessments for Instruction in Reading: Technical Manual.*

Based on the project's findings, the Florida Department of Education suggests that the assessments were an improvement over the DIBELS, as they include the pre-K population and evaluate both vocabulary and comprehension. These diagnostic inventories can be used to guide instruction and to accurately monitor the progress of students in the area of reading. Additionally, it has been suggested that FAIR can be used to assess standards in grades 3-12 and predict student end-of-year performance on standardized measures (Florida Center for Reading Research, 2008a). The Florida Department of Education (2009a) somewhat contradicts this in the technical manual when they state that "the purpose of the maze task is not to predict performance on the

FCAT, but rather to identify which students among those predicted to perform below grade level on the FCAT have difficulties with basic reading skills" (p. 42). However, it stands to reason that if basic reading skills are necessary to be successful on the FCAT then the maze task should be related.

The Florida Department of Education also suggests that while the maze task was constructed to be a measure of basic reading efficiency, there was also an emphasis on reading fluency (Florida Department of Education, 2009). This would indicate that the maze task is therefore related to measures of oral reading fluency. However, the state of Florida only provides this information for students in grades 6-10.

Data from the pilot study (N=828) regarding the maze task and performance on the FCAT for fourth graders is provided by the state. Generally, there is a decrease in the correlation from third (r=.66) to tenth grade (r=.44). The correlation in fourth grade (r=.57) is midway between the two cited correlations (Florida Department of Education, 2009). The relationship of FCAT performance level and maze score for students from the pilot study in grade 4 can be seen in Table 1. None of this data is specific to students with EBD. These correlations are significantly less than previous research on maze tasks and performance on reading comprehension measures (Wayman et al., 2007).

Table 1

Fourth Grade Maze Task Scores Corresponding to Various Levels of FCAT Performance

	N	Mean	S.D.
FCAT Level			
1	112	12.56	6.84
2	157	18.79	5.7
3	298	21.83	7.56
4	164	28.21	9.03
5	30	32.77	8.42

FCAT

The Florida Comprehensive Assessment Test is part of the Florida Department of Education's assessment and school accountability program. It is designed to measure student achievement against Florida's Sunshine State Standards and since 2000 has been administered in grades 3-10 (Harcourt, 2007). This study is concerned with the reading portion of the FCAT administered to all third and fourth grade students. FCAT reading is said to assess four cluster areas of reading; words and phrases in context; main idea, plot, and purpose; comparisons and cause/effect; and reference and research (Harcourt, 2007).

The FCAT reading assessment is made up of multiple-choice; gridded-response; and two types of performance tasks: short-response and extended-response questions. Measures of reliability are noted in the technical manual provided by the state of Florida. The Florida Department of Education reports that FCAT scores have reliabilities similar to other

standardized and statewide tests (3rd grade α=.890; 4th grade α=.853). Because reliabilities around 0.90 are typically viewed as positive and test scores can fluctuate randomly, the state suggests that the FCAT only be viewed as one indication of student achievement. Data are not reported for any subgroups, including students with EBD.

Summary

Chapter II provided an overview of the research literature that pertains to assessment in reading for students with emotional and behavioral disorders. Topics included the model of learning, characteristics and prevalence rates for students with emotional/behavioral disabilities, laws and mandates, and assessments. The review of the literature has provided the knowledge and understanding necessary to draw conclusions about the predictive value of reading assessments for students with emotional and behavioral disorders. A discussion of the research methodology, including a description of the research questions, study design, participants, ethical considerations, data collection procedures, instrumentation, and data analysis procedures is presented in Chapter III.

CHAPTER III

Methodology

This study examined the effectiveness of two types of assessments as predictors of success on the reading portion of the FCAT. The use of these assessments ultimately determined the effectiveness of curriculum-based measurements for monitoring the progress of students with and without EBD and at predicting their success on the FCAT. The determination of this information could be extremely beneficial for teachers as they work to design instruction to effectively meet the needs of students with and without disabilities.

The first purpose of this study was to investigate the effectiveness of the Dynamic Indicators of Basic Early Literacy Skills [DIBELS] (Good & Kaminski, 2002) at predicting success for third grade students with EBD on the Florida Comprehensive Assessment Test. These scores were compared against a matched group of general education peers. The study investigated the relationship between the oral reading fluency (ORF) score of the DIBELS with the developmental scale score in reading on the FCAT.

Secondly, this study investigated the effectiveness of the Florida Assessments for Instruction in Reading (FAIR) at predicting success for students with and without EBD on the Florida Comprehensive Assessment Test (FCAT). The relationship between the maze standard score of the FAIR with the developmental scale score in reading on the FCAT was examined to determine if the FAIR is a predictor of FCAT success for fourth grade students with EBD. Again, the scores were compared to a matched group of general education peers.

Lastly, through the examination of archival data, this study determined which tool is the better predictor of student success on the reading portion of the FCAT. The study

compared the performance of students with and without EBD on the ORF of the DIBELS and maze of the FAIR which were administered two times each with the scores with the students' scores on the FCAT.

Chapter three describes the research methodology used in the study and provides a description of the research questions investigated, the design of the study, and participants who were involved. Additionally, ethical considerations, data collection procedures, instrumentation, and data analysis procedures are discussed.

Research questions

The research questions that guided this study were as follows:

1. How do the mean scores for students with EBD compare to the mean scores for students without EBD on the ORF, maze, 3rd grade FCAT, and 4th grade FCAT assessments?
2. What is the relationship between ORF scores on the DIBELS and total reading scores on the FCAT for students with and without EBD?
    a. What is the relationship between AP1 ORF scores on the DIBELS and total reading scores on the 3rd grade FCAT for students with and without EBD?
    b. What is the relationship between AP2 ORF scores on the DIBELS and total reading scores on the 3rd grade FCAT for students with and without EBD?

   c. What is the relationship between AP1 ORF scores on the DIBELS and total reading scores on the 4th grade FCAT for students with and without EBD?
   d. What is the relationship between AP2 ORF scores on the DIBELS and total reading scores on the 4th grade FCAT for students with and without EBD?

3. What is the relationship between AP1 and AP2 ORF scores on the DIBELS and AP2 ORF scores on DIBELS for 3rd grade students with and without EBD?
4. What is the relationship between Maze Percentile Rank scores on the FAIR and the total reading scores on the 4th grade FCAT for students with and without EBD?
   a. What is the relationship between AP1 Maze Percentile Rank scores on the FAIR and the total reading scores on the 4th grade FCAT for students with and without EBD?
   b. What is the relationship between AP2 Maze Percentile Rank scores on the FAIR and the total reading scores on the 4th grade FCAT for students with and without EBD?
5. What is the relationship between AP1and AP2 Maze Percentile Rank scores on the FAIR for 4th grade students with and without EBD?
6. What is the relationship between ORF scores on the DIBELS and Maze Percentile Rank scores on the FAIR for students with and without EBD?

a. What is the relationship between AP1 ORF scores on the DIBELS and AP1 Maze Percentile Rank scores on the FAIR for students with and without EBD?
b. What is the relationship between AP2 ORF scores on the DIBELS and AP2 Maze Percentile Rank scores on the FAIR for students with and without EBD?

7. What is the relationship between total reading scores on the 3rd grade FCAT and total reading scores on the 4th grade FCAT for students with and without EBD?

Design

The study consisted of a correlational design whereby the linear relationships among the maze and oral reading fluency measures (predictor variables) and the Florida Comprehensive Assessment Test (criterion variable) were determined using archival data. Mertens (1998) suggests that this type of methodology is appropriate when exploring the relationships between variables in which there is no experimental manipulation of the treatment variables. Given that some of the variance may be explained by the various demographic variables, procedures will be used to determine if any differences exist across gender, ethnicity, socio-economic status, and disability. In an attempt to control for differential selection, a matched comparison group was used. The comparable group was matched to the sample group by gender, ethnicity, and SES.

Participants

Fifty-nine students classified as EBD and enrolled in grade 3 during the 2008-2009 and grade 4 during the 2009-2010 school-years in the Lee County School District (LCSD)

who participated in this study. These students had documented emotional and behavioral disorders (EBD) as defined in the Individuals with Disabilities Education Improvement Act (U.S. Department of Education, 2004) and were receiving special education services in the district. This population was then matched across ethnicity, socioeconomic status, and gender with students without EBD who were also enrolled in grade 3 during the 2008-2009 and grade 4 during the 2009-2010 school-years in the LCSD, with a total number of 118 participants. As will be discussed in Chapter IV, the sample was reduced through attrition to a final sample size of N=110.

The Lee County School District (LCSD) is a large public school district in Lee County, Florida; it is the ninth-largest school district in Florida and one of the 50th largest school districts in the United States (LCSD, 2009). The district is located on the southwest coast of Florida and educates approximately 82,000 students in grades PK-12. In 2009-2010 approximately 826 of the students were students with EBD (Florida Department of Education, 2009a). The LCSD uses a continuum of services to educate students with disabilities. The student characteristics of this population can be seen below in Table 2.

Forty-three (n=43) of the participants with EBD in this study were students currently enrolled in one of three self-contained classroom models in the district. Of these, forty-one (n=41) were currently placed in a supportive behavior classroom. In the supportive behavior model, students receive instruction in all academic areas and social skills in a highly structured classroom. Faculty and staff utilize a behavior program with the goal of returning the student to the general education setting on a partial or full-time basis (LCSD, 2009). Two (n=2) of the participants were receiving services in an intensive academics class. This class

is designed to provide specialized reading instruction with intensity beyond that which they would receive in an inclusive general education setting. Typically, these students are functioning more than two years below grade level in reading. Ten (*n*=10) of the participants were currently enrolled in a special center school for students with emotional disabilities. This special school focuses on meeting the individual needs of students in the areas of behavior, communication, socialization, and academics with the goal of successfully returning the student to a less restrictive environment within the individual's geographical school as soon as possible (LCSD, 2009). Six participants (*n*=6) were mainstreamed and receiving services in the general education setting through a consultative model. A summary of student characteristics can be seen in Table 2.

Table 2

Student Characteristics of EBD Population in LCSD

Characteristic	Male (*n*=52)	Female (*n*=7)	Total (*N*=59)
Ethnicity			
Black	36.5	42.9	37.3
Hispanic	09.6	28.6	11.9
Mixed	09.6	00.0	08.5
White	44.2	28.6	42.4
SES			
Free/reduced	80.8	100.0	83.1
Non free/reduced	19.2	00.0	16.9
Service Delivery Model			
Self-contained	75.0	57.1	72.9
Center school	13.5	42.8	16.9
Consultative	11.5	00.0	10.2

Ethical Considerations

Prior to implementation, permission to conduct this study was obtained from the Accountability, Research, and Continuous Improvement Department (ARC) of the School District of Lee County where the study took place and the author is currently employed. Additionally, permission was obtained from the Institutional Review Board (IRB) of Barry University where the author is pursuing a PhD degree. The study involved the collection and analyses of existing archival data that are available via public records request. In order to maintain confidentiality, all student demographic and assessment data were collected by an administrative designee of the School District of Lee County. An identification number was assigned to each student and data were recorded on a student data sheet (Appendix C) by the designee to protect the identity of students. Completed student data sheets were then given to the researcher for analysis. At no time did the researcher have direct access to student names or have the ability to identify any individual students. Throughout the study, all student data sheets were secured in a locked cabinet and accessible to the researcher. Upon completion of the study, all student data sheets were secured in a locked cabinet and will be kept for a period of five years. Data were reported in aggregate and at no time were individual student data reported. Upon completion of the study, all findings were shared with the Accountability, Research, and Continuous Improvement Department of the School District of Lee County.

Data collection procedures

In the state of Florida, DIBELS had been widely used to progress monitor the reading development of students for the last several years. Then in 2009-2010, the state introduced

the FAIR which uses a maze component to monitor the reading development of students. As a result, there are two different data sets from the different assessments administered in 2008-2009 and 2009-2010 school years. In conducting this study, the researcher adhered to the following procedures: (1) as mentioned previously, consent was obtained from the School District of Lee County and Institutional Review Board of Barry University to conduct the study; (2) DIBELS, FAIR, and FCAT assessment scores of the participants were acquired from the Accountability, Research, and Continuous Improvement Department of LCSD for the following:

a. ORF scores from 2008-2009
b. FCAT Reading scores from 2008-2009
c. FAIR scores from 2009-2010
d. FCAT reading scores from 2009-2010

(3) an identification number was assigned to each participant to protect the identity of the participants; (4) throughout the study, all data and associated records were secured in a locked cabinet that was only accessible to the researcher; (5) all student data sheets were secured and will be kept for a period of five years.

Instrumentation

This study reported the scores from three different assessments: DIBELS ORF, FAIR maze task, and FCAT Reading. These assessments are common assessments used by the state of Florida and the Lee County School District for the purposes of monitoring and measuring student progress. Reliability and validity data have been discussed previously and are provided by the state. These assessments have been approved for state and district-wide use

and are used regularly for instructional decision-making purposes in both the general and special education settings.

