Analyzing the Effects of Two Response to Intervention Tools, Oral Reading Fluency and Maze Assessments, in the Language Arts Classrooms of Middle School Students

A Dissertation

Presented to

The Faculty of the School of Education

Liberty University

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Education

By

Donna L. Jones

October 2009

Analyzing the Effects of Two Response to Intervention Tools, Oral Reading Fluency and Maze Assessments, in the Language Arts Classrooms of Middle School Students

by Donna Lyerly Jones

APPROVED:

COMMITTEE CHAIR	Dale Clemente, Ed.D.
COMMITTEE MEMBERS	Roy Bradley Brown, Ed.D.
	Catherine Chatmon, Ed.D.
ASSISTANT DEAN, ADVANCED PROGRAMS	Scott B. Watson, Ph.D.

ABSTRACT

Donna Jones. ANALYZING THE EFFECTS OF TWO RESPONSE TO INTERVENTION TOOLS, ORAL READING FLUENCY AND MAZE ASSESSMENTS, IN THE LANGUAGE ARTS CLASSROOM OF MIDDLE SCHOOL STUDENTS. (Under the direction of Dr. Dale Clemente) School of Education, September, 2009.

This quantitative study analyzed data to find a valid and reliable assessment for progress monitoring also having predictive power of a student's future reading performance on a state-mandated standardized reading achievement evaluation. The Response to Intervention (RTI) model was implemented in the language arts classrooms of a rural middle school in northeast Georgia to study the effectiveness of instruction for all students, the at-risk, general, and advanced population. This study used Oral Reading Fluency (ORF) and Maze fluency assessments to monitor student progress and to analyze data to drive instruction. The data were gathered weekly over a 23 week period, rotating the ORF and Maze assessments. Stratified random sampling was used to choose the students receiving ORF assessments. The Maze assessments were given to all students in the language arts classrooms. The data were analyzed through multiple regressions to find if there was a relationship between the ORF and Maze assessments and Georgia's reading portions of the Criterion Referenced Competency Test (CRCT) or the Ninth Grade Literature and Composition End of Course Test (EOCT). The data indicated that the ORF and Maze assessments were significant predictor variables for the CRCT and EOCT, and the ORF data indicated a stronger correlation.

ACKNOWLEDGEMENTS

To my husband and best friend of 28 years, Steve Jones, I give my gratitude, thanks, and unconditional love for his willingness to encourage me throughout my dissertation process and to listen and learn about reading and progress monitoring. He even found ways to apply it on the football field and in the physical education classroom.

To my children, Kristen and Danny (my new son-in-law), Ben, Sara Beth, and Nathan, I give thanks for helping me to keep real life front and center throughout the process from a wedding to band competitions to school to football. They add the spice to my life that gives balance, excitement, and joy.

To my parents, Rev. C. Donald and Rebecca Lyerly, I give thanks for teaching me how to persevere even in adversity and to do any task that is before me with the underlying passion to help others.

For my chair, Dr. Dale Clemente, I give my gratitude for her leadership and expertise in helping to make this study successful.

For Dr. Catherine Chatmon, I owe thanks for her expertise of reading.

For Dr. Brad Brown, I owe thanks for his knowledge of the dissertation process, his consistent availability, and his incredible leadership. He demonstrated why our school excels and his belief that all students can learn. The reliability and validity of this study was enhanced through his endeavor to keep the methodology sound.

To the teachers and students at North Hall Middle School, what a great learning community! The teachers carried out the study with integrity and became great data analyzers while the students took ownership of their learning process.

To John Hewitt, I want to thank for helping me along the way to keep the data valid and unbiased.

To the foundation of my life, my heavenly Father, I give thanks for being the truth that guides my life. “So that *my* trust may be in the Lord, I teach you today, even you….teaching you true and reliable words, so that you can give sound answers to him who sent you.” (Proverbs 22:19-21 New International Version)

TABLE OF CONTENTS

ABSTRACT…….....iii

ACKNOWLEDGEMENTS…...…….iv

LIST OF TABLES……..…….....x

LIST OF FIGURES……...……xii

I. INTRODUCTION……....1

Background of the Study…….1

Statement of the Problem……...…….….4

Purpose of the Study……4

Research Questions……..5

Professional Significance of the Study……...….8

Overview of the Methodology…….9

Definitions of Terms……..……10

Organization of the Study……..17

II. THEORETICAL REVIEW OF LITERATURE...……18

Introduction……...…….18

Historical Background of Reading…….....18

Identification of Crucial Findings in Reading……...18

Implementing Educational Reform……19

Important Components of Reading Instruction……...…20

Whole Language to Balanced Reading Approach……..…21

Accountability Issues…….22

Response to Intervention (RTI) Framework……..23

Measurement Tools……26

Norm-referenced and Criterion-Referenced Assessments…….26

Measuring Reading with Lexiles: The Lexile Framework……29

Progress Monitoring Tools: Curriculum Based Measurement (CBM)…33

The Construct of Reading……..36

The Definition of Reading and Identifying At-Risk Readers……36

The Literacy Puzzle……...38

Incorporating the Essential Components of Reading…….42

The Importance of Fluency……44

The Fluent Reader……..46

Using Fluency to Assess Reading Progress……...47

Directions for Curriculum Based Measurement (CBM)……48

Directions for Oral Reading Fluency…….……48

Directions for Maze Assessment……...52

Using CBM as a Progress Monitoring Tool……..53

Effects of Progress Monitoring Materials……..56

Effects of Curriculum on Progress Monitoring…….56

The Effects of the Difficulty Level of Progress Monitoring Probes……..60

Growth Rates and Progress Monitoring……...……..63

What the Research Means to this Study……65

III. METHODOLOGY AND RESEARCH PROCEDURES……67

Research Design, Context, and Access……..……67

Population……..68

Instrumentation.....68
The Validity of ORF and Maze Assessments…… 68
The Reliability of ORF……..69
The Reliability of the Maze Assessment……...…70
Overview of the CRCT……..……71
The Validity of the CRCT……..…….....71
The Reliability of the CRCT Scaled Score……72
The Reliability of the CRCT Lexiles…….……73
The Reliability of the CRCT Level……...…….73
The Validity of the EOCT……..74
The Reliability of the EOCT……..74
Procedures of Curriculum-Based Measurement……75
Preparation of Probes…….76
Procedures……..77
Procedures for ORF Assessment……...…77
Data Collection for ORF……....78
Procedures for the Maze Fluency Assessment……...82
Data Collection for the Maze…….84
Data Analysis.....84
Potential Findings……..85
IV. RESEARCH FINDINGS…….87
Demographic Data…….87
Demographics of the Maze Assessment……88
Demographics of the ORF Assessment…….89

Overall Explanation of Data……..90

Results of 6th Grade……93

Results of 7th Grade……..……...…97

Results of 8th Grade……..……..……..101

Research Questions…….…….109

Research Question 1……109

Research Question 2……112

Research Question 3……116

Summary of Findings……...……...120

V. DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS……...…121

Discussion of Findings…….121

Implications……. 129

Limitations……...…131

Recommendations…...…….133

Summary…..……135

REFERENCES…...…….138

APPENDIXES…….160

A. ORF PROBE……...162

B. MAZE PROBE……..……..164

C. TEACHER FLUENCY PACKET ……..…167

LIST OF TABLES

1. CRCT and EOCT Scaled Reading Scores Compared to Level...…….…….28

2. Determining Placement Level Based on ORF……...……49

3. Expected Growth Rates in ORF……..……...50

4. Reliability Coefficients (Cronbach's Alpha) and the Raw Score for the SEM for Middle School Students Who Took the 2008 and 2009 Georgia CRCT…… 73

5. The CSEMs at the Cut Scale Score that Define the Performance Levels of the 2008 and 2009 CRCT……..……...……...74

6. Summary of Coefficient Alpha and SEMs across Administrations for 2007-08 Ninth-Grade Lit EOCT……..…….……..…….75

7. Summary of CSEM for 2007-08 Ninth-Grade Lit EOCT…..……...75

8. Correlation of Lexiles to Proficient to Proficient and Exemplary Based on 2007-08 CRCT Scores…….79

9. Number of Students Assessed with Maze Assessment…….…….87

10. Students with Complete Data Monitored with Correct Methodology…… 89

11. Totals for the 2008-09 School Year and Those Randomly Selected for ORF..……...90

12. Averages of Assessments, Words Correct Per Minute (WCPM), Correct Responses……...……91

13, 6th Grade ORF Summary Output……...……..…93

14. 6th Grade Maze Summary Output……94

15. 7th Grade ORF Summary Output…….97

16. 7th Grade Maze Summary Output……....98

17. 8th Grade ORF Summary Output……...…101

18. 8th Grade Maze Summary Output……...…102

19. 8th Grade ORF Summary Output for Students Taking EOCT……...…105

20. 8th Grade Maze Summary Output for Students Taking EOCT……..……106

21. Comparison of All Students and Those Monitored with Correct Methodology……129

LIST OF FIGURES

1. Typical Reader and Text Measures by Grade……30

2. Predicted Comprehension of Books at Various Lexile Measures……..…….. 30

3. 6th Grade ORF and Maze Line Fit Plots……95

4. 6th Grade ORF and Maze Residual Plots…….…..96

5. 7th Grade ORF and Maze Line Fit Plots…….……...99

6. 7th Grade ORF and Maze Residual Plots…….……100

7. 8th Grade ORF and Maze Line Fit Plots..……103

8. 8th Grade ORF and Maze Residual Plots…….……104

9. 8th Grade ORF and Maze Line Fit Plots from Students Taking EOCT……...107

10. 8th Grade ORF and Maze Residual Plots from Students Taking EOCT……108

11. Regression Equations to Predict the Future CRCT Score……...……. 115

CHAPTER ONE: INTRODUCTION

Finding a valid and reliable means for teachers to assess the performance and progress of students' skills in reading and ensuring that students can read and be successful has become a daunting task for teachers (Hosp & Hosp, 2003). Many of the states across America have adopted standardized reading assessments as a way to measure that *all* students can read based on the federal government's guidelines under the *No Child Left Behind Act of 2001* (U.S. Dept. of Ed, 2003). The phrase, *all students,* has especially created more concern among educators because of the students who do not meet the reading standards of their particular grade level creating a label of being at-risk in reading. What about the students who read above the standard level? Should they be left behind or held to a lower standard? Should the progress of advanced readers be monitored as well? Realizing the importance of progress monitoring data in determining the effectiveness of an intervention through the Response to Intervention (RTI) model and assessing every student's growth are vital in the learning process (Deno, Fuchs, Marston, & Shin, 2001). With the plethora of reading assessments available, finding the most effective one can be a difficult task for an educator. Finding a valid and reliable assessment to help a teacher monitor a student's progress and guide instructional practice will make progress monitoring more effective and efficient by providing a tool to identify that each student is following a normal academic progression in reading as well as providing an indicator of a student's future performance on the summative evaluation.

Background of the Study

"Reading ability is central to students' learning, to their success in school, and

ultimately to their success in life" (Salinger, 2003, p. 79). Children and adults that cannot read have a disadvantage (Salinger, 2003). Regrettably, children entering kindergarten with poor literacy and language skills seldom catch up in school and have a greater chance of being referred for special education services (Whitehurst & Oningan, 1998, 2001). Snow, Burns, and Griffin (1998) found that a child who cannot read by the end of second grade has only a 25% chance of reading at grade level by the end of his or her elementary years. Most of the children in fourth grade with reading problems will continue to have reading difficulties by the end of their high school years and also have a higher probability of dropping out of school (Scarborough, 2001). More than 75% of students who drop out of school credit it to difficulties in learning to read (Lyon, 2001). Statistics show that approximately eight million young people between fourth and twelfth grade struggle to read at grade level. About 70% of older readers require some form of remediation (Biancarosa & Snow, 2004). Surveys of adolescents and young adults who have criminal records show that at least half have problems in reading (Lyons, 2001). In some states, the size of prisons a decade in the future can be predicted by fourth grade reading failure rates (Lyons, 2001). About half of children and adolescents with a background of substance abuse have reading problems (Lyons, 2001). Because of these statistics, the National Institute of Child Health and Human Development (NICHD) considers reading failure to show a national public health problem (Lyons, 2001). These findings triggered educational reform.

