EXAMINING THE VALIDITY OF CURRICULUM-BASED MEASUREMENT FOR STUDENTS WITH EMOTIONAL AND BEHAVIOR DISORDERS IN MIDDLE SCHOOL

by

CHRISTOPHER DARRELL MARTIN

(Under the Direction of Cecil Fore, III)

ABSTRACT

There has been a recent emphasis on improving the academic performance of students with disabilities (NCLB, 2002). Improving the academic performance of students with Emotional and Behavior Disorders is especially important in the current accountability era in which there is much emphasis placed on performance of standardized tests. The purpose of this study was to investigate the validity of curriculum-based measurement for students with Emotional and Behavior Disorders. The curriculum-based measures Maze, and Oral Reading Fluency were compared to the Woodcock-Johnson III subtests (reading fluency, passage comprehension). The assessments were administered to fifty-five students in a middle school setting. Results indicated significant correlations between the curriculum-based measures (Maze, Oral Reading Fluency) and the Woodcock-Johnson subtests (reading fluency, passage comprehension). Results of the study are discussed along with limitations, implications for practice, and implications for future research.

INDEX WORDS: Maze, Oral Reading Fluency, curriculum-based measurement

EXAMINING THE VALIDITY OF CURRICULUM-BASED MEASUREMENT FOR STUDENTS WITH EMOTIONAL AND BEHAVIOR DISORDERS IN MIDDLE SCHOOL

by

CHRISTOPHER DARRELL MARTIN

B.S. Psychology, Columbus State University, 1996

M.Ed., Columbus State University, 1999

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2006

EXAMINING THE VALIDITY OF CURRICULUM-BASED MEASUREMENT FOR STUDENTS WITH EMOTIONAL AND BEHAVIOR DISORDERS IN MIDDLE SCHOOL

by

CHRISTOPHER DARRELL MARTIN

Major Professor:	Cecil Fore, III
Committee:	William N. Bender Jonathan M. Campbell Tom J. Clees John Langone

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2006

DEDICATION

It is a great honor for me to dedicate all of my hard work and motivation to the completion of this program to my parents. Also, I want to thank my brothers for the inspiration and guidance to always excel in life.

ACKNOWLEDGEMENTS

I would like to thank my special friends from the great city of Chicago, IL that gave the support to allow me to embark on this educational journey and complete this program. Further, I would like to thank my professors, educational administrators, and friends that, at one time or another, contributed to my current knowledge and guidance in pursuit of excellence.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS v

LIST OF TABLES ix

CHAPTER

1 INTRODUCTION 1

Rationale 3

Purpose 6

Research Questions 6

2 REVIEW OF THE LITERATURE 8

Introduction 8

Related Research 10

Oral Reading Fluency 22

Maze 24

Conclusion 26

3 METHOD 34

Participants 34

Setting 35

Instruments and Materials 36

Procedures 43

Reliability 44
Design 45
4 RESULTS 49
Introduction 49
Research Questions 49
Analysis of Research Questions 50
Summary 70
5 DISCUSSION 72
Introduction 72
Summary of Results 74
Limitations of Study 78
Implications for Practice 79
Implications for Future Research 80
Conclusion 81
REFERENCES 83
APPENDICES 91
A PARENTAL CONSENT FORM 91
B ORAL READING FLUENCY PASSAGE 1 94
C ORAL READING FLUENCY PASSAGE 2 96
D ORAL READING FLUENCY PASSAGE 3 98
E MAZE PASSAGE 1 100
F MAZE PASSAGE 2 102
G MAZE PASSAGE 3 104

H PROCEDURE CHECKLIST 106

LIST OF TABLES

Page

Table 1: Research Summary Table 28

Table 2: Student Demographics 35

Table 3: ANOVA for Maze Testing 50

Table 4: Means for Maze Testing by Ethnicity 51

Table 5: ANOVA for Oral Reading Fluency Testing 51

Table 6: Means for Oral Reading Fluency Testing by Ethnicity 52

Table 7: ANOVA for Woodcock-Johnson Passage Comprehension Testing 53

Table 8: Means for Woodcock-Johnson Passage Comprehension Testing by Ethnicity 53

Table 9: ANOVA for Woodcock-Johnson Reading Fluency Testing 54

Table 10: Means for Woodcock-Johnson Reading Fluency Testing by Ethnicity 54

Table 11: ANOVA for Maze Testing 55

Table 12: Means for Maze Testing by Setting 55

Table 13: ANOVA for Oral Reading Fluency Testing 56

Table 14: Means for Oral Reading Fluency Testing by Setting 56

Table 15: ANOVA for Woodcock-Johnson Passage Comprehension Testing 57

Table 16: Means for Woodcock-Johnson Passage Comprehension Testing by Setting 57

Table 17: ANOVA for Woodcock-Johnson Reading Fluency 58

Table 18: Means for Woodcock-Johnson Reading Fluency Testing by Setting 58

Table 19: Correlation Matrix 60

Table 20: ANOVA for Maze Testing 61

Table 21: ANOVA for Oral Reading Fluency Testing 61

Table 22: ANOVA for Maze and Oral Reading Fluency Testing 62

Table 23: ANOVA for Maze Testing 63

Table 24: ANOVA for Oral Reading Fluency Testing 63

Table 25: ANOVA for Oral Reading Fluency Testing 64

Table 26: ANOVA for Maze Testing 64

Table 27: ANOVA for Maze Testing 65

Table 28: ANOVA for Oral Reading Fluency Testing 65

Table 29: ANOVA for Maze and Oral Reading Fluency Testing 66

Table 30: ANOVA for Maze Testing 67

Table 31: ANOVA for Oral Reading Fluency Testing 67

Table 32: ANOVA for Oral Reading Fluency Testing 68

Table 33: ANOVA for Maze Testing 68

Table 34: Correlation Matrix 69

Table 35: ANOVA for Oral Reading Fluency Testing 70

Chapter 1

Introduction

Current academic reforms and accountability changes are rapidly occurring with the recent No Child Left Behind legislation (NCLB, 2002). Improved student achievement as measured through statewide achievement tests has become a focal point for educational progress in our schools. The widespread adoption of statewide assessments along with accountability procedures has placed a great demand on our schools and the classroom teachers for the students to perform at the optimal level of performance during assessment. The demand for increased accountability and academic improvement has not been singularly exclusive to regular education, but special educators also find themselves urged to demonstrate the effects of their programs and practices (Yssledyke, Thurlow, & Shriner, 1992). Unfortunately, statewide achievement tests fail to provide teachers with diagnostic information related to student attainment of specific instructional goals (Crawford, Tindal, & Stieber, 2001). Many times, the statewide assessments are administered on an infrequent basis, providing teachers with limited information about students' ongoing progress toward mastering academic benchmarks.

Among the many debates currently centered on the education of students with disabilities, the need to incorporate empirically derived educational practices into teachers' instructional repertoires is frequently cited as a critical concern (Foegen, Espin, Allinder, Rose, Markell, 2001). Fortunately, research conducted in recent years has yielded a profusion of educational interventions shown to be beneficial for students with disabilities (Forness, Kavale, Blum, & Lloyd, 1997). However, translation of this research into routine educational practice has

not always been observed (Carnine, 1997; Stone 1998). Explanations for the poor track record for translating research into practice are many and include such considerations as: (a) trustworthiness, (b) usability, and (c) accessibility (Carnine, 1992, Richardson, Anders, Tidwell, & Lloyd, 1991). Several special education researchers have responded to these demands by increasing their efforts in working collaboratively with educators on the development and integration of interventions into classrooms for students with disabilities (e.g., Fuchs & Fuchs, 1998; Gersten, Morvant, & Brengelman, 1995). One intervention that has been useful for monitoring the progress of special education students has been the usage of curriculum-based measurement (Fuchs & Fuchs, 1992; Shin, Deno, & Espin, 2000).

Curriculum-based measurement was developed to address the effectiveness of a special education intervention model referred to as databased program modification (Deno & Mirkin, 1977). The model was based on the notion that special educators could use repeated measurements to evaluate classroom instruction and improve effectiveness. Stanley Deno, along with several doctoral students (Doug Martson, Steve Robinson, Mark Shinn, Jerry Tindal, Caren Wesson, and Lynn Fuchs) set out launch a systematic program of research on the technical aspects, logistics, and instructional effectiveness of progress monitoring (Fuchs, 2004).

Curriculum-based measurement was developed from a broad set procedure known as curriculum-based assessment. Fuchs and Fuchs (1992) mentioned that curriculum-based measurement is a standardized methodology for measuring academic performance in the school's curriculum. Research indicates curriculum-based measurement provides accurate information about a student's academic standing and progress, which can be used for a variety of psycho-educational decisions (Fuchs & Fuchs, 1992). Curriculum-based measurement has been used for (a) screening and (b) identifying students for special services (Martson, Mirkin, &

Deno, 1984; Shinn, 189), (c) formulating goals and objectives for Individualized Educational Plans (IEP's) (Deno, Mirkin, & Wesson, 1983), (d) monitoring student progress and improving educational programs (Fuchs, Deno, & Mirkin, 1984), (e) transitioning students to less restrictive environments (Allen, 1989; Fuchs, Fuchs, Hamlett, Phillips, & Bentz, 1994), and for (f) summatively evaluating school programs (Germane & Tindal, 1985, Martson, 1988).

Curriculum-based measurement has been used to predict how well students will perform on statewide competency tests of achievement (Crawford, Tindal, & Stieber, 2001). Helwig, Anderson, and Tindal (2002) indicated that predicting how students will perform on statewide competency tests of achievement is critical. More efficient measures that can provide similar information can be extremely valuable tools for teachers. Measures that give teachers snapshots of a student's conceptual understanding of academic concepts at their grade level can fill the need for formative progress monitoring. In addition, justification for predicting achievement scores can be found in the school accountability movement that has put a premium on educators' providing evidence of student learning (Erickson, Ysseldyke, Thurlow, & Elliot, 1998).

Rationale

Recent national programs such as Reading First and No Child Left Behind (NCLB, 2002) have been put into place by the federal government, requiring educators to deliver a well-planned curriculum using evidence-based instructional methods. To assist professionals, expert panels were assembled to review instructional, intervention, and assessment products and processes to determine whether specific procedures qualify as scientifically based. The panels developed lists of procedures from which professionals can select for use in schools (Kame'enui, 2002). As an example of one effort, the creation of a Reading First Assembly Academy Assessment Committee reviewed a large number of assessment practices in various areas of reading (i.e.,

screening, diagnosis, progress monitoring, and outcome assessment; Francis et al., 2002). Many approved instruments are appropriate for summative evaluation of various reading components (e.g., phonemic, awareness, comprehension). A smaller number is appropriate for formative evaluation or assessment that is sensitive to instructional growth and intervention effects. With general education being mostly affected by NCLB, curriculum-based measurement is generally recognized in special education as a valid and reliable indicator of student performance, but has been used relatively rarely in general education despite encouragement (Shinn, Shinn, Hamilton, & Clarke, 2002).

Curriculum-based measurement was judged by the Reading First Assessment Panel as having sufficient evidence for use in progress monitoring or formative evaluation of reading fluency (Francis et al., 2002). Curriculum-based measurement can be used to enhance the integration of services between special and general education by providing common measurement processes that can be applied to assess normal progress through a curriculum or to distinguish a child's performance from the norm of the peers (i.e., the child's failure to progress is not due to lack of exposure to quality instruction; Ardon et. al., 2004). Curriculum-based measurement can be used to conduct universal screening of all children in a school, and this has been advocated as a way to detect problems early and to begin intervention for children who are at-risk (Donovan & Cross, 2002).

Evaluating the effectiveness of curriculum-based measurement has been used across a variety of assessment activities including: screening, pre-referral evaluation, placement in remedial and special education programs, formative evaluation, and evaluation of reintegration and inclusion (Deno, 2003). Procedures for measurement that are drawn directly from the instructional materials used by the classroom teacher are referred to as curriculum-based.

Evidence has shown that the same procedures can be used successfully with materials drawn from other generic sources that are commonly referred to as general outcome procedures (Fuchs & Deno, 1994), or dynamic indicators of basic skills (Shinn, 1998). Curriculum-based measures refer to specific standard procedures (e.g., technical adequacy, standard tasks, prescriptive materials, administering and scoring, performance sampling, multiple samples, and time efficiency). The material for use in curriculum-based measurement may be obtained from instructional materials used by local schools with key factors being the selection process of the materials to increase the utility of the procedures for making instructional decisions (Shinn, 1989). Curriculum-based measurement procedures include specifics of sample duration, administration, student directions, and scoring procedures. The prescriptive selection of materials, along with the standardization of the procedures, is necessary to ensure reliability and utility of the data for individual and group comparisons (Shinn, 1995). Also, curriculum-based measurement monitors academic performance through the use of direct observation procedures. All of the scores for curriculum-based measurement are obtained by counting the number of correct and incorrect responses made in a fixed time period (Deno, 2003). Another important aspect of curriculum-based measurement is that performance is repeatedly sampled across time. Task difficulty is held constant and inferences can be drawn regarding the generalizations of student proficiency at reading unfamiliar but comparable text. Finally, curriculum-based measurement is designed for efficiency, and the samples are one to three minutes in duration depending on the skill being measured and the number of samples necessary to maximize reliability.

Purpose

The purpose of this study is to examine the validity of curriculum-based measurement for students with Emotional and Behavior Disorders in a middle school setting. While the literature base in this research is extensive, there is a lack of research using curriculum-based measurement with students at the secondary level. Further, previous research lacks in the use of curriculum-based measurement for students with Emotional and Behavior Disorders. In this study, examining the validity of curriculum-based measurement will hopefully validate the technical adequacy of the curriculum-based measures in providing evidence-based instructional methods to add to the increased demands of accountability and data-based decision making in the classroom.

Research Questions

Proceeding in the development of effective curriculum-based measurement strategies requires careful examination of the research. The literature reviewed and rationale for evidence-based instructional methods herein provides a foundation for an examination of questions in curriculum-based measurement techniques, and their use with students in a middle school setting. These answers will allow more detailed and future investigations to refine aspects of curriculum-based measures for use with students at the secondary level to provide the most effective instruction. The questions that will be answered by this study include:

1. Are there differences by ethnicity across the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?

2. Are there differences by setting (self-contained vs. resource) across the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?
3. What is the concurrent validity of the assessment measures (Maze, Oral reading fluency)?
4. What is the predictive-criterion related validity of the assessments (Maze, oral reading fluency)?

Chapter 2

Review of the Literature

Introduction

Curriculum-based measurement has been used across a variety of assessment activities including, screening, pre-referral evaluation, placement in remedial and special education programs, formative evaluation, and evaluation of reintegration and inclusion (Deno, 2003). Procedures for measurement that are drawn directly from the instructional materials used by the classroom teacher are referred to as curriculum-based. Evidence has shown that the same procedures can be used successfully with materials drawn from other generic sources that are commonly referred to as general outcome procedures (Fuchs & Deno, 1994), or dynamic indicators of basic skills (Shinn 1998). In contrast to curriculum-based assessment, which has been used to refer to a wide range of informal assessment procedures, curriculum-based measures refer to specific standard procedures (e.g., technical adequacy, standard tasks, prescriptive materials, administering and scoring, performance sampling, multiple samples, time efficiency). The reliability and validity of curriculum-based measurement have been achieved through using standardized procedures for sampling performance on reading, writing, and math skills. The standard tasks identified for use in curriculum-based measurement involve reading aloud from text and selecting words deleted from text (maze) in reading, writing word sequences when given a story starter or picture in writing, writing letter sequences, and writing correct answers/digits in solving problems in math computation (Deno, 2003). Next, the material for use in curriculum-based measurement may be obtained from instructional materials used by local

school with key factors being the selection process of the materials to increase the utility of the procedures for making instructional decisions (Shinn, 1989). Curriculum-based measurement procedures include specifics of sample duration, administration, student directions, and scoring procedures. The prescriptive selection of materials along with the standardization of the procedures is necessary to ensure reliability and utility of the data for individual and group comparisons (Shinn, 1995). Also, curriculum-based measurement monitors academic performance through the use of direct observation procedures. All of the scores for curriculum-based measurement are obtained by counting the number of correct and incorrect responses made in a fixed time period (Deno, 2003). Another important aspect of curriculum-based measurement is that performance is repeatedly sampled across time. Task difficulty is held constant and inferences can be drawn regarding the generalizations of student proficiency at reading unfamiliar but comparable text. Finally, curriculum-based measurement is designed for efficiency and the samples are 1 to 3 minutes in duration depending on the skill being measured and the number of samples necessary to maximize reliability.