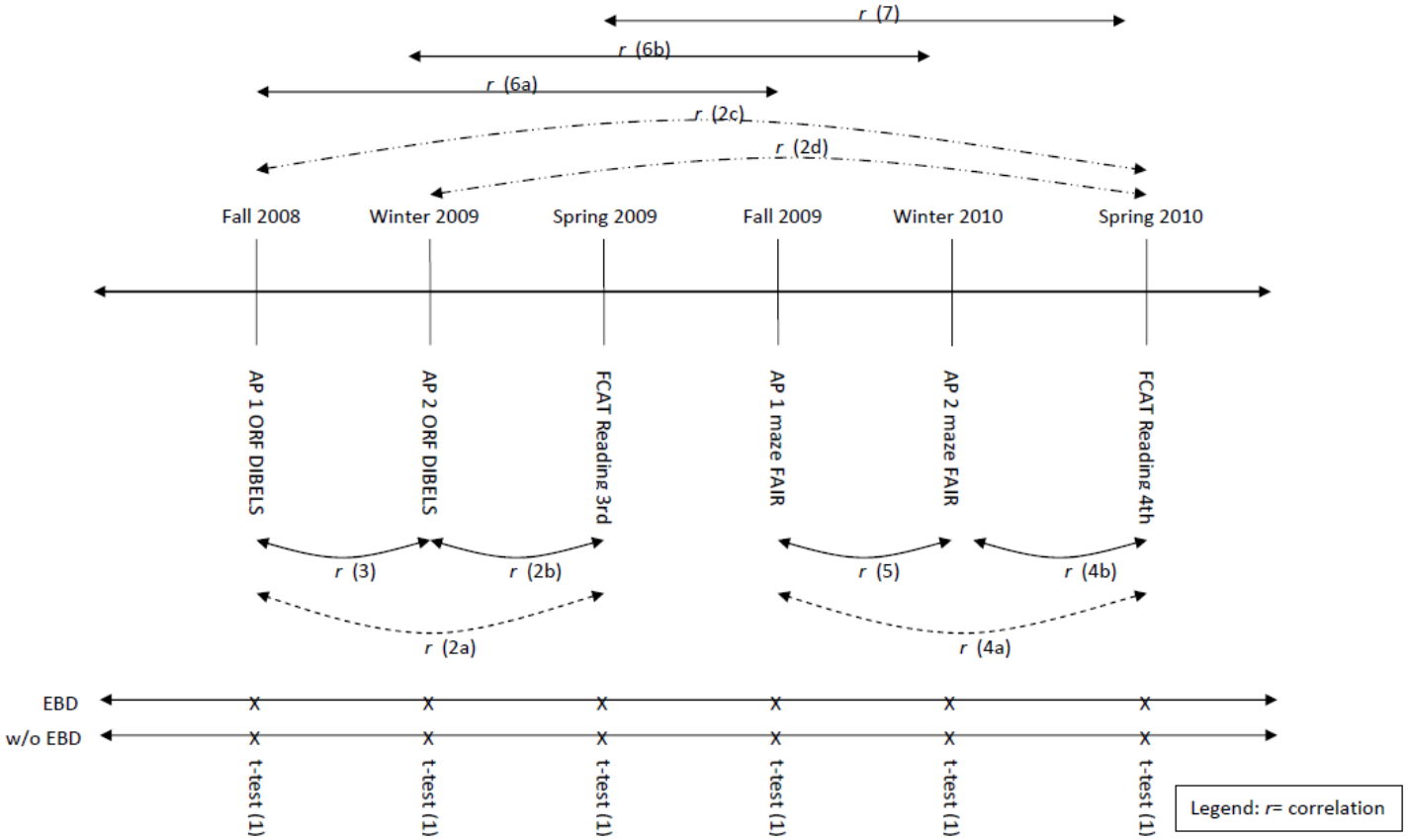
Data analysis

For this quantitative study, descriptive statistics were used to describe the groups; and tables, figures, and graphs were used to display the data. Pearson correlation coefficients were used to examine the relationships among archival data of common assessments used in the state of Florida. These scores included: the maze score of the Florida Assessments for Instruction in Reading [FAIR] (Florida Department of Education, 2009), the oral reading fluency (ORF) score of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS, Good & Kaminski, 2002), and the developmental scale score in reading on the Florida Comprehensive Assessment Test. Additionally, multiple regression analyses were performed to control for demographic variables (disability, gender, ethnicity, and socioeconomic status) identified in the research questions. The Statistical Package for the Social Sciences (SPSS) 18 was used to analyze the data and prepare charts for the purpose of data reporting. A summary of the analyses completed as related to the research questions investigated can be seen in Figure 4.

Summary

Chapter III described the research methodology used in the study and provided a description of the research questions, study design, and participants. Additionally, ethical considerations, data collection procedures, instrumentation, and data analysis procedures were addressed. Chapter IV addresses the results of the study and Chapter V explains the findings, limitations, and recommendations for further study.

Figure 4 *Statistical Analyses*



Note: Statistical analyses model showing the assessment variables and analyses to be performed. The corresponding research questions investigated are shown in parenthesis.

CHAPTER IV

Results

In this section, a description of the analyses used and the results of the analyses are presented. These results provide the foundation for the discussion in Chapter V.

Sample attrition and missing scores

After completion of data collection, the first part of the data analysis involved the investigation for completeness of scores. Multiple students from the sample population had incomplete sets of assessment data across the two year period. Four students with EBD, however, were missing both third and fourth grade FCAT scores so these students were excluded entirely from the analysis. As a result, four students with similar demographics from the matched group (students without EBD) were also excluded. This left a sample of 110 students (N = 110) for analysis. It should be noted that other students had missing scores or scores of zero which greatly reduced the sample size for the various analyses which may have impacted results. This will be discussed further in the limitations section of Chapter V. Generally, the maze assessments were most problematic with as many as 17 students with EBD and 12 students without EBD having missing or incomplete maze scores. For the purposes of data analysis, these missing assessment scores and scores of zero were assumed invalid and were coded 666.

Several data categories were coded for correlation and regression analysis as follows: gender was coded as: males = 1, and females = 2; ethnicity as: White = 1, Black = 2, Hispanic = 3, and mixed = 4; disability as: students with EBD = 1, and students without EBD = 2; and socioeconomic status as: non-free and reduced lunch = 1, free and reduced lunch = 2.

Descriptive statistics

Student characteristics of the resultant sample (N = 110) can be seen below in Table 3. The table offers a comparison of the experimental and matched groups. Interestingly enough, while the matched group was chosen by demographic characteristics (ethnicity, gender, SES) the groups are very similar in age with the mean age of students with EBD only slightly higher than the mean age of those students without EBD.

Table 3

Student Characteristics as a Percentage of the Sample

Characteristic	w/ EBD (n=55)	w/o EBD (n=55)	Total (N=110)
Gender			
Male	89.1	89.1	89.1
Female	10.9	10.9	10.9
Ethnicity			
Black	38.2	36.4	37.3
Hispanic	12.7	12.7	12.7
Mixed	9.1	10.9	10.0
White	40.0	40.0	40.0
SES			
Free/reduced	85.5	81.8	83.6
Non free/reduced	14.5	18.2	16.4
Age			
minimum	8.67	8.58	8.58
maximum	11.42	11.00	11.42
mean	9.72	9.61	9.66

Comparisons between the mean scores for students with and without EBD for all assessments (DIBELS ORF, FAIR maze, and FCAT Reading) were then completed. Descriptive statistics can be seen below in Table 4. Students with EBD scored significantly lower than the students without EBD on all assessments. Students with EBD read an average of 30 words per minute less than the students without EBD. Similarly, they also scored an average of 13.3 less on the maze and 43.3 points less on the FCAT than the students without EBD. These are similar to results shown in the literature. However, as discussed earlier it should be noted that the sample size for each analysis varied due to the incomplete and/or missing assessment scores within the data.

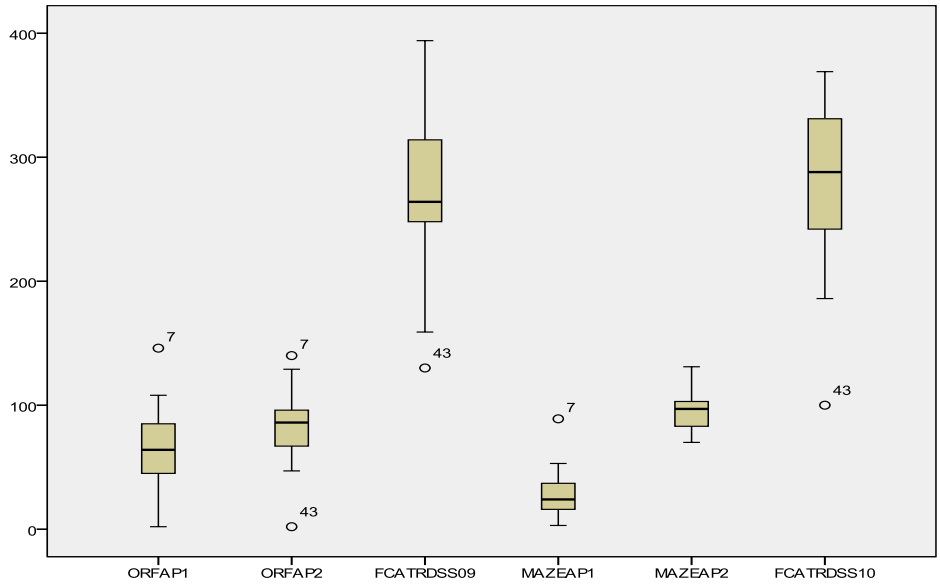
Box plots were then constructed to check for outliers and examine the distribution of assessment scores. As expected, scores for students with EBD (Figure 5) were less evenly distributed than the assessment scores of students without EBD (Figure 6). A few outliers were noted; however, given the already reduced sample size, these scores were included for further analysis.

Table 4

Comparison of Means of Assessments for Students With and Without EBD

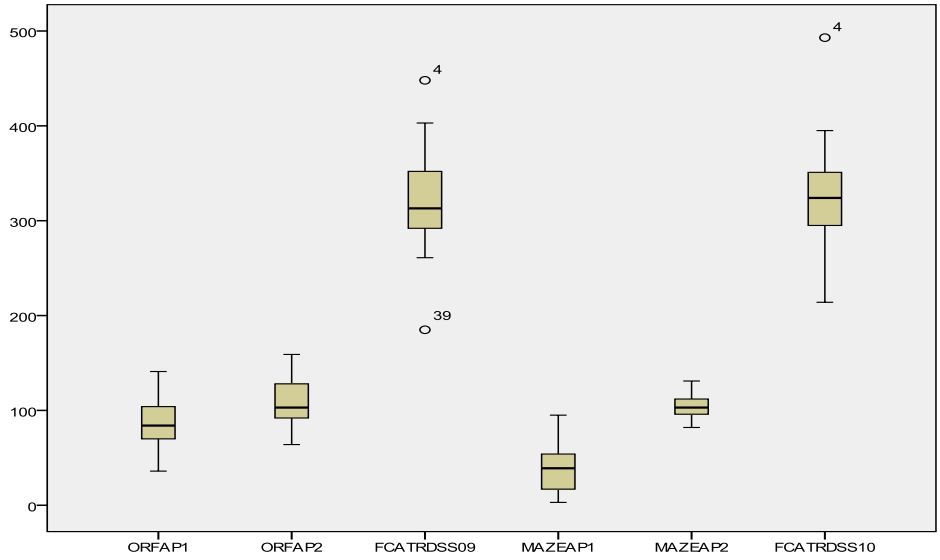
	n	Mean	S.D.
ORF AP1			
with EBD	55	59.70	30.823
without EBD	53	86.05	25.452
ORF AP2			
with EBD	53	75.89	32.963
without EBD	53	106.29	23.128
3rd Grade FCAT RD SS			
with EBD	55	271.02	60.876
without EBD	53	315.35	44.288
Maze AP1			
with EBD	38	21.66	18.387
without EBD	43	36.58	23.992
Maze AP2			
with EBD	38	93.26	13.640
without EBD	46	104.83	11.885
4th Grade FCAT RD SS			
with EBD	55	275.26	65.072
without EBD	50	317.47	49.320

Figure 5 *Boxplot of Assessment Scores for Students With EBD*



Note: Boxplot showing the distribution of assessment scores for students with EBD.

Figure 6 *Boxplot of Assessment Scores for Students Without EBD*



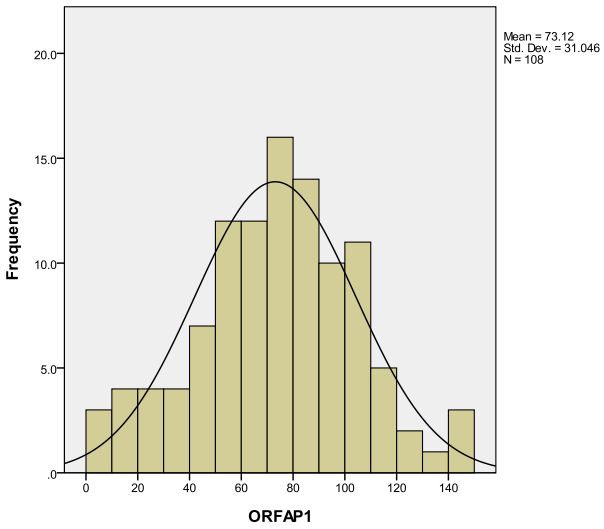
Note: Boxplot showing the distribution of assessment scores for students without EBD.

Assumptions for regression analysis

Regression analysis assumes that variables are normally distributed. To determine this, histograms were produced to visually determine if the sample was normally distributed for ORF AP1, ORF AP2, 3rd Grade FCAT, Maze AP1, Maze AP2, and 4th Grade FCAT scores. Although the study was not a randomized study, the histograms show a fairly normal distribution for the ORF assessments and the maze assessment scores. The FCAT assessments scores also show a fairly normal distribution among the sample. These can be seen in Figures 7-12 below.

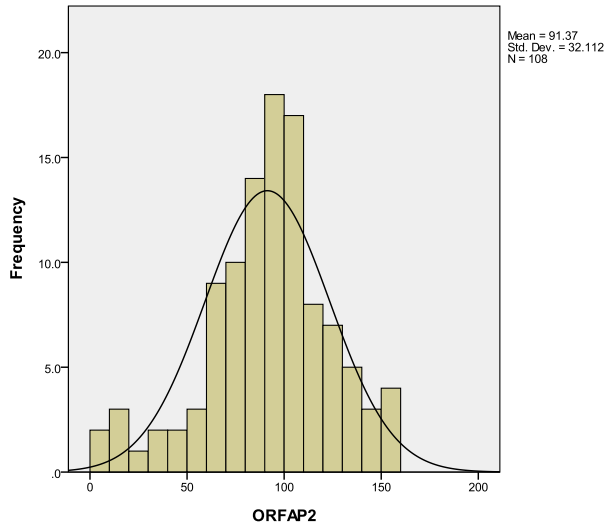
To check further for a normal distribution, all assessment scores were then graphed using a probability-probability plot to once again check for a normal distribution of scores. Again, while the distribution for all of the assessments is not perfect, the scores appear to generally follow the ideal diagonal line. These can be seen in Appendix D.