In the 90s, the federal government commissioned Congress to form a panel to study effective reading programs and research-based reading practices that were vital to a reading program. The National Reading Panel (NRP, 2000) was formed and found that evidence-based practices, such as phonemic awareness, phonics, fluency, vocabulary and

comprehension strategies, can make a meaningful difference with at-risk or struggling readers. Implementing these reading components can sufficiently close the gap in their reading skills (Marchand-Martella, Martella, & Przychodzin-Havis, 2007).

> Given the importance of reading and the overwhelming number of students who struggle with reading beyond Grade 3, we are left with the conclusion that with strong literacy skills, doors open for individuals; with poor literacy skills, doors close for them. Focused and intensive reading intervention is the key to unlock these doors and allow individuals to access the working world more successfully. (Marchand-Martella, Martella, & Przychodzin-Havis, 2007, p. 2)

Without effective assessments, there will be a lack of proper curricula and interventions (Ardoin, 2006). Ineffective assessments may be detrimental to the at-risk and also to the gifted students because they may be left under-challenged (Volker, Lopata, & Cook-Cottone, 2006). Educators need assessments that provide answers and help them to identify a student's academics strengths and weaknesses to be able to make educational decisions that will benefit the student and not just take up more time (Olinghouse, Lambert, & Compton, 2006). Based on the use of effective progress monitoring, a teacher may choose to increase the amount of instruction, slow the pace, or change the methods used completely (Olinghouse, Lambert, & Compton, 2006). The use of effective progress monitoring can help a teacher to identify a need in a particular student for further testing or other interventions, and it may even be used to help in making special education eligibility decisions (Deno, Fuchs, Marston, & Shin, 2001). The data can evaluate students' maintenance of intervention effects or utilize the Response to Intervention (RTI) model analyzing the effectiveness of an intervention providing an objective way to use data to differentiate instruction (Ardoin, 2006). The differentiated

instruction may identify the teacher's need to increase the level of knowledge that students need or to provide remedial instruction to review problem areas (Barnes & Harlacher, 2008).

Statement of the Problem

Curriculum-based measurement (CBM) is a way to help students and teachers to evaluate what has been learned, what needs to be learned, and what needs to be revisited, but it must be used and analyzed to help and not hinder the student and the educational process (Ardoin, 2006). Two CBM tools that were found to be effective and efficient in indicating reading achievement were oral reading fluency (ORF) and Maze assessments (Fuchs, Fuchs, Hosp, & Jenkins, 2001; Good, Simmons, & Kame'enui, 2001; Shinn, 1989, 1998; Tindal & Marston, 1990). Studies of the ORF and Maze assessments suggested that these assessments have predictive power on state-mandated high-stakes assessments (Good et al., 2001; McClinchey & Hixon, 2004; Schulte, Villwock, Whichard, & Stallings, 2001; Stage & Jacobsen, 2001). Future studies of these assessments may hold the keys to a progress monitoring tool that follows a student's reading progress from kindergarten to 12th grade and may be a vital tool in identifying effective intervention strategies utilizing the Response to Intervention (RTI) model. There is a current need for additional studies that focus on the effects of ORF and Maze assessment as progress monitoring tools for middle school students. Future studies are needed to see if these tools are reliable in identifying effective intervention strategies for reading within the RTI model (Wayman, Wallace, Wiley, Tichá, & Espin, 2007).

Purpose of the Study

The purpose of this quantitative study was to find a valid and reliable assessment, Oral Reading Fluency or Maze assessment, for progress monitoring in the general

population of a rural middle school in northeast Georgia that incorporated the Response to Intervention (RTI) model to promote the development of literacy skills for all students. The data were analyzed to find the assessment that had predictive power of a student's future reading performance on the state-mandated summative evaluation.

Research Questions

1. Will using an oral reading fluency (ORF) or Maze assessment be a valid and reliable progress monitoring tool in predicting reading achievement for a middle school student?

Hypothesis (H_1): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for 6th grade students.

Hypothesis (H_2): There is a statistically significant relationship between the scores from the Maze assessment and the reading scores from the Georgia Criterion-Referenced Competency Test for 6th grade students.

Hypothesis (H_3): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for 7th grade students.

Hypothesis (H_4): There is a statistically significant relationship between the scores from the Maze assessment and the reading scores from the Georgia Criterion-Referenced Competency Test for 7th grade students.

Hypothesis (H_5): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for 8th grade students.

Hypothesis (H_6): There is a statistically significant relationship between the scores from the Maze assessment and the reading scores from the Georgia Criterion-Referenced Competency Test for 8th grade students.

Hypothesis (H_7): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the scores from the End of Course Test for 8th grade students taking 9th Grade Literature and Composition.

Hypothesis (H_8): There is a statistically significant relationship between the scores from the Maze assessments and the scores from the End of Course Test for 8th grade students taking 9th Grade Literature and Composition.

2. Will the ORF or the Maze be able to predict a middle school student's reading performance on the Georgia Criterion Referenced Competency Test (CRCT) or the Georgia Literature End of Course Test (EOCT)?

Hypothesis (H_9): The oral reading fluency assessment is a predictor variable for reading performance of 6th grade students on the Georgia Criterion-Referenced Competency Test.

Hypothesis (H_{10}): The Maze assessment is a predictor variable for reading performance of 6th grade students on the Georgia Criterion-Referenced Competency Test.

Hypothesis (H_{11}): The oral reading fluency assessment is a predictor variable for reading performance of 7th grade students on the Georgia Criterion-Referenced Competency Test.

Hypothesis (H_{12}): The Maze assessment is a predictor variable for reading performance of 7th grade students on the Georgia Criterion-Referenced Competency Test.

Hypothesis (H_{13}): The oral reading fluency assessment is a predictor variable for reading performance of 8th grade students on the Georgia Criterion-Referenced Competency Test.

Hypothesis (H_{14}): The Maze assessment is a predictor variable for reading performance of 8th grade students on the Georgia Criterion-Referenced Competency Test.

Hypothesis (H_{15}): The oral reading fluency assessment is a predictor variable for reading performance of 8th grade students taking the Georgia End of Course Test for Ninth Grade Literature and Composition.

Hypothesis (H_{16}): The Maze assessment is a predictor variable for reading performance of 8th grade students taking the Georgia End of Course Test for Ninth Grade Literature and Composition.

3. Will the ORF and/or the Maze provide valid and reliable data for students scoring below standard, meeting standard, or exceeding standard on the summative evaluations?

Hypothesis (H_{17}): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students scoring below standard.

Hypothesis (H_{18}): There is a statistically significant relationship between the scores from the Maze assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students scoring below standard.

Hypothesis (H_{19}): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the

Georgia Criterion-Referenced Competency Test for middle school students meeting standard.

Hypothesis (H_{20}): There is a statistically significant relationship between the scores from the Maze assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students meeting standard.

Hypothesis (H_{21}): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students exceeding standard.

Hypothesis (H_{22}): There is a statistically significant relationship between the scores from the Maze assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students exceeding standard.

Professional Significance of the Study

The professional significance of the study was identified in the following ways. To find an efficient and effective reading assessment tool to use that:

- Improved reading achievement for all students
- Identified academic strengths and weaknesses
- Evaluated the Response to Intervention (RTI) model
- Was used for progress monitoring
- Assessed instructional effectiveness
- Helped in instructional planning
- Predicted performance on the CRCT and EOCT

- Began to develop a continual and flexible reading profile of a student's progress from elementary to high school.

Overview of the Methodology

This quantitative study analyzed the relationship between the predictor variables [mean of progress monitoring data from two fluency assessments, oral reading fluency (ORF) and Maze fluency, scaled scores and level from 2008 CRCT] and a dependent variable in reading, the 2009 CRCT or 2009 Georgia's 9th Grade Literature and Composition EOCT. The researcher trained language arts teachers in one rural middle school in northeast Georgia regarding how to give the ORF and Maze assessments to their students and how to use this data as progress monitoring tools. After establishing a baseline point, the teachers implemented the progress monitoring tool, Maze, twice monthly to establish data for the Response to Intervention (RTI) model and to monitor student growth. The ORF was given once each quarter for regular education students and twice monthly, opposite weeks of Maze, to the at-risk students to be able to monitor the effectiveness of instruction and academic growth. For the study, [5%] of the students were randomly selected from the principal from all levels to be given the ORF twice monthly as well to compare to the Maze data. The data were used to see which assessment was a valid and reliable predictor variable for the CRCT and the EOCT. All the middle school students in the study took the CRCT, and the 8th grade advanced readers took the EOCT as well as the CRCT based on the 9th grade Literature and Composition class that they took during the school year. The CRCT and the EOCT were state-mandated high-stakes tests required by the state of Georgia, identifying if a student does not meet, meets, or exceeds the Georgia Performance Standards (GPS) for the grade level or course (GADOE, 2008a; 2008b). The students were given a score based on their

reading performance. Students taking the CRCT who did not meet the standards, or scored below a scale score of 800, received a Level 1. The students who met the standards, or scored a scale score from 800-849, received a Level 2. The students who exceeded standards, or scored 850 or above, received a Level 3 (GADOE, 2008a; 2008b). These scores provided the subgroups that the researcher analyzed. Students with a Level 1, 2, or 3 were identified respectively in the following subgroups, Level 1, Level 2, and Level 3. Each group was monitored based on their level of reading performance using the ORF and Maze assessments, following the RTI model. The data were used for two purposes: to identify effective or ineffective interventions and instruction and to identify a valid and reliable predictor variable to predict the student's performance on the state-mandated assessment.

Definition of Terms

The following terms are operational definitions of a number of terms that were used in this study.

Achievement levels are levels that are based on recommendations from panels of educators and members of the public based on performance standards in different subject areas. The levels are measured as Basic, Proficient, and Advanced and provide a measure of what a student should know and understand at each grade (National Assessment of Educational Progress (NAEP), 2008).

Advanced readers are students who generally score 90% or above on reading achievement tests recognized as superior achievement (NAEP, 2008).

Alphabetics is reading instruction that includes phonemic awareness and phonics instruction (NRP, 2000).

Analytic phonics instruction is when students are taught whole word units first and then systematic instruction that links the specific letters in the word(s) with the respective sounds (NRP, 2000).

At-risk readers are students who fall below grade level on the standardized and grade level assessments.

Constant is "a characteristic that takes on the same value for all individuals in a study" and is the opposite of a variable (Ary, Jacobs, Razavieh, & Sorensen, 2006, p. 630).

Construct is an image, idea, or theory that is abstract that formed from a combination of simpler observable elements (NAEP, 2008).

Correlational research is "research that attempts to determine the extent and the direction of the relationship between two or more variables" (Ary et al., 2006, p. 631).

Correlation coefficient is "a statistic that shows the degree of relationship between two variables. Its value ranges between -1.00 and +1.00" (Ary et al., p. 631).

Credits earned is the term used on a high school transcript that is changed to standardized Carnegie units. One Carnegie unit in Georgia requires 150 seat hours which is fifty minutes a day, or 135 hours with a block schedule waiver and gives the student if he or she passes it, one high school credit. This credit is provided to middle school students who have demonstrated an advanced level of knowledge in a particular subject (NAEP, 2008).

Criterion Referenced Competency Test (CRCT) is the mandated summative evaluation given in the state of Georgia to measure how well the student's achievement demonstrates knowledge of the Georgia Performance Standards (GPS). These data were

also used to identify students' academic strengths and weaknesses according to the GPS and to measure the quality of education in Georgia (GADOE, 2008c).

Decodable text is controlled vocabulary within a text that is usually taught before the text is read (NRP, 2000).