Curriculum-based measurement as a Problem-Solving Model

Problem-solving models are used as an important component at the pre-referral level. Many of the common features of problem solving models include the definition of the problem, direct measure of the problem, design of the interventions, and the monitoring progress with the revisions for interventions and evaluating the outcomes. However, most research tends to measure outcomes of students in the pre-referral process based on perceptions of progress by professionals in education, as opposed to the direct tracking of the student progress (Weishaar, Weishaar, & Budt, 2002).

Curriculum-based measurement and continuous monitoring of student progress has been well established by research (Fuchs & Fuchs, 1986; Deno, 1986). When using curriculum-based measurement, students respond directly on brief probes that are taken from the current curriculum used in the classroom. Curriculum-based measurement is being used by problem-solving teams with local norms to help identify students' difficulties and to monitor the effectiveness of pre-referral interventions for reading, spelling, mathematics, and written expression (Weishaar, Weishaar, & Budt, 2002; Deno, 2002, Fuchs, 2002).

Weishaar, Weishaar, & Budt (2002) stated how the problem solving team using the model of curriculum-based measurement was useful in assisting the team to identify the interventions and focus the team to four areas (i.e., modifying teaching strategies within the regular classroom, referral to counseling, involving the parents, and behavioral modification). The problem-solving team was successful in working toward the vision of providing assistance to children experiencing difficulty in the regular classroom. Utilizing curriculum-based measurement along with local norms allowed the team to compare students to their local peer group and to begin tying assessments to intervention. Furthermore, this led the team to effectively plan, promote, and implement new services that ultimately benefited all children and contributed to a positive impact of decreasing the number of referrals to special education.

Related Research

Curriculum-based measurement can be thought of as a type of curriculum-based assessment. Three key features define curriculum-based assessment. It is a vehicle for (a) focusing instruction on the tasks of evaluating what needs to be taught and how best to teach it, (b) monitoring individual students' responses to instruction, and (c) using those performance data to make decisions about individually appropriate decisions for students with learning

problems (Jones, Southern, Brigham, 1998). Fuchs and Deno (1991) have identified the two models of mastery measurement and general outcome measurement. They define mastery measurement as the breaking down of global skills into a set of sub-skills, which are used as short-term objectives. The sub-skills are then taught and measured in sequence identifying short-term progress. Fuchs and Fuchs (1997) define general outcome measurement as using standardized procedures and long-term goals, which are similar to the same testing procedures remaining constant over a long period of time. Based on these distinctions, most curriculum-based assessment models fall under mastery measurement because they include teacher-made tests that incorporate task analysis which requires that different items and tests be developed for each skill that is taught (Deno, 1992). However, curriculum-based measurement falls under general outcome measurement because it uses repetitious measures of global skills. These measures use different but equivalent forms of representation of sub-skills embedded in the global skills (Fuchs, Fuchs, Hamlett, & Stecker, 1990).

Curriculum-based measurement is a standardized process which allows for valid and reliable information to be obtained as well as providing guidelines and specific procedures on how to select testing material (i.e., the probes; Deno, 1992). However, curriculum-based assessment reliability and validity data for mastery are unknown because they are not standardized and oftentimes designed by teachers (Hosp & Hosp, 2003). Furthermore, curriculum-based measurement is more distinct from other models of curriculum-based assessments by (a) short time to administer, (b) short time to score, (c) suitable for multiple forms for progressing monitoring, (d) inexpensive to develop and produce, and (e) sensitivity to student improvement over time (Jenkins, Deno, & Mirkin, 1979).

Curriculum-based measurement and Norm-referenced test

In the United States, recent national programs such as Reading First and No Child Left Behind (NCLB, 2002) have been put into place by the federal government requiring educators to deliver a well-planned curriculum using evidence-based instructional methods. To assist professionals, expert panels were assembled to review instructional, intervention, and assessment products and processes to determine whether specific procedures qualify as scientifically based. The panels developed lists of procedures from which professionals can select for use in schools (Kame'enui, 2002). As an example of one effort, the creation of a Reading First Assembly Academy Assessment Committee reviewed large number of assessment practices in various areas of reading (i.e., screening, diagnosis, progress monitoring, and outcome assessment). Many approved instruments are appropriate for summative evaluation of various reading components (e.g., phonemic, awareness, comprehension). A much smaller number are appropriate for formative evaluation or assessment that is sensitive to instructional growth and intervention effects. With general education being mostly affected by NCLB, curriculum-based measurement is generally recognized in special education as a valid and reliable indicator of student performance, but has been used relatively rarely in general education despite encouragement (Shinn, Shinn, Hamilton, & Clarke, 2002).

Curriculum-based measurement has many advantages (Shinn, 1989) in that data can be compared to local norms, and assessment can be conducted repeatedly to evaluate intervention effects and curriculum modifications. Curriculum-based measurement in reading has been shown to be a reliable, valid, and unbiased measure for screening, monitoring performance, and instructional decision making with criterion-related validity coefficients ranging from .73 to .81 (Hintze, Callahan, Mathews, Williams, & Tobin, 2002; Martson, 1989). As stated in a study

conducted by Ardoin et. al. (2004), significant correlations were discovered between Curriculum-based measurement-Maze and the Woodcock-Johnson III, and Iowa Test of Basic Skills subtest in conducting a universal screening in reading. The results of the study suggested that a single probe is sufficient for purposes of universal screening when being compared to norm-referenced assessments. The use of curriculum-based measurements by schools was considered to be a quick, and cost-efficient screening device for immediate use by schools for identifying early intervention for students with reading difficulties. Furthermore, the study stated how most group-administered norm-referenced tests are costly, require a substantial time to administer, and it is often several months before schools are provided with the results. For example, students that were administered the Iowa Test of Basic Skills in this study had to wait several months before the results were returned. District school psychologists often reported that decisions were usually made without the norm-referenced data because of the late arrival in the school year of the performance results.

Implementation of curriculum-based measurement at the secondary level

Curriculum-based measurement was initially developed to help teachers at the primary level to increase the achievement of students struggling to learn basic skills in reading, writing, and math (Shinn, 1989). As the process of developed has moved forward, educators at the secondary level have become interested in the application of similar evaluation approaches with their students. Due to the demand, technical work has proceeded on establishing curriculum-based measurement progress monitoring techniques for assessing student growth both in advanced academic skills and in content area learning (Espin, Scierka, Skare, & Halvorson, 1999). The technical aspects of developing curriculum-based measurement methods to assess growth in reading and writing at the secondary level have generated outcomes that appear both

promising and tentative. Attempts to establish the criterion validity of the same writing and reading measures that have been used at the elementary level have revealed that same measures correlate with the same criteria (e.g., test scores, averages), however, the correlations are not as strong for the secondary levels. However, a recent study conducted by Martson (2002) revealed high correlations for students at the middle school grades. However, more research is needed at the secondary level (middle and high school).

Approach model of Curriculum-based measurement

Traditional special education models and service delivery has led to problems of overrepresentation of minority students as well as prompted discussion and controversy over the process by which children are identified for special education (Twenty-Second Annual Report to Congress, 2000; Larry P. v. Riles, 1986; Reschly, 1984; Turlington, 1986). Also, the quality and value of traditional special education service delivery models have prompted procedural changes and modifications in different categories (Grossman, 1983; Hodapp, 1995; Polloway & Smith, 1983). Curriculum-based measurement integrates key concepts from traditional measurement theory and classroom-based observational methodology to create an approach to assessment. All curriculum-based measurement assessments sample a broad range of skills that are representative of the curriculum for a particular classroom and grade level. Curriculum-based measurements rely on broad samplings and standardized administration and scoring procedures to provide a performance indication for a particular individual. The performance score represents an individual's global level of competence within a specific domain (Fuchs, 2004). Professionals and practitioners can use this performance indicator to identify discrepancies in performance levels between individuals and peer groups, which can provide informed decisions about the need for special education services or identify the amount of support needed (Fuchs & Fuchs,

1997). Curriculum-based measurement provides repeated performance sampling, fixed time recording, graphic displays, and qualitative descriptions of student performance. These classroom-based observational methods permit estimates of slope for different time periods and alternative interventions for the same individual. Furthermore, this creates the data that is necessary for testing the effectiveness of different treatments for a certain student. Also, research states that these time-series analytic methods result in better instructional interventions, higher expectations, revisions to instructional programs, and better achievement for the students (Fuchs, Fuchs, & Hamlett, 1989; Fuchs, Fuchs, Hamlett, & Stecker, 1991).

In addition, each curriculum-based measurement measure samples the multiple skills embedded within the curriculum, providing descriptions of student performance to supplement the graphed, and quantitative analysis of performance (Fuchs & Fuchs, 1997). This data provides profiles of the individual and offer the advantage of being embedded within the local curriculum and provide strategies for improving student programs. Also, this information an prove to be useful for the teachers resulting in planning more varied, specific, and responsive instruction to meet individual student needs (Fuchs, Hamlett, & Fuchs, 1991).

Curriculum-based measurement and intelligence discrepancy model

Concerns about the use of Intelligence Quotient discrepancy formula to determine Learning Disability eligibility has been voiced recently by leading scholars and policymakers in the discussion of the reauthorization of Individuals with Disabilities Education Act (National Academy of Sciences, 2002; National Association of School Psychologists, 2002; President's Commission on Excellence in Special Education, 2002). For example, administering IQ tests to all students considered for LD was viewed as problematic by Minneapolis special education department (Ysseldyke & Marston, 1999). Many of the education leadership staff advocated that

intelligence tests be administered only when deemed appropriate and essential for the referred student. The staff and department were guided by a philosophy of (a) the appropriateness of IQ tests for eligibility determination, (b) discriminatory procedures and outcome bias, (c) allocation of school psychologists time, and (d) linking assessment to instruction. The Martson, Muyskens, Lau, and Canter (2003) study stated that using CBM measures instead of the regular discrepancy model kept identification numbers constant of students being identified for special education services. The use of curriculum-based measurement improved the assessment and decision-making process of special education along with assisting the regular education's mission of accelerating the learning of all students. The use of the CBM model, (a) improved screening of all students struggling in academic and behavior in general education settings, (b) improved the academic and behavioral interventions for those students struggling in general education, (c) increased the range of academic and behavioral interventions and cultural diversity training for all staff, and (d) improved special education evaluations.

Progress monitoring with Curriculum-based measurement

Although curriculum-based measurement is used for screening and categorical classification in some school districts (Shinn, 1989), its primary use is to measure the outcome of instruction. Fuchs, Fuchs, and Stecker (1989) discovered that teacher usage of curriculum-based measurement was related to the use of more objective data sources for determining the adequacy of student progress and more frequent decisions to modify the instruction for the students.

Stecker and Fuchs (2000) examined the importance of designing student programs based on individual progress-monitoring data. The study consisted of 22 special education teachers monitoring the math progress of 42 students in grades 2 through 8 with mild to moderate learning disability and emotional/behavior disability using curriculum-based measurement. The

teachers made instructional adjustments for the 42 students based on curriculum-based measurement data along with instructional adjustments for a matched group of students. The results indicated that students for whom teachers tailored instructional adjustments based on those students' own CBM data performed significantly better on a global achievement test as compared to the students whose instructional adjustments were not based on their own data.

Shinn, Gleason, and Tindal (1989) examined the effects of passage difficulty level on the performance of students receiving special education or remedial reading instruction. The progress of 30 students in grades 3 through 8 was monitored in one of two different measurement conditions. The first condition was one level of the curriculum below and one level above instructional placement. The second condition was two and four curriculum levels above instructional placement. The curriculum-based measurement reading data were collected four days per week for 4 weeks. The results suggested no significant differences in the slope of improvement within condition as a function of difficulty level or from those sections in the curriculum from which progress-monitoring reading probes were selected. This study recommended that materials should be drawn from curricular levels at which students should be expected to be placed in approximately one year.

When used to index student progress, curriculum-based assessment has been shown to be sensitive to student change over time (Fuchs, 1986, 1993; Fuchs & Fuchs, 1986). Additionally, in being sensitive to the effects of instruction, curriculum-based measurement has shown to be influenced by variables other than instruction. For example, the basic curriculum-based measurement datum in reading (i.e., oral reading fluency) can be affected by variables such as who administers the reading passages and where the reading passages are administered (Derr & Shapiro, 1989; Derr-Minneci & Shapiro, 1992) and the level in the curriculum used for probe

development (Dunn & Eckert, 2002; Hintze, Daly, & Shapiro, 1998). When used in a time-series manner, formative decision making and evaluation may be affected by how many data points are available for inspection or the context in which students are being evaluated (e.g., is progress being judged within individual or according to some group standard; Shinn, Powell-Smith, & Good, 1996). Also, decision-making and evaluation may be affected by the nature of the curriculum used for assessment (Hintze, Shapiro, & Lutz, 1994), or by the number of data points used for determining slope (Shinn, Good, & Stein, 1989).

Another important consideration when using curriculum-based measurement for progress monitoring is the manner in which the actual probes are developed (Hintze & Christ, 2004). For example, Hintze, Shapiro, and Lutz (1994) found that the type of curriculum used in the sampling process could significantly change the type of growth that might be observed over time for a student. Reading curricula that were characterized by uncontrolled readability and vocabulary proved too difficult for students and insensitive to growth over time. In another study, Hintze and Shapiro (1997) found that by selecting text and controlling for readability and vocabulary content from otherwise uncontrolled material, sensitive progress monitoring growth information could be obtained that mirrored the type of growth that would be seen in controlled text. The results from this study suggested that when selecting reading material for use in curriculum-based measurement progress monitoring, that assessors closely monitor the difficulty of the chosen text making sure that the readability and vocabulary were appropriate for the given grade to help ensure that slope estimates are sensitive to instruction and growth over time.

The effects of material selection on the amount of measurement error that is observed over time when monitoring progress is another concern for curriculum-based measurement (Christ, 2002). Some studies have found low amounts of measurement error relative to slope

over a progress monitoring time period of 1 to 2 years (Deno, Fuchs, Marston, & Shin, 2001; Fuchs, Fuchs, Hamlett, Walz, & Germann, 1993), whereas other studies conducted over a shorter period of time (e.g., 10-12 weeks) have found significant amounts of measurement error relative to slope or growth data (Hintze & Shapiro, 1997). Shorter progress monitoring periods lead to less stable growth estimates and increased measurement error (Christ, 2002). Large amounts of measurement error relative to slope can have effects on the precision of educational decision-making. With error estimates that are as large or greater than slope estimates, practitioners have means of knowing whether decisions are based on the product of the actual construct of interest (i.e., slope), or of measurement error caused by short monitoring periods and variations of the material used for assessment (Hintze, Owen, Shapiro, & Daly, 2000).

Curriculum-based measurement and learning disabilities

Through the usage of curriculum-based measurement in formulating eligibility decisions, learning problems can be viewed as the discrepancy between actual and expected levels of academic competence demonstrated within the classroom (Fuchs & Fuchs, 1997). The curriculum-based measurement approach offers several advantages over traditional learning disability identification models. First, the curriculum-based measurement model is less expensive than traditional models of identification. Poland, Thurlow, Ysseldyke, and Mirkin (1982), mentioned how the cost estimates for the typical special education eligibility assessment are often higher than reimbursement for educating a child with a mild disability for an entire year. Also, school psychologists typically devote most of their time to formal testing on commercial norm-referenced instruments (Goldwater, Meyers, Christenson, & Garden, 1984), and school districts allocate a lot of money to purchase those tests (Marston, 1989). However,

CBM eligibility process is time efficient and cost efficient because commercial materials are not required (Fuchs & Fuchs, 1997).

In addition, curriculum-based measurement is linked closely to and generates information about the problems that prompted the assessment initially (i.e., academic difficulty), as well as it simultaneously informs instructional planning as it identifies students for services by pointing out the academic deficits (Fuchs, & Fuchs, 1997). The repeated administration of several curriculum-based measurement tests provides educators with information about the instructional level of the student. Also, the measures create descriptions of the students' strengths and weaknesses in the local school curriculum. Marston et al.(1984) stated that CBM referrals as compared to teacher referrals were more consistent with the district criteria for LD eligibility.