Figure 7 *Oral Reading Fluency Distribution Curve (AP1)*



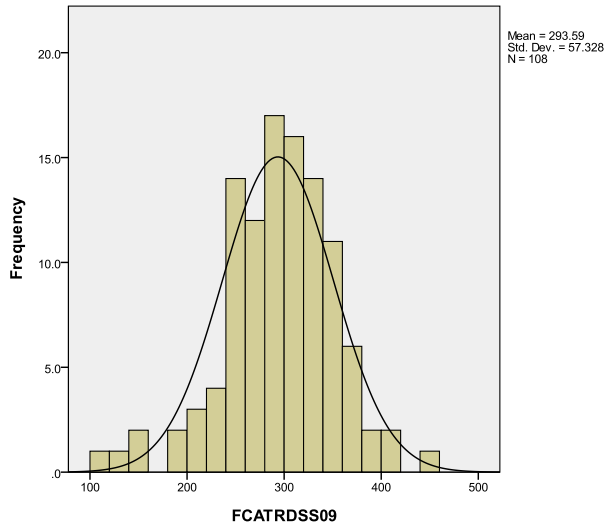
Note: Histogram with distribution curve showing frequency of oral reading fluency scores from the fall administration of DIBELS.

Figure 8 *Oral Reading Fluency Distribution Curve (AP2)*



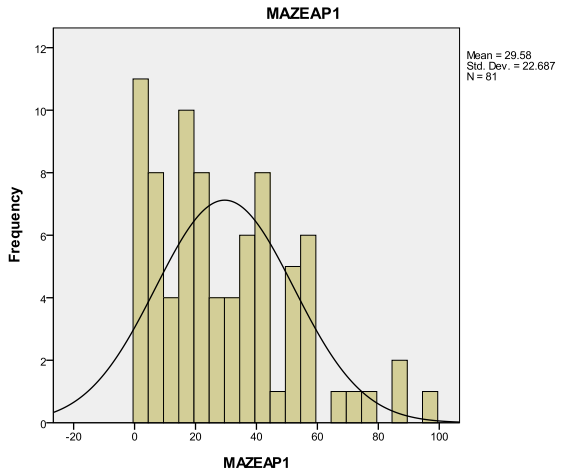
Note: Histogram with distribution curve showing frequency of oral reading fluency scores from the spring administration of DIBELS.

Figure 9 *3rd Grade FCAT in Reading Distribution Curve*



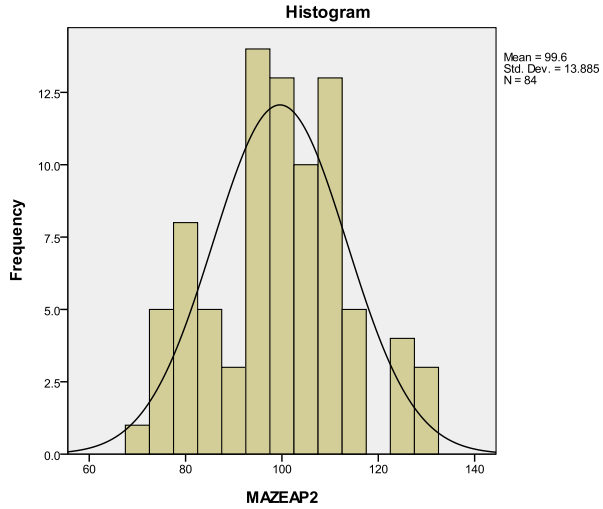
Note: Histogram with distribution curve showing frequency of 3rd grade FCAT reading scores.

Figure 10 *Maze Distribution Curve (AP1)*



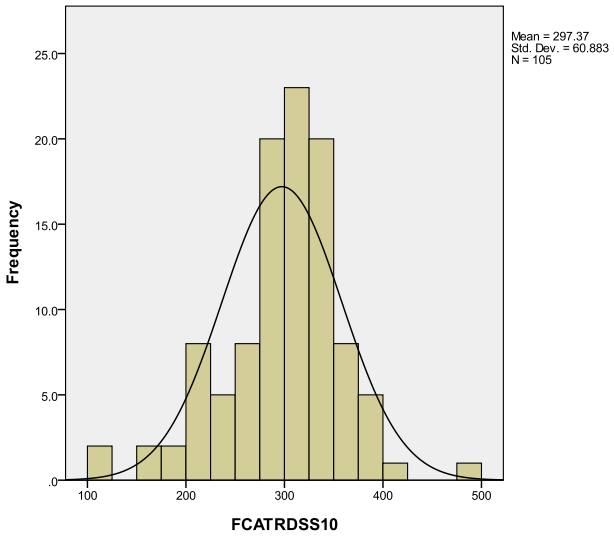
Note: Histogram with distribution curve showing frequency of maze task scores from the fall administration of FAIR.

Figure 11 *Maze Distribution Curve (AP2)*



Note: Histogram with distribution curve showing frequency of maze task scores from the spring administration of FAIR.

Figure 12 *4th Grade FCAT in Reading Distribution Curve*



Note: Histogram with distribution curve showing frequency of 4th grade FCAT reading scores.

Sample size also plays an important role in obtaining a reliable regression model as smaller samples are more susceptible to bias. Field (2009) suggests that in order to find a large effect then a sample of size of *N*=80 should be sufficient for most instances, while Maas and Hox (2004) suggest that a minimum of 50 are needed to obtain correct estimates of the standard errors. With the attrition of scores as discussed earlier, the overall sample size for this study of *N*=110 and all sub-samples (*n*= 108, *n*= 81, *n*=84, *n*= 105) for the separate analyses meet this criteria.

Independent *t* tests

An independent-samples *t* test was calculated comparing the mean scores for students with EBD to the mean scores for students without EBD on the ORF, maze, 3rd grade FCAT, and 4th grade FCAT assessments. A summary of the results can be seen in Table 5.

Table 5

Independent Samples t-tests for Assessments for Students With and Without EBD

	n	*t*	Sig. (2-tailed)
ORF AP1	108	-4.853	.000
ORF AP2	108	-5.565	.000
3rd Grade FCAT RD SS	108	-4.339	.000
Maze AP1	81	-3.110	.003
Maze AP2	84	-4.154	.000
4th Grade FCAT RD SS	105	-3.767	.000

Note $n = 55$

An independent-samples *t* test comparing the mean AP1 ORF scores for students with EBD to the mean AP1 ORF scores for students without EBD on the DIBELS assessment found a significant difference between the means of the two groups (t (106) = -4.853, p = .000). The mean ORF score for students with EBD was significantly lower (M =59.7, SD = 30.823) than the mean ORF score for students without EBD (M = 86.05, SD = 25.452). This was also true when comparing mean AP2 ORF scores. There was a significant difference between the means of the two groups (t (106) = -5.565, p =.000). The mean ORF score for students with EBD was significantly lower (M =75.89, SD = 32.963) than the mean ORF score for students without EBD (M = 106.29, SD = 23.128).

Next, an independent-samples *t* test was calculated to compare the mean scale score for students with EBD to the mean scale score for students without EBD on the reading portion of the third grade FCAT and found a significant difference between the two groups (*t* (106) = -4.339, *p* = .000). The mean scale score for students with EBD (*M* =271.02, *SD* = 60.876) was lower than the mean scale score for students without EBD (*M* = 315.35, *SD* = 44.288).

Independent-samples *t* tests were also calculated to compare the mean maze scores for students with EBD to the mean maze scores for students without EBD on the FAIR assessment. For the AP1 FAIR, the mean maze score for students with EBD was 21.66 (*SD* = 18.387), and the mean maze score for students without EBD was 36.58 (*SD* = 23.992). Statistically a significant difference between students with and without EBD was not found (*t* (79) = -3.110, *p* =.003). For the AP2 FAIR, the mean maze score for students with EBD was 93.26 (*SD* = 13.640), and the mean maze score for students without EBD was 104.83 (*SD* = 11.885). Again, a significant difference between the groups was found (*t* (82) = -4.154, *p* = .000).

Again, an independent-samples *t* tests comparing the mean scale scores for students with EBD to the mean scale scores for students without EBD on the reading portion of the fourth grade FCAT found a significant difference between the two groups (*t* (103) = -3.767, *p* = .000). The mean scale score for students with EBD (*M* =275.26, *SD* = 65.072) was lower than the mean scale score for students without EBD (*M* = 317.47, *SD* = 49.320).

Correlations

The next analysis involved an investigation into the linear relationship between the student characteristics of ethnicity, SES and gender for the resulting sample (*N*=108) and

scaled score of the reading portion of the third grade FCAT using Pearson correlation coefficients. These can be seen below in Table 6. The student characteristics of gender and SES did not show a correlation with student achievement on the reading portion of the 3rd grade FCAT. Interestingly however, ethnicity showed a weak correlation ($r = .221$, $p < .05$) with achievement, whereas disability demonstrated a stronger correlation ($r = .388$, $p < .01$). It is not surprising that results would suggest a relationship between achievement and disability as this is well documented in the literature. However, the weaker correlation between ethnicity and achievement may suggest that there may be some cultural bias inherent in the assessment.

Table 6

Correlations Among Student Characteristics and 3rd Grade FCAT Reading

	Disability	Ethnicity	Gender	SES
Mean 3rd Grade FCAT Reading Score	.388**	-.221*	.037	-.129
Disability	1	.019	.000	-.049
Ethnicity		1	.118	.350**
Gender			1	.155
SES				1

Note N=108 *$p < .05$ **$p < .01$

Another analysis investigated the linear relationship between the student characteristics of disability, race, gender, and SES for the resulting sample (n=105) and the scaled score of

the reading portion of the fourth grade FCAT using Pearson correlation coefficients. These can be seen below in Table 7. Once again, the student characteristics of gender and SES did not show a correlation with student achievement on the reading portion of the 4th grade FCAT. Interestingly however, ethnicity showed a weak correlation ($r = .201$, $p < .05$) with achievement, whereas disability demonstrated a stronger correlation ($r = .348$, $p < .01$). Again it is not surprising that results would suggest a relationship between achievement and disability; however, again the weaker correlation between ethnicity and achievement may suggest that there may be some cultural bias in the assessment.

Table 7

Correlations Among Student Characteristics and 4th Grade FCAT Reading

	Disability	Ethnicity	Gender	SES
Mean 4th Grade FCAT Reading Score	.348**	-.171	-.065	-.201*
Disability	1	.019	.000	-.049
Ethnicity		1	.118	.350**
Gender			1	.155
SES				1

Note N=105 *$p < .05$ **$p < .01$

Pearson correlation coefficients were then computed among all of the assessments (3rd Grade FCAT, ORF AP1, ORF AP2, 4th Grade FCAT, Maze AP1, and Maze AP2) first for students with EBD and then for students without EBD. Strong positive correlations were found

among all of the assessments for students with EBD and ranged in value from, $r = .447$, $p < .01$ to $r = .920$, $p < .01$. While slight, there generally was a stronger correlation between the ORF and FCAT scores than between the Maze and FCAT scores for students with EBD. These are summarized below in Table 8.

Table 8

Correlations Among Assessments for Students With EBD

	3rd Grade FCAT Reading Score	ORF AP1	ORF AP2	Maze AP1	Maze AP2
Mean 4th Grade FCAT Reading Score	.628**	.640**	.666**	.632**	.540**
Mean 3rd Grade FCAT Reading Score	1	.661**	.695**	.513**	.447**
ORF AP1		1	.920**	.729**	.831**
ORF AP2			1	.719**	.702**
Maze AP1				1	.815**
Maze AP2					1

Note $n = 55$ **$p < .01$

For students without EBD, again strong positive correlations were found among all of the assessments. Values ranged from $r = .512$, $p < .01$ to $r = .868$, $p < .01$. Generally there was a stronger correlation between the ORF and FCAT scores than between the Maze and FCAT scores for students without EBD. Results are summarized below in Table 9.

Table 9

Correlations Among Assessments for Students Without EBD

	3rd Grade FCAT Reading Score	ORF AP1	ORF AP2	Maze AP1	Maze AP2
Mean 4th Grade FCAT Reading Score	.791**	.596**	.658**	.512**	.550**
Mean 3rd Grade FCAT Reading Score	1	.549**	.575**	.617**	.553**
ORF AP1		1	.868**	.664**	.818**
ORF AP2			1	.585**	.761**
Maze AP1				1	.822**
Maze AP2					1

Note n= 55 **$p < .01$

Regression analyses

Regression analyses were completed in an attempt to predict success for students with and without EBD on the reading portion of the 3rd and 4th grade FCAT using ORF and maze assessment scores while controlling for disability, gender, ethnicity, and socio-economic status. In the literature, these variables have been shown to correlate to assessment scores.

The first analyses performed were simple linear regression analyses. These can be seen below in Table 10. For these analyses, forced entry was used to enter the predictors into the model simultaneously so as to reduce the influences of random variation in the data (Studenmund & Cassidy, 1987 as cited in Field, 2009).

Table 10

Regression Analysis for Students With and Without EBD

	3rd Grade FCAT Reading Score (R^2)	4th Grade FCAT Reading Score (R^2)
ORF AP1/AP2		
with EBD	.498***	.467***
without EBD	.341***	.436***
ORF AP1/AP2/FCAT 3		
with EBD		.526***
without EBD		.687***
Maze AP1/AP2		
with EBD		.471***
without EBD		.323***

***$p < .001$

A linear regression was calculated to predict 3rd grade FCAT scores based on the ORF AP1 and ORF AP2 scores for students with and without EBD. For students with EBD, a significant regression equation was found (F (2, 47) = 23.302, $p < .001$), where $R^2 = .498$. For students without EBD, a significant regression equation was also found (F (2, 52) = 13.437, $p < .001$), where $R^2 = .341$. While both results are significant, they seem to indicate that ORF fluency may be a better predictor for students with EBD than for students without EBD.