Distractor is the incorrect choices on a multiple-choice test item (NAEP, 2008).

End of Course Test (EOCT) is the summative evaluation given in eight subject content areas for grades 8 through 12 that measures specific content knowledge and skills of the specific area. The subject areas are Ninth Grade Literature and Composition, American Literature and Composition, Mathematics I, Mathematics II, Physical Science, Biology, U.S. History, and Economics. The EOCT also provides diagnostic information regarding students' academic strengths and weaknesses and the effectiveness of classroom instruction (GADOE, 2008c).

English Language Learners (ELL) are students who are learning the English language concepts and knowledge. Some schools use the term Limited English Proficient (LEP) (NAEP, 2008).

Explicit phonics instruction is teaching phonics with direct instruction and giving the student an opportunity to apply the skills in decodable text formats which usually have a controlled vocabulary. For instance, the sound and word is taught before the student reads the passage, and then as the student reads the passage, the word is used in the text (NRP, 2000).

Fluency is the ability of the reader to read with speed, accuracy, and proper expression (NRP, 2000).

Implicit phonics instruction is an approach that teaches phonics incidentally. The teacher looks for opportunities to teach the elements as they appear embedded in the text. It also uses decodable text less frequently (NRP, 2000).

Intervention is a strategy implemented in the instructional process to help a student who has a problem in the learning process.

Lexile scales are scales that measure both reader ability and text difficulty (MetaMetrics, Inc., 2008).

Lexiles are the most widely accepted reading measure in use today because they provide a tool for measuring students' reading abilities like a thermometer measures temperature (MetaMetrics, Inc., 2008). Lexiles are tools that link assessment with instruction across the curriculum. The letter "L" is used as the abbreviation for Lexile. For example, a student reads at 690 L.

Maze assessment is a fluency assessment given to the student. The student reads a probe of about 400 words for two and one half minutes to three minutes in which every seventh word has been replaced with three choices, one correct choice and two distracters. The number of correct choices is counted and used for progress monitoring.

Median is the middle number in a list of scores.

Microlevel is on an individual level.

Multiple regression is "the prediction of a criterion using two or more predictor variables in combination. Each predictor is weighted in proportion to its contribution to prediction accuracy" (Ary et al., p. 635).

Multiple regression equation is "the equation showing the weights assigned to each predictor" (Ary et al., p. 635).

National Assessment of Education Progress (NAEP) is a national assessment that represents what students know and are able to do in different subjects.

National Institute of Child Health and Human Development (NICHD) was created in 1962 at the request of the President and the support of Congress to research human development from preconception to adulthood, especially focusing on developmental disabilities, such as mental retardation, and studying events occurring during pregnancy (Shriver, 2007).

National Reading Panel (NRP) is the panel formed by the NICHD to investigate the effectiveness of various approaches to teaching children to read (NRP, 2000).

Negative correlation is "a correlation with high scores on one variable associated with low scores on the other variable" (Ary et al., 2006, p.635).

Ninth Mental Measurements Yearbook is a reference source that list tests and provides a critique of the test (Ary et al., 2006).

Oral Reading Fluency assessment (ORF) is an assessment given to students individually. The student reads a probe of about 200 words orally while the evaluator times the student for one minute while listening and marks the errors in reading. The evaluator records the words correct per minute (WCPM).

Phonemes are the smallest units or sounds that make up a spoken language (NRP, 2000).

Phonemic awareness (PA) is a part of reading instruction that includes instruction on how to manipulate phonemes or sounds in spoken syllables and words (NRP, 2000).

Phonics instruction is a way of teaching one to read emphasizing the understanding of letter-sound correspondences and their use in reading and spelling (NRP, 2000). Its primary focus is to help readers understand how letters are connected to

sounds making letter-sound correspondences and patterns in spelling and then applying phonics in their reading (NRP, 2000).

Positive correlation is "a correlation with high scores on one variable associated with high scores on the other variable, and low scores associated with low scores" (Ary et al., p. 637).

Positive predictive power is the probability of a student with a score below the cut score who truly does fail (Ary et al.).

Predictive validity evidence is "the relationship between scores on the measure and criterion scores available at a future time" (Ary et al., 2006, p. 247). In other words, the relationship between the students' scores on a test and the grades that will be eventually earned in a future criterion or standard that will be analyzed, and if a relationship is found, the future performance can be predicted.

Predictor is "a variable from which predictions are made in a prediction study" (Ary et al., 2006, p. 637).

Proficient is the level on a standardized evaluation which identifies a student as having solid academic performance for each grade level (Quatroche, 1999).

Progress monitoring is a means of assessing students and monitoring throughout the school year regarding how they are responding to instruction or the interventions provided and then using the data to make decisions creating an instructional match (Coyne & Harne, 2006; NASDSE, 2006).

Reading First is a program at the national level under *No Child Left Behind* (2001) which is designed to make sure that every student is a successful reader.

Reliability is when a measurement is consistent in measuring (Ary et al., 2006).

Response to Intervention (RTI) is a model in which all students are provided an appropriate level of evidence-based instruction based on the students' needs. The different aspects of the RTI model are frequent assessments of a student's academic progress, decision-making based on data analysis, and placement of students within their appropriate instructional level using different levels of support (Barnes & Harlacher, 2008).

Scale scores represent a range of scores representing a summary of a group of students of what they know and can do in particular subject areas. The CRCT and the EOCT data are presented as scale scores (GADOE, 2008c).

Statistically significant is the result of a statistical test to identify if the relationship between two numbers is statistically significant. Significant means that the observed relationship is not likely to be associated to sampling and measurement error, but it identifies that the relationship observed is less than a previous given possibility of being due to a function of chance (Ary et al., 2006).

Struggling readers are usually behind in grade level or sometimes called at-risk.

Students with disabilities (SWD) are students with a learning disability who may need special instruction to meet his or her educational goals (NAEP, 2008).

Systematic phonics instruction is teaching phonics with a sequential set of phonics elements along with a dimension of explicitness depending on the type of method used (NRP, 2000).

Synthetic phonics instruction is a direct instruction teaching method where the student is taught the individual letter or letter combinations and its appropriate sound, and then he/she is taught to blend the sounds into words (NRP, 2000).

Treatment intervention is the strategy used if a student is struggling academically (Barnes & Harlacher, 2008; Grossen, 1997).

Validity is when an instrument that is used to measure actually measures what it claims to measure (Ary et al., 2006).

Whole Language is a constructivist approach to teaching reading with a holistic perspective where readers use their knowledge of language and symbolic relationships and try to make sense of the text (Goodman, Watson, & Burke, 1987). For example, in the phonics approach, a student learns to sound out a word explicitly or implicitly, but in the whole language approach, the student does not learn to sound out words but uses contextual clues to figure out the word, such as the picture in the book or shape of the word (Dudley-Marling & Paugh, 2004).

Words Correct per Minute (WCPM) is the specific type of data that is collected in an Oral Reading Fluency assessment. The student is timed for one minute and the total words read minus the number of errors is recorded as the WCPM.

Organization of the Study

This study is organized into five specific chapters. Chapter I includes the introduction, background of the study, statement of the problem, purpose of the study, research questions, professional significance, and organization of the study. Chapter II provides a comprehensive review of literature about reading and Curriculum-Based Measurement (CBM) of reading as a progress monitoring tool. Chapter III provides methodology and procedures. Chapter IV gives a summary of findings and Chapter V provides a summary, discussion, and implications of the study as well as discusses the limitations and possibilities for future studies.

CHAPTER 2: THEORETICAL REVIEW OF LITERATURE

Introduction

The purpose of this chapter is to review literature that relates to progress monitoring and reading. The chapter is divided into the following sections: historical background of reading, accountability issues, Response to Intervention (RTI) model, accountability issues and measurement tools, the construct of reading, the importance of fluency, directions for curriculum based measurement (CBM), using CBM as a progress monitoring tool, the effects of progress monitoring materials, and what the research means to this study.

Historical Background of Reading

Identification of Crucial Findings in Reading

During the late 1990s, several crucial findings created a change in America's educational system. In 1997, The National Institute of Child Health and Human Development (NICHD) was asked by Congress to "convene a national panel to assess the status of research-based knowledge, including the effectiveness of various approaches to teaching children to read" (National Reading Panel, 2000). The panel formed was called the National Reading Panel (NRP) which conducted a two-year meta-analysis study of experimental and quasi-experimental research literature.

In 1999, the National Assessment of Educational Progress (NAEP) identified through the NAEP reading report card problems in America's schools. The NAEP (1999) reading report card summarized the results of reading achievement in grades four, eight, and twelve by identifying students at three different levels, basic, proficient, and

advanced. At a basic reading level, students have only partial mastery of the knowledge and skills that are essential to succeed in reading at each grade level. Proficient level means that a student has a solid academic performance and competence as they are given challenging material, and advanced means that they have superior reading performance (Quatroche, 1999). Even though the average reading scores increased for all levels, for 4th, 8th, and 12th graders, the percentages for those who performed at or above proficient levels were 31, 33, and 40 percent and for those who performed at the highest percent, the percentages were 7, 3, and 6 percent, respectively (Quatroche, 1999). At 4th grade level, there were no significant changes from the 1994 and 1992 assessments.

These findings caused concern because only one third or less of the students showed that they could read above the basic reading level (Quatroche, 1999). The National Assessment of Educational Progress on Reading led to the birth of the *No Child Left Behind Act of 2001 (NCLB),* and the findings of the NRP became significant in the development of Reading First, the literacy component of the *NCLB* which created a drastic change in the educational systems across America.

Implementing Educational Reform

On December 13, 2001, the House approved President Bush's education reform act, *NCLB,* and began to redefine the role of the federal government in regards to the education of all students in kindergarten through twelfth grade by focusing on accountability, teacher quality, parental notification, and resources (Miller, 2001). Because of the lawmakers' analysis of the data from the multiple assessments regarding the reading achievement of the students across the nation, they created *NCLB* as a response. *NCLB* represented four major principles, one being

> H.R. 1 will result in the creation of assessments in each state that measure what children know and learn in reading and math in grades 3-8. Student progress and achievement will be measured according to tests that will be given to every child, every year. (U.S. Dept. of Ed, 2003)

Because of the disturbing literacy situation across the United States and the implementation of *NCLB,* federal and state policymakers took initiatives to fix the reading problem (Brynildssen, 2002). These initiatives were two-fold. One was to be in compliance with NCLB, and the other was to find ways to mend the problem of illiteracy. *NCLB* (2001) required that each state had to:

1. prepare an annual report showing the greatest gains in reading achievement
2. reduce the number of children in grades 1-3 who read below grade level
3. increase the percentage of children who read at grade level or above.

To demonstrate effectiveness of reading instruction, *NCLB* (U. S. Dept. of Education, 2001) required states to measure progress annually in reading in grades three through eight. This mandate caused educators to revisit their reading instruction and find ways to reach all students by finding effective instructional strategies that help students become proficient readers (NRP, 2000).

Important Components of Reading Instruction

The NRP (2000) found the following items to be the most important components of reading instruction: alphabetics (including phonemic awareness and phonics instruction), fluency, comprehension, teacher education, and computer technology. These components in reading instruction became debated issues throughout educational systems (Dudley-Marling & Paugh, 2004; Foertsch, 2003; Smith, 2003). Some students were caught in the middle of the whole language or systematic direct instruction debate

creating gaps in their reading instruction. Treatment intervention research suggested that appropriate early direct instruction seemed to be a good strategy for reading problems, but without the appropriate assessments that monitored a student's reading progress to identify academic strengths and weaknesses, students received continuous interventions that were ineffective (Grossen, 1997). Therefore, effective progress monitoring assessments became important in identifying the students who were at-risk and finding an intervention that worked (Taylor, 2004).