Curriculum-based measurement and Emotional/Behavior disorders

Curriculum-based measurement is a procedure that was developed by individuals involved in special education for usage in special education. All of the procedures involve the direct observation of behavior and use the single case analytical procedures that are characteristic of applied behavior analysis (Deno, 1997). Applied behavior analysis is a system developed for use with any behavior in any setting. Curriculum-based measurement is based on procedures derived from sources outside of special education, however, the most extensive applications of applied behavior analysis have been in special education, and the early applications of applied behavior analysis occurred to academic instruction most often within special education (e.g., Lovitt, 1976; Fuchs, 2003). Crawford, Tindall, and Stieber (2001) mentioned more studies are clearly needed to examine the technical adequacy of curriculum-based measurement for students with emotional and behavior disorders (EBD). Students with emotional and behavior disorders are variable populations with many unique characteristics that may affect the technical adequacy

of the measures. Having additional studies will assist in buttressing the proposition that curriculum-based measurement can be used to accurately inform progress for students with emotional and behavior disorders, especially those who have academic deficits. Moreover, some students with emotional and behavior disorders may be displaying their problem behaviors to escape and avoid difficult academic tasks. Once the technical adequacy of curriculum-based measurement measures have been established for instructional problem-solving, research will be needed to examine how these measures may be used within a function-based approach to intervention

planning (Espin & Tindal, 1998).

Curriculum-based measurement and technical adequacy

Shin, Deno, and Espin (2000) conducted a study to observe the technical adequacy of curriculum-based measurement for assessing students' growth over time on curriculum-based measurement-Maze tasks. The results of their study indicated that maze had alternate-form reliability with a mean coefficient of .81 between testing during a 1 to 3 month interval. Alternate-form reliability estimates in this study are compatible with those in early research on alternate-form reliability (Bradley, Ackerson, & Ames, 1978; Marston, 1989; Parker, Tindal, & Hasbrouck, 1989).

Martson (1989) conducted four studies that assessed the validity of curriculum-based measurement measures total words written, words correct, and correct letter sequences using the Test of Oral and Written Language (TOWL), and Stanford Achievement Test (SAT). He found that correlations between total words written and criterion measures ranged from .41 to .84 for words correct. Correlations of words spelled correctly and the criterion measures ranged from

.45 to .92 (Martson, 1989). Correct letter sequences were between .57 and .87 (Naquin & Slider, 2002).

For several years, studies examining the validity of curriculum-based measurement oral reading fluency have been conducted. The findings have provided strong support for oral reading fluency as a valid measure of student's general reading skills (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Martson & Deno, 1882). Numerous studies have examined curriculum-based measurement's relations with published norm-referenced reading achievement tests such as the Stanford Achievement Test, the Stanford Diagnostic Reading Tests, and the Woodcock Reading Mastery Tests. Also, there is evidence of criterion-related validity including relations between reading fluency and criterion-referenced basal reading mastery tests (Fuchs & Deno, 1981; Tilly, 1989). Typical correlation coefficients in these studies have ranged from .60 to .90 with most correlations around .80 (Shinn & Good, 1992).

Oral Reading Fluency

The curriculum-based measurement oral reading fluency is a short fluency-based measure of oral reading performance. It is based on a one-minute timed sample of reading behavior. The student is asked to read out loud and the examiner marks the number of errors that are made (Martston, 1989). Using standardized procedures, students typically read orally from a level of their reading series for a set of repeated, one-minute timings. This is rich quantitatively counting the number of words read correctly per minute. This metric has been used to make special education problem identification (screening) and problem certification (eligibility) decisions, write individualized education plans objectives and monitor the effectiveness of reading instructional intervention, and determine when reading problems have been resolved (Shinn & Good, 1992).

Research on Oral Reading Fluency

Fewster and Macmillan (2002) conducted a study using curriculum-based measurement oral reading fluency and writing for the screening and placement of 465 middle and high school students. Students in grades 6 and 10 were compared with their year-end English and social studies grades received in each subject area. The students' grades from their permanent records were compared with the curriculum-based measurement measures. Curriculum-based measurement probes were developed according to procedures outlined by Tilly and Carlson (1992) for words read correctly and words spelled correctly. All of the results indicated a significant correlation between the results on the curriculum-based measurement measures and the grade level performance for the students. Words read correctly showed a significantly higher correlation with students' grades than words spelled correctly in the written expression. There were higher correlations with words read correctly and English than with the social studies scores. Also, the correlations were the highest for the students in the eighth grade as compared with the other grade levels in middle and high school.

Hintze (1997) conducted a study that assessed the effects of the curriculum on the technical features of curriculum-based measurement in reading. The study compared the association or criterion-related validity of survey-level curriculum-based measurement using literature-based basal reading material and authentic trade books. A total of 57 students enrolled in second, third, and fourth grades from one elementary school located in a suburban school district in the Northeast. There were 32 males and 25 females of whom 82% were white, 9% were African-American, 4 % Latino, and 5% were Asian. Eighty-six percent of the students received their reading instruction within the regular classroom, while 14% received either remedial or special education services. All of the students were primarily instructed in the

literature-based basal series with authentic trade books used as supplementary material. Three passages of at least 200 words were used for every grade level using curriculum-based measurement oral reading fluency. The results indicated that the concurrent validity of curriculum-based measurement oral reading measures were strong with a .665 for the authentic trade book series and a .655 for the literature-based basal series. The measure was similar regardless of the reading material that was used. In addition, developmental fluency rates were also similar across the two curricula.

Madelaine and Wheldall (1998) conducted a study using a curriculum-based passage reading test for monitoring the performance of low readers. This study examined the criterion validity of Wheldall Assessment of Reading Passages a curriculum-based measurement measure (Wheldall, 1996) against other established standardized reading tests (Neal Analysis of Reading-Revised, Neale, 1988; Multilit Word Attack Skills Placement Test, Macquarie University Special Education Center, 1996). This study was conducted with 50 students from a Sydney independent school with 25 males and 25 females. The students were given five passages from the WARP and their scores were compared with two other standardized reading measures. The results indicated criterion validity with reading accuracy of .83 to .87 when comparing phonic word attack skills test, .67 to .72 on reading comprehension, and .75 to .78 word attack skills. Inter-correlations between the five passages were shown to be very high at .94 to .96 demonstrating high alternate forms reliability.

Maze

The MAZE procedure of curriculum-based measurement requires the selection of a grade-level passage of at least 250-words. The first and last sentences of the passage are left intact. Next, a group of words is inserted for every fifth word. The student is asked to select the

original word by circling it from among three to five distracters within a three to five minute time period. The difficulty of maze will vary according to the difficulty of the passage and the difficulty of the distracters selected (Howell & Nolet, 2000).

Research on MAZE

Shin, Deno, and Espin (2000) conducted a study using 43-second graders from a large urban area in the Midwest. The study used ten different MAZE passages to assess students' reading performance over a school year. Passages were randomly selected from generic grade-level reading materials. The curriculum-based measurement-MAZE was found to have adequate technical characteristics. Results indicated that the MAZE task had an alternate-form reliability of .81. The MAZE task also sensitively reflected improvement of student performance over a school year and revealed inter-individual differences in growth rates. Additionally, growth rates estimated on repeated MAZE scores were positively related to later reading performance on a standardized reading test.

Faykus, McCurdy, and Barry (1998) conducted a study using six students with special needs in suburban Philadelphia. The study used two curriculum-based measurement procedures (ORF and MAZE) to measure student progress in reading. The students were examined twice weekly for 12 weeks. The results indicated that oral reading rates might be a more efficient indicator of reading progress than maze within the 12 weeks that data was recorded. It appears that oral reading rates resulted in fewer data points falling at or above the progress goal line than maze with this particular group of students. However, it was mentioned that one reason for the higher results with oral reading fluency might have been attributed to the amount of components. Oral reading fluency (fluent decoding, fluency) uses fewer components than maze (fluent

decoding, fluency, and comprehension) in evaluating student progress in reading. Teachers found both procedures (oral reading fluency and maze) to be useful measures for reading.

Conclusion

Frequent assessment is essential for the measurement of student performance and learning. The research findings presented provide a basis for using curriculum-based measurements to set standards, evaluate programs, and student performance. Curriculum-based measurements provides a validated set of procedures that allow classroom teachers flexibility in measuring a student's performance and growth in math, reading, and writing. In addition, curriculum-based measurement provides immediate and accurate feedback on achievement and performance to teachers, students, and parents. Also, any negative performance trends are immediately identified to allow educators essential time to make quick response changes in a students program. Furthermore, the flexibility of curriculum-based measurements enables educators to align the assessment with the current curriculum and allows for a range of response modes. Curriculum-based measurements allow educators to set growth standards for both general and special education students using curriculum-based measurements as the repeated measure of students' performance.

The studies presented in this literature review demonstrates that curriculum-based measurements are sensitive enough to be used in comparison with standardized and state assessments in assessing a student's reading ability. Teachers must recognize curriculum-based measurement as a general indicator and interpret its results in combination with various other types of assessments and data (Fewster & McMillan, 2002). Curriculum-based measurement results can be helpful in identifying problems that warrant future investigations for the students in the classroom. The information from curriculum-based measurement measures can be

extended into the development of appropriate intervention plans directly related to a student's current curriculum.

Researchers may want to focus further investigation into considerations with the technical adequacy of curriculum-based measurement when using different curricula as it relates to growth rate. For example, having the research data to show reading growth patterns during the year, as well as, individual growth rates during different academic years. Another cause for a further research investigation is whether or not curriculum-based measurement reading strategies mentioned in some of these studies would have the same benefits if used with students that have intense reading problems and deficits.

Another implication to be considered for research is that standards must be established, defined, and replicated for the best rules and scoring guidelines for curriculum-based measurement measures along with the predictive value of the alternate assessments. Finally, more research is warranted to explore further the use of curriculum-based measurements in predicting how students will perform on standardized assessments at the primary and secondary levels. The proposed study will build on current data concerning curriculum-based measurement to indicate the development of appropriate intervention plans directly related to a student's current curriculum and having data to show reading growth patterns and progress monitoring during the academic year. The data collected from this investigation will assist practitioners with selecting the most efficient and effective strategies for monitoring student progress. Furthermore, this study will build upon the foundation for additional inquiries into the various characteristics of the effective uses for curriculum-based measurement for classroom teachers.

Table 1

Summary of Research

Citation	Participants	Assessments	Results
Espin, Scierka, Skare, & Halverson (1999)	147 high school students (7 special education students)	CBM-written expression California Achievement Test	The results indicated that sentences and characters per word were the most strongly and consistently correlated with the writing performance measures. Also, a consistent pattern of relations was also found for correct word sequences and mean length of the correct word sequences. With the exception of the sentences and correct word sequences, all measures had a consistent and reliable pattern of relations with the other measures of writing proficiency.
Fuchs, & Fuchs (1991)	59 students (special education)	CBM-spelling S-MAIRS (Spelling-Modified Accuracy of Implementation Rating Scale-Revised)	The results indicated that the achievement of the curriculum-based assessment groups were comparable, but greater than that of the control group. The CBM teachers increased goals for approximately 90% of their students.

Citation	Participants	Assessments	Results
Shinn, Gleason, & Tindal (1989)	30 students in grades 3-8 (special education students)	CBM-oral reading fluency	Results suggested no significant differences in the slope of improvement within condition as a function of difficulty level or from those sections in the curriculum from which progress-monitoring reading probes were selected.
Fewster, & Macmillan (2002)	465 students in grades 6-10	CBM-oral reading fluency CBM-writing	Results indicated words read correctly showed a significantly higher correlation with students' grades than words spelled correctly in the written expression. There were higher correlations with words read correctly and English than with the social studies scores. Also, the correlations were the highest for the students in the eight grade as compared with the other grade levels in middle and high school.
Hintze (1997)	57 students in grades 2-4 (14%special education)	CBM-oral reading fluency	The results indicated that the concurrent validity of CBM oral reading measures were strong with a .665 for the authentic trade book series and a .655 for the literature-based basal series. The measure was similar regardless of the reading material that was used.

Citation	Participants	Assessments	Results
Madelaine, & Wheldall (1998)	50 students (elementary school)	CBM-WARP (Wheldall Assessment of Reading Passages) NAR-R (Neal Analysis of Reading-Revised)	The results indicated a criterion validity with reading accuracy of .83 to .87 when comparing phonic word attack skills test, .67 to .72 on reading comprehension, and .75 to .78 word attack skills. Inter-correlations between the five passages were shown
Shin, Deno, & Espin (2000)	43-second graders	CBM-MAZE	Results indicated that the MAZE task had an alternate-form reliability of .81. The MAZE task also sensitively reflected improvement of student performance over a school year and revealed inter-individual differences in growth rates. Additionally, growth
Faykus, McCurdy, and Barry (1998)	6 students (special education)	CBM-MAZE CBM-oral reading fluency	The results indicated that oral reading rates might be a more efficient indicator of reading progress than maze within the 12 weeks that data was recorded. It appears that oral reading rates resulted in fewer data points falling at or above the progress

Citation	Participants	Assessments	Results
Gansle, Noell, VanDerHeyden, Naquin, & Slider (2002)	179 elementary students	CBM-written expression ITBS (Iowa Test of Basic Skills)	Results of the comparison between the criterion measure (ITBS) and the CBM measures, the highest correlation with a reliability of .59 occurred with the language usage/expression and total writing subscale score. Words in correct sequence had a correlation of .46 with the ITBS, while total words written appeared not to be useful for predicting skill in written language as measured by criterion measures for third and fourth graders.
Naquin, & Slider (2002)	57 students in grades 2-4 (14%special education)	CBM-writing Louisiana Educational Assessment ITBS (Iowa Test of Basic Skills)	The results indicated that total words written were not perceived to be meaningful or instructionally useful indicators for students at the elementary level.
Shinn, & Good (1992)	238 elementary students (4% special education)	CBM-written retell Stanford achievement test	The results of the correlations coefficients for the 3rd graders were .56 and .50. The correlations for the 5th graders were .43 and .51. The correlation coefficients for the 3rd and 5th graders were significantly different with making comparisons to the Stanford achievement test demonstrating a high alternate forms reliability.

Citation	Participants	Assessments	Results
Allinder et. al. (2001)	50 student in middle school (11% special education)	CBM-oral reading fluency CBM-MAZE Woodcock Reading Mastery-Revised Clinical Evaluation of Language Fundamentals-Third Ed.	Results indicated that all students improved on a standardized norm-referenced reading measure, but students who used a specific oral reading strategy made significantly greater progress in reading, as measured by curriculum-based measurement maze procedure.
Hintze, & Conte (1997)	57 students in elementary school (14% special education)	CBM-oral reading fluency Degrees of Reading Power Test	Results indicated that the correlation between survey-level CBM and reading comprehension was similar regardless of the material used for assessment.
Hintze, Daly III, & Shapiro (1998)	80 students in elementary school (12% special education)	CBM-oral reading fluency	Results indicated that the amount of progress observed (i.e., slope of improvement) varied as a function of grade and whether student progress was monitored in grade or goal level material.
Hintze, & Christ (2004)	99 students in elementary school (2 % special education)	CBM-oral reading fluency	The results suggested that controlled reading passages significantly reduced measurement error as compared to uncontrolled reading passages, leading to increased sensitivity and reliability of measurement.

Citation	Participants	Assessments	Results
Tindal, & Parker - 2001	172 students in middle school (30% special education)	CBM-written expression California Achievement Test	The results suggested that regression of holistic ratings on objective scores produced moderately strong results for two production independent indices (percentage of words correctly spelled and percentage of words correctly sequenced.
McGlinchey, & Hixson (2004)	1,362 students in elementary school (40%special education)	CBM-oral reading fluency Michigan Educational Assessment Program	The results indicated a positive correlation between the two measures. The positive and negative predictive power of the reading sample was higher than the base rate of failing and passing the MEAP.
Stage, & Jacobson (2001)	173 elementary students (18% special education)	CBM-oral reading fluency Washington Assessment of Student Learning	The results indicated that slope in oral reading fluency across the school year and the oral reading fluency probes administered in Sept., Jan., and May reliably predicted May WASL reading performance.

Chapter 3

Method

Participants

The participants were students from a suburban middle school in the Southeast who were receiving services for emotional and behavior disorders. Fifty-five students receiving services for emotional and behavior disorders (EBD) were used in this study. The total population of the middle school was 2000 students. The total special education student population was 170 students. The middle school had 86 regular education teachers and 20 special education teachers. The middle school had 8 self-contained special education teachers and 12 resource special education teachers. Students from both classroom settings (self-contained and resource) were used in this study. Twenty-two percent of the special education population was on free and reduced lunch. The ethnicity of the school was approximately 70% Caucasian, 28% African American, and 2 % other. The poverty rate for the school was less than 12%.