A linear regression was also calculated to predict 4th grade FCAT scores based on the ORF AP1 and ORF AP2 scores for students with and without EBD. For students with EBD, a significant regression equation was found (F (2, 47) = 19.706, $p < .001$), where $R^2 = .467$. For students without EBD, a significant regression equation was found (F (2, 52) = 20.071, $p < .001$), where $R^2 = .436$. Once again, although slight, ORF fluency may be a better predictor for students with EBD than for students without EBD.

Adding 3rd grade FCAT scores to ORF AP1 and ORF AP2 improved the model for students with and without EBD. However, there appeared to be a greater improvement for students without EBD. This is not surprising given the difficulty many students with EBD have completing standardized assessments. For students with EBD, a significant regression equation was found (F (3, 42) = 15.564, $p < .001$), where $R^2 = .526$. For students without EBD, a significant regression equation was also found (F (3, 51) = 37.303, $p < .001$), where $R^2 = .687$.

A linear regression was also calculated to predict 4th grade FCAT scores based on the Maze AP1 and Maze AP2 scores for students with and without EBD. For students with EBD, a significant regression equation was found (F (2, 23) = 10.244, $p < .001$), where $R^2 = .471$. For students without EBD, a significant regression equation was found (F (2, 34) = 8.105, $p < .001$), where $R^2 = .323$. Surprisingly, the maze appears to be a better predictor for students with EBD than for students without EBD.

Lastly, a multiple linear regression model was created to predict 4th grade FCAT scores based on ORF scores, 3rd grade FCAT scores, and maze scores for students with and without EBD while controlling for disability, ethnicity, socio-economic status and gender. Stepwise

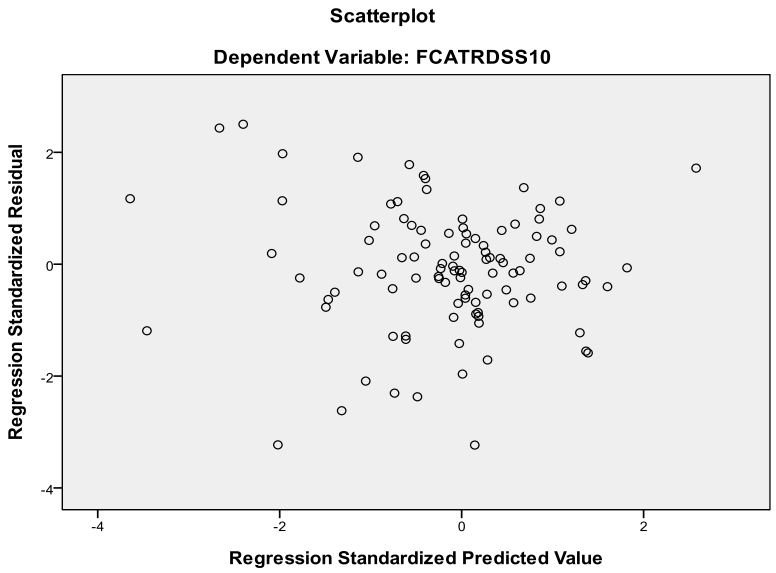
entry was used to determine which variables explained the most variance. Predictor variables were entered into the model using the mathematical criterion of .05 for entry and .10 for removal for the probability of *F*. This allowed the computer to only retain those variables that improved the predictive power of the model and excluded those that did not.

Disability, ORF AP2, and 3rd grade FCAT scores were found to be significant predictors of success on the 4th grade FCAT for students with and without disabilities. A significant regression was found (F (3, 58) = 40.821, $p < .001$), where $R^2 = .679$. Gender, ethnicity, socio-economic status, and maze scores were all excluded from the model as they did not add any significant contribution to the model. The regression analyses results support the literature that suggests that ORF may be more sensitive to reading growth and a better predictor of success on standardized assessments than maze measures for students with and without disabilities (Ardoin et al., 2004; Jenkins & Jewell, 1993).

To check the assumptions of the model, a scatter plot of standardized residual against the standardized predicted value was created. As can be seen in Figure 13 below, there is a random array of points, fairly evenly dispersed around zero. This suggests that the assumptions of linearity and homoscedasticity have been met (Osborne & Waters, 2002).

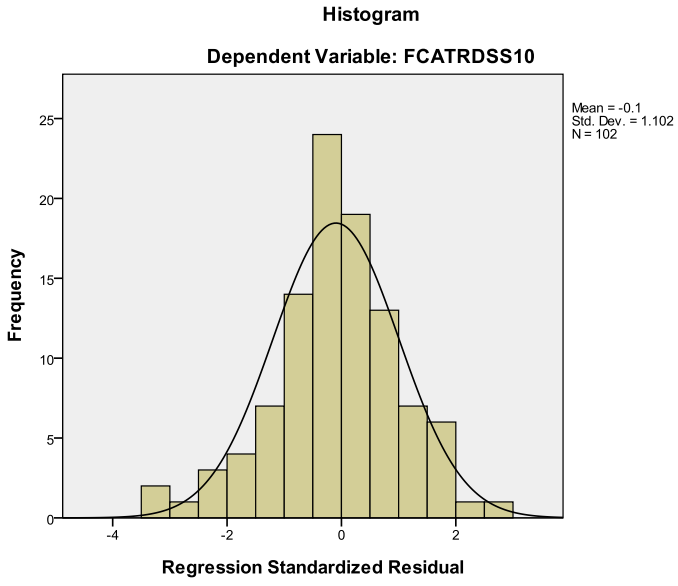
A histogram and normal probability plot were created to check the normality of residuals. As can be seen below in Figures 14 and 15, while not perfect, both show a fairly typical distribution suggesting that the model appears to be accurate for the study sample and therefore could be generalized to the population (Field, 2009).

Figure 13 *Plot of Standardized Residual Against Standardized Predicted Value*



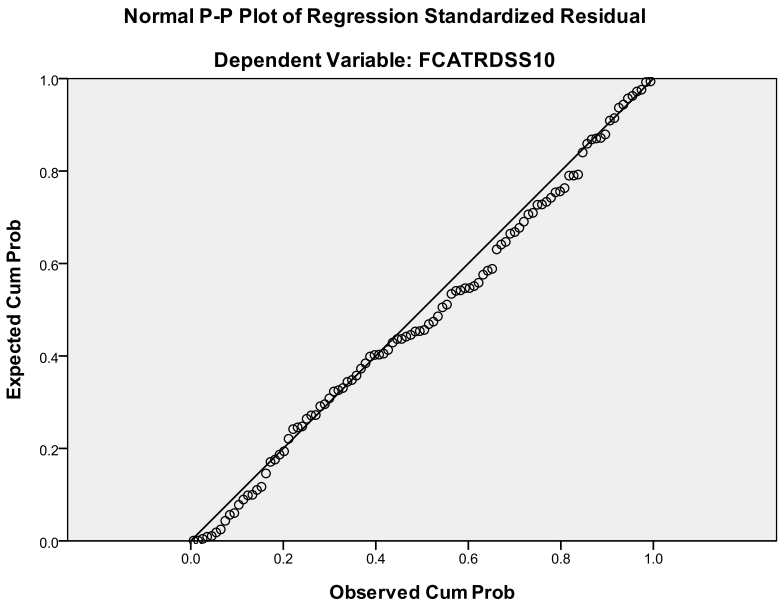
Note: Scatter plot showing standardized residuals graphed against standardized predicted value

Figure 14 *Regression Standardized Residual Distribution Curve*



Note: Histogram with distribution curve showing frequency of the standardized residuals

Figure 15 *Normal P-P Plot of Regression Standardized Residual*



Note: Probability-probability plot showing the cumulative probability of standardized residuals plotted against the expected normal distribution of scores.

Chapter IV presented the results of the study. Descriptive statistics were included as well as the results of data analyses conducted. Chapter V will explain the findings, limitations, and recommendations for further study.

CHAPTER V

Discussion

Summary of findings

The purpose of this study was to extend the research on the use of the ORF and maze measures with students with and without EBD by investigating the relationship of the Dynamic Indicators of Basic Early Literacy Skills (DIBELS, Good & Kaminski, 2002) and the Florida Assessments for Instruction in Reading (FAIR) with the FCAT. Of particular interest was the attempt to determine which curriculum based measure is a better assessment for predicting success on the FCAT for students with and without EBD. The DIBELS uses timed oral reading passages to measure fluency, whereas the FAIR utilizes a maze passage to measure a student's reading growth. The use of these assessments is a statement or critical assumption about the effectiveness of curriculum-based measurements for monitoring the progress of students with and without EBD and at predicting their success on the FCAT. The determination of this assumption is extremely beneficial for teachers of students with and without EBD for the purposes of monitoring the progress of their students and determining the amount of time necessary for instruction to ensure success on high stakes assessments.

The results of the study suggest that there is a significant relationship between oral reading fluency measures and the FCAT and between maze measures and the FCAT for students with and without EBD; however, the levels of these relationships varied. Additionally, the results seem to indicate that both maze and oral reading fluency measures can be used with some accuracy with both groups of students to predict success on the FCAT. There are several implications that merit further discussion.

Implications and summary of research questions

In 1963, John Carroll proposed a model that defined school learning as a function of time. Further, he characterized the learning rate for students as a function of the opportunity and perseverance to learn (or time spent) and a student's aptitude, ability to understand instruction, and the quality of instruction (time needed). While there is no definitive way to measure aptitude, Carroll felt that it could be estimated by measures of past performance (Carroll, 1963). Understanding the role that the assessment of student performance has on learning and the subsequent monitoring of student learning using curriculum based measures are therefore critical to improving student learning. Additionally, when a teacher understands the amount of time needed for a student to learn, then he/she is able to manipulate the classroom environment to increase the time allocated for learning and improve the amount of time a student spends academically engaged. As Carroll suggests, aptitude determines the amount of time needed which can be useful to educators when defining the present levels of their students in order to effectively guide instructional decision making. This is especially true for students with EBD given the academic challenges they face. It is equally important that a teacher understand how a student is progressing. When coupled with the accurate determination of a student's present levels, progress monitoring can assist in effectively establishing the amount of time that a student needs to be successful within the curriculum.

A review of the literature suggested that students with EBD generally lag behind their general education peers in the area of reading (i.e., Barton-Arwood et al., 2005; Forness, Bennett, & Tose, 1983; Kauffman, 2005). This study investigated how the mean

scores for students with EBD compared to the mean scores for students without EBD on the ORF, maze, 3rd grade FCAT, and 4th grade FCAT assessments. This study found similar results as independent samples *t* tests indicated that for all assessments (oral reading fluency, maze, and FCAT), students with EBD scored lower than those students without EBD (see Table 5). This is not surprising given the number of studies that have demonstrated that students with EBD are at-risk for academic failure, especially in the area of reading. What a review of the literature did not reveal, however, was the use of curriculum-based assessments like oral reading fluency and maze tasks specifically for students with EBD. Herein lies the significance of what this study adds to the current body of literature; most notably that CBM measures, like ORF and maze measures, can be used successfully for students with and without EBD.

This study also investigated the relationship between ORF scores on the DIBELS and total reading scores on the third and fourth grade FCAT for students with and without EBD. Results indicated that there is a significant correlation between oral reading fluency measures and the FCAT (ranging from $r = .661$ to $r = .695$) for students with EBD. For students without EBD, a significant correlation between oral reading fluency measures and the FCAT (ranging from $r = .549$ to $r = .658$) was also found. These correlations are significantly lower than those found by Fuchs, Fuchs, & Maxwell in 1988 (ranging from $r = .80$ to $r = .91$) and Hosp and Fuchs in 2005 (ranging from $r = .82$ to $r = .88$), but similar to those found by Buck and Torgeson (2003) who found a significant correlation between ORF scores and the reading portion of the FCAT ($r = .70$) for third graders. Again, all of

these studies included students with disabilities; however, none of the aforementioned studies were specific to students with EBD.

Another interesting finding was the existence of a higher correlation between ORF (AP1, AP2) and maze (AP1) and the FCAT for students with EBD than for students without EBD (See Tables 8 and 9.). Due to the difficulty many students with EBD have demonstrated in the educational environment coupled with their poor academic performance (Epstein & Cullinan, 1983; Lane, Little, Redding-Rhodes, Phillips, & Welsh, 2007; Nelson, Benner, Lane, & Smith, 2004), this is somewhat surprising. Given that this study used archival data, there is no way of knowing what the assessment environments looked like. However, it may be that the assessment environment included accommodations that were used with the students with EBD. These accommodations may have impacted the results.

Interestingly, only a slight difference exists between the first administration (r = .549 to r = .661) and the second administration (r = .575 to r =.695) of ORF when correlated with the FCAT for both students with and without EBD. This suggests that ORF is a fairly stable measure over time and a good predictor of success, which mirrors the findings of Wood (2006) who determined that ORF was a better predictor of success on the statewide reading test than the previous year's performance on the same test. All of this suggests that regardless of past performance, early on in the year, teachers can have access to the information necessary to predict with some accuracy how well all of their students will perform on the FCAT regardless of the student's disability. Given that Carroll proposed that learning is a function of time and that the aptitude (past performance) of a

student determines the amount of time needed for the student to be successful, it stands to reason that the earlier a teacher can make this determination, the more likely the student will get the time they need and therefore experience educational success. While this is important for all at-risk students, this is essential for students with EBD.

Next, the study sought to investigate the relationship between Maze Percentile Rank scores on the FAIR and the total reading scores on the 4th grade FCAT for students with and without EBD. Results show a significant correlation between the maze tasks and the FCAT ranging from r = .447 to r = .632 for students with EBD and from r = .512 to r = .617 for those students without EBD. Again these are lower than results achieved in previous studies; Fuchs and Fuchs (1990) (as cited in Wayman, et al., 2007) found a correlation of .77 between a maze task and the Stanford Achievement Test, while Jenkins and Jewell (1993) showed correlations between a maze task and the Metropolitan Achievement Test ranging from .65 to .75. This may be due to differences in reliability and validity between the more traditional and established assessments (SAT and MAT) and the newer FCAT. In addition, as discussed previously, the published reliability and validity coefficients of the FAIR are not as high as some of the more established instruments used in the field of education. It may also be important to note that neither study was specific to students with EBD. The results of this study may suggest then that CBM can be as useful for students with EBD as it is with their general education peers. However, as was discussed earlier, given the slightly higher correlations between the ORF and the FCAT than the correlations between the maze and the FCAT, teachers may choose to use ORF rather than maze measures as it may be a better tool for students with and without EBD.