Whole Language to Balanced Reading Approach

The National Reading Panel's report triggered a new philosophy toward reading, a balanced systematic approach. As the students witnessed these changes in their elementary years, the pendulum began to swing from a whole language approach that had its roots in the educational philosophy of John Dewey (Gutek, 2005) to a more balanced systematic research-based approach which was rooted in *NCLB* (U. S. Dept. of Education, 2001). Teachers had different philosophies of education due to previous training (Foertsch, 2003). Some teachers continued to use the whole language approach while others made changes. The different philosophies of reading caused gaps in the reading progress of some students.

Different philosophies regarding phonetics instruction was and still is one of the most debated issues that affect the progress of reading achievement in students. The National Reading Panel (2000) found that "systematic phonics instruction makes a bigger contribution to children's growth in reading than alternative programs providing unsystematic or no phonics instruction" (p. 2-92). The NRP stated,

> …it is important to emphasize that systematic phonics instruction should be integrated with other reading instruction to create a balanced reading program.

> Phonics instruction is never the total reading program...By emphasizing all the processes that contribute to reading growth, teachers will have a better chance of making every child a reader. (p. 2-136)

With the NRP's findings, schools looked at each component and searched for a valid and reliable means to make sure that all students could read. Educational systems sought to find out how students could be assessed effectively in reading comprehension. How does a teacher know that a student has reached the capabilities to read and comprehend appropriate grade level material? In other words, if a student is in seventh grade, can the student read and comprehend seventh grade material, or if a student is in eighth grade, can he or she read and comprehend eighth grade material? If the answer is yes, then the student should be successful on a standardized assessment, meeting the federal government's requirements under *NCLB* (U. S. Dept. of Education, 2001). If the student cannot read at the appropriate level, then the student may struggle on the standardized summative evaluation. Trying to find a way to identify academic strengths and weaknesses, many educational systems focused on the students' reading abilities.

Accountability Issues

Learning to read and reading to learn is a process that affects every academic subject (Marcon, 1995). Some children easily master the process, and some children struggle. As the mandates of *No Child Left Behind* (NCLB) were signed into law on January 8, 2002 by President Bush, the educational systems increased their search for ways to reach the established guidelines of *NCLB.* The Individuals with Disabilities Education Act of 2004 (IDEA; 2004) included many of the elements of *NCLB* (Byrnes, 2008). U.S. Secretary of Education Margaret Spellings praised the final regulations of IDEA 2004 with these words, "No Child Left Behind and the Individuals with

Disabilities Education Act have put the needs of students with disabilities front and center. We now have a laser-like focus on helping these kids" (Byrnes, 2008, p. xxx).

One of the components in the new IDEA 2004 was the issue that schools face in identifying students with disabilities. Questions were raised due to increasing amounts of criticisms about the large number of students who were being labeled as having a specific learning disability. As a result, IDEA 2004 encouraged schools to stop using the aptitude/achievement gap in favor of determining first if a student demonstrates improved learning through response to intervention. IDEA 2004 stated that if students did not receive "experienced sound instruction in the basic elements of reading" (Byrnes, 2008, p. xxxi), schools could not identify them as eligible for special education. In 2007 in the State of the Union Address, President Bush (2007) summarized what the law established and where the law would take America's educational system.

> Five years ago, we rose above partisan differences to pass the *No Child Left Behind Act*, preserving local control, raising standards, and holding those schools accountable for results.... Now the task is to build on the success, without watering down standards, without taking control from local communities, and without backsliding and calling it reform….We know what works: high standards, accountability, more choices for parents, and sound, proven methods of instruction. These principles have yielded real and sustainable results. (¶ 13)

With these words, explicit guidelines were set. The meaning of the individual terms from President Bush's speech became important to understand, and the Response to Intervention model was implemented in many states.

Response to Intervention (RTI) Model

The RTI model provided a multi-tiered hierarchy model as a choice for each

school district to use in making eligibility decisions (Ardoin, 2006). This hierarchy of learning ensured that students had high-quality literacy practices available in their learning process (Gettinger & Stoiber, 2007). RTI was a more proactive and preventative approach to be used by educators (National Association of State Directors of Special Education (NASDSE), 2006). RTI was a model that provided a means of identifying if students were being taught an appropriate level of evidence-based instruction based on the students' needs. The components of the RTI model were frequent assessments of a student's academic progress, decision-making based on data analysis, and placement of students within their appropriate instructional level utilizing different levels of support (Barnes & Harlacher, 2008). Barnes and Harlacher defined five principles of RTI as follows: "(1) a proactive and preventative approach to education, (2) ensuring an instructional match between student skills, curriculum, and instruction, (3) a problem-solving orientation and data-based decision making, (4) use of effective practices, and (5) a systems-level approach" (p. 419). The features and principles overlap within the RTI model. These features were identified as: "(1) multiple tiers, (2) assessment system, (3) protocol, and (4) evidence-based instruction" (Barnes & Harlacher, 2008, p. 420).

Gettinger and Stoiber (2007) describe a Three Tier approach which was the approach that this study implemented. In this approach, Tier 1 includes providing all students with a literacy-rich environment, a research student-based decision making process, placement of students within a range of instructional supported curriculum, and instructional activities to support the students' literacy development. In other words, the main goal is establishing "exemplary, scientifically based literacy practices within each classroom by focusing on environmental quality…a comprehensive curriculum and research-based literacy strategies" (Gettinger & Stoiber, 2007, p. 202). At this level,

teachers are involved in professional development, coaching sessions, and collaborative planning to ensure high-quality, scientifically-based instruction that creates a literacy-rich classroom (Gettinger & Stoiber, 2007).

Tier 2 becomes more directed to students who may exhibit the need for extra assistance, and it includes instruction that is repeated daily, teacher-directed, smaller instructional groups with more exposure to language and print, extra practice with literacy skills that fills in gaps in the student's reading process, and/or activity adaptations for the smaller groups of students based on individual learning needs (Gettinger & Stoiber, 2007). In Tier 2, teachers and assistants receive training and coaching to be able to implement this instruction as well (Gettinger & Stoiber, 2007).

Tier 3 becomes more intensive instruction incorporating individualized tutoring meeting specific learning needs. Tier 3 instructors or tutors are trained to provide explicit and highly focused instruction (Gettinger & Stoiber, 2007). There are three general views on when a student can be referred for Special Education services. According to the RTI model of Fuchs and Fuchs (2005), the evaluation can be after a student receives Tier 2 instruction; therefore, in their model, Tier 3 is special education. Another view included the referral process as part of Tier 3 because students may or may not qualify for special education (Marston, Lau, & Carter, 2003).

Barnes and Harlacher (2008) discussed a Tier 4 model which gave individualized instruction at Tier 3 and then referred for Special Education at Tier 4 (Ikeda, Grimes, Tilly, Allison, Kruns, & Stumme, 2002; Reschly, 2005). The four-tier model of Ikeda and others (2005) was as follows: Tier 2 (standard intervention), Tier 3 (individualized problem-solving), and Tier 4 (special education). In the RTI model for younger children, Tier 2 and Tier 3 include stronger dosages of the same reading content of Tier 1;

whereas, for older students, Tier 2 and 3 interventions are provided with different concepts as those in Tier 1.

Regardless of which RTI model is implemented, all have the same basic goal—to link data as the basis for instructional decisions. Within the RTI model, assessments or progress monitoring tools must be incorporated to establish the need for a student to move from Tier 1 or 2 and to establish the effectiveness of interventions. The progress monitoring tools should be valid and reliable tools with good predictive power (Gettinger & Stoiber, 2007). Some measurement tools that are valuable under the RTI model are norm-referenced, criterion-referenced, and informal assessments.

Measurement Tools

States began to implement different assessments as ways to demonstrate accountability, high standards, and implementation of the RTI model. Some still debated the issues, but the stakes were usually too high not to conform. One of the major guidelines was that by 2014, all states would be held accountable for making sure that all students could read at grade level; therefore, most states searched for evidence-based answers to help them achieve this goal. (U.S. Department, 2003). To identify if a state is meeting accountability standards, the law provided the following guidelines: Each state will disaggregate their test scores. Each state will take part in the National Assessment of Educational Progress (NAEP). Each state will report state and NAEP findings to parents (U. S. Department, 2004). The focus turned to students of all ages who were at-risk in reading leading many states to implement standardized assessments based on adopted standards to measure reading.

Norm-referenced and Criterion-Referenced Assessments

To be able to identify strengths and weaknesses in students and their instruction in

reading, school systems used norm-referenced tests to demonstrate how the individual student's performance on the reading component compared to other students of the same age across the United States. They also implemented criterion-referenced assessments that demonstrated what a student can do without a relation to others (Ary, Jacobs, Razavieh, & Sorensen, 2006). The performance of the student on the criterion assessment was reported as the mastery of level of a defined content or skill domain, such as standards or benchmarks adopted by the state (Ary et al., 2006). These standards include specific reading content for each student to learn from kindergarten to twelfth grade.

An example of using criterion to measure reading performance is the researcher's state, Georgia. The End of Course Test (EOCT) is the standardized assessment for 9th grade literature and the Criterion Referenced Competency Test (CRCT) is the standardized summative assessment to measure reading success according to the reading Georgia Performance Standards (GPS) adopted by the state (Georgia Department of Education [GADOE], 2008c). According to Georgia rules, a student must pass the reading portion of the CRCT in third, fifth, and eighth grade to be eligible for promotion to the next grade. The EOCT is given to any student completing the 9th Grade Literature and Composition course. This test is a measurement of a student's performance based on the GPS (GADOE, 2008c). The EOCT is not a pass/fail test but makes up a certain percentage of a student's grade for the course. Both tests were created using standards-based test development steps to establish validity and reliability (GADOE, 2008d). The steps began with the development of test specifications, content domain specifications, preliminary performance level descriptors, and test item specifications. The final steps were the writing of test items, the review and approval or revision of newly developed test items, field testing of new items, review of field-tested items with statistics,

development of the test form, the standard setting, and finally the technical documentation detailing the reliability and validity of the tests. The researcher participated in the CRCT Data Analysis and Item Review. The EOCT and CRCT have demonstrated through statistics reliability and validity (GADOE, 2008d).

The Georgia Department of Education (GADOE) has adopted grade level specific criterion or standards to identify specific elements of reading. The Georgia Department of Education (2008c) states,

> Testing measures student achievement of the state mandated curriculum, to identify students failing to achieve mastery of content, to provide teachers with diagnostic information, and to assist school systems in identifying strengths and weaknesses in order to establish priorities in planning educational programs.

Table 1 identifies the scaled scores for measuring the mastery of reading content according to the Georgia Department of Education (2008c).

Table 1

CRCT and EOCT Scaled Reading Scores Compared to Levels

Georgia Performance Standards (GPS)	Does Not Meet Expectations Basic Level	Meets Expectations Proficient Level	Exceeds Expectations Advanced Level
Reading1-8 CRCT	Below 800 Score 1	800 to 849 Score 2	850 or Above Score 3
9th Grade Literature and Composition EOCT	Below 400	400 to 449	450 or Above

According to an analysis of the data from the Georgia 2007-08 CRCT, students scoring a scaled score of 800 typically get about 50% of the questions correct on the standardized evaluation, and those scoring 850 generally have about 90% correct. After the state mandated assessment is given, the state provides a report to the school showing

a summary of the student's progress in reading upon completion of instruction (Cobb, 2003, GADOE, 2008a). They present a final analysis of the reading instruction and assessments that the students received throughout the school year. The CRCT complies with the mandates of *NCLB* (2001) and provides the federal government with the data used to grade the school's and system's performance. Each school receives a report card from the federal government identifying if they have made Adequate Yearly Progress (AYP) according to the guidelines of *NCLB* (2001).

Measuring Reading with Lexiles: The Lexile Model

Along with the CRCT Reading scores, a Lexile score is reported as well that provides a vital measurement of reading. The Lexile is the most widely accepted reading measure in use today because it gives a "thermometer for measuring students' reading abilities" (MetaMetrics, Inc., 2008, ¶8). Lexiles provide a measurement tool that links assessment with instruction across the curriculum. The Lexiles connect student to text.