The students that participated in the study were administered four different types of assessments (Maze comprehension, Oral Reading Fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency). The Maze comprehension relates to the accuracy and speed at which a student selects a word from a multiple-word group to properly complete sentences within a passage. The Maze comprehension procedure was administered to the students within large groups during a time period of three minutes per group. The Oral Reading Fluency relates to the speed and accuracy with which a student read words. The Oral Reading Fluency was administered to each student individually

during a one-minute time period across three different reading passages. The Woodcock-Johnson III subtests reading comprehension and passage comprehension were administered individually to each student during a five- minute time period. The Woodcock-Johnson III subtest reading fluency measures the person's ability to quickly read simple sentences to a moderate level. The Woodcock-Johnson III subtest passage comprehension requires the person to read a short passage and identify a missing key word that makes sense in the context of the passage.

Table 2

Race, N, and Percent of Total Sample

Race	N	Percent	Grade levels 6th	7th	8th
White American	25	45	10	8	7
African American	21	38	7	9	5
Latino American	5	9	1	3	1
Other*	4	8	2	1	1

Note. * The Other category consisted of Native American (2), Indian American (2).

Setting

The study took place in a suburban middle school (6th, 7th, and 8th grades) in the Southeast. The reading measures were administered in two special education classrooms (self-contained and resource). The self-contained room was composed of eight to ten students at one time. The resource room was composed of nine to fifteen students at one time. Each room consisted of individual desks along with a teacher's desk and two tables. Also, a conference room adjacent to the self-contained room was used for the purposes of this study to administer

the individual assessments. The conference room consisted of a teacher's desk, a round table, and four chairs. The Maze comprehension procedure was administered to the entire class, whereas the Oral Reading Fluency, Woodcock-Johnson III subtest reading fluency, and the Woodcock-Johnson III subtests passage comprehension were administered individually in a smaller adjacent room. Students that did not participate in the individual assessments were engaged in their regularly scheduled classroom activities working either individually, with the paraprofessional, or another teacher. The students that participated were engaged in the individual assessment process for approximately 30 minutes.

Curriculum

All of the students received the same format of instruction from the general education curriculum. The students received classroom instruction for all content areas (Language Arts, Science, Social Studies, and Math) as outlined in the performance-based standards mandated by the local school district and state department of education.

Instruments and Materials

All of the students were administered the two curriculum-based measures (Maze comprehension, Oral Reading Fluency) and the Woodcock-Johnson III Test of Achievement subtests (Woodcock-Johnson III reading fluency, Woodcock-Johnson III subtest passage comprehension). The Maze comprehension procedure was administered to the students within large groups during a time period of three minutes per group. The Oral Reading Fluency was administered to each student individually during a one-minute time period across three different reading passages. The Woodcock-Johnson III subtest reading comprehension was administered individually to each student during a three-minute time period. The Woodcock-Johnson III subtest passage comprehension is a non-timed assessment that was administered individually to

each student. The students' results for each assessment (Maze comprehension, oral reading fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency) were compared.

Woodcock-Johnson Test of Achievement (WJ-III). The Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001) is a standardized assessment that consists of two co-normed batteries (Woodcock-Johnson III Test of Cognitive Abilities and the Woodcock-Johnson III Tests of Achievement). The batteries together comprise a wide range and comprehensive system for measuring general intellectual ability, cognitive abilities, oral language, and academic achievement. The norms for the Woodcock-Johnson III are based on data from the same sample of subjects. This feature allows for direct comparisons among and within a subject's scores that have a degree of accuracy not possible when comparing scores from separately normed tests.

The Woodcock-Johnson batteries were designed to provide the most valid methods for determining patterns of strengths and weaknesses based on actual discrepancy norms. The Woodcock-Johnson Test of Achievement (Mather & Woodcock, 2001) provides a complete battery of co-normed tests for measuring general intellectual ability, specific cognitive abilities, scholastic aptitude, oral language, and academic achievement. Test-retest and concurrent validity of the Woodcock-Johnson III Test of Achievement are high. The test-retest coefficient of the Woodcock-Johnson III Test of Achievement is .93 (McGrew & Woodcock, 2001). The Woodcock-Johnson III Test of Achievement provides concurrent validity by presenting correlations between the Woodcock-Johnson III Test of Achievement and the Kaufman Test of Educational Achievement (.76; Kaufman & Kaufman, 1992 ; McGrew & Woodcock, 2001). Also, the Woodcock-Johnson III Test of Achievement technical manual provides concurrent

validity correlations with the Wechsler Individual Achievement Test (.67; McGrew & Woodcock, 2001; Wechsler, 1992).

Woodcock-Johnson-III subtest Ready Fluency. The Woodcock-Johnson Reading Fluency (Mather & Woodcock, 2001) measures the person's ability to quickly read simple sentences to a moderate level. The individual attempts to complete as many items as possible within a three-minute time limit. The Woodcock-Johnson III Test of Achievement subtest reading fluency has a test-retest reliability of .86 (McGrew & Woodcock, 2001). The concurrent validity for the Woodcock-Johnson III Test of Achievement subtest reading fluency is not specifically provided in the technical manual. The Woodcock-Johnson III Test of Achievement subtest reading fluency is included in a cluster (basic reading skills, reading comprehension) when providing correlations for validity. The Woodcock-Johnson III Test of Achievement reading cluster provides concurrent validity correlations with the Kaufman Test of Educational Achievement (.62; Kaufman & Kaufman, 1992; McGrew & Woodcock, 2001).

Woodcock-Johnson-III subtest Passage Comprehension. The Woodcock-Johnson Passage Comprehension (Mather & Woodcock, 2001) initial items involve symbolic learning, or the ability to match pictures to an object. Next, the items are presented in a multiple-choice format and require the person to point to the picture represented by a phrase. The remaining items require the person to read a short passage and identify a missing key word that makes sense in the context of the passage. The items become increasingly difficult by removing the pictorial stimuli and by increasing passage length, level of vocabulary, and complexity of syntactic and semantic cues. The Woodcock-Johnson III Test of Achievement subtest passage comprehension has a test-retest reliability of .88 (McGrew & Woodcock, 2001). The concurrent validity for the Woodcock-Johnson III Test of Achievement subtest passage comprehension is not specifically

provided in the technical manual. The Woodcock-Johnson III Test of Achievement passage comprehension is included in a cluster (basic reading skills, reading comprehension) when providing correlations for validity. The Woodcock-Johnson III Test of Achievement reading cluster provides concurrent validity correlations with the Kaufman Test of Educational Achievement (.62; Kaufman & Kaufman, 1992; McGrew & Woodcock, 2001).

Oral Reading Fluency. Oral reading fluency (Shinn, 1998) relates to the speed and accuracy with which a student reads. The students are given a one-minute passage to read through a 350 word reading passage correctly pronouncing words in each sentence. This process is repeated for three different reading passages. The students' performance is assessed by words read correctly minus the number of errors. An error is considered as any mispronunciation of the word or substitutions, omissions, or three-second pauses or struggles. If a passage is too difficult and the student reads less than ten words correctly in one minute, discontinue administration of any passages from that grade level and use that number as the words read correctly score.

Studies examining the validity of curriculum-based measurement Oral Reading Fluency have been conducted for several years. The findings have provided strong support for Oral Reading Fluency as a valid measure of students' general reading skills (Deno, Mirkin, & Chiang, 1982; Fuchs, Fuchs, & Maxwell, 1988; Martson & Deno, 1882). Numerous studies have examined CBM's relations with published norm-referenced reading achievement tests such as the Stanford Achievement Test, The Stanford Diagnostic Reading Tests, and the Woodcock Reading Mastery Tests. Also, there is evidence of criterion-related validity including relations between reading fluency and criterion-referenced basal reading mastery tests (Fuchs & Deno, 1981; Tilly, 1989). Typical correlation coefficients in these studies have ranged from .60 to .90 with most correlations around .80 (Shinn & Good, 1992).

For example, Hintze (1997) conducted a study that assessed the effects of the curriculum on the technical features of curriculum-based measurement in reading. The study compared the association or criterion-related validity of survey-level curriculum-based measurement using literature-based basal reading material and authentic trade books. A total of 57 students enrolled in second, third, and fourth grades from one elementary school located in a suburban school district in the Northeast participated in the study. There were 32 males and 25 females of whom 82% were white, 9% were African-American, 4 % Latino, and 5% were Asian. Eighty-six percent of the students received their reading instruction within the regular classroom, while 14% received either remedial or special education services. All of the students were primarily instructed in the literature-based basal series with authentic trade books used as supplementary material. Three passages of at least 200 words were used for every grade level using CBM oral reading fluency. The results indicated that the concurrent validity of curriculum-based measure Oral Reading Fluency was strong with a .665 for the authentic trade book series and a .655 for the literature-based basal series. The measure was similar regardless of the reading material that was used. In addition, developmental fluency rates were also similar across the two curricula.

Madelaine and Wheldall (1998) conducted a study using a curriculum-based passage reading test for monitoring the performance of low readers. This study examined the criterion validity of the curriculum-based measure WARP (Wheldall Assessment of Reading Passages, Wheldall, 1996) against other established standardized reading tests (Neal Analysis of Reading-Revised, Neale, 1988; Multilit Word Attack Skills Placement Test, Macquarie University Special Education Center, 1996). This study was conducted with 50 students from a Sydney independent school with 25 males and 25 females. The students were given five passages from the WARP, and their scores were compared with two other standardized reading measures. The results

indicated a criterion validity with reading accuracy of .83 to .87 when compared to phonic word attack skills test, .67 to .72 on reading comprehension, and .75 to .78 on word attack skills. Inter-correlations between the five passages were shown to be very high at .94 to .96, demonstrating a high alternate-forms reliability.

Maze comprehension. Maze (Shinn, 1998) relates to the accuracy and speed at which a student selects a word from a multiple-word group to properly complete sentences within a passage. The students are given a three-minute time limit. The first sentence of a passage remains intact; however, every nth word is deleted, and replaced with three choices. Under the time limit, the student selects an alternative that meaningfully replaces the blank. A three-item multiple-choice format with 50 word sets is employed with only one choice representing a semantically meaningful replacement. The students' performance is assessed by subtracting the number of incorrect answers from the total number of items attempted. If a student completes the passage ahead of time, the student's packet is collected, the time is noted on the paper, and the student's score is prorated. Prorating involves converting the student's time to seconds, dividing by the number correct, and dividing that number into 180 (representing 180 seconds= 3 minutes) to receive the student's score.

The Maze measurement has many advantages (Shinn, 1989) in that data can be compared to local norms, and assessment can be conducted repeatedly to evaluate intervention effects and curriculum modifications. Studies in curriculum-based measurement maze reading have been shown to be a reliable, valid, and unbiased measure for screening, monitoring performance, and instructional decision-making with criterion-related validity coefficients ranging from .73 to .81 (Ardoin et al., 2004; Shin, Deno, Espin, 2000).

For example, in a study conducted by Ardoin et.al. (2004), significant correlations were discovered between CBM-Maze and the Woodcock-Johnson III, and ITBS subtest in conducting a universal screening in reading. The results of their study suggested that a single probe is sufficient for purposes of universal screening when being compared to norm-referenced assessments. The use of curriculum-based measurements by schools was considered to be a quick, and cost-efficient screening device for identifying early interventions for students with reading difficulties. Furthermore, the study stated how most group-administered norm-referenced tests are costly and require substantial time to administer. In addition, it is often several months before schools are provided with the results. For example, students that were administered the ITBS in this study had to wait several months before the results were returned. District school psychologists often reported that decisions were usually made without the norm-referenced data because the performance results arrived so late in the school year.

Shin, Deno, and Espin (2000) conducted a study to observe the technical adequacy of curriculum-based measurement for assessing students' growth over time on CBM-Maze tasks. Forty-three second graders from a Midwestern city were used in this study. Pearson product-moment correlations and hierarchical linear regression models were used to test for the validity and reliability of the curriculum-based measures with the California Achievement Test. The results of their study indicated that maze had alternate-form reliability with a mean coefficient of .81 between testing during a one to three month interval. Alternate-form reliability estimates in this study are compatible with those in early research on alternate-form reliability (Bradley, Ackerson, & Ames, 1978; Marston, 1989; Parker, Tindal, & Hasbrouck, 1989).

The curriculum-based measures (Maze, Oral Reading Fluency) that were used for this study were commercially developed assessments developed by Aimsweb (Shinn & Shinn, 2002).

Aimsweb is a formative assessment system that provides performance data and reporting information. Aimsweb utilizes the curriculum-based measures to assess student performance.

Procedures

The students were administered four different types of assessments (Maze comprehension, Oral Reading Fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency). The assessments (Maze comprehension, Oral Reading Fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency) were administered twice (time 1 and time 2) during this study. The time period between the time one and time two administrations were nine weeks. The Maze comprehension was administered as a group to the students identified as having an emotional/behavior disorder within the self-contained classroom and the resource classroom. The Maze comprehension procedure was administered to the students within large groups during a time period of three minutes per group. Individual sheets with the reading passages were placed face down on each student's desk and a stopwatch was used to monitor the three-minute time limit. When the students were ready to begin, they were asked to turn over the sheet of paper containing the passage and the timing was started for the Maze comprehension assessment. At the end of the three-minute time limit, the students were asked to place their pencils down. The students' performance was assessed by subtracting the number of incorrect answers from the total number of items attempted.

The Oral Reading Fluency was administered to each student individually. The individually administered items were given in the conference room adjacent to the self-contained classroom. The Oral Reading Fluency assessment took approximately ten minutes per student. The students were asked to enter the conference room where they were presented with a sheet of

paper with a reading passage placed face down on the table. The students were asked to begin, and they were timed for one minute while reading aloud through a passage. The students were given a one-minute passage to read through a 350 word reading passage, correctly pronouncing words in each sentence. This process was repeated for three different reading passages. The students' performance was assessed by words read correctly minus the number of errors. An error was considered any mispronunciation of the word, substitutions, omissions, or three-second pauses or struggles.

The administration of the Woodcock-Johnson III subtests reading fluency was administered to each student individually. The testing procedure as outlined by the Woodcock-Johnson III Test of Achievement administration manual were followed for each student during the administration of the Woodcock-Johnson III subtest reading fluency. The administration of the Woodcock-Johnson III subtests passage comprehension was administered to each student individually. The testing procedure as outlined by the Woodcock-Johnson III Test of Achievement administration manual were followed for each student during the administration of the Woodcock-Johnson III subtest passage comprehension.

Reliability

The data examiners were trained on the use of the procedures for conducting the curriculum-based measure assessments as outlined in research on the curriculum-based measurement process (Fuchs & Fuchs, 1997; Shinn, 1998). The observer for the data collector checked procedural integrity. Using a checklist, an independent and trained observer noted whether procedures were being followed on a step-by-step basis. The procedures included whether all materials were present, time kept accurately, instructions given specifically, and whether students were following the procedures correctly. The examiners used a checklist to

ensure that examiners were consistent in the administration and scoring of the curriculum-based measures (see Appendix H).

Both scorers collected data on the student responses. In order to ensure consistency, the scorers practiced scoring five (5) students randomly chosen from the population of students that are participating. This data was compared using a point-by-point comparison in which agreements are divided by agreements plus disagreements and multiplied by 100 to compute a percentage. The same formula was used to calculate agreements during all of the curriculum-based measures. Equation 1 represents the formula to be used to calculate inter-rater reliability:

$$\left(\frac{Number\ of\ Agreements}{Number\ of\ Agreements + Disagreements} \right) X\ \ 100$$

After the competence level was reached, the scorers numbered the remaining students and divided them between the two scorers to be scored individually. However, both scorers rated every subsequent 5th student independently in order to determine the extent to which satisfactory rater reliability was sustained throughout the scoring process. Data collected during the procedural checklist and scoring of the assessment measures was 100% due to the sensitivity and threat to skewing the results. If there was a discrepancy in the scoring, the measures were rescored by the researcher and observer until 100% agreement was reached and then the data was entered.

Design

Repeated Measures analysis of variance (ANOVA) were used to compare differences by ethnicity and settings across the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension). Next, Pearson product moment correlations were conducted to determine to examine the validity of the curriculum-based measures. Finally, regression analyses were conducted to determine the

curriculum-based measures (Maze, Oral Reading Fluency) as predictors while being measured to the Woodcock-Johnson III subtests (reading fluency, passage comprehension).

Research Questions

Research question 1**:** Are there differences by ethnicity across the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?

A 2X2 Repeated Measures analysis of variance (ANOVA) was conducted for each ethnic group on each assessment (Maze, oral reading fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency) to address this research question. Also, repeated measures analyses were used to test for an interaction of main effects for the ethnic groups and the assessments (Maze, oral reading fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency).