The results of investigation into the relationship between ORF scores on the DIBELS and Maze Percentile Rank scores on the FAIR for students with and without EBD indicate that there is a significant correlation between the oral reading fluency scores (AP2) in 3rd grade and the maze scores (AP2) in 4th grade for students with EBD ($r = .702$, $p < .01$) and without EBD ($r = .761$, $p > .01$). A significant correlation also exists between the oral reading fluency scores (AP1) in 3rd grade and the maze scores (AP1) in 4th grade for students with EBD ($r = .729$, $p < .01$) and for students without EBD ($r = .664$, $p < .01$). This suggests that both types of CBM are useful for monitoring the progress of students with and without EBD. These results mirror those found in the literature (i.e., Fuchs, Fuchs, & Maxwell, 1988; Hintze, Owen, Shapiro, & Daly, 2000; Jenkins, Fuchs, van den Broek, Espin, & Deno, 2003; Shinn et al., 1992) that suggest that CBM in general is a valid way of monitoring student progress. So while students with EBD generally may score lower than their general education peers on assessments, both groups of students are alike and respond equally to strategies and tools designed to meet the needs of diverse learning groups (Nebraska Department of Education, 1996).

In trying to answer which is a better predictor of future success on standardized assessments, linear regression analyses were completed. Results suggest that theORF may be a better predictor for students with and without EBD, with an R^2 value of .498 for students with EBD and .341 without. For the maze measures, the R^2 values ranged from .471 for students with EBD to .323 for those without. Interestingly, ORF scores from the previous year ($R^2 = .436$) were a better predictor for students without EBD than the maze

measures (R^2 = .323). For students with EBD, they were very similar (R^2 = .467, R^2 = .471).

Another interesting finding was that while the addition of the FCAT score from the previous year to the regression model did improve the model's ability to predict success on the FCAT for both students with and without EBD, for students with EBD, it was improved to a lesser extent. ORF scores alone (R^2 = .467) accounted for almost as much variability as ORF scores and the 3rd grade FCAT (R^2 = .526). However, there was significant improvement in the regression model for students without EBD, whereby ORF scores alone (R^2 = .436) accounted for less variability as ORF scores and the 3rd grade FCAT (R^2 = .687). Once again, this may suggest that CBM are a more useful assessment for use with students with EBD than standardized assessments like the FCAT.

The results indicate that assessments from the previous school year can be as useful as those from the current school year when trying to predict the success of students on a standardized assessment, and that the addition of the maze assessment in 4th grade did not improve the predicted success of those same students on the standardized assessment regardless of disability. This may suggest that CBM measures may become less useful for predicting success as a student gets older. This is similar to the study of Jenkins and Jewell (1993) who compared the technical adequacy of both a read-aloud measure and a maze task and found correlations ranging from .58 to .88 across grades two through six. Results found that correlations between both the read-aloud and maze measures and standardized assessments declined from grades two to six. Guthrie found the maze measure to have good stability and a high correlation with standardized measures of reading proficiency for

students with and without disabilities (Guthrie, 1973; Guthrie et al., 1974), while the results of Ardoin et al. (2004) indicted that the addition of a maze task did not increase the predictability of third grade students' success on a standardized achievement test in reading. Regardless, knowing that assessment scores obtained earlier in the year are almost as good an indicator to teachers as those assessments administered later on in the year is useful to educators and policy makers alike as they plan and revise the current assessment windows.

Lastly, this study examined the relationship between total reading scores on the 3rd grade FCAT and total reading scores on the 4th grade FCAT for students with and without EBD. There is a significant correlation between reading scores on the 3rd grade FCAT and the reading scores on the 4th grade FCAT for students with EBD ($r = .628$, $p < .01$) and for students without EBD ($r = .791$, $p < .01$). These results suggest that a student's previous year score can be a predictor of current achievement on a standardized assessment like the FCAT. However, it is interesting that once again, for students with EBD, there was a lower correlation across assessment years than there was for their general education peers.

Although not a research question, data analysis revealed some other interesting results that may have some implications for teachers, administrators, and policy makers alike in the state of Florida. Correlations were calculated between administrations for oral reading fluency and maze tasks. Oral reading fluency measures showed significant correlations between the fall and spring administrations for students with EBD ($r = .920$, $p < .01$) and without EBD ($r = .868$, $p < .01$). Significant correlations also existed between the spring and fall administrations of the maze task for students with EBD ($r = .815$, $p < .01$) as well as for students without EBD

($r = .822$, $p < .01$). This suggests that ORF may be a slightly more stable tool for students with EBD than the maze assessment. This may be useful to educators and policy makers alike as the state of Florida moves forward with its use and refinement of the Florida Assessments in Reading that include maze tasks.

Correlations were also computed among demographic variables and assessments. Interestingly, ethnicity showed a weak correlation ($r = .221$, $p < .05$) with achievement, whereas disability demonstrated a stronger correlation ($r = .388$, $p < .01$). It is not surprising that results would suggest a relationship between achievement and disability as this is well documented in the literature. However, the weaker correlation between ethnicity and achievement may suggest that there may be some cultural bias inherent in the FCAT assessment. This indicates the need for further investigation into cultural differences that may exist among students that impact student achievement. More importantly however, the results highlight the fact that the creators of high-stakes assessments like the FCAT and policy makers who dictate the use of such assessments, must be aware of the existence of cultural differences when using these tools to make important educational decisions for students.

Given the current educational environment and the pressure placed on students, teachers, administrators, and schools for all students to be successful, the results of the study are promising. As discussed in previous chapters, it is imperative that educators be able to not only determine base line levels of achievement but also to accurately monitor student progress. This assists with determining the amount of time needed for a student to meet benchmarks as well as with measuring a student's response to instruction. This study adds to the literature base as it specifically investigated the use of CBM with students with EBD for whom very

little research has been completed. While much focus and emphasis may be placed on curriculum choices, instructional delivery, methods, and classroom environments, the results certainly suggest that an equally important educational decision revolves around how we monitor progress and measure success. Choosing the type of assessment used could have some serious implications for the teachers of students with EBD as they work to identify which students are at-risk, make instructional decisions, and monitor their students' progress. This is something that may often be overlooked as many times these decisions are made by people outside of the classroom (ie. administrators, district personnel, legislators). Choosing the wrong measuring tool could theoretically have a negative impact on the success of students. This is especially important for students with EBD who already are at-risk for school failure and for those educators who teach in the state of Florida where the use of the DIBELS ORF assessments has recently been discontinued and replaced with the FAIR, which includes the maze tasks.

Limitations

The results of this study included several limitations, including but not limited to the following discussion.

First, the study used a convenience sample of current students. This limited the sample size which has been shown to impact results when attempting to determine correlations and creating prediction models.

Secondly, the study was limited to those students in third and fourth grade. The literature suggests that the use of CBM may give varying results across grade levels and age may have impacted the results. Hops and Fuchs (2005) found that the relationship

between reading aloud and reading proficiency changed with age. While the results of this study do not specifically refute or support these findings, it is important to note that the study did occur over a two-year period.

Furthermore, the study was limited to a single school district in the state of Florida so the sample did not have characteristics representative of the entire district and/or state. The state of Florida is very large state and demographics are known to vary from one geographical area to another.

Another important limitation is that this was the first year of administrations of the FAIR maze assessment in the state of Florida. It was well documented that many schools and districts had difficulties with the online computer format. Students were reportedly kicked off the system or scores invalidated for unexplained reasons. This is likely the reason that there were so many students with a score of zero or with missing data altogether. All of this may have impacted scores and thereby the statistical results obtained.

Additionally, the assessments were not administered concurrently. ORF scores from one year were compared to maze scores from the next year. There is no way to control for other variables that may have impacted student achievement from year to year. Also, all of the data were archival so there is no way of knowing if assessments were administered correctly as designed. In order to get accurate scores, proctors must follow the specific administration guidelines of each assessment tool.

Lastly, while every effort was made to match both sample populations (students with and without EBD) across gender, ethnicity, and SES, a perfect matched group was not obtained.

This is one of the limitations when using a convenience sample and attempting to match one group to another.

Recommendations for further study

Keeping in mind the limitations previously mentioned, the analysis of the data suggests several areas for further study as well as ways to improve the current study. Several recommendations are discussed below.

Further studies should attempt to include a larger sample size, including but not limited to increasing the representation of the sample across gender, ethnicity, SES, grade level, and geographic areas. A larger sample size would improve the ability of researchers to draw conclusions about the usefulness of various CBM measures for the progress monitoring of student success for those students with and without disabilities. It would also allow researchers to determine if the use of one progress monitoring is better suited for a specific grade level or demographic group than another. This is important for teachers for many reasons, but the main emphasis should be on obtaining an accurate and meaningful assessment of student ability.

Additionally, further studies should take pains to include procedures in the study design to insure the correct administration of the assessments, including but not limited to training and fidelity checks. This is especially important when using newer technologies for the administration of said assessments. This too would help ensure that researchers could draw more accurate conclusions about any analysis completed.

Ideally, future studies should also include multi-year assessment data. This would ensure the stability of results over time and lead to a much clearer picture of the types of assessment

data that are most useful for educators and administrators alike. This would help ensure a more intricate system of progress monitoring that is sensitive to fluctuations in student achievement.

It would be beneficial to include the concurrent administrations of both ORF and maze assessments to the same group of students in future study designs. This would ensure a more equitable comparison among assessments as individual student achievement is known to vary from year to year.

Additionally, further studies should investigate the reliability and validity of ORF, maze tasks, and the FCAT for all students, but more specifically, for the population of students with EBD, as there is very little research specific to this population. As more and more decisions are made outside of the individual classroom regarding what to teach, how to teach, and how to assess, it will be important that the most effective tools are used with each population, especially those students with EBD. This is particularly important in the state of Florida with its adoption of the FAIR with its maze measures and the discontinued use of the ORF assessments within DIBELS.

Lastly, future study design should measure the effects of certain accommodations that are commonly provided to students with disabilities. It is important that researchers and practitioners alike understand the impact that the provision of these accommodations have on the reliability and validity of educational assessments. This would be especially useful for students with EBD since very little research exists for this group.

Conclusion

The purpose of this study was to extend the research on the use of ORF and maze measures with students with and without EBD through the relationship of the Dynamic

Indicators of Basic Early Literacy Skills (DIBELS, Good & Kaminski, 2002) and the Florida Assessments for Instruction in Reading (FAIR) with the FCAT. While results indicated that a relationship does exist, the extent to which it does needs to be examined further. The literature suggests that reading failure in general is associated with an increased risk for negative developmental outcomes including teen pregnancy, dropping out of school, substance abuse, unemployment, and antisocial behavior (McGill-Frazen & Allington, 1991; Stanovich, 1993/1994). Additionally, many important decisions are being made that not only impact the future educational experience of students but also have large economic and social consequences for individuals and schools alike. It is very important that educators become responsible decision makers and use assessment procedures that do not discriminate against certain groups of students. Hopefully, this study can serve as a starting point for further studies about the use of CBM with all students; and especially for students with EBD. Given the challenges that students with EBD face both academically and behaviorally, it is of the utmost importance that researchers assist teachers of this disadvantaged population in finding ways to accurately monitor their progress and effectively measure their success.

References

Adams, M. J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Annie. E. Casey Foundation, The. (2009). Early warning! Why reading by the end of third grade matters. Retrieved from http://www.aecf.org/~/media/Pubs/Initiatives/KIDS%20COUNT/123/2010KCSpecReport/Special%20Report%20Executive%20Summary.pdf.

Ardoin, S. P., Witt, J. C, Suido, S. M., Connell, J. E., Koenig, J. L., Resetar, J. L., et al. (2004). Examining the incremental benefits of administering a maze and three versus one curriculum-based measurement reading probes when conducting universal screening. *School Psychology Review*, *33*(2), 218-233.

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

Baker, S. K., & Smith, S. (2001). Linking school assessments to research-based practices in beginning reading: Improving programs and outcomes for students with and without disabilities. *Teacher Education and Special Education, 24*(4), 315-332. doi: 10.1177/088840640102400406.

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

Barriga, A. Q., Doran, J. W., Newell, S.B., Morrison, E. M. Barbetti, V., & Robbins, B.R. (2002). Relationships between problem behaviors and academic achievement in adolescents: The unique role of attention problems. *Journal of Emotional and Behavioral Disorders, 10*(4), 233-240. doi: 10.1177/10634266020100040501.

Barton-Arwood, S. M., Wehby, J. H., & Falk, K. B. (2005). Reading instruction for elementary-age students with emotional and behavioral disorders: Academic and behavioral outcomes. *Exceptional Children, 72*(1), 7-27.

Berkeley, S., Bender, W. N., Gregg Peaster, L., & Saunders, L. (2009). Implementation of response to intervention: A snapshot of progress. *Journal of Learning Disabilities, 42*(1), 85-95. doi: 10.1177/0022219408326214.

Brown-Chidsey, R., & Steege, M. W. (2005). *Response to intervention: Principles and strategies for effective practice*. New York, NY: The Guilford Press.

Buck, J., & Torgesen, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test*. Tallahassee, FL: Florida Center for Reading Research (FCRR Technical Report #1).

Buckley, J. A. (2009). Implementing evidence-based interventions in elementary schools for students with and at-risk for severe behavior disorders. *Journal of Emotional and Behavioral Disorders, 17*(4), 195-196. doi: 10.1177/1063426609345866.

Carlson, J. F., Römhild, A., McCormick, C., Chin, K., Geiseinger, K. F., Shaw, L., & Foley, B. P. (2010). Evaluating reading tests for the state of Florida: Dynamic Indicators of Basic Early Literacy Skills, 6th Edition (DIBELS) and Florida

Assessments for Instruction in Reading (FAIR). Lincoln, NE: Buros Center for Testing.