The Lexile scale ranges from 200L (beginning readers) to 1700L (advanced readers) (MetaMetrics, 2007). The Lexile Model was created to use the Lexile measure and scale to match the reader with the text. The purpose of the Lexile Model is to connect readers to text in which they will have 75% comprehension. The Lexile measures are founded on two reliable and valid predictors of difficulty of a text in regards to comprehension: semantic difficulty or word frequency and syntactic complexity or sentence length (MetaMetrics, 2007). When a book, text, or assessment is given a Lexile, the following process takes place: the text is divided into 125 word sections. Each section is compared to about 600-million word Lexile corpus which are taken from different sources and genres, and each word in the sentence is counted (MetaMetrics, 2007). Figure 1 shows the range of Lexiles within each grade level reader and text.

Figure 1

Typical Reader and Text Measures by Grade (MetaMetrics, 2008, ¶ 27)

1	Up to 300L	200L to 400L
2	140L to 500L	300L to 500L
3	330L to 700L	500L to 700L
4	445L to 810L	650L to 850L
5	565L to 910L	750L to 950L
6	665L to 1000L	850L to 1050L
7	735L to 1065L	950L to 1075L
8	805L to 1100L	1000L to 1100L
9	855L to 1165L	1050L to 1150L
10	905L to 1195L	1100L to 1200L
11 and 12	940L to 1210L	1100L to 1300L

Figure 2

Predicted Comprehension of Books at Various Lexile Measures (MetaMetrics, 2008, ¶ 5)

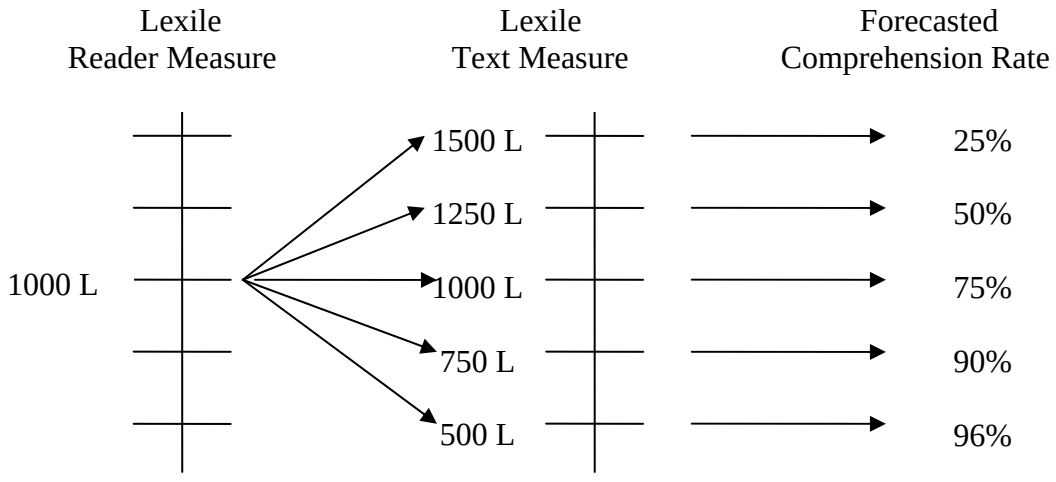


Figure 2 demonstrates the relationship between Lexile reader measure and Lexile text measure. If a reader has a Lexile of 1000 and is reading a text that measures 1000L, then one can predict that the student will comprehend 75% of the book, which is the

targeted reading rate. The targeted reading rate is where the reader reads enough to understand the text, but the text still presents some challenge. At this rate, the reader will not be bored nor will he or she be frustrated which will create a beneficial reading experience (MetaMetrics, 2008). On the other hand, if the same reader chooses to read a text of 1500L, he or she will have about 25% comprehension and be on a frustrational level. A reader with a Lexile measure of 1000L would have a Lexile range from 900L-1050L, about 100 below and 50 above (Stenner, Sanford, Burdick, & Burdick, 2006; MetaMetrics, 2008).

Because tens of thousands of books and tens of millions of newspapers and magazine articles, more than 450 publishers, all major standardized reading tests, and many popular reading programs have Lexile measures, teachers are afforded a very valuable tool in enabling them to personalize reading instruction and to better communicate reading strengths and weaknesses to parents (MetaMetrics, 2008). This finding means that since many assessments provide Lexiles with their reading assessments, such as Georgia CRCT, North Carolina's End of Course Test, Texas Assessment of Knowledge and Skill, Utah Core Assessment Series, to name a few, educators can look at the Lexiles as a measurement tool and determine the progress of a student's reading ability (MetaMetrics, 2008). If a student's Lexile does not grow over time and meet grade level expectations, then supplemental instruction may be needed.

Examples of reading material with Lexile measures can be found in many sources. Books within the Lexile range of a typical second grader are *Arthur Goes to Camp* (380L) and *Arthur, Clean Your Room!* (370L) (MetaMetrics, 2008). Harry Potter books have a Lexile measure in the range of 880L to 950L. *Little Women* has a Lexile of

1300L, and *Don Quixote* measures 1410L. The higher Lexile levels contain longer sentences and longer, more difficult words.

The Lexile measures of some newspapers are as follows: New York Times (1380L), Washington Post (1350L), Wall Street Journal (1320L), Associated Press (1310L), and USA Today (1200L) (Fulton, 2004). Examples of Lexile-based texts are *Psychology: An Introduction* published by Prentice Hall (1300L), *Word97* published by Glencoe/McGraw-Hill (800L), and *Science (Grade 4)* published by Addison-Wesley (600L). There are Lexile measures attached to personal use: Aetna Health Care Discount Form (1360L), Medical Insurance Benefit Package (1280L), Application for Student Loan (1270L), Federal Tax Form W-4 (1260L), and Installing Your Child Safety Seat (1170L) (Stenner, 2006). The Lexile of the SAT, a college entrance exam, is 1330L, and the ACT measures 1210L (Horry County Schools, 2008).

Why is a Lexile score a good measurement of reading? The *Ninth Mental Measurements Yearbook* (Mitchell, 1985) examined 97 reading assessments by providing a conceptual rationale and a scale. The result was 97 reading metrics that were nonexchangeable which caused confusion among educators, researchers, policy makers, and parents. Because of the multiple reading scales and their confusion, the Lexile Model of Reading was created to help relieve the confusion and to unite the measurement of reading by providing a common, supplemental metric shared by test publishers, book publishers, and text aggregators, such as EBSCO, Proquest, SAT, and many others (Stenner, Burdick, Sanford, & Burdick, 2006).

The Lexile Model provides a unit of reading measure, Lexiles, which provide a common scale for matching reader ability and text difficulty (MetaMetrics, 2006). The Lexile Model of Reading measures comprehension as the difference between a reader

measure and a text measure. The Lexile Model uses a 2-variable equation to predict text difficulty, word frequency, and sentence length. Word frequency and sentence length combine to produce a regression equation (Stenner et al., 2006). The difference in Lexiles between the reader and the text is what determines comprehension. If the reader measure is greater than the text measure, the comprehension rate will be greater than 75%. If the reader measure is less than the text measure, the comprehension rate will be less than 75% (Stenner, Sanford, Burdick, & Burdick, 2006).

Using the Lexile Model, a teacher can identify to what extent the student is able to comprehend a text and to identify what would be a frustrational, instructional, or an independent level of learning. For example, if a student takes a test or reads a book based on Lexiles, and he/she is reading on the Lexile level of the test or book, the student should have 75% comprehension. If the student makes 90%, he or she is reading above the Lexile level of the test. If the student makes below 75%, the student is reading below the Lexile of the test. This data can be used as a reference point to identify the students who are at-risk or advanced in reading. Statistically looking at the definition of Lexile scores, if a student comprehends below 75% (±3%), he or she would be at-risk in reading because the student would be below grade level in reading (MetaMetrics, 2006).

Progress Monitoring Tools: Curriculum Based Measurement (CBM)

While the summative evaluation by a state may provide valuable information about the student's year long progress and the effectiveness of instruction, assessments throughout the year offer a different tool that should be utilized to monitor progress. Assessments are frequent measures that are diagnostic and formative. Assessments are important considerations of any educational system because they are a critical part of effective teaching and learning (Cobb, 2003; Stecker, Lembke, & Foegen, 2008). An

effective classroom demonstrates a connection between curriculum, assessment, and instruction (Cobb, 2003).

Guskey (2003) made three recommendations to change the way that educators should approach assessment. First of all, assessments should be useful for students as well as teachers which is accomplished by informing the students of the concepts and skills needed to achieve and to give them criteria that must be reached to be successful. Secondly, assessments should be a checkpoint of learning, not an end, and they should be followed by a reflection to determine what the next step should be. Finally, as a teacher gives an assessment, feedback, and corrective instruction, then students must be able to demonstrate their new level of understanding and achievement (Guskey, 2003).

To be able to identify the students who are at-risk and finding interventions that work, effective progress monitoring of the students' data were implemented to guide instruction (Hintze, Daly, & Shapiro, 1998; Taylor, 2004). Teachers and school systems utilized progress monitoring by using assessments to see if the students were making progress in a particular subject according to the standards or benchmarks of the grade level (GADOE, 2008c). It is vital for teachers of students who are at-risk to provide progress monitoring to measure the effects of an intervention (Ardoin, 2006). Using ineffective intervention for extended time periods could create problems for the student; therefore, implementing a supplemental assessment procedure to help in deciding if an intervention is ineffective within a shorter time period could greatly benefit the student. Ardoin (2006) explored "the utility of monitoring maintenance of direct intervention effects as a supplement to using standard Curriculum-Based Measurement in reading (R-CBM) progress monitoring procedures" (p. 715). Results of Ardoin's study (2006) are important because his study suggested that using data to evaluate students' maintenance

of intervention effects helps to see if the intervention is effective and affords an objective way to use data to differentiate instruction. The procedure helped in making special education eligibility decisions. Ardoin summarized Elliot and Fuch's study (1997) of progress monitoring.

> Progress monitoring measures used to evaluate a student's academic progress need to have the following characteristics: (a) be quick to administer (b) have adequate reliability and validity, (c) be representative of what the student is learning, (d) aid in intervention development, and (e) be sensitive to gains in academic performance so that intervention effectiveness can be evaluated. (p. 716)

As the evaluations and assessments are implemented in the educational process, at-risk readers or those students who fall below the standards set forth by each state can be identified and monitored more effectively and efficiently, but using effective progress monitoring assessments to determine the needs of gifted readers is crucial as well in determining the curriculum for the gifted reader (Dooley, 1993). It should never be assumed that gifted students have mastered all skills and concepts, such as phonemic awareness, phonics, fluency, comprehension and vocabulary skills (NRP, 2000). If the skills and concepts are mastered, then less direct instruction is needed and effective instruction should be planned accordingly, such as curriculum compacting (Dooley, 1993). Curriculum compacting is a systematic process which allows for one of the greatest needs of the gifted readers—mastering basic skills at an appropriate level and pace (Feldhusen & Wyman-Robinson, 1980). Effective assessment and progress monitoring should lead to effective reading instruction for all students (Ardoin, 2006).