Research question 2: Are there differences by setting (self-contained vs. resource) across the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?

A 2X2 Repeated Measures analysis of variance (ANOVA) was conducted for each setting (resource vs. self-contained) on each assessment (Maze, oral reading fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency) to address this research question. Also, the repeated measures analyses were used to test for an interaction of main effects for the setting (resource vs. self-contained) and the assessments (Maze, Oral Reading Fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency).

Research question 3: What is the concurrent validity of the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?

Pearson product-moment correlations were used to address this research question. A correlation matrix constructed on the performance results on the CBM measures (Oral reading fluency, Maze) as compared to the performance results on the Woodcock-Johnson III subtests (reading fluency, passage comprehension) was established. Next, regression analyses were conducted using the CBM measures (Oral Reading Fluency, Maze) as individual predictors while being compared to the Woodcock-Johnson III subtests (reading fluency, passage comprehension) as the criterion. Also, regression analyses were conducted using the CBM measures (Oral reading fluency, Maze) together as predictors while being compared to the Woodcock-Johnson III subtests (reading fluency, passage comprehension) as the criterion. Finally, analyses were conducted using the CBM measures (oral reading fluency, Maze) as the first and second variable interchangeably within a hierarchical model.

Research question 4: What is the predictive-criterion related validity of the assessments (Maze, oral reading fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency) from pre-test to post-test?

Pearson product-moment correlations were used to address this research question. A correlation matrix constructed on the performance results on the curriculum-based measures (Oral reading fluency, Maze) as compared to the performance results on the Woodcock-Johnson III subtests (reading fluency, passage comprehension) was established. Next, regression analyses were conducted using the curriculum-based measures (Oral reading fluency, Maze) together as a predictor while being compared to the Woodcock-Johnson III subtests (reading fluency, passage

comprehension) as the criterion. Finally, analyses were conducted using the CBM measures (oral reading fluency, Maze) as the first and second variable interchangeably within a hierarchical model.

Chapter 4

Results

Introduction

The purpose for the study was to examine the validity of curriculum-based measurement for students with Emotional and Behavior Disorders in a middle school setting. The curriculum-based measures (Maze, Oral Reading Fluency) and the Woodcock-Johnson III Test of Achievement subtests (Passage Comprehension, Reading Fluency) were administered to students with Emotional and Behavior Disorders. Comparisons were made between the curriculum-based measures curriculum-based measures (Maze, Oral Reading Fluency) and the Woodcock-Johnson III Test of Achievement subtests (Passage Comprehension, Reading Fluency) to answer various questions.

Research Questions

1. Are there differences by ethnicity across the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?
2. Are there differences by setting (self-contained vs. resource) across the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?
3. What is the concurrent validity of the assessment measures (Maze, Oral reading fluency)?
4. What is the predictive-criterion related validity of the assessments (Maze, oral reading fluency)?

Analysis of Research Questions

Are there differences by ethnicity across the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?

A set of 2 (time of assessment: pre vs. post) X 2 (ethnicity: Caucasian vs. African American) repeated measures analyses of variance (ANOVAs) were conducted for the four measures (Maze, oral reading fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency). For the Maze, table 3 presents the main effect of the time of assessment was significant F (1, 44) = 8.798, p = .005. There was no significant main effect for the Maze procedure F (1,44) = .120, p = .731. Table 4 shows the student performance at time two was higher than the scores for time one for both ethnic groups. African-American students scores for time two (M = 18.84, SD = 8.54) were higher than the scores for time one (M = 18.80, SD = 8.52). Caucasian students scores for time two (M = 22.60, SD = 8.84) were higher than the scores for time one (M = 18.84, SD = 4.72).

Table 3

ANOVA for Maze Testing

Source	SS	df	MS	F	*p*
Maze	597.705	1	597.705	8.798	.001*
Maze*race	8.139	1	8.139	0.12	0.731
Error	2989.263	44	67.938		

Note. race = black vs. white, *p < .001.

Table 4

Means for Maze Testing by Ethnicity

Measure	Ethnicity	N	*M*	*SD*
Maze time1	White	25	18.08	4.72
	Black	21	18.80	8.54
Maze time2	White	25	22.60	8.84
	Black	21	24.52	10.07

For the Oral Reading Fluency, table 5 shows the main effect of time was significant F (1,44) = 36.886, p = < .005. There was no significant main effect of ethnicity on student performance with the Oral Reading Fluency F (1, 44) =. 708, p = .405. Table 6 shows African-American students scores were lower at time one (M = 109.76, SD = 51.72) than time two (M = 158.52, SD = 46.09). Also, Caucasian students scores were lower at time one (M = 62.00, SD = 25.02) than time two (M = 126.44, SD = 60.98).

Table 5

ANOVA for Oral Reading Fluency Testing

Source	SS	df	MS	F	*p*
Orf	73127.2	1	73127.2	36.886	.001*
orf*race	1402.678	1	1402.678	0.708	0.405
Error	87229.985	44	1982.5		

Note. oral reading fluency = orf, race = black vs. white, *p < .001.

Table 6

Means for Oral Reading Fluency Testing by Ethnicity

Measure	Ethnicity	N	*M*	*SD*
orf time1	White	25	62.00	25.05
	Black	21	109.76	51.72
orf time2	White	25	126.44	60.98
	Black	21	158.52	46.09

Note. Oral Reading fluency = orf.

For the Woodcock-Johnson III Passage Comprehension, table 7 shows the main effect of time of assessment was not significant F (1,44) =2.547, p =. 118. There was no significant main effect of ethnicity for the Woodcock-Johnson III subtest Passage Comprehension F (1, 44) =. 088, p = .768. Table 8 shows Caucasian and African-American students scored approximately equally on the Woodcock-Johnson III Passage Comprehension between time one and time two. The Caucasians scores for time one (M = 75.96, SD = 9.61) and time two (M = 78.76, SD = 10.13) were similar. The African-American scores for time one (M = 77.76, SD = 8.20) and time two (M = 80.41, SD = 11.31) were similar.

Table 7

ANOVA for Woodcock-Johnson Passage Comprehension Testing

Source	SS	df	MS	F	*p*
Wjpc	247.43	1	247.43	2.547	0.118
wjpc*race	8.56	1	8.56	0.088	0.768
Error	4274.625	44	97.151		

Note. Woodcock-Johnson Passage Comprehension= wjpc, race = black vs. white.

Table 8

Means for Woodcock-Johnson Passage Comprehension Testing by Ethnicity

Measure	Ethnicity	N	*M*	*SD*
wjpc time1	White	25	75.96	9.61
	Black	21	77.76	8.20
wjpc time2	White	25	78.76	10.13
	Black	21	80.14	11.89

Note. Woodcock-Johnson Passage Comprehension = wjpc.

For the Woodcock-Johnson III Reading Fluency, table 9 shows the main effect of time of assessment was not significant F (1,44) = 2.222, p =. 143. There was no significant main effect of ethnicity F (1,44) = .082, p = .775. Table 10 shows Caucasian and African-American students' scores were similar between time one and time two. Caucasians scores at time one (M = 76.89, SD = 5.40) and time two (M = 75.76, SD = 5.40) were similar. African-American students scores at time one (M = 75.84, SD = 6.09) and time two (M = 82.00, SD = 11.31) were similar.

Table 9

ANOVA for Woodcock-Johnson Reading Fluency Testing

Source	SS	df	MS	F	*p*
Wjrf	163.364	1	163.364	2.222	0.143
wjrf*race	6.06	1	6.06	0.082	0.775
Error	3235.299	44	73.53		

Note. Woodcock-Johnson Reading Fluency = wjrf, race = black vs. white.

Table 10

Means for Woodcock-Johnson Reading Fluency Testing by Ethnicity

Measure	Ethnicity	N	*M*	*SD*
wjrf time1	White	25	76.89	5.40
	Black	21	75.84	6.09
wjrf time2	White	25	75.76	12.16
	Black	21	82.00	11.31

Note. Woodcock-Johnson Reading Fluency = wjrf.

Research Question Two

Are there differences by setting (self-contained vs. resource) across the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?

A set of 2 (time of assessment: pre vs. post) X 2 (setting: Resource vs. Self-contained) repeated measures analyses of variance (ANOVAs) were conducted for the four measures (Maze, oral reading fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III

subtest reading fluency). For the Maze procedure, table 11 shows the main effect of time was significant F (1, 44) = 7.217, p = .005. Also, the main effect of setting was significant F (1,53) = 9.036, p = .004. Table 12 shows students' scores for self-contained time one (M = 20.30, SD = 6.33) and time two (M = 26.88, SD = 9.63) increased more than students' scores for resource time one (M = 18.14, SD = 7.21) and time two (M = 20.26, SD = 8.72).

Table 11

ANOVA for Maze Testing

Source	SS	df	MS	F	*p*
Maze	512.387	1	512.387	7.217	.001*
Maze*setting	139.296	1	139.296	1.962	.001*
Error	3763.104	53	71.002		

Note. setting = self-contained vs. resource,*$p < .001$.

Table 12

Means for Maze Testing by Setting

Measure	Setting	N	*M*	*SD*
Maze time1	Resource	29	18.14	7.21
	Self-contained	26	20.3	6.33
Maze time2	Resource	29	20.26	8.72
	Self-contained	26	26.88	9.63

For the Oral Reading Fluency, table 13 shows the main effect of time was significant F (1,53) = 53.872, p = .005. The main effect of setting was not significant F (1, 53) =. 089, p =

.767. Table 14 shows the students' scores for both settings increased during time two. The students' scores in resource increased from time one (M = 74.72, SD = 36.30) to time two (M = 133.41, SD = 53.68). The students' scores in self-contained increased from time one (M = 91.50, SD = 47.78) to time two (M = 91.50, SD = 47.78).

Table 13

ANOVA for Oral Reading Fluency Testing

Source	SS	Df	MS	F	*p*
Orf	102598.372	1	102598.372	53.872	.001*
orf*setting	168.918	1	168.918	0.089	0.767
Error	100938.046	53	1904.491		

Note. setting = self-contained vs. resource ,Oral Reading Fluency = orf, *$p < .001$

Table 14

Means for Oral Reading Fluency Testing by Setting

Measure	Setting	N	*M*	*SD*
orf pre	Resource	29	74.72	36.30
	Self-contained	26	91.50	47.78
orf post	Resource	29	133.41	53.68
	Self-contained	26	155.15	57.83

Note. Oral Reading fluency = orf.

For the Woodcock-Johnson III Passage Comprehension, table 15 shows the main effect of time was not significant F (1,53) =2.689, p = .107. Also, the main effect of setting was not significant F (1, 53) = 2.116, p = .152. Table 16 shows students' scores were similar for both

settings. The students' scores in resource at time one (M = 75.96, SD = 9.61) and time two (M = 76.31, SD = 10.11) were similar. The students' scores in self-contained at time one (M = 77.76, SD = 8.20) and time two (M = 83.53, SD = 11.86) were similar.

Table 15

ANOVA for Woodcock-Johnson Passage Comprehension Testing

Source	SS	df	MS	F	*p*
Wjpc	256.235	1	256.235	2.689	0.107
wjpc*setting	201.689	1	201.689	2.116	0.152
Error	5050.584	53	95.294		

Note. Woodcock-Johnson Passage Comprehension = wjpc, setting = self-contained vs. resource.

Table 16

Means for Woodcock-Johnson Passage Comprehension Testing by Setting

Measure	Setting	N	*M*	*SD*
wjpc time1	Resource	29	75.96	9.61
	Self-contained	26	77.76	8.20
wjpc time2	Resource	29	76.31	10.11
	Self-contained	26	83.53	11.86

Note. Woodcock-Johnson Passage Comprehension = wjpc.

For the Woodcock-Johnson III Reading Fluency, table 17 shows the main effect of time was not significant F (1,53) =2.466, p = .122. Also, the main effect of setting was not significant F (1, 53) =. 082, p = .773. Table 18 shows the resource and self-contained students' scores were

similar. The resource students' scores at time one (M = 76.89, SD = 5.40) and time two (M = 76.48, SD = 6.09) were similar. The self-contained students' scores at time one (M = 75.84, SD = 6.09) and time two (M = 81.30, SD = 11.85) were similar.

Table 17

ANOVA for Woodcock-Johnson Reading Fluency

Source	SS	df	MS	F	*p*
Wjrf	174.652	1	174.652	2.466	0.122
wjrf*setting	236.616	1	236.616	3.341	0.773
Error	3753.748	53	70.825		

Note. Woodcock-Johnson Reading Fluency = wjrf, setting = self-contained vs. resource.

Table 18

Means for Woodcock-Johnson Reading Fluency Testing by Setting

Measure	Setting	N	*M*	*SD*
wjrf time1	Resource	29	76.89	5.40
	Self-contained	26	75.84	6.09
wjrf time2	Resource	29	76.48	11.25
	Self-contained	26	81.30	11.85

Note. Woodcock-Johnson Reading Fluency = wjrf.

Research Question Three

What is the concurrent validity of the assessment measures (Maze, Oral reading fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest passage comprehension)?

Pearson correlations were calculated to determine the pattern of associations between the performance results on the CBM measures (Oral reading fluency, Maze) at time one as compared to the performance results on the Woodcock-Johnson III subtests (reading fluency, passage comprehension) at time one. Table 19 shows the correlation matrix for the measurements. A correlation was found with the CBM measure (Maze) and the Woodcock-Johnson III subtest Passage Comprehension(r = .888, p < .05). Next, a correlation was found with the CBM measure (Oral Reading Fluency) and the Woodcock-Johnson III subtest Passage Comprehension (r = .452, p < .05). Also, a correlation was found with the CBM measure (Oral Reading Fluency) and the Woodcock-Johnson III subtest Reading Fluency (r = .473, p < .05).

Table 19

Correlation Matrix

	1	2	3	4
1. Maze	1	.462*	.888*	0.216
P	-	0.001	0.001	0.114
N	55	55	55	55
2. Orf		1	.452*	.473*
P		-	0.001	0.001
N		55	55	55
3. WJPC			1	0.262
P			-	0.053
N			55	55
4. WJRF				1
P				-
N				55

Note. Oral Reading Fluency=ORF, Woodcock-Johnson III Passage Comprehension = WJPC, Woodcock-Johnson III Reading Fluency = WJRF, *$p < .05$, two-tailed.

Regression analyses were conducted using the CBM measures (Oral reading fluency, Maze) as single predictors and the pre-test scores of Woodcock-Johnson III subtests Passage Comprehension as the criterion. First, Maze was entered as a single predictor. Table 20 shows that the model did account for significant portion of the variance in the Woodcock-Johnson III

subtest Reading Fluency R-square = .046; F (1,53) = 2.58, p = .114, and Maze was a significant predictor (Beta = .216, p = .114). Next, Oral Reading Fluency was entered as a single predictor. Table 21 shows that the model account for 20.4% of the variance in the Woodcock-Johnson III Passage Comprehension F (1,53) = 14.27, $p < .001$. Oral Reading Fluency was a significant predictor (Beta = .473, p = .001).

Table 20

ANOVA for Maze Testing

Source	SS	df	MS	F	R^2	*p*
Regression	81.822	1	81.822	2.582	0.046	0.114
Residual	1679.378	53	31.686			
Total	1761.2	54				

Table 21

ANOVA for Oral Reading Fluency Testing

Source	SS	df	MS	F	R^2	*p*
Regression	393.856	1	393.856	15.266	.224*	0.001
Residual	1367.344	53	25.799			
Total	1761.2	54				

Note. *$p < .001$.

For the next analysis, Maze and Oral Reading Fluency were entered together as predictors of the Woodcock-Johnson III Reading Fluency. Table 22 shows that the model was significant, accounting for 22.4% of the variance in Woodcock-Johnson III Reading Fluency F

(2,52) = 98.535, p = .005. As expected from the single predictor analyses, Maze was a significant predictor (Beta = .745, p = .001), while Oral Reading Fluency was not (Beta = -.004, p = .978).

Table 22

ANOVA for Maze and Oral Reading Fluency Testing

Source	SS	df	MS	F	R^2	p
Regression	393.877	2	196.939	7.49	.224*	0.001
Residual	1367.323	52	26.295			
Total	1761.2	54				

Note. *p < .001.