Carroll, J. B. (1963). A model for school learning. *Teacher's College Record, 64,* 723-733.

Carroll, J. B. (1989). The Carroll Model: A 25 year retrospective and prospective view. *Educational Researcher, 18*, 26-31.

Clark, C. M. (1987). The Carroll model. In C. Fisher & D. C. Berliner (Eds.), *International Encyclopedia of Teaching and Teacher Education* (pp. 36-40). Oxford, England: Pergamon.

Cooley, W. W., & Lohnes, P. R. (1976). *Evaluation research in education*. New York: Irvington Publishers.

Coyne, M. D., Kame'enui, E. J., & Simmons, D. C. (2004). Improving beginning reading instruction and intervention for students with LD: Reconciling "all" with "each." *Journal of Learning Disabilities*, *37*(3), 231-239. doi:10.1177/00222194040370030801.

Cranney, A. G. (1972-1973). The construction of two types of cloze reading tests for college students. *Journal of Reading Behavior, 5*, 60-64.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52*, 219-232.

Deno, S. L. (1990). Individual differences and individual difference: The essential difference for special education. *The Journal of Special Education, 24*(2), 160-173.

Deno, S. L., Maruyama, G., Espin, C. A., & Cohen, C. (1989). *The basic academic skills samples (BASS).* Minneapolis, MN: University of Minnesota.

Deno, S. L., Maruyama, G., Espin, C., & Cohen, C. (1990). Educating students with mild disabilities in general education classrooms: Minnesota alternatives. *Exceptional Children, 57*(2), 150-161.

Deno, S., Mirkin, P. K., & Chiang, B. (1982). Identifying valid measures of reading. *Exceptional Children*, *49*, 36-45.

Elliot, J. L. (2003). IDEA 2003: Reauthorization or retrofit? *School Administrator, 60*(3), 28-30.

Epstein, M., & Cullinan, D. (1983). Academic performance of behaviorally disordered and learning-disabled pupils. *The Journal of Special Education, 17*(3), 303-307. doi: 10.1177/002246698301700306.

Espin, C. A., & Deno, S. L. (1995). Curriculum-based measures for secondary students: Utility and task specificity of text-based reading and vocabulary measures for predicting performance on content-area tasks. *Diagnostique, 20,* 121-142.

Field, A. (2009). *Discovering statistics using SPSS (3rd ed.).* Los Angeles, CA: Sage Publications.

Fleming, C. B., Harachi, T. W., Cortes, R. C., Abbott, R. D., & Catalano, R. F. (2004). Level and change in reading scores and attention problems during elementary school as predictors of problem behavior in middle school. *Journal of Emotional and Behavioral Disorders, 12*(3), 130-144. doi: 10.1177/10634266040120030101.

Fletcher, J. M., Shaywitz, S. E., Shankweiler, D. P., Katz, L., Liberman, I. Y., Stuebing, K. K., Francis, D. J., et al. (1994). Cognitive profiles of reading disability: Comparisons of

discrepancy and low achievement definitions. *Journal of Educational Psychology*, *86*(1), 6-23. doi: 10.1037/0022-0663.86.1.6.

Florida Center for Reading Research (2008a). *The Florida Assessments for Instruction in Reading overview.* Retrieved from http://www.fcrr.org/staffpresentations/Foorman/FL%20Assessment%20District%20Contact%20Training%.

Florida Center for Reading Research (2008b). *Presentation for Florida Organization of Instructional Leaders 11/05/2008*. Retrieved from http://www.justreadflorida.com/pdf/FloridaAssessmentsReading-RegMeet.pdf.

Florida Center for Reading Research (2008c). *Florida Assessments for Instruction in Reading brochure*. Retrieved from http://www.justreadflorida.com/pdf/FLAiRBrochureVer3.pdf.

Florida Department of Education. (2009). *Florida assessments for instruction in reading: Technical manual 2009-2010 edition.* Tallahassee, FL. Author.

Florida Department of Education, Bureau of Exceptional Education and Student Services (2009a). *2009 LEA Profile*. Retrieved from http://www.fldoe.org/ese/pdf/2009LEA/lee.pdf.

Florida Department of Education, Bureau of Exceptional Education and Student Services (2009b). *Rule 6A-6.03016 Exceptional student education eligibility for students with emotional/behavioral disabilities.* Retrieved from https://www.flrules.org/gateway/ruleNo.asp?id=6A-6.03016.

Florida Department of Education, Bureau of Exceptional Education and Student Services (2009c). *2009 SEA Profile*. Retrieved from http://www.fldoe.org/ese/pdf/2009LEA/SEAProfile.pdf.

Florida Department of Education, Bureau of Exceptional Education and Student Services (2007). *Students with disabilities enrollment, Ages 6-21, 2007-2008*. Retrieved from www.fldoe.org/ese/datapage.asp.

Florida Department of Education. (2011). *Florida Response to Intervention/instruction.* Retrieved from www.florida-rti.org/flMod/definition.htm.

Florida Department of Education, Just Read, Florida! (2009). *Resources for sustainability of Reading First.* Retrieved from http://www.justreadflorida.com/reading_first.asp.

Forness, S. R., Bennett, L., & Tose, J. (1983). Academic deficits in emotionally disturbed children revisited. *Journal of the American Academy of Child Psychiatry, 22,* 140-144.

Fuchs, L., & Deno, S. L. (1994). Must instructionally useful performance assessment be based in the curriculum? *Exceptional Children, 61*(1), 15-24.

Fuchs, L. S. (1999). Connecting assessment to instruction. *Schools in the Middle, 9*(4), 18-21.

Fuchs, L., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53*(3), 199-208.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review, 21*(1), 45-58.

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

Fuchs, L. S., & Fuchs, D. (1999). Monitoring student progress toward the development of reading competence: A review of three forms of classroom-based assessment. *School Psychology Review, 28*(4), 659-671. doi: 10.1598/RRQ.41.1.4.

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

Fuchs, L.S., & Fuchs, D. (2009a). On the importance of a unified model of responsiveness to intervention. *Child Development Perspectives, 3*(1), 41-43. doi: 10.1111/j.1750-8606.2008.00074.

Fuchs, D., & Fuchs, L. S. (2009b). Responsiveness to intervention: Multilevel assessment and instruction as early intervention and disability identification. *The Reading Teacher, 6*3(3), 250-252. doi:10.1598/RT.63.3.10.

Fuchs, L. S., Fuchs, D., & Maxwell, L. (1988). The validity of informal reading comprehension measures. *Remedial and Special Education, 9*(2), 20−28. doi: 10.1177/074193258800900206.

Gardner, R. R, Rudman, H. C, Karlsen, B., & Merwin, J. C. (1982). *Stanford Achievement Test*. Iowa City: Harcourt Brace Jovanovich.

Good, R. H., & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly, 11*(4), 326-336. doi: 10.1037/h0088938.

Good, R. H., & Kaminski, R. A. (2002). *DIBELS Oral Reading Fluency Passages for First through Third Grades (Technical Report No. 10).* Eugene, OR: University of Oregon.

Greenbaum, P., & Dedrick, R. (1996). National adolescent and child treatment study (NACTS): Outcomes for children with serious emotional and behavioral disturbance. *Journal of Emotional and Behavioral Disorders, 4*(3), 130-147. doi:10.1177/106342669600400301.

Gresham, F. M., Lane, K. L., McMillan, D. L., & Bocian, K. M. (1999). Social and academic profiles of externalizing and internalizing groups: Risk factors for emotional and behavioral disorders. *Behavioral Disorders, 24*, 231-245.

Guskey, T. R. (2007). Closing achievement gaps: Revisiting Benjamin S. Bloom's "learning for mastery." *Journal of Advanced Medicine*, *19*(1), 8-31.

Guthrie, J. T. (1973). Reading comprehension and syntactic responses in good and poor readers. *Journal of Educational Psychology, 65*(3), 294-299. doi: 10.1037/h0035643.

Guthrie, J. T, Seifert, M., Burnham, N., & Caplan, R. I. (1974). The maze technique to assess, monitor reading comprehension. *The Reading Teacher, 28*, 161-168.

Hagans, K. S. (2008). A Response-to-Intervention approach to decreasing early literacy differences in first graders from different socioeconomic backgrounds: Evidence for the intervention validity of the DIBELS. *Assessment for Effective Intervention, 34*(1), 35-42. doi:10.1177/1534508408314170.

Hamilton, C., & Shinn, M. R. (2003). Characteristics of word callers: An investigation of the accuracy of teachers' judgments of reading comprehension and oral reading skills. *School Psychology Review, 32*(2), 228-240.

Harcourt Assessment, Inc. (2007). *FCAT reading and mathematics technical report for 2006 FCAT test administrators*. Retrieved from http://fcat.fldoe.org/pdf/fc06tech.pdf.

Hasbrouck, J., & Tindal, G. A. (2006). Oral reading fluency norms: A valuable assessment tool for reading teachers. *The Reading Teacher*, *59*(7), 636-644.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, *34*(3), 372–386.

Hintze, J.M., Shapiro, E. S., Conte, K. L., & Basile, I.M. (1997). Oral reading fluency and authentic reading material: Criterion validity of the technical features of CBM survey-level assessment. *School Psychology Review*, *26*(4), 535−553.

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

Hintze, J. M., Owen, S. V., Shapiro, E. S., & Daly, E. J. (2000). Generalizability of oral reading fluency measures: Application of G theory to curriculum-based measurement. *School Psychology Quarterly*, *15*(1), 52−68. doi: 10.1037/h0088778

Hosp, M. K., & Fuchs, L. S. (2005). Using CBM as an indicator of decoding, word reading, and comprehension: Do the relations change with grade? *School Psychology Review*, *34*(1), 9-26.

Hosp, M. K., Hosp, J. L., & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: The Guilford Press.

Jacobson, R. (1998). Teachers improving learning using metacognition with self-monitoring learning strategies. *Education, 118*(4), 579-589.

Jenkins, J., Fuchs, L., van den Broek, P., Espin, C, & Deno, S. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology, 95*(4), 719-729.

Jenkins, J. R., & Jewell, M. (1993). Examining the validity of two measures for formative teaching: Reading aloud and maze. *Exceptional Children, 59*(5), 421-432.

Johnson, E., Mellard, D. F., Fuchs, D., & McKnight, M. A. (2006). *Responsiveness to intervention (RTI): How to do it*. Lawrence, KS: National Research Center on Learning Disabilities.

Kaminski, R., Cummings, K.D., Powell-Smith, K.A., & Good, R.H.III (2008). Best practices in using Dynamic Indicators of Basic Early Literacy Skills for formative assessment and evaluation. In A. Thomas and J. Grimes (eds.) *Best practices in school psychology V*. Bethesda, MD: National Association of School Psychologists.

Kauffman, J. M. (2005). *Characteristics of emotional and behavioral disorders of children and youth* (8th ed.) Upper Saddle River: Pearson.

Kauffman, J.M., Cullinan, D., & Epstein, M.H. (1987). Characteristics of students placed in special programs for the seriously emotionally disturbed. *Behavioral Disorders, 12*, 175-184.

Kingston, A. J., & Weaver, W. W. (1970). Feasibility of cloze techniques for teaching and evaluating culturally disadvantaged beginning readers. *The Journal of Social Psychology*, *82*(2), 205-214.

Kranzler, J. H., Brownell, M. T., & Miller, M. D. (1998). The construct validity of curriculum-based measurement of reading: An empirical test of a plausible rival hypothesis. *Journal of School Psychology*, *36*(4), 399-415. doi: 10.1016/S0022-4405(98)00018-1.

Landrum, T. J., Tankersley, M., & Kauffman, J. M. (2003). What is special about special education for students with emotional or behavioral disorders? *Journal of Special Education, 37*(3), 148-156. doi: 10.1177/00224669030370030201.

Lane, K., Little, A. L., Redding-Rhodes, J., Phillips, J. R., & Welsh, M. (2007). Outcomes of a teacher-led reading intervention for elementary students at-risk for behavioral disorders. *Exceptional Children, 74*(1), 47-70.

Lassen, S. R., Steele, M. M., & Sailor, W. (2006). The relationship of school-wide positive behavior supports to academic achievement in an urban middle school. *Psychology in the Schools, 43*(6), 701-712. doi: 10.1002/pits.20177.

LaBerge, D., & Samuels, S. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293-323.

Lee County School District, The. (2009). About us. Retrieved from http://www.leeschools.net/info/about.htm.

Maas, C. J. M., & Hox, J. J. (2004). Robustness issues in multilevel regression analysis. *Statistica Neerlandica, 58*(2), 127-137.

Marcotte, A. M., & Hintze, J. M. (2009). Incremental and predictive utility of formative assessment methods of reading comprehension. *Journal of School Psychology, 47*(5), 315-335. doi: 10.1016/j.jsp.2009.04.003.

MacGinitie, W. H., Kamons, J., Kowalski, R. L., MacGinitie, R. K., & McKay, T. (1978). *Gates-MacGinitie Reading Tests (2nd ed.).* Chicago, IL: Riverside.

Markell, M. A., & Deno, S. L. (1997). Effects of increasing oral reading: Generalization across reading tasks. *The Journal of Special Education, 31*(2), 233-250. doi: 10.1177/002246699703100205.

Marston, D. (1989). Curriculum-based measurement: What is it and why do it? In M. R. Shinn (Ed.), *Curriculum-based measurement: Assessing special children* (pp. 18-78). New York, NY: Guilford Press.

McCardel, P., & Chhabra, V. (2004). *The voice of evidence in reading research*. Baltimore, MD: Brookes.

McGill-Frazen, A., & Allington, R. (1991). The gridlock of low reading achievement: Perspectives on practice and policy. *Remedial & Special Education, 12*, 20-30.

Mehrens, W. A., & Clarizio, H. F. (1993). Curriculum-based measurement: Conceptual and psychometric considerations. *Psychology in the Schools, 30*, 241-254.