To establish effective reading instruction, studies have found a valid and reliable progress monitoring tool, a curriculum-based measurement in reading (R-CBM; Fuchs & Fuchs, 1997; Hintze & Shapiro, 1997; Marston; 1989). Understanding the effectiveness of progress monitoring in identifying a student's progress in reading and the potential in correlating the assessments with Lexiles, this study investigated using literature-based reading assessments based on Lexiles to monitor reading progress of the students. Because Lexiles are reported in ranges throughout the grade levels and most assessments are reported with a Lexile score (MetaMetrics, 2008), a teacher should be able to monitor the progression of the student's Lexiles through Lexile-based assessments and literature, and predict those who can achieve 75% comprehension on the summative evaluation. The Lexile Model of Reading adds more information for the teacher but not more time. It is not just another test, but a thermometer that provides a measure that "ties day-to-day work in the classroom to critical high-stakes tests" (MetaMetrics, 2008). The Lexile Model gives an overall picture of the growth of a student's reading ability from preschool to graduate school (MetaMetrics, 2008). Utilizing progress monitoring to assess a student's reading progress with valid and reliable measurements are vital in meeting each student's needs and establishing effective strategies to improve a student's reading ability whether the student is at-risk or advanced in reading abilities, and Lexiles provide a means to connect the reader and the assessments (Stenner, Horabin, Smith, & Smith, 1988; MetaMetrics, 2008).

The Construct of Reading

The Definition of Reading and Identifying At-Risk Readers

To understand the problems and effective strategies that are needed, as well as identifying a student's strengths and weaknesses in reading, there needs to be an

established definition of reading (Anderson, Hiebert, Wilkinson & Scott, 1985). . Some have suggested that reading is the ability to pronounce words. Others say that reading is learning to bring meaning to text. Research supports the meaning that reading is both. The complete definition of reading includes both, decoding and the construction of meaning (Anderson et al., 1985). The NRP (2000) defined reading as including several behaviors: reading real words in isolation or in context, reading pseudowords that can be pronounced but have no meaning, reading text aloud or silently, and comprehending text that is read silently or orally. The NRP (2000) found that a complete reader must utilize the components of phonemic awareness, phonics, fluency, and comprehension.

By establishing the definition of reading, it becomes easier to identify at-risk or struggling readers. Reading disabled or at-risk readers are students who score below the 25^{th} percentile under traditional standards (Foorman, Francis, Beeler., Winikates, & Fletcher, in press). The difference between a student who has a learning disability in reading and a child who is a struggling reader without a disability is in the severity of the problem (Grossen, 1997). Not being able to decode single words is the most consistent indicator of a reading problem (Grossen, 1997). Lyon (1994, 1995) suggests that using traditional methods, such as comparing IQ-achievement scores to identify a struggling reader is not reliable. Lyon (1994, 1995) also suggests that the most effective way to tell if the student's reading inability does not meet expectations is to compare the student's performance with other students of the same age and to compare the student's reading performance to academic performance in other domains, such as listening comprehension, verbal expression, mathematics, and written expression.

The primary ability component that children with reading difficulties differ from other children is in the area of phonological processing (Grossen, 1997). After at-risk

readers have been identified either through low reading achievement scores, below grade level reading performance as compared to other students his or her age, or through poor performance on other reading domains, the learning communities must identify the most effective strategies to improve reading achievement (Lyons, 1994, 1995). A successful strategy is to implement preventive interventions at an early age (Pikulski, 1994).

The Literacy Puzzle

Marcon (1995) explained the transitional process from primary to upper elementary and middle school readers,

> Through the primary grades, children are learning to read. An academically directed approach typically emphasizes the act of reading over comprehension. Beginning in fourth grade, children are reading to learn; comprehension is critical. In fourth grade, they encounter more abstract concepts that do not necessarily match up with their everyday experiences. Additionally, fourth-grade teachers expect children to be more independent in the learning process, to assume more responsibility for their learning, and to show greater initiative. (¶ 35)

Older students who continue to struggle to read need the same research-based strategies identified by the NRP as essential for success in reading because these areas make up pieces of a "cognitive puzzle" of literacy that help struggling readers to gain the strategies that they need to grow as an effective reader (Salinger, 2003, p. 80).

The first piece of the puzzle is phonemic awareness (Salinger, 2003). The smallest units that create the spoken language are called phonemes (NRP, 2000). Teaching phonemic awareness involves teaching students to focus on and manipulate phonemes into spoken syllables and words (NRP, 2000). Students who lack phonemic awareness have problems with learning to read (Share & Stanovich, 1995). If students have

phonemic awareness experience, they have the ability to interpret the following: what they hear represents sounds, sentences and phrases have individual words, words can be divided into individual sounds, and individual sounds can be blended into new words (Salinger, 2003). The reason that phonemic awareness is the foundation of literacy is because it shows that children can recognize and manipulate the sounds of language which is a vital connection between verbal play and translating print to sound (Salinger, 2003). In short, when children struggle in the area of phonemic awareness, they have difficulty turning spelling into sounds or phonics (Grossen, 1997).

The next important piece of the literary puzzle is phonics (Salinger, 2003). The National Reading Panel's (2000) meta-analysis found that "systematic phonics instruction makes a bigger contribution to children's growth in reading than alternative programs providing unsystematic or no phonics instruction" (p. 2-92). The analysis disclosed that systematic phonics instruction provides significant benefits for students in kindergarten through 6th grade and for struggling readers (NRP, 2000). There are different types of approaches to phonics instruction, such as synthetic phonics instruction (teaching links between individual letter or letter combinations with the correct sound and then forming words) and analytic phonics instruction (teaching the whole word unit first and then following in systematic instruction linking the specific letters in the word with their respective sounds) (NRP, 2000). Students with disabilities and students without disabilities improved their ability to read substantially when they were taught using systematic synthetic phonics instruction (NRP, 2000). Phonics instruction should help students to see and implement the relationships between letters and the sounds in spoken words, the alphabetic principle (Salinger, 2003). As students develop this dependency on

the relationships between letters and sounds, they are able to read some words “on sight” or to decode words by applying the knowledge of phonics (Salinger, 2003, p. 80).

Another important piece of the research-based literacy puzzle is fluency (Salinger, 2003). According to Bell, Ziegler, and McCallum (2004), fluency is being able to read with speed, ease, and accuracy and the correct intonation and rhythm. Fluency enables students to group words into meaningful units while making connections between the ideas that they are reading (Salinger, 2003). Fluency is critical for comprehension, but it is often neglected in classroom instruction. If a student cannot read a passage efficiently and fluently, he or she will have difficulty understanding the meaning of the passage (NRP, 2000). To improve fluency, there must be reading practice. The National Reading Panel found that guided repeated oral reading procedures that included guidance from teachers, peers, or parents had a significant and positive impact on word recognition, fluency, and comprehension in all grade levels for all students, with or without reading problems (NRP, 2000). The Panel (2000) could not find any multi-year research showing a relationship between guided oral reading and the emergence of fluency.

Vocabulary instruction is another piece of the literacy puzzle (Salinger, 2003). Teaching vocabulary strategies include building up their reading, writing, speaking, and listening vocabularies (Salinger, 2003). Vocabulary knowledge consists of different kinds of knowledge, such as the need to learn synonyms for words they already know and the meanings of new words and their ideas and concepts (Salinger, 2003). The NRP’s (2000) report found that vocabulary instruction leads to gains in comprehension as long as the methods are appropriate for the age and ability. Using computers in vocabulary instruction was more effective than some traditional methods. Vocabulary can be taught incidentally in the context of oral reading or listening to others. It is also helpful to learn

the vocabulary before reading a text. Other techniques that enhance vocabulary development were those that restructured tasks and provided repeated exposure, such as having the student use the word in various contexts (NRP, 2000). As students become interested in new words, they are motivated to learn new words and gain new purposes for reading (Salinger, 2003).

The last piece of the literacy puzzle that is the ultimate goal of all reading instruction is comprehension (Salinger, 2003). Struggling readers have comprehension problems because they cannot identify words accurately or fluently while reading. They lack background knowledge, or they do not have the strategies to tap into what they know (Salinger, 2003). While some students get comprehension strategies informally, they suggested that explicit or formal instruction in the application of comprehension strategies is highly effective in increasing understanding (NRP, 2000).

Thinking of these literacy components as pieces of a puzzle will help one to understand the struggling middle school readers (Salinger, 2003). While older students may have had pieces of this literacy puzzle, they have not been able to fit the puzzle together into a "smooth, dependable array of strategies and skills for gaining meaning from text" (Salinger, 2003, p. 81). As a result, these students are generally not successful academically which has led to a low self-esteem, lack of motivation, and an apathy toward school (Johnston &Winograd, 1985; Torgesen, 1982). If a student is still struggling in reading in middle school, the gaps become wider (Salinger, 2003).

While it becomes increasingly difficult to close the widening gaps in reading for middle school students, it is not too late to help (Salinger, 2003). Educators must carefully choose effective and efficient interventions that will help these students put the pieces together. Some may need specialized intense intervention depending on the

deficiency in the basic reading skills (Torgesen, Rashotte, Alexander, & MacPhee, 2002). While planning instruction for the struggling readers, educators need to realize that if the interventions did not work in the early years of elementary school, keeping the same interventions will not help them after fourth grade (Salinger, 2003). When teaching struggling middle school readers, a teacher must deliberately plan what they teach and find ways that most effectively help students to practice and use all of the literacy skills (Salinger, 2003).

Incorporating the Essential Components of Reading

Some students receive intervention at the microlevel of phonemic awareness and phonics skills to teach them how to decode words (Salinger, 2003). Stanovich (1986) found that students who struggle to identify words cannot begin to comprehend the text; therefore, if they learn how to improve basic word attack strategies, the reading progress will follow. If a student is a complete nonreader, then the intervention must go back to phonemic awareness (Salinger, 2003). Most struggling middle school readers have some concept of phonemic awareness, but they have not made the connection to broad reading and writing tasks (Salinger, 2003). Research suggests that students try to find meaningful clusters of letters as they decode words. They also think of clusters as they learn to spell (Treiman, 1992). This process leads to the concept of teaching older students to look to the units of sound that are bigger than phonemes but smaller than words. As students learn new letter-sound combinations and clusters of sounds, they need extensive practice applying the new knowledge to the task of reading as they are learning (Grossen, 1997). This is called using decodable text.

> A "decodable" word is a familiar one that a learner has been prepared ahead of time to sound-out (attach speech sounds to each of) its letters. Decodable texts

> thus are ones that contain only familiar words that students have previously been prepared to decode through the application of phonics rules. It is discovered empirically that beginning readers are more successful in accurately reading decodable texts than they are in reading texts that contain words students have had no prior *d*irect, *i*ntensive, *s*ystematic, *e*arly, and *c*omprehensive (DISEC) phonics instruction on how to identify. (Groff, 2001, ¶ 4).

Another valuable tool in identifying word meaning is instructing how a student can identify syllables for decoding unfamiliar words which leads to the word meaning (Salinger, 2003). One way to help struggling older readers is to teach them many high frequency rhymes or root words and to learn how to use these pronounceable word parts to help them read and write (Fry, Kress, & Fountoudidis, 1993; Gunning, 1995). To teach learners power over language, an activity would be to have students to use root words and make new words and to use them as building blocks (Salinger, 2003). This instruction is similar to the concept of teaching younger students using phonemes to make words or make up rhymes (Salinger, 2003).

Struggling readers may also have problems in overlooking important differences in words such as, *pacific/specific, flesh/flash, scalding/scolding* (Salinger, 2003). They may not hear or see the differences as they read. These miscues can result in a lack of comprehension (Salinger, 2003). A way to increase the students' vocabulary is to clarify their attention to speech and their ability to connect speech sounds to letter symbols by teaching them to pay close attention to small differences in consonant and vowel patterns (Salinger, 2003). These concepts of reading develop into the fluency component. Fluency is the component that is directly linked to a student's comprehension (Allington, 1983; Johns, 1993; NRP, 2000; Pinnell, Pikulski, Wixson, Campbell, Gough, & Beatty, 1995;

Rasinski, Padak, McKeon, Wilfong, Fiedauer, & Heim, 2005; Samuels, 1988; Schreiber, 1980).

The Importance of Fluency

Fluency is a construct of reading that can be useful, effective, and efficient as a measurement intervention tool for students and teachers because it provides a major indicator of a student's reading ability (Hudson, Lane, & Pullen, 2005). Fluency provides teachers with the knowledge that an intervention is working and demonstrates to the student that he or she can read (Hudson et al., 2005). Fluency studies have suggested a strong correlation to reading comprehension (Allington, 1983; Johns, 1993; Samuels, 1988; Schreiber, 1980). Therefore, oral-reading fluency has the ability to be a key indicator of overall reading achievement (Stecker, Lembke, & Foegen, 2008).