Two hierarchical regressions were then conducted, entering Maze and Oral Reading Fluency interchangeably as the first and second predictors of Woodcock-Johnson III Reading Fleuncy. In the first hierarchical regression, Maze was entered at Step 1 while Oral Reading Fluency was entered in at Step 2. With Maze as the only predictor, table 23 shows that the model was significant R-square = .216; F (1, 53) = 2.582, p = .114, and Maze was significant as a predictor in the model (Beta = .216, p = .114). Table 24 shows that adding Oral Reading Fluency resulted in a significant model R-square = .473; F (2, 52) = 7.490, p = .001. The change in R-square was not significant F change (1,52)= 11.868, p = .001, indicating that adding Oral Reading Fluency to the model did not significantly increased the prediction of Woodcock-Johnson III Reading Fluency.

Table 23

ANOVA for Maze Testing

Source	SS	df	MS	F	R^2	*p*
Regression	81.822	1	81.822	2.582	0.046	0.114
Residual	1679.378	53	31.686			
Total	1791.2	54				

Table 24

ANOVA for Oral Reading Fluency Testing

Source	SS	df	MS	F	R^2	*p*
Regression	393.877	2	196.939	7.49	.224*	0.001
Residual	1367.323	52	26.295			
Total	1761.2	54				

Note. *$p < .001$.

The order of entry of the two predictors was then switched, and the hierarchical model was repeated. In Step 1, Oral Reading Fluency was the only predictor. Table 25 shows that this model was significant R-square = .224; F (1, 53) = 15.266, p = .001, and the Oral Reading Fluency was a significant as a predictor (Beta = .473, p = .001). In Step 2, Maze was added. Table 26 shows that this model was significant R-square = .001; F (2, 52) = 7.490, p = .001, but Maze was the only significant predictor (Beta = -.004, p = .978).

Table 25

ANOVA for Oral Reading Fluency Testing

Source	SS	df	MS	F	R^2	*p*
Regression	393.856	1	393.856	15.266	.224*	0.001
Residual	1367.344	53	25.799			
Total	1761.2	54				

Note. *$p <$. 001.

Table 26

ANOVA for Maze Testing

Source	SS	df	MS	F	R^2	*p*
Regression	393.877	2	196.939	7.49	.224*	0.114
Residual	1367.323	52	26.295			
Total	1761.2	54				

Note. *$p <$.05.

Regression analyses were conducted using the CBM measures (Oral reading fluency, Maze) as single predictors and the time one scores of Woodcock-Johnson III subtests Passage Comprehension as the criterion. First, Maze was entered as a single predictor. Table 27 shows that the model did account for significant portion of the variance in the Woodcock-Johnson III subtest Reading Fluency R-square = .888; F (1,53) = 198.262, p = .001, and Maze was a significant predictor (Beta = .888, p = .001). Next, Oral Reading Fluency was entered as a single predictor. Table 28 shows that the model account for 20.4% of the variance in the Woodcock-

Johnson III Passage Comprehension F (1,53) = 13.594, $p < .001$. Oral Reading Fluency was a significant predictor (Beta = .452, $p = .001$).

Table 27

ANOVA for Maze Testing

Source	SS	df	MS	F	R^2	*p*
Regression	3405.746	1	3405.746	198.262	.789*	0.001
Residual	910.436	53	17.178			
Total	4316.182	54				

Note. *$p < .001$.

Table 28

ANOVA for Oral Reading Fluency Testing

Source	SS	df	MS	F	R^2	*p*
Regression	881.073	1	881.073	13.594	.204*	0.001
Residual	3435.109	53	64.813			
Total	4316.182	54				

Note. *$p < .001$.

For the next analysis, Maze and Oral Reading Fluency were entered together as predictors of the Woodcock-Johnson III Passage Comprehension. Table 29 shows that the model was significant, accounting for 79.1% of the variance in Woodcock-Johnson III Passage Comprehension F (2,52) = 98.535, $p = .001$. As expected from the single predictor analyses, Maze was a significant predictor (Beta = .864, $p = .001$), while Oral Reading Fluency was not (Beta = .052, $p = .467$).

Table 29

ANOVA for Maze and Oral Reading Fluency Testing

Source	SS	df	MS	F	R^2	*p*
Regression	3415.067	2	1707.533	98.535	.791*	0.001
Residual	901.115	52	17.329			
Total	4316.182	54				

Note. *$p < .001$.

Two hierarchical regressions were then conducted, entering Maze and Oral Reading Fluency interchangeably as the first and second predictors of Woodcock-Johnson III Passage Comprehension. In the first hierarchical regression, Maze was entered at Step 1 while Oral Reading Fluency was entered in at Step 2. With Maze as the only predictor, table 30 shows that the model was significant R-square = .888; F (1, 53) = 198.262, p = .001, and Maze was significant as a predictor in the model (Beta = .888, p = .001). Table 31 shows that adding Oral Reading Fluency resulted in a significant model R-square = .890, F (2, 52) = 98.535, p = .001, and Oral Reading Fluency was not significant as a predictor in the model (Beta = 052, p = .467). The change in R-square was not significant F change (1,52)= .538, p = .467, indicating that adding Oral Reading Fluency to the model did not significantly increased the prediction of Woodcock-Johnson III Passage Comprehension.

Table 30

ANOVA for Maze Testing

Source	SS	df	MS	F	R^2	*p*
Regression	3405.746	1	3405.746	198.262	.789*	0.001
Residual	910.436	53	17.178			
Total	4316.182	54				

Note. *$p < .001$.

Table 31

ANOVA for Oral Reading Fluency Testing

Source	SS	df	MS	F	R^2	*p*
Regression	3415.067	2	1707.533	98.535	.791*	0.001
Residual	901.115	52	17.329			
Total	4316.182	54				

Note. *$p < .001$.

The order of entry of the two predictors was then switched, and the hierarchical model was repeated. In Step 1, Oral Reading Fluency was the only predictor. Table 32 shows that this model was significant R-square = .204; F (1, 53) = 13.594, p = .005, and the Oral Reading Fluency was a significant as a predictor (Beta = .452, p = .005). In Step 2, Maze was added. Table 33 shows that this model significant R-square = .587; F (2, 52) = 98.535, p = .005, but Maze was the only significant predictor (Beta = .864, p = .005). The change in R-square was significant F change (1,52) = 146.227, p = .001, indicating that adding Maze to the model significantly increased the prediction of Woodcock-Johnson III Passage Comprehension.

Table 32

ANOVA for Oral Reading Fluency Testing

Source	SS	df	MS	F	R^2	*p*
Regression	881.073	1	881.073	13.594	.204**	0.001
Residual	3435.109	53	64.813			
Total	4316.182	54				

Note. *$p < .001$.

Table 33

ANOVA for Maze Testing

Source	SS	df	MS	F	R^2	*p*
Regression	3415.067	2	1707.533	98.535	.791*	0.001
Residual	901.115	52	17.329			
Total	4316.182	54				

Note. *$p < .001$.

Research Question Four

What is the predictive-criterion related validity of the assessments (Maze, oral reading fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency) from time one to time two over a nine-week period?

Pearson correlations were calculated to determine the pattern of associations between the performance results on the CBM measures (Oral reading fluency, Maze) as compared to the performance results on the Woodcock-Johnson III subtests (reading fluency, passage comprehension). Table 34 shows the time two correlation matrix for the measurements. A

correlation was found with the CBM measures (Oral Reading Fluency) and the Woodcock-Johnson III subtest Passage Comprehension (r = .300, $p < .05$).

Table 34

Correlation Matrix

	1	2	3	4
1. Maze	1	0.13	-0.011	0.11
P	-	0.343	0.936	0.425
N	55	55	55	55
2. Orf		1	0.114	.300*
p		-	0.406	0.026
N		55	55	55
3. WJPC			1	0.082
p			-	0.551
N			55	55
4. WJRF				1
p				-
N				55

Note. Oral Reading Fluency = ORF, Woodcock-Johnson III Passage Comprehension = WJPC, Woodcock-Johnson III Reading Fluency= WJRF, *$p < .05$, two-tailed.

Regression analyses were conducted using the CBM measures (Oral reading fluency, Maze) time one scores as single predictors and the time two scores of Woodcock-Johnson III

Reading Fluency as the criterion. Oral Reading Fluency was a single predictor. Table 35 shows that Oral Reading Fluency did account for a significant portion of variance with the Woodcock-Johnson III Reading Fluency R-square = .090; F (1,53) = 5.226, p = .025.

Table 35

ANOVA for Oral Reading Fluency Testing

Source	SS	df	MS	F	R^2	*p*
Regression	662.59	1	662.59	5.226	.090*	0.025
Residual	6719.337	53	126.78			
Total	7381.927	54				

Note. *p <.05.

Summary

The data presented above showed the main effect of time was significantly greater for all students with Emotional and Behavior Disorders time two performance for both curriculum-based measures (Oral Reading Fluency, Maze). Oral Reading Fluency data indicated that Caucasian students' scores increased more than African-American students' scores between time one and time two. Next, the data indicated that African-American students' scores increased significantly more between time one and time two for the Woodcock-Johnson III subtest Reading Fluency. Also, the data showed a significant difference with main effect of time for the curriculum-based measures (Maze, Oral Reading Fluency) and setting (self-contained vs. resource). All students with Emotional and Behavior Disorders scores were significantly greater at time two as opposed to time one for both curriculum-based measures for settings (self-contained, resource). Also, the data indicated the main effect of setting was significant for the self-contained students' scores for the Maze and Woodcock-Johnson Passage Comprehension.

Self-contained students' scores increased significantly more from time one to time two on the Maze assessment. Next, self-contained students' scores increased more from time one to time two on the Woodcock-Johnson III Passage Comprehension. As noted in the correlation matrices, correlations were shown to be significant between the curriculum-based measures (Maze, Oral Reading Fluency) at time one and the Woodcock-Johnson III subtests (Passage Comprehension, Reading Fluency) at time two. Regression analyses indicated that the curriculum-based measure (Maze) was a predictor for student performance on the Woodcock-Johnson III Passage Comprehension. Regression analyses indicated that curriculum-based measure (Oral Reading Fluency) was a predictor for student performance on the Woodcock-Johnson Reading Fluency. Also, regression analyses indicated the curriculum-based measure (Oral Reading Fluency) at time one was a predictor of student performance on the Woodcock-Johnson III subtest (Reading Fluency) at time two.

Chapter 5

Discussion

Introduction

The following chapter provides a summary of the results of the study along with summative and evaluative statements regarding the data. Each research question is discussed in further detail with generalized statements that may be made concerning the results. Further, a discussion of limitations will continue in the overall evaluation of the study. Next, Implications for the classroom will be discussed. Finally, recommendations for future research regarding curriculum-based measurement are presented.

This study was designed to evaluate the validity of curriculum-based measurements for students with emotional and behavior disorders in a middle school setting. All of the students were administered two curriculum-based measures (Maze comprehension, Oral Reading Fluency) and the Woodcock-Johnson III Test of Achievement subtests (Woodcock-Johnson III reading fluency, Woodcock-Johnson III subtest passage comprehension). The Maze comprehension procedure was administered to the students within large groups. The Oral Reading Fluency measure was administered to each student individually across three different reading passages. The Woodcock-Johnson III subtest reading comprehension was administered individually to each student. Also, the Woodcock-Johnson III subtest passage comprehension was administered individually to each student. The students' results for each assessment (Maze comprehension, Oral Reading Fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency) were compared. The data examiners were

trained and used procedures for conducting the curriculum-based measures as outlined in research on the curriculum-based measurement process (Fuchs & Fuchs, 1997; Shinn, 1998).

The design allowed for comparisons of student performance between the curriculum-based measures and the subtests of the Woodcock-Johnson III Test of Achievement. Also, comparisons of concurrent validity and predictive validity were explored for the assessment measures given at different time periods. Further, differences were determined between the subgroups of students with emotional and behavior disorders by setting (resource versus self-contained) and ethnicity (Caucasian versus African-American).

The results of this study were informative with information on the students' reading performance resulting from the administration of the assessments to students with emotional and behavior disorders. The information obtained from the results can be applicable to best practices and data-based decisions for progress monitoring of students in the classroom. Overall, the results contributed to research findings suggesting the use of curriculum-based measures as a valid measure for monitoring the progress of students (Francis et al., 2002). Each research question will be evaluated in order to determine if any summative statements can be made after data evaluation. The four research questions were:

1. Are there differences by ethnicity across the assessment measures (Maze, Oral Reading Fluency, Woodcock-Johnson III subtest Reading Fluency, Woodcock-Johnson III subtest Passage Comprehension)?
2. Are there differences by setting (self-contained versus resource) across the assessment measures (Maze, Oral Reading Fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest Passage Comprehension)?

3. What is the concurrent validity of the assessment measures (Maze, Oral Reading Fluency)?
4. What is the predictive-criterion related validity of the assessments (Maze, Oral Reading Fluency)?

Summary of Results

The first research question asked if there were differences by ethnicity across the assessment measures (Maze, Oral Reading Fluency, Woodcock-Johnson III subtest Reading Fluency, Woodcock-Johnson III subtest Passage Comprehension). The results indicated there were significant differences by ethnicity (black versus white) and time (time one versus time two) across the assessment measures. Caucasian students' scores increased more than African-American students' scores between time one and time two for the Oral Reading Fluency measure. Next, African-American students' scores increased more between time one and time two on the Woodcock-Johnson III subtest reading fluency. The results of the African-American students scoring higher may suggest that curriculum-based measurement is an emerging alternative to traditional testing for African-American students (Fore III, Burke, & Martin, 2006).

The examination of the main effect of time indicated significant differences for both ethnic groups. The students' scores at time one were higher than the scores at time two for both ethnic groups (blacks and whites) on the Maze assessment. The same results occurred for the Oral Reading Fluency with a significant increase in the students' scores from time one to time two. The results indicated no significant differences for the main effect of time by the ethnic groups for the Woodcock-Johnson III subtests (passage comprehension, reading fluency).

Research Question Two

The second research question asked if there were differences by setting (self-contained versus resource) across the assessment measures (Maze, Oral Reading Fluency, Woodcock-Johnson III subtest reading fluency, Woodcock-Johnson III subtest Passage Comprehension). The results indicated there were differences by setting (self-contained versus resource) and time (time one versus time two) across the assessment measures. The students' scores in the self-contained setting increased more over time then the students in the resource setting for the Maze assessment. The same results occurred for the Woodcock-Johnson III passage comprehension assessment with student scores increasing more for the self-contained students. There were no significant differences for setting (resource versus self-contained) for Oral Reading Fluency and the Woodcock-Johnson III subtest oral reading fluency.

The examination of the main effect of time (time one versus time two) indicated significant differences for both settings. Self-contained and resource students' scores significant increased between time one and time two for the Maze and Oral Reading Fluency assessment. The results indicated no significant differences for the main effect of time by the setting for the Woodcock-Johnson III subtests (passage comprehension, reading fluency).

Overall, the self-contained students performed better on the assessments than the students in the resource setting. The findings were interesting because the self-contained setting is a more restrictive environment than resource within the local school. Students in the self-contained setting usually have more severe behaviors that interfere with the ability to learn within a classroom setting. Findings may suggest that students placed in the self-contained setting may have been over-identified within the school setting where the study was conducted.

Research Question Three

The third research question asked for the concurrent validity of the assessment measures (Maze, Oral Reading Fluency). Pearson correlations were calculated to determine the pattern of associations between the performance results on the CBM measures (Oral Reading Fluency, Maze) at time one as compared to the time one performance results on the Woodcock-Johnson III subtests (reading fluency, passage comprehension). A Correlation was found with the curriculum-based measure (Maze) and the Woodcock-Johnson III subtest passage comprehension. Also, a correlation was found with the curriculum-based measure (Oral Reading Fluency) and the Woodcock-Johnson III subtest passage comprehension. The significant correlation between the curriculum-based measures (Maze and Oral Reading Fluency) and the Woodcock-Johnson III subtest passage comprehension aligned with previous research documentation of correlations existing between these measures and standardized assessments (Fewster & McMillan, 2002; Shin, Deno, & Espin, 2000). Next, regression analyses were conducted using the curriculum-based measures (Oral Reading Fluency, Maze) as single predictors and the time one scores of Woodcock-Johnson III subtests (reading fluency, passage comprehension) as the criterion. Oral Reading Fluency was found to be a significant predictor of student performance on the time one scores of the Woodcock-Johnson III subtests reading fluency and passage comprehension. Maze was found to be a predictor of performance on the time one scores of the Woodcock-Johnson III subtest passage comprehension. Maze was not found to be a predictor of performance on the time one of the Woodcock-Johnson III subtest reading fluency.