Mertens, D. M. (1998). *Research methods in education and psychology: Integrating diversity with quantitative & qualitative approaches*. Thousand Oaks, CA: Sage Publications, Inc.

Morris, R. D., Shaywitz, S. E., Shankweiler, D. R., Katz, L., Stuebing, K. K., Fletcher, J. M., Lyon, R.G., et al. (1998). Subtypes of reading disability: Variability around a phonological core. *Journal of Educational Psychology, 90*(3), 347-373.

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

Nebraska Department of Education. (1996). *Teaching strategies for students with diverse learning needs.* Retrieved from http://www.nebraskasocialstudies.org/pdf/tsfswdln.pdf.

Nelson, J. R., Benner, G., Lane, K., & Smith, B. (2004). Academic achievement of K-12 students with emotional and behavioral disorders. *Exceptional Children, 71*(1), 59-73.

Nelson, J. R., Johnson, A., & Marchand-Martella, N. (1996). Effects of direct instruction, cooperative learning, and independent learning practices on the classroom behavior of students with behavioral disorders: A comparative analysis. *Journal of Emotional and Behavioral Disorders, 4*(1), 53-62. doi: 10.1177/106342669600400106.

Nelson, J. M., & Machek, G. R. (2007). A survey of training, practice, and competence in reading assessment and intervention. *School Psychology Review, 36*(2), 311-327.

Osborne, J. & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, 8(2). Retrieved from http://PAREonline.net/getvn.asp?v=8&n=2

Parker, R., Hasbrouck, J. E. , & Tindal, G. (1992). The maze as a classroom-based reading measure: Construction methods, reliability, and validity. *The Journal of Special Education, 26*(2), 195-218. doi: 10.1177/002246699202600205.

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In Solso RL. *Information processing and cognition: The Loyola symposium*. Hillsdale, N.J: L. Erlbaum Associates.

Prescott, G. A., Balow, I. H., Hogan, T. P., & Farr, R. C. (1984). *Metropolitan Achievement Tests (MAT-6).* San Antonio, TX: Psychological Corp.

Quenemoen, R., Thurlow, M., Moen, R., Thompson, S., & Morse, A. B. (2004). *Progress monitoring in an inclusive standards-based assessment and accountability system* (Synthesis Report 53). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Reeves, T. C. (1999). *A model to guide the integration of the world wide web as a cognitive tool in K-12 education*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.136.5351&rep=rep1&type=pdf

Riedel, B. W. (2007). The relation between DIBELS, reading comprehension, and vocabulary in urban first-grade students. *Reading Research Quarterly, 42*(4), 446-567.

Rock, E. E., Fessler, M. A., & Church, R. P. (1997). The concomitance of learning disabilities and emotional/behavioral disorders: A conceptual model. *Journal of Learning Disabilities, 30*(3), 245-263. doi: 10.1177/002221949703000302.

Roehrig, A. D., Petscher, Y., Nettles, S. M., Hudson, R. F., & Torgesen, J. K. (2008). Accuracy of the DIBELS Oral Reading Fluency measure for predicting third grade reading comprehension outcomes. *Journal of School Psychology, 46*(3), 343-366. doi:10.1016/j.jsp.2007.06.006 .

Rones, M., & Hoagwood, K. (2000). School-based mental health services; A research review. *Clinical Child and Family Psychology review, 3*(4), 223-241.

Ruhl, K. L., & Berlinghoff, D. H. (1992). Research on improving behaviorally disordered students' academic performance: A review of the literature, *Behavioral Disorders*, *17*, 178-190.

Schilling, S. G., Carlisle, J. F., Scott, S. E., & Zeng, J. (2007). Are fluency measures accurate predictors of reading achievement? *The Elementary School Journal, 107*(5), 429-448. doi: 10.1086/518622.

Scott, T. M., & Shearer-Lingo, A. (2002). The effects of reading fluency instruction on the academic and behavioral success of middle school students in a self-contained EBD classroom. *Preventing School Failure*, *46*(4), 167-173.

Shankweiler, D., Crain, S., Katz, L., Fowler, A E., Liberman, A. M., Brady, S. A., Thornton, R., … Shaywitz, B. A. (1995). Cognitive profiles of reading-disabled children: Comparison of language skills in phonology, morphology, and syntax. *Psychological Science, 6*(3), 149-156. doi: 10.1111/j.1467-9280.1995.tb00324.x.

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading

and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*(1), 19-35. doi: 10.1177/0734282905285237.

Shaw, R., & Shaw, D. (2002). *DIBELS oral reading fluency-based indicators of third grade reading skills for Colorado State Assessment Program (CSAP)*. Eugene: University of Oregon.

Shaywitz, S. (2003). *Overcoming dyslexia*. New York, NY: Guilford Press.

Shelton, N. R., Altwerger, B., & Jordan, N. (2009). Does DIBELS put reading first? *Literacy Research and Instruction, 48*(2), 137-148. doi: 10.1080/19388070802226311.

Shin, J., Deno, S. L., & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. The *Journal of Special Education, 34*(3), 164-172. doi: 10.1177/002246690003400305.

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

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, *30*(3), 407–419.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*(1), 32-71.

Stanovich, K. E. (1993/1994). Romance and reality. *The Reading Teacher, 47*, 280-291.

Stanovich, K. E. (1999). The sociopsychometrics of learning disabilities. *Journal of Learning Disabilities*, *32*(4), 350-361. doi:10.1177/002221949903200408.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers.* New York, NY: Guilford Press.

Stanovich, K. E., & Siegel, L. S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology*, *86*(1), 24-53. doi: 10.1037/0022-0663.86.1.24.

Sutherland, K. S., & Wehby, J. H. (2001). The effect of self-evaluation on teaching behavior in classrooms fro students with emotional and behavioral disorders. *Journal of Special Education*, *35*(3), 161-171. doi: 10.1177/002246690103500306.

Torgeson, J. K. (2002). The prevention of reading disabilities. *Journal of School Psychology*, *40*(1), 7-26. doi: 10.1016/S0022-4405(01)00092-9.

Torgeson, J. K., Morgan, S., & Davis, C. (1992). The effects of two types of phonological awareness training on word learning in kindergarten children. *Journal of Educational Psychology*, *84*(3), 364-370. doi: 10.1037/0022-0663.84.3.364.

University of Oregon Center on Teaching and Learning. (2009). *DIBELS Information.* Retrieved from https://dibels.uoregon.edu/docs/dibelsinfo.pdf

U.S. Department of Education. (2002a). *A new era: Revitalizing special education for children and their families.* Retrieved from http://www.ed.gov/inits/commissionsboards/whspecialeducation/reports/index.html.

U. S. Department of Education. (2002b). *Twenty-fourth annual report to Congress on the implementation of the handicapped act*. Washington, DC: U.S. Government Printing Office.

U.S. Department of Education. (2004). *Individuals with Disabilities Education Act.* Washington, DC: Author.

U.S. Department of Education, National Center for Education Statistics. (2006). *Digest of Education Statistics, 2005*. Retrieved from http://nces.ed.gov/fastfacts/display.asp?id=64.

U.S. Department of Education, Office of Special Education Programs. (2007). *IDEA regulations alignment with the No Child Left Behind (NCLB) Act*. Retrieved from http://idea.ed.gov/explore/view/p/%2Croot%2Cdynamic%2CTopicalBrief%2C3%2C.

U.S. Department of Education, Office of Special Education Programs. (2008). *Part B—Trend data report for states and outlying areas, 2003-2004 through 2007-2008*. Retrieved from https://www.ideadata.org/docs/Part%20B%20%20Trend%20Data%20Report%20for%20States%20and%20Outlying%20Areas,%202003-04%20through%202007-08.pdf.

U.S. Department of Health and Human Services. (1999). *A report of the surgeon general.* Rockville, MD: U.S Public Health Service.

Valencia, S. W., Smith, A. T., Reece, A. M., Li, M., Wixson, K. K., & Newman, H. (2010). Oral reading fluency assessment: Issues of construct, criterion, and consequential validity. *Reading Research Quarterly*, *45*(3), 270-291.

Wagner, M. M., Kutash, K., Duchnowski, A. J., Epstein, M. H., & Sumi, C. (2005). The children and youth we serve: A national picture of the characteristics of students with emotional disturbances receiving special education. *Journal of Emotional and Behavioral Disorders, 13*(2), 79-96. doi: 10.1177/10634266050130020201.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education, 41*(2), 85-121. doi: 10.1177/00224669070410020401.

Wiley, H. I., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26*(4), 207-214. doi:10.1177/07419325050260040301.

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11*(2), 85-104. doi: 10.1207/s15326977ea1102_1.

Wright, P. W. D., Wright, P. D., & Heath, S. W. (2004). *No Child Left Behind.* Hartfield, VA: Harbor House Law Press, Inc.

Yell, M. L. (1992). Barriers to implementing curriculum-based measurement. *Diagnostique, 18,* 99-112.

Appendix A

Sample of 3rd Grade DIBELS (6th Ed.) ORF Passage

ORF Progress Monitoring 20

Firefighters

Firefighters are always busy, even when there are no fires to	11
put out. After each fire, the firefighters have to dry their fire	23
hoses. They also clean their trucks and inspect their tools. The	34
air tanks and the water tanks must be refilled. Even the boots,	46
hats, and coats must be cleaned and checked. Everything must be	57
ready for the next call.	62
When they are not at fires, firefighters spend time getting	72
ready for the next call. They also provide training in fire	83
prevention and fire safety. They go to schools and businesses to	94
make people aware of all they can do to prevent fires. They	106
know the best way to respond to a fire is to keep it from	120
happening in the first place.	125
When there is a call, the firefighters really get busy. The	136
firefighters must get to the fire as quickly as possible. If the fire	149
is in a building, the first job is to make sure everyone is safe and	164
out of the building. They also try to prevent the fire from	176
spreading. The firefighters hook up their hoses in front of the	187
building and quickly put out the fire. Everything must be fast and	199
organized.	200
After the fire is out, firefighters must check the building	210
carefully and wait to make sure there is no fire left. Even one	223
live ember can get the fire started all over again. After each call	236
the firefighters go back to their station and clean up all over	248
again.	249

Total words: _____ – errors: _____ = words correct: _____

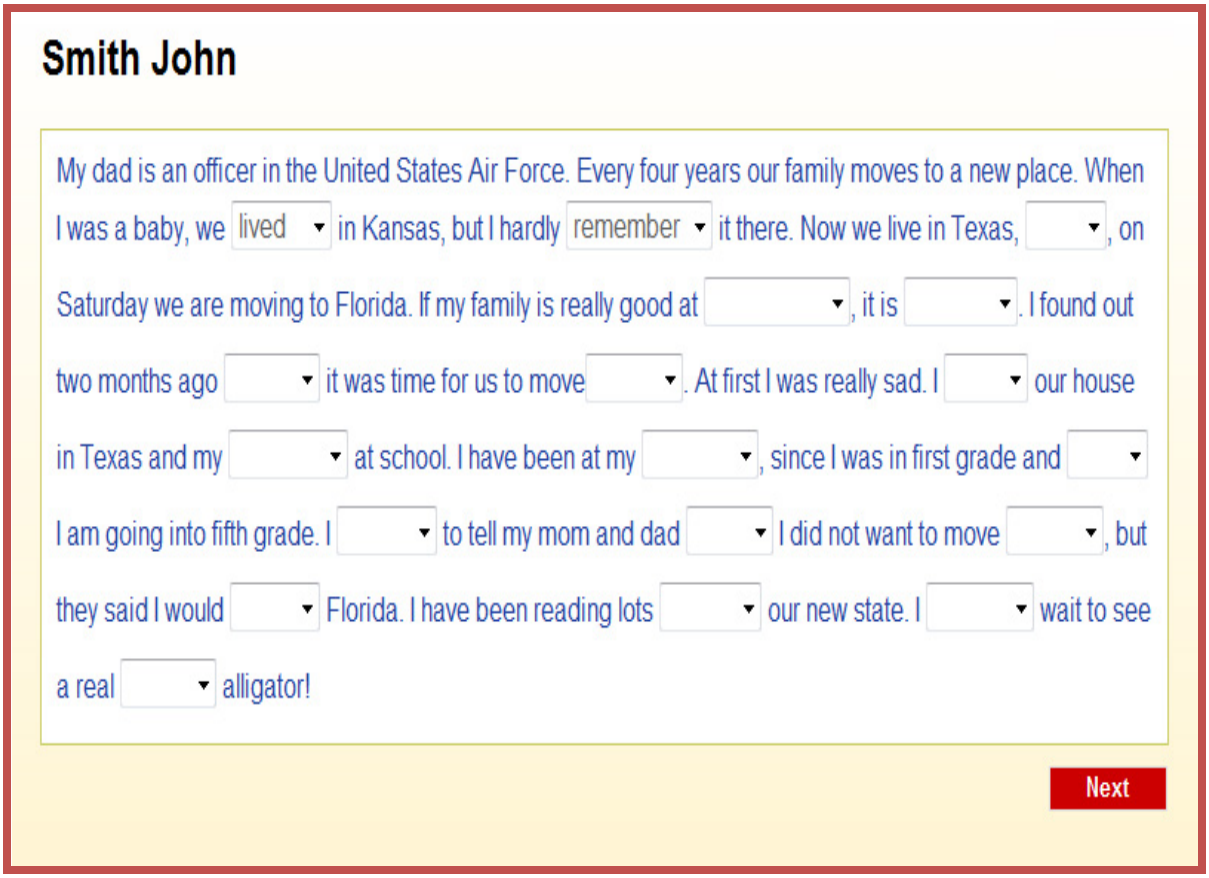
Retell: ORF Total:_____

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48		
49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71		
72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94		

Retell Total:_____

Appendix B

Sample FAIR Grades 3-12 Maze Task Item (FCRR, 2008)



Appendix C

Archival Data Form

Participant # _____

Student age:

_____ years _____months

Gender:

_____ Male _____ Female

Ethnicity:

_____Black _____Hispanic

_____Mixed _____White

Socioeconomic Status:

_____free/reduced _____non free/reduced

Service Delivery Model:

_____self-contained _____center school

_____consultative _____gen. ed.

Assessment Scores:

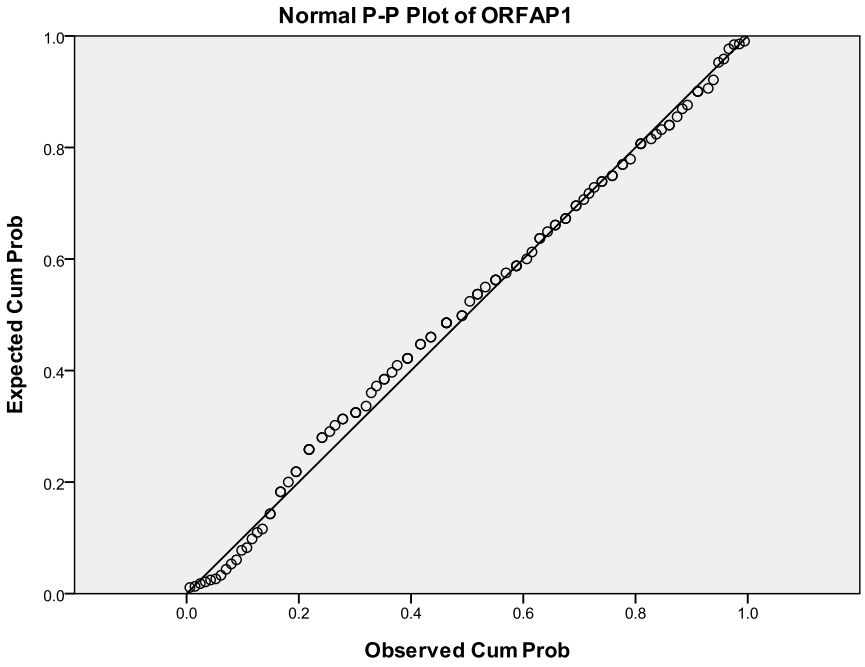
ORF AP1_____ ORF AP2_____

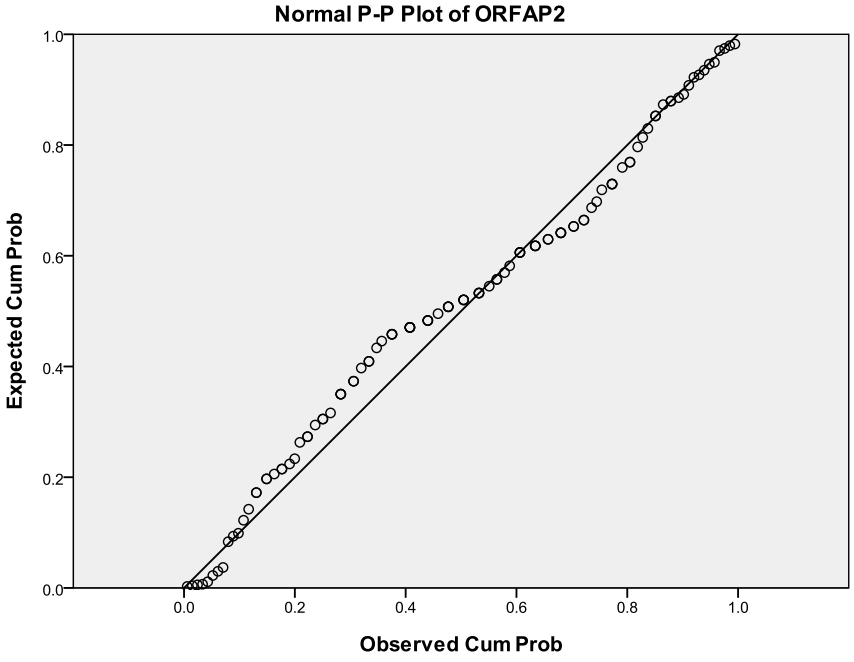
Maze AP1_____ Maze AP2_____

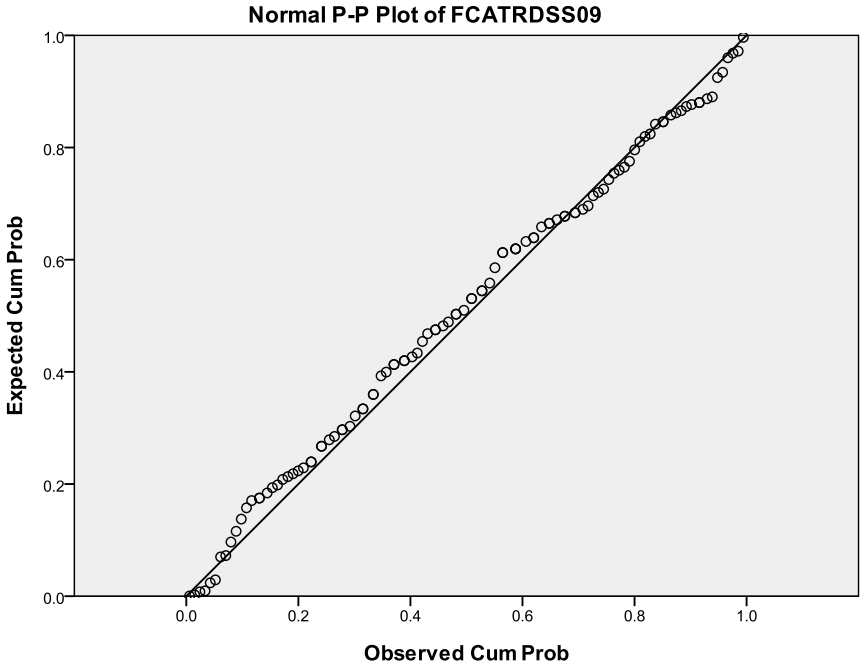
FCAT RD 3rd_____ FCAT RD 4th_____

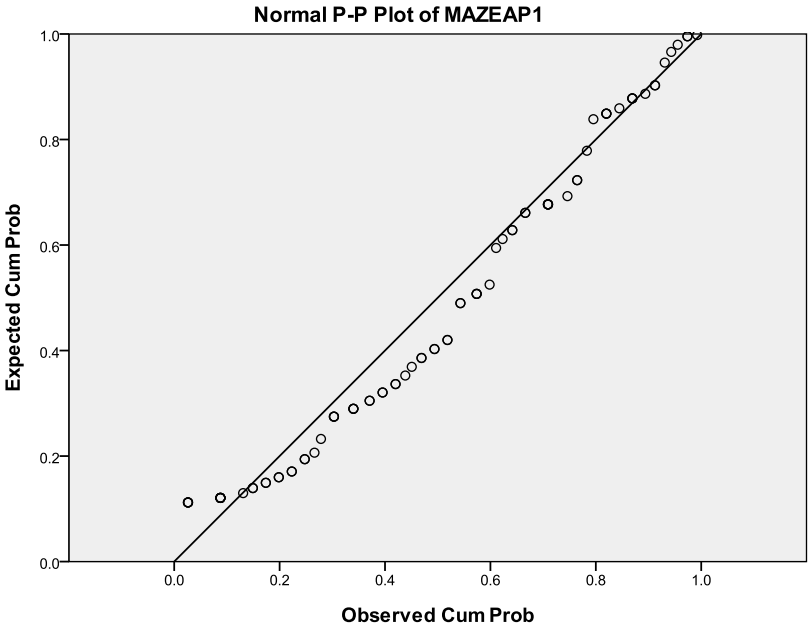
Appendix D

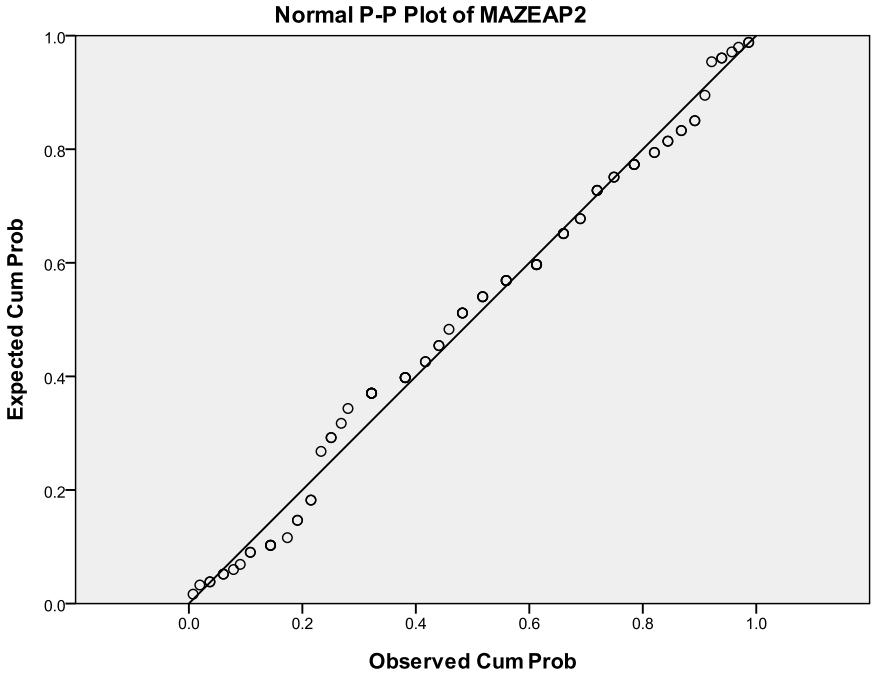
Probability-Probability Plots

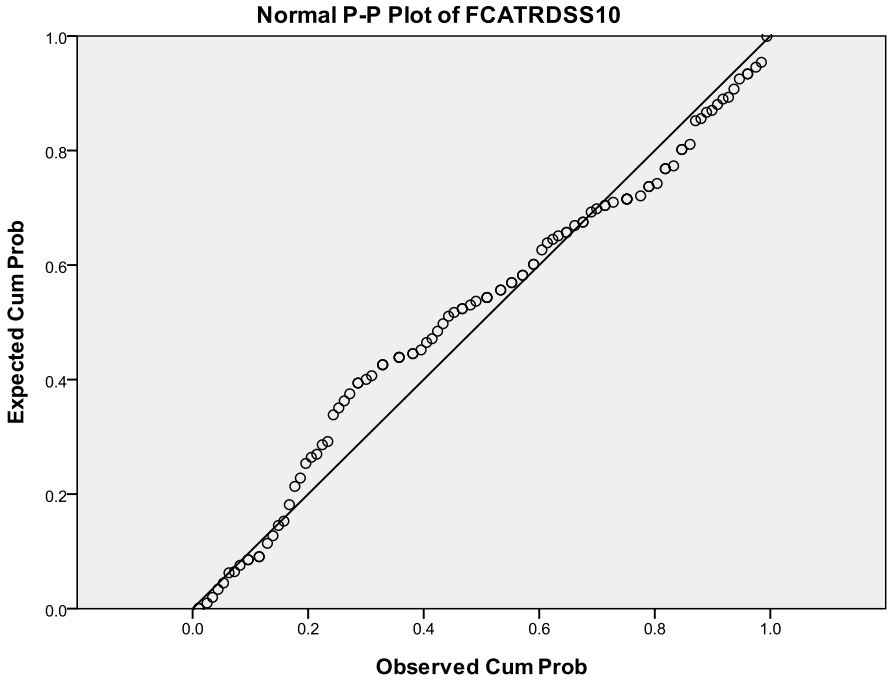





Normal P-P Plot of FCATRDSS09
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Normal P-P Plot of MAZEAP1
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Normal P-P Plot of MAZEAP2
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0


Normal P-P Plot of FCATRDSS10
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

Appendix E

IRB Approval Page

BARRY UNIVERSITY

11300 NE Second Avenue
Miami Shores, FL 33161-6695
phone 305-899-3020
800-756-6000, ext. 3020
fax 305-899-3026
www.barry.edu

OFFICE OF THE PROVOST
INSTITUTIONAL REVIEW BOARD

Research with Human Subjects
Protocol Review

Date:	March 21, 2011
Protocol Number:	110309
Title:	The Predictive Value of Maze and ORF Measures for Students with and without EBD
Approval Date:	3/21/11
Name:	Jason W. Kurtz
Address:	372 Prather Drive Ft. Myers, FL 33919
Sponsor:	Dr. Catherine Roberts
School:	ADSOE

Dear Mr. Kurtz:

On behalf of the Barry University Institutional Review Board (IRB), I have verified that the specific changes requested by the IRB have been made. Therefore, I have granted final approval for this study as exempt from further review.

As principal investigator of this protocol, it is your responsibility to make sure that this study is conducted as approved by the IRB. Any modifications to the protocol or consent form, initiated by you or by the sponsor, will require prior approval, which you may request by completing a protocol modification form.

It is a condition of this approval that you report promptly to the IRB any serious, unanticipated adverse events experienced by participants in the course of this research, whether or not they are directly related to the study protocol. These adverse events include, but may not be limited to, any experience that is fatal or immediately life-threatening, is permanently disabling, requires (or prolongs) inpatient hospitalization, or is a congenital anomaly cancer or overdose.

The approval granted expires on March 21, 2012. Should you wish to maintain this protocol in an active status beyond that date, you will need to provide the IRB with and IRB Application for Continuing Review (Progress Report) summarizing study results to date.

If you have questions about these procedures, or need any additional assistance from the IRB, please call the IRB point of contact, Mrs. Barbara Cook at (305)899-3020 or send an e-mail to

dparkhurst@mail.barry.edu . Finally, please review your professional liability insurance to make sure your coverage includes the activities in this study.

Sincerely,

Doreen C. Parkhurst, MD, FACEP

Doreen C. Parkhurst, M.D., FACEP
Chair Institutional Review Board
Associate Dean, SGMS &
Program Director, PA Program
Barry University
Box SGMS
11300 NE 2nd Avenue
Miami Shores, FL 33161

Cc: Dr. Catherine Roberts

Note: The investigator will be solely responsible and strictly accountable for any deviation from or failure to follow the research protocol as approved and will hold Barry University harmless from all claims against it arising from said deviation or failure.