"Reading fluency is one of the defining characteristics of good readers, and a lack of fluency is a common characteristic of poor readers" (Hudson, Lane, & Pullen, 2005, p. 702). Fluency is a combination of the "*accurate* reading of connected text at a conversational *rate* with appropriate *prosody* or expression" Hudson, Mercer, & Lane (2000) as cited by (Hudson et al., p. 702). Struggling readers usually do not learn fluency incidentally or automatically. Many times they need direct instruction in how to read fluently and focused practice within the regular reading program (Salinger, 2003).

Fluency is a reliable predictor of problems with reading comprehension (Stanovich, 1991). Each aspect of fluency (accuracy, rate, and prosody) connects to text comprehension (Hudson et al., 2005). Without accuracy, the reader will not be able to access the author's intended meaning. With inaccurate reading of words, many times the text is misinterpreted. "Poor automaticity in word reading, or slow, laborious movement through the text taxes the reader's capacity to construct an ongoing interpretation of the

text" (Hudson, et al., p. 703). As a struggling reader uses poor prosody or expression, he or she can become confused "through inappropriate or meaningless grouping of words or through inappropriate applications of expressions" (Hudson et al., p. 703). Identifying a dysfluent reader and planning appropriate instruction is an important factor in developing his or her comprehension as well (Hudson et al., 2005). On the other hand, identifying a fluent reader at a higher level provides a clear indication of the level at which a student can be instructed in reading. Pinnell, Pikulski, Wixson, Campbell, Gough, and Beatty (1995) conducted a study that found a close relationship between fluency and reading comprehension.

Harris and Hodges (1995) defined fluency in *The Literacy Dictionary* as being free from word identification problems, but fluency definitions in research began to go beyond word recognition problems to include comprehension processes also (Thurlow & van den Broek, 1997). "Fluency helps enable reading comprehension by freeing cognitive resources for interpretation, but it is also implicated in the process of comprehension as it necessarily includes preliminary interpretive steps" (NRP, 2000, p. 3-6).

The hindrance of cognitive resources is why reading accuracy, speed, and expression interfere with comprehension (NRP, 2000). During the reading process, there are two basic cognitive tasks, decoding and comprehension (NRP, 2000). Both of these tasks require cognitive resources, and at any given moment, memory can be limited by the amount of cognitive resources that the student may have (NRP, 2000). If the student's cognitive resources are used for decoding, then he or she may have nothing left for comprehension which will slow down a nonfluent reader's reading process. For example, Rasinski, Padak, McKeon, Wilfong, Fiedauer, and Heim (2005) found that fluency is a factor that correlates to comprehension even among high school students, especially the

struggling readers. Rasinski and others (2005) found in their study that even though at-risk high school students read with a high degree of accuracy, they used so much of their cognitive energy to read that they drained cognitive energy from their ability to comprehend.

The Fluent Reader

Reading for the fluent reader is much easier than the nonfluent reader because the fluent reader recognizes the printed word with ease and speed utilizing few cognitive processes (NRP, 2000). A fluent reader is able to do multiple tasks at the same time, such as word recognition, comprehension, and drawing references (NRP, 2000). While a skilled reader may not automatically recognize all words, such as uncommon, low-frequency words, they usually have different options or strategies that they have learned for word recognition (NRP, 2000).

Research on eye movement during reading provides a description of a fluent reader's characteristics. Rayner (1998) summarized the differences in eye movements of good and poor readers:

> There are well-known individual differences in eye movement measures as a function of reading skill: Fast readers make shorter fixations, longer saccades [the jump of the eye from one fixation to another], and fewer regressions than slow readers (Everatt, Bradshaw, & Hibbard, 1998; Everatt & Underwood, 1994; Rayner, 1978; Underwood, Hubbard, & Wilkinson, 1990)…In characterizing the eye movement patters of dyslexic readers, Olson, Kliegl, Davidson, & Foltz (1985) categorized such readers as plodders, and explorers' plodders made relatively short forward saccades, and more regressions, whereas explorers

showed more frequent word skipping, longer forward saccades, and more regressions. (p. 392)

Using Fluency to Assess Reading Progress

Fluency is a key to valuable progress monitoring tools (Wayman, Wallace, Wiley, Tichá, & Espin, 2007). Progress monitoring of the strengths and weaknesses in reading can be done through two informal reading assessments, an oral reading fluency (ORF) assessment that measures rate and accuracy and a silent reading assessment called Maze fluency (Wayman et al., 2007). Both assessments are known as Curriculum-Based Measurement (CBM).

Curriculum-based measurement of oral reading fluency is a way of monitoring a student's progress in reading achievement and even of predicting potential success on the summative evaluation (Christ, 2006; Christ & Silberglitt, 2007; Deno, 1985; Graney & Shinn, 2005; Hasbrouck & Tindal, 2006; Silberglitt, 2006; Silberglitt & Hintze, 2007; Speece, 2005; Stecker et al., 2008; Wiley & Deno, 2005). Deno (1985) found that the effectiveness of CBM was in the ability to be administered frequently and its ability to identify student growth. CBM is inexpensive and easily scored (Wiley & Deno, 2005). CBM in reading is usually made of 1-minute reading probes that are created from the curriculum of the school and are given regularly over the school year (Hintze, Owen, Shapiro, & Daly, 2000; Wiley & Deno, 2005). This measurement of reading has been found to be highly reliable for showing reading growth and making decisions regarding placement (Hintze et al., 2000). Researchers have found that the standardized procedures of CBM are valid indicators of a student's academic performance (Espin & Deno, 1994-1995; Fuchs & Deno, 1994). Fuchs and Deno (1994) refer to this approach as general outcome measurement (GOM). GOM oral reading measures fluency by having a student

read aloud a passage for one minute and counting the number of words read correctly which measures rate and accuracy (Fuchs & Deno, 1994). Baker and Good (1995) found that using CBM of oral reading fluency was a reliable and valid assessment for bilingual and English-only students as well. CBM allows educators and psychologists to measure and compare individual student growth to prior levels of performance and peer growth rates, to measure the effectiveness of instruction, and to measure the student's response to intervention (RTI) (Fuchs, L. S., 1986; Fuchs, L.S., & Fuchs, D., 1998). Researchers have found that there is a direct relationship between oral reading fluency and the potential for passing the state summative assessment (Crawford, Tindal, & Stieber, 2001; Good, Simmons, & Kameenui, 2001; McGlinchey & Hixon, 2004; Sibley, Biwer, & Hesch, 2001; Stage & Jacobsen, 2001). Therefore, CBM of oral reading fluency has predictive power (Wiley & Deno, 2005).

Directions for Curriculum Based Measurement (CBM)

Directions for Oral Reading Fluency

Hosp and Hosp (2003) provided directions regarding how to conduct CBM for reading. The first step was to gather the materials: different but equivalent reading passages or probes, a timer, a writing utensil, a graph of equal intervals to plot the data, and directions for administering and scoring the passages. The three passages should be from the same grade level made up of about 200 words preferably passages that the student has not read (Wesson, 1992). Two copies of each passage will be needed, one for the teacher and student (Hosp & Hosp, 2003). At the end of each line on the teacher passage, a running total of the word count should be written to help the teacher in scoring, and the student copy should be laminated for multiple uses. These three passages will determine the student's instructional level. The teacher needs to administer the three

probes one after the other or over three consecutive days. The median score is taken and that score was compared to Table 2 as shown by Hosp and Hosp (2003). If the data did not match the information in Table 2, then a below grade level or above grade level passage is given from the original passages. The process is repeated until the student's instructional level meets the criteria in Table 2 noting the number of errors which can denote a frustrational level (Fuchs & Deno, 1982). Once the student's instructional level is determined, then 30 equivalent passages should be used to progress monitor the student's growth throughout the year at the student's goal level which is based on what skills and level of curriculum the student should reach in one year, generally one year above instructional level (Shinn, Gleason, & Tindal, 1989). Hintze, Daly, and Shapiro (1998) recommended for grades 1 and 2 using 30 passages from the student's instructional level, and for grades 3 and above using goal level material or passages one year above instructional level because the passages could be more sensitive to growth and allow for frequent change in the early grades but provide other diagnostic information.

Table 2

Determining Placement Level Based on ORF

Grade Level	Placement Level	Correct Words per Minute (CWPM)	Errors per Minute
1-2	Frustration	<40	>4
	Instructional	40-60	4 or less
	Mastery	>60	4 or less
3-6	Frustration	<70	>6
	Instructional	70-100	6 or less
	Mastery	>100	6 or less

Note. Adapted from *Developing goals and objectives for educational programs by* Fuchs, L. S., & Deno, S. L. (1982). Washington, DC: American Association of Colleges for Teacher Education.

The scoring and directions for instructional level or goal level passages are the same (Hosp & Hosp, 2003). Errors are words that are mispronounced, omitted, substituted, or reversed. If the student hesitates on a word more than 3 seconds, he/she is given the word, and it is marked as an error. Repetitions and insertions of words are ignored, and self corrections within 3 seconds are counted correct (Hosp & Hosp, 2003).

Table 3

Expected Growth Rates in ORF

Grade	Realistic Growth Rates		Ambitious Growth Rates
1	2		3
2	1.5		2
3	1		1.5
4	.85		1.1
5	.5		.8
6	.3		.65
Grade		Growth Rates	
1		1.80	
2		1.66	
3		1.18	
4		1.01	
5		.58	
6		.66	

Note. Adapted from "Formulative evaluation of academic progress: How much growth can we expect? By Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993), *School Psychology Review, 22*, 27-48, and "Using curriculum-based measurement to establish growth standards for students with learning disabilities" by Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). *School Psychology Review, 30*, 507-524.

In Table 3, Hosp and Hosp (2003) adapted and compiled data from two studies. Fuchs, Fuchs, Hamlett, Walz, and Germann (1993) show realistic and ambitious growth rates for typically developing students, and Deno and others (2001) identify growth rates. The passages used to monitor growth should correspond for the expected growth for that

grade (Hosp & Hosp, 2003). For example, a fifth grade student reading at seventh grade instructional level would have expected growth corresponding to eighth grade. If the student's median score on eighth grade material is 95 words read correctly, and the student is monitored for 34 weeks, the expected ambitious growth would be .8 x 34 weeks = 27.2 which is added to 95 words read correctly or 27.2 + 95 = 122.2. Thus, 122.2 words read correctly is the determined goal (Hosp & Hosp, 2003). These scores are plotted on the interval graph in which the vertical axis is the number of words read correctly, and the horizontal axis is the number of weeks monitored, allowing for data to be plotted one to two times a week. A goal line or regression line is drawn on the graph connecting the beginning instructional level to the expected ambitious growth allowing for a reference point that demonstrates the effectiveness of instruction. The data points are plotted each week. If at any time there are four consecutive scores below the goal line, instruction probably needs to be changed. If the student is consistent in scoring above the line, then the goal is raised (Fuchs, Fuchs, & Hamlett, 1989). This analysis of data becomes effective progress monitoring helping to meet all students' needs.

The methodology used in various research all found CBM of oral reading fluency to be valid measurements of reading, and all followed the same basic procedures except for varying the time between passage assessments (Christ & Silberglitt, 2007; Graney & Shinn, 2005; Silberglitt, 2006; Silberglitt & Hintze; 2007; Stecker, et al., 2008; Wiley & Deno, 2005). The results of the studies identify several benefits of CBM of oral reading fluency: the ability to define academic strengths and weaknesses (the frustrational, instructional, or mastery level), to evaluate response to intervention, to use for progress monitoring, to assess instructional effectiveness, to help teachers in instructional planning, and to predict performance on the state's summative evaluation (Christ &

Silberglitt, 2007; Graney & Shinn, 2005; Silberglitt, 2006; Silberglitt & Hintze; 2007; Stecker, et al., 2008; Wiley & Deno, 2005). Research supports the theoretical basis regarding the strength of reading aloud and the dependability of the measures especially with second to fifth grade. As students get older, such as middle school age, the correlations between reading aloud and criterion variables may decrease (Wayman, Wallace, Wiley, Tichá, & Espin, 2007). Research suggests that more studies are needed for older, middle, and high school students, and that the Maze selection seems to be more appropriate for older students (Wayman et al., 2007).