The next analyses were hierarchical regressions were then entering Maze and Oral Reading Fluency interchangeably as the first and second predictors of Woodcock-Johnson III

Passage Comprehension. In the first hierarchical regression, Maze was entered at Step 1 while Oral Reading Fluency was entered in at Step 2. The hierarchical model indicated that Maze would have been sufficient as a predictor without including Oral Reading Fluency. Next, the order of entry of the two predictors was then switched, and the hierarchical model was repeated with Oral Reading Fluency at step 1 and Maze entered at step 2. The hierarchical model indicated that adding Maze to at step 2 increased the significance with both measures as predictors together.

The same hierarchical regressions were conducted analyses were conducted entering Maze and Oral Reading Fluency interchangeably as the first and second predictors of Woodcock-Johnson III Reading Fluency. The hierarchical model indicated that adding Oral Reading Fluency to the model at step 2 increased the significance with both measures as a predictor. Next, the order of entry of the two predictors was then switched, and the hierarchical model was repeated and the results indicated that both measures together were significant as predictors.

Research Question four

The fourth research question asked for the predictive-criterion related validity of the assessments (Maze, Oral Reading Fluency). Pearson correlations were calculated to determine the pattern of associations between the performance results on the curriculum-based measures (Oral Reading fluency, Maze) as compared to the time two performance results on the Woodcock-Johnson III subtests (reading fluency, passage comprehension). A correlation was found with the curriculum-based measures (Oral Reading Fluency) and the Woodcock-Johnson III subtest passage comprehension. The significant correlation between the Oral Reading Fluency and the Woodcock-Johnson III subtest passage comprehension aligned with previous

research documentation of correlations existing between Oral Reading Fluency and standardized assessments (Crawford, Tindal, & Stieber, 2001; Fewster & McMillan, 2002; Shin, Deno, & Espin, 2000).

Limitations of the Study

The results of this study provide support for curriculum-based measurement to assess the reading progress of students with emotional and behavior disorders. However, there were some limitations to the study and caution should be used when making comparisons and generalizing statements about the results.

First, the sample population of this study was relatively small. There were only 55 students included in this study. The small sample size increased the standard error of measurement and prohibits the ability to make generalizations to a larger population.

Second, the study had a small sample size when comparing ethnicity and student performance on the on the four different types of assessments (Maze comprehension, oral reading fluency, Woodcock-Johnson III subtest passage comprehension, Woodcock-Johnson III subtest reading fluency) prohibiting more generalized statements about ethnicity and student performance. Therefore, results and comparisons in regard to performance were only limited to the African American and Caucasian ethnic groups.

Third, the current study encountered difficulty in controlling for the effect of classroom instruction on student performance between time one and time two in both classroom settings. Additionally, in being sensitive to the effects of instruction, the results of the measures could have been influenced by variables other than instruction (i.e., where administration occurred, severity of disability), so generalizing the results should be done with caution.

Implications for Practice

The study expanded on the conditions of use for curriculum-based measurement being compared with standardized measurements further supporting curriculum-based measurement use as a valid measurement tool for comparisons with student performance (Fuchs & Fuchs, 1997; Crawford, Tindall, & Stieber, 2001). The data presented in this study demonstrates that curriculum-based measures are sensitive enough to be used in comparison with standardized assessments in assessing a student's reading ability. Teachers must recognize curriculum-based measurement as a general indicator and interpret its results in combination with various other types of assessments and data (Crawford, Tindal, & Stieber, 2001; Fewester & McMillan, 2002; Fuchs & Fuchs, 1997). Curriculum-based measurement results can be helpful in identifying problems that warrant future investigations for the students in the classroom. The information from curriculum-based measures can be extended into the development of appropriate intervention plans directly related to students with Emotional and Behavior Disorders in the classroom. These applications make curriculum-based measures an attractive component for classroom teachers and school personnel in monitoring student performance for students' with emotional and behavior disorders.

Results of the current study are consistent with those of previous studies that have suggested that assessing reading comprehension may be similar to assessing oral reading fluency (Deno et al., 1982; Fuchs & Fuchs, 1997). The students that exhibit poor oral reading skills have comprehension skills that are often equal to or lower than their reading fluency levels (Hinzte & Conte, 1997). The students that may be referred for reading problems and found to have oral reading fluency difficulties implies that a separate assessment of comprehension may not be warranted. A screening for reading comprehension may be incorporated by selecting a simple

Maze passage and procedure for students who may exhibit adequate decoding and fluency skills. Another practical use for curriculum-based measurement would be to formulating goals and objectives for Individual Education Plans (Deno, Mirkin, & Wesson, 1983; Fuchs & Fuchs, 1997). Curriculum-based measures can provide immediate feedback for a teacher to see if an intervention is having a positive effect and if not, to make a modification in a plan that was previously established. The measures can be instructional in pointing out areas of weakness to be addressed specifically for each student through the goals and objectives of the IEP. Furthermore, curriculum-based measurement procedures involve the direct observation of student behavior and use single case analytic procedures that are similar to characteristics of applied behavior analysis (Deno, 1997).

The assessment process can be an important aspect to monitor the progress of student performance and assessing the effectiveness of a particular education program. This would suggest that these components are what make curriculum-based measurement such an attractive component for classroom teachers and school personnel in educating students with Emotional and Behavior Disorders. As stated in a study conducted by Ardoin et. al. (2004), the use of curriculum-based measurements by schools was considered to be a quick, and cost-efficient screening device for immediate use by schools for identifying reading difficulties. Furthermore, the study stated how many norm-referenced tests are costly and require a substantial amount of time to administer.

Implications for Future Research

There are several implications for future research with the technical adequacy of curriculum-based measurement for students with Emotional and Behavior Disorders. More studies are clearly needed examining the validity and technical adequacy of curriculum-based

measurement for students with Emotional and Behavior Disorders. Students with Emotional and Behavior Disorders are variable populations with many unique characteristics that may affect the technical adequacy of the measures. Having additional studies will assist in buttressing the proposition that curriculum-based measurement can be used to accurately inform progress for students with Emotional and Behavior Disorders, especially those who have academic deficits.

Moreover, some students with Emotional and Behavior Disorders may be displaying their problem behaviors to escape and avoid difficult academic tasks. Once the validity and the technical adequacy of curriculum-based measures have been established for instructional problem-solving, research will be needed to examine how these measures may be used within a function-based approach to intervention planning (Espin & Tindal, 1998; Fuchs, 2004).

Conclusion

This study was designed to evaluate the validity of curriculum-based measurement for students with Emotional and Behavior Disorders in a middle school setting. The findings from this study were consistent with previous research findings documentation of correlations existing between curriculum-based measures and standardized assessments (Crawford, Tindal, & Stieber, 2001; Fewster & McMillan, 2002; Hintze, Owen, Shapiro, & Daly, 2000; Shin, Deno, & Espin, 2000). Also, results from this study were consistent with previous research findings documenting the predictability of curriculum-based measures for student performance on standardized assessments (Fuchs, 2004; Crawford, Tindal, & Stieber, 2001). The use of curriculum-based measurement as a source of information in screening and eligibility decisions meets current functional assessment requirements for classroom teachers. The information from the curriculum-based measures can provide some objective data that can be incorporated with other assessments in developing the appropriate intervention plans directly relating to the current

academic curriculum for students' with Emotional and Behavior Disorders. Also, the information from the curriculum-based measures can be useful in determining pre-referral strategies and other curriculum instructional interventions for students' with Emotional and Behavior Disorders.

References

Allinder, R. M. (2001). Improving fluency in at-risk readers and students with learning disabilities. *Remedial & Special Education*, 22(1), 48-55.

Allinder, R., Bolling, R., Oats, R., & Gagnon, W. (2000). Effects of teacher self-monitoring on implementation of curriculum-based measurement and mathematics computation achievement of students with disabilities. *Remedial & Special Education*, 21(2), 219-226.

Clarke, B., & Shinn, M. R. (2004). A preliminary investigation into the identification and development of early mathematics curriculum-based measurement. *School Psychology Review*, 33, 234-248.

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

Deno, S. L., & Fuchs, L. S. (1987). Developing curriculum-based measurement systems for data-based special education problem solving. *Focus on Exceptional Children*, 19(8), 1–16.

Deno, S. L., Fuchs, L. S., Marston, D. & Shin, J. (2001). Using curriculum-based measurements to establish growth standards for students with learning disabilities. *School Psychology Review*, 30(4), 507-525.

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

Deno, S., Mirkin, P. & Wesson, C. (1984). How to write effective data-based IEPs. *Teaching Exceptional Children*, 12, 99-104.

Elliott, S. N. & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. *School Psychology Review*, 26, 224-236.

Espin, C., Scierka, B., Skare, S., & Halverson, N. (1999). Criterion-related validity of curriculum-based measures in writing for secondary school students. *Reading & Writing Quarterly*, 15(1), 5-28.

Espin, C. A., & Tindal, G. (1998). Curriculum-based measurement for secondary students. In M. R. Shinn (Ed.), Advanced applications of curriculum based measurement (pp. 214-253). New York: Guilford Press.

Faykus, S., & McCurdy, B. (1998). Evaluating the sensitivity of the maze as an index of reading proficiency for students who are severely deficient in reading. *Education & Treatment of Children*. 21(1), 1-21.

Fewster, S. & Macmillan, P. (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial & Special Education*, 23(3), 149-158.

Fore III, C., Burke, M., & Martin, C. (2006). Curriculum-based measurement and problem-solving model: An emerging alternative to traditional assessment for african american children and youth. *Journal of Negro Education, 99, 1-10.*

Forness, S., Kavale, K., Blum, I., & Lloyd, J. (1997). What works in special education and related services: Using meta-analysis to guide practice. *Teaching Exceptional Children*, 29(6), 4-9.

Fuchs, L. (2004). The past, present, and future of curriculum-based measurement research. *School Psychology Review*, 33(2), 188-192.

Fuchs, L. S., & Deno, S. L. (1991). Paradigmatic distinctions between instructionally relevant measurement models. *Exceptional Children*, 58, 232-243.

Fuchs, L., & Fuchs, D. (1990). The role of skills analysis in CBM in math. *School Psychology Review*, 19(1), 6-23.

Fuchs, L., Fuchs, D. (1991). Effects of expert system advice within curriculum-based measurement on teacher planning and student achievement in spelling. *School Psychology Review*, 20(1), 49-60.

Fuchs, L., Deno, S., & Mirkin, P. (1984). Effects of frequent curriculum-based measurement and evaluation on pedagogy, student achievement, and student awareness of learning. *American Educational Research Journal*, 21, 449-460.

Fuchs, L. & Fuchs, D. (1984). Criterion-referenced assessment without measurement: How accurate for special education. *Remedial and Special Education*, 5, 29-32.

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

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Thompson, A., Roberts, H., Kubec, P., & Stecker, P. (1994). Technical features of a mathematics concepts and applications curriculum-based measurement *s*ystem. *Diagnostique*, 19(4), 23-49.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Phillips, N. B., & Bentz, J. (1994). Class wide curriculum-based measurement: Helping general educators meet the challenge of student diversity. *Exceptional Children*, 60, 518-537.

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

Fuchs, D., Roberts, P. H., Fuchs, L. S., & Bowers, J. (1996). Reintegrating students with learning

disabilities into the mainstream: A two-year study. *Learning Disabilities Research & Practice*, 11, 214-229.

Fuchs, L.S., & Fuchs, D. (1997). Use of curriculum-based measurement in the identification of students with disabilities. *Focus on Exceptional Children*, 30(3), 1-16.

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, 659–671.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal*, 28, 617-641.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P.(1990). The role of skills analysis in curriculum-based measurement in math. *School Psychology Review*, 19, 6-22.

Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review* , 22 (1), 27-48.

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

Fuchs, D. & Fuchs, L. S. (1997). Use of curriculum-based measurement in identifying students with disabilities. *Focus on Exceptional Children*, 30, 1-16.

Germann, G., & Tindal, G. (1985). Applications of direct and repeated measurement using curriculum based assessment. *Exceptional Children*, 51 (2), 110-121.

Good, R. H. III &Jefferson, G. (1998). Contemporary perspectives on Curriculum-Based

Measurement Validity. In M. R. Shinn (Ed.), Advanced applications of curriculum-based measurement (pp. 61–88). New York: Guilford.

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

Good, R. H., Simmons, D. C.,& Smith, S. (1998). Effective academic interventions in the UnitedStates: Evaluating and enhancing the acquisition of early reading skills. School *Psychology Review*, 27, 45–56.

Gresham, F. M., MacMillan, D. L., & Bocian, K. M. (1996). Learning disabilities, low achievement, and mild mental retardation: More alike than different? *Journal of Learning Disabilities*, 29 (6), 570-581.

Grimes, J. & Tilly, W.D., (1996). Policy and process: Means to lasting educational change. *School Psychology Review*, 25, 465-476.

Hargrove, L., Church, K., Yssel, N., & Koch, K. (2002). Curriculum-based: Reading and state academic standards. *Preventing School Failure*, 46(4), 148-152.

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

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

Hintze, J. M., & Shapiro, E. S. (1997). Curriculum-based measurement and literature-based

reading: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology, 35*, 351-375.

Hintze, J., Shapiro, E., & Lutz, J. (1994). The effects of curriculum on the sensitivity of curriculum-based measurement in reading. *The Journal of Special Education,* 28, 188-202.

Howell, K. W. & Nolet, V. (2000). *Curriculum-based evaluation: Teaching and decision making (Thrid Edition).* Belmont, CA: Wadsworth Publishing Company.

Helwig, R., Anderson, L., & Tindal, G. (2002). Using a concept-grounded, curriculum-based measure in mathematics to predict statewide test scores for middle school students with LD. *The Journal of Special Education*, 36(2), 102-112.

Kaminski, R. A., & Good, R. H., III. (1996). Toward a technology for assessing basic early literacy skills. *School Psychology Review*, 25, 215–227.

Marston, D. B. (1989). *A curriculum-based measurement approach to assessing academic performance: What it is and why do it.* In Mark R. Shinn Curriculum-Based Measurement: Assessing Special Children. New York: The Guilford Press.

Marston, D., & Magnusson, D. (1985). Implementing curriculum-based measurement in special and regular education settings. *Exceptional Children*, 52 , 266-276.

Marston, D. B., Mirkin, P.K., & Deno, S. L. (1984). Curriculum-based measurement: An alternative to traditional screening, referral, and identification. *The Journal of Special Education*, 18, 109-117.

Martson, D., Muyskens, P., Lau, M., Canter, A. (2003). Problem-solving model for decision making with high-incidence disabilities: The minneapolis experience. *Learning Disabilities Research & Practice*, 18, 197-200.

Madeline, A., & Wheldall, K. (1998). Towards a curriculum-based passage reading test for monitoring the performance of low-progress readers using standardized passages: A Validity study. *Educational Psychology*, 18(4), 471-479.

Naquin, G., & Slider, N. (2002). Moving beyond total words written: The reliability, criterion validity, and time cost of alternate measures for curriculum-based measurement in writing. *School Psychology Review*, 31(4), 477-497.

Scott, V., & Weishaar, M. (2003). Curriculum-based measurement for reading progress. *Intervention In School and Clinic*, 38(3), 153-159.

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

Shinn, M.R. (1998). *Advanced applications of curriculum-based measurement*. New York, NY: the Guilford Press.

Shin, J., Deno, S., & Espin, C. (2000). Technical adequacy of the maze task for curriculum-based measurement of reading growth. *Journal of Special Education*, 34(3), 164-173.

Shinn, M. R., & Good, R. H. (1992). Curriculum-based measurement of oral reading fluency: A confirmatory analysis of its relation to reading. *School Psychology Review,* 21(3), 459-480.

Shinn, M., Good, R., & Parker, C. (1999). *Noncategorical special education services with students with severe achievement deficits.* In D. J. Reschly, W.D. Tilly, & J. P. Grimes (Eds.), *Special Education in Transition* (pp. 81-105). Longmont, CO: Sopris West.

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

Shinn, M., Good, R., & Stein, S. (1989). Summarizing trend in student achievement: A comparison of models. *School Psychology Review*, 18, 356- 370.

Shinn, M. & Hubbard, D. (1992). Curriculum- based measurement and problem solving assessment: Basic procedures and outcomes. *Focus on exceptional children*, 24(5), 1-20.

Tilly, W., & Carlson, S. (1992). Creating measurement materials. In M. R. Shinn, N. Knuston , & W.D. Tilly (Eds.), Curriculum-based measurement training modules. Eugene: University of Oregon.

Tindal, G. A., & Marston, D. B. (1990). *Classroom-based Assessment: Evaluating Instruction Outcomes.* Columbus, OH: Merrill Publishing Company.

Tindal, G., McDonald, M., Tedesco, M., Glasgow, A., Almond, P., Crawford, L., & Hollenbeck, K. (2003). Alternate assessments in reading and math: Development and validation for students with significant disabilities. *Exceptional Children*, *69*(4), 481-494.

U.S. Department of Education. (2001). 22nd Annual Report to Congress on the Implementation of IDEA. Washington, D.C.

U.S. Department of Education. (2002). No Child Left Behind Education Act. Washington, D.C.

Appendix A

PARENT CONSENT FORM

The University of Georgia
College of Education
537 Aderhold Hall
Athens, GA 30602
678-407-5185; FAX 706-542-2929

(parent)

Parental Permission Form

I agree to allow my child, _____, to take part in a research study titled, "Examining the Validity of Curriculum-based Measurement for students with Disabilities in Middle School," which is being conducted by Christopher Martin and Cecil Fore, from the Special Education Department at the University of Georgia (678-407-5185). I do not have to allow my child to be in this study if I do not want to. My child can stop taking part at any time without giving any reason, and without penalty. I can ask to have the information related to my child returned to me, removed from the research records, or destroyed.

The following points have been explained to me:

The purpose of this study is to investigate the use of curriculum-based measures as valid measures for assessing a child's reading ability at the middle school level.

As a part of the regular curriculum, my child will be assessed using reading measures for reading fluency and reading comprehension, and these methods are based on best practices.

If I allow my child to participate in this research project, my child will be asked to complete such tasks as taking pretests and posttests, which will take about 25 minutes, and allowing researchers to observe and record results during the instructional period. If I do not want my child to take part in the project, their scores on the pre- and posttest measures will not be recorded and included in the project. These procedures will take about 25 minutes of my child's time and will

not interfere with any classroom activities. These tests will include Maze comprehension, oral reading fluency, Woodcock-Johnson III subtest reading fluency, and Woodcock-Johnson III subtest passage comprehension.

Benefits: While some students may improve their ability to read and comprehend information as a result of curriculum-based measurement, there is no guaranteed benefit to participation in this research study.

Foreseeable Risks: There are no foreseeable risks or discomforts to participants.

Voluntary: My child's participation is voluntary, and they may withdraw from the study at any time and for any reason and will not affect their grade. There is no penalty for not participating or withdrawing.

Confidentiality: All data collected in this study will be confidential; all person-identifiable data will be coded so that no one, including individual students, parents, teachers, schools, or districts can be identified.

The researcher will answer any questions about the research, now or during the course of the project, and can be reached by telephone at: 770-554-4003. I may also contact the professor supervising the research, Dr. Cecil Fore, Special Education Department, at (678) 407-5185.

- I understand the study procedures described above. My questions have been answered to my satisfaction, and I agree to allow my child to take part in this study. I have been given a copy of this form to keep.

_____ _____ _____

Name of Researcher **Signature** **Date**

Telephone: _____ Email: _____

_____ _____ _____

Name of Parent or Guardian *Signature* *Date*

Please sign both copies, keep one and return one to the researcher.

Additional questions or problems regarding your child's rights as a research participant should be addressed to The Chairperson, Institutional Review Board,

Human Subjects Office, University of Georgia, 612 Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone (706) 542-3199; E-Mail Address IRB@uga.edu

Appendix B

ORAL READING FLUENCY PASSAGE 1

One of the reasons Becky loved traveling on the weekends to her grandparents' house in Connecticut was her Grandpa Bob. As soon as she and her parents pulled into the winding drive and stopped in front of her grandparents' house, she and Grandpa Bob would go for a walk.

Becky lived in New York City in an apartment she shared with her parents and a golden retriever named Ralph. She and Ralph walked in the city, and she saw lots of interesting things. They saw yellow taxicabs, men in suits, and women in high heels. But never in New York did she see the fantastic things she saw with Grandpa Bob on their walks.

Grandpa Bob knew just about everything there was to know about the forest and the animals that lived there. Once, when she and Grandpa Bob were walking, a blue bird landed on his shoulder. When it flew away, Becky remembered her grandpa had looked down at her and winked.

"It was just telling me some secrets, that's all, Becky," he had told her. "That bird just told me there would be a frost tonight and that there is a herd of deer nibbling on grass just beyond those maple trees."

Becky followed with her eyes where her grandpa was pointing and saw a pelt of brown fur and the long legs and the velvet nose that did indeed belong to a white-tailed deer.

She couldn't believe a blue bird was smart enough to tell her grandpa all that. At the same time, she wished one of those critters would land on her shoulder and sing secrets to her.

Later that evening, when Grandpa Bob was dozing in front of the fire with his pipe hanging out of his mouth and Becky and her mom and dad were playing a game of cards with Grandma, Becky leaned in close to her mom and whispered in her ear.

"When I grow up Mom, I think I’m going to be like Grandpa Bob."

Appendix C

ORAL READING FLUENCY PASSAGE 2

I live in the tiny town of Peanut, Pennsylvania, in the second to the last house at the end of a dead-end road. There isn't a whole lot to do in Peanut, but we do have a grocer, a baker, a hairdresser, a mechanic, AND an inventor. It's Tommy O'Connor who put our little village on the map. He lives just down the road from me. His house is right before the dead end. A lot of people go back there, turning their cars around when they realize they're lost.

That's how he acquires many of his customers for his bizarre inventions. People pull onto our road, thinking it'll go on forever, but it doesn't. It stops dead in its tracks at Tommy's front door. That means booming business for Tommy.

Tommy has signs for his inventions posted in the ditches up and down our road and out on the freeway. Gigantic signs are nailed to telephone poles and dead tree trunks. They hang from tree branches and other people's mailboxes. Tourists often stop at Tommy's house, and once they're there, they exclaim over his strange inventions. Then they pull out their checkbooks and spend big bucks.

Last spring, Tommy crossed his lawn mower with his snowmobile and used it to both clear snow and cut grass. Just a month ago, he found an old hot air balloon in the dump, hooked it up to his own car, and now he no longer has to fight traffic on the way to work. He transformed his wife's hairdryer into a miniature rocket engine and his son's Nintendo into a toaster. Last week he

attempted to sell me a calculator that he'd turned into a cell phone, but I had to refuse because we didn't need any more cell phones in the house.

He tuned up my car for me about a week ago. The repairs it required were simple—an oil change and a refill on windshield wiper fluid. When I got it back it could go from zero to ninety in a second flat and the radio could pick up stations from around the world.

Appendix D

ORAL READING FLUENCY PASSAGE 3

In the field of geology, there isn't a scientist with more skill, determination or motivation than Dr. Isabel Rodriguez. Dr. Rodriguez started her collection of rocks at an early age. When she was a child, she would take rocks home to her family's cattle ranch at the edge of the desert. As a grown woman, she turned her childhood interest into a career and now teaches geology to college students from the United States and other countries. As a college instructor, she trains young men and women to be skilled geologists. She shows them how to read maps of major landforms and how to tell the difference between a diamond and a lump of glass. She provides her students knowledge they can use while looking for rocks and fossils in the field.

For example, her students learn that the Red Mountains in Colorado are tinted red because of iron compounds and that the best source for gems are rivers flowing from volcanoes. One afternoon, as part of her lecture, Dr. Rodriquez held up her pencil. "The graphite in the lead of this pencil is chemically identical to diamonds," she said.

"But because they have different crystal structures, they have very different physical properties. You can write with graphite in the pencil, but it is basically worthless.

On the other hand, the diamond comes in a variety of colors and is priceless." In the field, Dr. Rodriguez is a rock-finding whiz, amazing her students during their outings across the sun-baked desert. Although many fossils and semi-precious stones lie in plain view, they are clear only to her keen eyes.

"An opal has a blue-green glow," she tells her students as they walk across the desert. "When you find one, notice how it reflects sunlight."

Dr. Rodriguez is constantly digging up new treasures. Finding a million-year-old carbon imprint of a fern frond trapped in a sandstone wall is not unusual for Dr. Rodriquez. On outings with her students, she is frequently heard saying, "This stone is amazing. Students, come and look at this find!"

Appendix E

MAZE PASSAGE 1

Mr. Lee thought his dog, Little Lee, was the smartest dog on the block. Little Lee brought Mr. Lee his **(lights, every, paper)** and slippers every morning and his **(lose, meat, glasses)** and word find book every evening. **(Little, Trusty, Go)** Lee could sit, speak, stay, and **(shake, years, stop**). He could open doors and turn **(walk, lights, mornings)** on and off. Little Lee could **(down, even, was)** howl Mr. Lee's favorite songs.

When **(he, the, Mr.)** Lee began to lose his sight **(a, and, dog)** few years ago, Little Lee began **(leading, scratched, evening)** Mr. Lee around whenever Mr. Lee **(let, park, put)** him on his leash. Little Lee **(cracked, watched, few)** Mr. Lee's every step. He knew **(howl, red, little)** lights meant stop and green lights **(book could, meant)** go. He helped Mr. Lee cross **(streets, neighborhood, walk**), get on elevators, and shop for **(groceries, songs, always**).

He had a knack for picking **(had, even, out)** the more tender pieces of steak and **(chocolate, little, pork)** chops at the meat market.

In **(all, the, a)** mornings, Mr. Lee always took Little **(Lee, smartest, park)** for a walk. Sometimes they would **(follow, go, steak)** to the park and Mr. Lee **(night, scratched, would)** let Little Lee off his leash. **(In, After, His)** walking in the park, they would walk down **(and, the, cross)** street a bit further to the **(ice, meat, furry)** cream stand. Mr. Lee always had **(began, chocolate, green)** chip, and Little Lee always had **(French, behind, red)** vanilla. Then Mr. Lee

would follow **(he, over, his)** trusty dog all the way home. **(Entire, Bone, One)** night over a dinner of steak **(and, understood, the)** baked

potatoes, Mr. Lee spoke to **(little, his, that)** dog. "Little Lee, you are not **(only, all, watched)** the smartest dog on the block, **(and, but, right)** I think you are the smartest **(couch, dog, with)** in the neighborhood. In fact you're **(pieces, all, probably)** the smartest dog in the entire **(elevators, city, cross**)." Little Lee wagged his tail as **(he, let, they)** cracked a bone between his teeth. **(You, He, Open)** barked twice to let Mr. Lee **(would, the, know)** that he understood him.

That night **(even, lights, while)** Mr. Lee watched a little television, **(howl, tail, Little)** Lee slept snuggled on the couch **(beside, further, could)** him. Mr. Lee scratched Little Lee **(in, twice, at)** all the right places. He scratched **(up, Little, favorite)** Lee behind his ears, between his **(thought, shoulder, doors)** blades, and up and down his **(trusty, furry, leash)** tummy. As always, Little Lee fell **(park, for, asleep)** with his head on Mr. Lee's **(lap, steak, entire**).

Appendix F

MAZE PASSAGE 2

David was always the first person in his family at the breakfast table. While his sisters were primping their **(could, hair, girls)** in front of the bathroom mirror, **(David, they, fresh)** was already halfway through his bowl **(in, of, the)** cereal and thinking about what he **(could, was, also)** going to eat next. David would **(pink, tried, eat)** anything— as long as it was **(students, breakfast, poached**). One of David's favorite meals was **(sausage, find, piles**), hash browns, and eggs. He also **(already, adored, served)** French toast, blueberry pancakes, and oatmeal **(of, egg, with)** brown sugar and fresh cream. He **(one, going, would)** eat poached eggs at the drop **(of, in, and)** a hat, and he could devour **(a, have, the)** half-foot pile of flapjacks in **(with, one, seventy**)-eight seconds flat.

David would rather **(brown, starve, tell)** than eat the spaghetti and meatballs **(a, the, his)** school cooks served the students for **(snacks, blueberry, lunch**). What David WOULD eat was a **(bacon, oatmeal, drop)** and egg sandwich and piles of **(fresh, favorite, place)** fruit.

A pink grapefruit with sugar **(would, him, was)** one of David's more frequent snacks.

"(Someday, One, Taking) of these days you're going to **(already, have, could)** to broaden your tastes, David," his **(first, sisters, mother)** would tell him as she packed **(him, than, you)** hardboiled eggs for lunch. "Someday you're **(going, thinking, students)** to find a place that doesn't **(with, eat, serve)** breakfast, and you're going to have **(at, the, to)** try something new."

"But just last **(new, week, lunch)** I tried that onion, mushroom, and **(shrimp, tell, food)** omelet at the restaurant," David told **(your, his, him)** mother. "They're always coming up with **(students, hardboiled, new)** things for breakfast." That day at **(breakfast, school, that**), David ate his hardboiled egg while **(nothing, all, bite)** the other students had pizza and **(cooked, rolled, hesitantly)** broccoli for lunch. Then, across the **(coming, table, pile**), David saw a girl pick up **(something, rest, nothing)** interesting and take a bite out **(for, is, of)** it. She rolled her eyes with **(guarantee, at, delight)** taking another bite. Then she **(started, these, served)** taking bites so fast that the **(try, fascinating, frequent)** piece of food was rapidly disappearing. **(Suddenly, Interesting, But)** she looked up. "You want some?" **"(That, What, Would)** is it?" David said, hesitantly.

"It's **(the, a, for)** cookie. Try it. I guarantee you'll **(she, feel, love)** it."

David took a tiny bite. **(And, A, His)** wonderful taste landed on his tongue **(and, so, the)** made him feel happy. David felt **(then, he, she)** could eat cookies and nothing but **(food, bites, cookies)** for the rest of his life.

Appendix G

MAZE PASSAGE 3

Mr. Mooney is an expert at his occupation. In fact, he is one of **(has, and, the)** few experts left in an occupation **(this, that, right)** is slowly dwindling and lacking well**-(finished, trained, can)** professionals.

Mr. Mooney shampoos animals. He **(were, finished, is)** known to boast that there is **(in, one, no)** animal too filthy, too big or **(too, is, all)** wide, and no animal too ferocious **(to, an, for)** him. He can get them all **(polished, top, clean)** and contrite. Mr. Mooney has the **(right, bristly, week)** to boast. Once, when Mr. Mooney **(top, is, was)** younger, he was called upon to **(shampoo, polished, beauty)** the walruses at the local zoo. **(That, When, If)** Mr. Mooney was finished with those **(heads, walruses, next**), their bristly coats were gleaming, their **(remained, roof, tusks)** were polished to perfection, and the **(next, gleaming, few)** hairs they possessed on the top **(of, the, for)** their heads were fit for a **(assistant, younger, beauty)** pageant.

The next week, the zookeeper **(were, that, was)** heard whispering to his assistant that **(the, for, and)** walruses that were usually crabby and **(impolite, terrible, even)** were unusually courteous and kind. According **(to, their, at)** the zookeeper, they remained that way **(start, from, for)** an entire week after Mr. Mooney **(with, had, have)** soaped them up and washed them **(over, heard, down**). Rumor has it that whenever they **(give, see, for)** someone coming at them with a **(scrub, filthy, even)** brush and a bar of rose**-(coated, scented, animal)** soap, they start to squeal with

(glory, everywhere, delight), flip over on their backs, and **(wag, fit, squeal)** their tails in excitement.

Whenever the **(perfection, walruses, zookeeper)** are brought up, Mr. Mooney smiles. **(Yes, Fact, Arrived**), they were a triumph for him, **(he, but, been)** if asked what his most memorable **(help, to, job)** was, he will tell you about **(them, the, impolite)** time he was asked to groom **(her, Mrs., rose)** Richman's peacocks. Even to this day, **(those, when, were)** Mr. Mooney recalls the sight that **(greeted, spilled, brush)** him when he arrived at Mrs. **(Mooney, flip, Richman's)** mansion, he shudders. Apparently there had **(been, at, spilled)** a terrible accident while some workers **(polished, were, was)** repairing Mrs. Richman's roof and tar **(coats, been, had)** spilled everywhere. As Mr. Mooney drove **(finished, up, smiles)** to her large house, he couldn't **(help, boast, sight)** but gasp at the flock of **(walruses, polished, peacocks)** coated in a thick, black mess. **(Mansion, Remained, Their)** beautiful plumage was black and their **(hairs, eyes, large)** were tragic and downcast. Never one **(to, at, coated)** give up or despair, Mr. Mooney **(greeted, feathers, climbed)** out of his van and went **(on, to, they)** work.

Appendix H

Procedural Checklist

Examiner: _____

Observer: _____

Steps	Observation
Place student copy in front of reader.	
Seated appropriate distance from reader.	
Places examiner copy out of view of reader.	
Says standardized directions.	
Starts stopwatch at the appropriate time.	
Monitors student for accuracy.	
Says “STOP”	
Monitors student to ensure students stop	
Collects Materials.	
Total Steps Correct	