Directions for Maze Assessment

The Maze assessment has an advantage over the reading aloud assessment because it does not need to be given individually (Wayman et al., 2007). The Maze for older students appears to be more of a reading comprehension measure (Wayman et al., 2007). The Maze selection is similar to oral reading fluency but utilizes the measurement of silent reading in goal-level material and is representative of overall reading achievement (Fuchs & Fuchs, 1992; Fuchs, Fuchs, Hamlett, & Ferguson, 1992; Shin, Deno, & Espin, 2000). In a Maze probe, a blank containing three choices is put in place of every *nth* word. The student is given two and one half to three minutes to read and circle the correct choice for the blank. The teacher counts the number of correct choices and stops when the student makes three incorrect choices in a row (Stecker et al., 2008). The Maze task is useful with upper elementary, middle, and high school students who are fluent with oral reading (Espin, Wallace, Lembke, & Campbell, 2004; Jenkins & Jewell, 1993). Fuchs and Fuchs (1992) studied the Maze selection giving the Maze twice weekly for eighteen weeks via the computer analyzing technical adequacy and teacher acceptance, and they found that the Maze task was sensitive to performance change over

time, and the difference in it and other measures was its small ratio of slope to standard error of estimate. Teachers reported that they liked the Maze more because they believed that it demonstrated multiple dimensions of reading, such as decoding, comprehension, and fluency, and students reported enjoying the Maze more (Fuchs & Fuchs, 1992).

Using CBM as a Progress Monitoring Tool

At first as CBM was extended to the secondary-school level, the focus was using the reading measures to predict content-area performance (Espin & Deno, 1993a, 1993b; Espin & Deno, 1995; Fewster & MacMillan, 2002; Yovanoff, Duesbery, Alonzo, & Tindal, 2005). Recently, research at the middle school level has turned to using CBM to predict general reading performance (Espin & Foegen, 1996; Espin, Wallace, Lembke, Campbell & Long,, 2007; Muyskens & Marston, 2006; Tichá, Espin, & Wayman, 2007). These researchers found that oral reading fluency and Maze demonstrated strong alternate-form reliability and moderate to strong criterion-related and predictive validity. Espin et al. (2007) and Tichá et al. (2007) found Maze selection reflected change in performance over time, but reading aloud did not; therefore, growth on Maze related to the student's performance on a state reading test and to changes on standardized reading achievement tests.

Through studies of CBM on students from diverse backgrounds, the researchers found mixed results (Wayman et al., 2007). Hixson and McGlinchey (2004) and Kranzler, Miller, and Jordan (1999) found CBM reading aloud overestimated reading performance of African American students and underestimated performance for Caucasian students. On the other hand, Hintze, Callahan, Matthews, Williams, and Tobin (2002) found the results of CBM reading aloud to be neither over nor underestimation of performance for either African American or Caucasian. For English learners (EL), the

results generally found moderate to strong reliability and criterion related validity coefficients for reading-aloud measures for EL students (Baker & Good, 1995; Wiley & Deno, 2005). Graves, Plasencia-Peinado, Deno, and Johnson (2005) found that gains on reading-aloud measures for EL students were similar to gains for non-EL students. Klein and Jimerson (2005) found that CBM reading aloud measures for the reading proficiency of Hispanic students whose home language was Spanish was systematically overpredicted which could potentially lead to systematic underidentification for services, and the reading proficiency for Caucasian students whose home language was English was systematically underpredicted leading to possibly systematic overidentification.

In recent studies, CBM reading measures have been used to examine their relationship to the performance on a state-mandated standardized test (Wayman et al., 2007). Crawford, Tindal,and Stieber (2001) and Good, Simmons, and Kameenui (2001) focused their research on establishing benchmark scores to predict passing or failing a state reading test. Other studies were correlational that examined the correlations between CBM reading-aloud measures and performance on state standards tests which reported diagnostic efficiency statistics, such as sensitivity, specificity, positive and negative predictive power (Hintze & Silberglitt, 2005; McGlinchey & Hixson, 2004; Stage & Jacobsen, 2001). Sensitivity is the percentage of students below a cut score who fail a test. The percentage of students who score above a cut score to pass a test is specificity. Positive predictive power is the probability of a student with a score below the cut score that truly fails, and negative predictive power is the probability of a student who scores above a cut score that really passes the test (Wayman et al., 2007).

The correlations between the CBM reading aloud measures and the states standard reading tests ranged from .60 to .80 except for the study by Stage and Jacobsen

(2001) which found the range to be .43 to .44 between the CBM reading aloud measure and the Washington state reading test. This difference has been attributed to the factor that Washington's reading test required short-answer and extended writing; therefore, students' scores reflect reading and writing measures. The other state tests did not use written responses. The diagnostic efficiency across the four studies was generally consistent. Sensitivity values were from 65% to 76%. Specificity values were from 74% to 82%. For all but Stage and Jacobsen (2001), the positive predictive power of the studies were 55% to 77% (Stage and Jacobsen was 41%), and the negative predictive power values were 83% to 90%, except for the McGlinchey and Hixon (2004) study where the value was 46%. Across the studies, CBM increased significantly to positive and negative predictive power above the base rates of prediction (Wayman et al., 2007).

Through Wayman's et al. (2007) synthesis of CBM in reading, a strong relationship existed between CBM reading aloud and reading proficiency. Reading aloud for elementary-school students was found to be a better indicator of reading comprehension than other comprehension measures. The studies also demonstrated that reading aloud was not just a speed-of-processing measure. On the other hand, reading aloud has not always indicated to the strongest relationship for very young and older students (Wayman et al., 2007).

For early readers, examining with an alternate measure, word ID, seemed to be a promising alternative. The correlations between reading-aloud and criterion measures stayed moderate to strong across the elementary school grades but decreased at the intermediate grades. There was no decrease in correlations for the Maze which remained fairly stable across all grades. While reading aloud seemed to be the strongest CBM measure for elementary students, reading aloud and Maze selection were both appropriate

for intermediate-grade students (Wayman et al., 2007). For secondary students, the Maze seems the most reliable choice (Wayman et al., 2007). Research was limited regarding middle school students. The available research suggests that reading aloud does not reflect growth for middle school students, but the Maze does (Wayman et al., 2007).

In general, the Maze seemed to be the most effective and efficient tool to monitor progress across the grades, at least from Grades 2 through 8, and the growth rates across grades seemed to be more consistent that for reading aloud (Wayman et al., 2007). Fuchs and Fuchs (1992) may have identified a key concept in progress monitoring of reading, maybe Maze reflects multiple aspects of reading proficiency to a greater extent than reading aloud. More studies need to examine the relationship between reading aloud and reading comprehension at the individual level, and there is also a need to repeat the research studying the relationship of the Maze and comprehension focusing on the results in regards to individual progress monitoring.

Markell and Deno (1997) suggested that large gains in reading aloud data might be necessary before one can assume gains in reading comprehension. Another issue to study is if word ID is used for beginning readers, reading aloud for elementary-grade students, and Maze for intermediate and secondary-school readers, how can the measures be connected to create a picture of growth across school years? If a way is found to link different measures across years, there is the potential that a seamless and flexible system of progress monitoring would be found (Wayman et al., 2007).

Effects of Progress Monitoring Materials

Effects of Curriculum on Progress Monitoring

Through the synthesis of literature from Wayman and others (2007), the issue of where the probes should be based was analyzed. Questions were in two categories. The

first question was of curriculum source and if technical adequacy of the measures differed according to the curriculum used to create the measures. This question included studies that compared reading passages created from different curricula and studies that created the generated passages from an instructional versus a “generic” or noninstructional curriculum (Wayman et al., 2007). Research was almost exclusive to reading-aloud measures, and few have studied word ID or Maze measures.

As Wayman and others (2007) analyzed the research involving these questions, three general themes emerged. First, the level of the source differs greatly with curriculum source. Secondly, even though the technical adequacy does not differ with curriculum source, the rates of growth may. Thirdly, it is not necessary to equate instructional and progress-monitoring material.

In regards to the first theme, differences in levels of performance, the results consistently show mean level differences in scores on probes from different curricula starting with the study of Tindal, Marston, Deno, and Germann (1982). Wayman and other summarized the findings,

> Studies have shown higher levels of performance on instructional materials than on mainstream materials (Tindal, Flick, & Cole, 1992), on literature-based materials than on authentic materials (Hintze, Shapiro, Conte, & Basile, 1997), on basal materials than on literature-based materials (Bradley-Klug, Shapiro, Lutz, & DuPaul, 1998), and on generic materials than on basal materials (Powell-Smith & Bradley-Klug, 2001). (p. 110)

Regarding Maze measures, higher levels of performance have been found on materials taken from materials controlled for difficulty than on literature-based materials (Brown-Chidsey, Johnson, & Fernstrom, 2005). Mean level differences were found to be

important to use in progress monitoring if they were used to compare students across classes, schools, districts, or a student's performance from one point in time to another. When used for this purpose, the source of material should be constant. Performance-level differences do not take into account the technical adequacy of the measures as indicators of reading performance or growth because measures can produce differences in levels of performances but demonstrate a good indicator of general reading proficiency (Wayman et al., 2007).

Secondly, the theme that emerged was from technical adequacy. Several studies found that few differences exist in the technical adequacy of reading-aloud measures selected from different curricula (Wayman et al., 2007). Fuchs and Deno (1992) compared two passages from two published basal curriculum series and compared them to the *Woodcock Reading Mastery Test* (WRMT; Woodcock, 1973) in first grade and sixth grade, and they found no differences in the magnitude of the correlations. Hintze, Shapiro, Conte, and Basile (1997) found no differences in alternate-form reliability or criterion-related validity for passages selected from authentic and literature-based curricula and no differences in the probes used for classifying students into subgroups. Hartman and Fuller (1997) found similar test-retest reliability and criterion-related validity coefficients for passages from literature-based materials and passages from basal-series in first through third grades. Brown-Chidsey and others (2005) found a high correlation between student performances on Maze probes that were from controlled and literature-based material.

Research was found on the developmental growth rates, such as changes in scores across grade levels. The results of these studies revealed mixed results. Some studies found that growth rates measured by CBM reading measures suggested few differences

relating to curriculum source (Fuchs & Deno, 1992; Hintze et al., 1997). Other studies found significant differences in growth rates (Hintze. Shapiro, & Lutz, 1994). They examined growth rates measure by literature-based and basal-series probes for two groups of third graders.

On the other hand, other researchers used literature-based and basal-series probes and found that even though differences in growth rates were evident relating to curriculum source, the results were inconsistent across grades (Hintze & Shapiro, 1997). Some researchers found higher growth rates in literature-based probes and others found higher growth rates for basal-series probes, and Wayman and others (2007) through the analysis of all the studies hypothesized that factors other than the curriculum source contributed to the differences in the growth rates. For example, it is difficult to identify equivalency of “parallel probes” used to monitor progress even if the probes are taken from the same curriculum and have the same readability level. Also, one can easily bunch difficult or easy probes near the beginning or end of the progress-monitoring session creating an effect on the slope values (Wayman et al., 2007).

One can deal with the potential effects of unequal passages by counterbalancing the order of passages across students so that the students do not read the probes in the same order. Another way to attend to the issue of using readability is to establish the equivalence of the passages (Wayman et al., 2007). In summary, even though growth rates are not affected by the curriculum source, educators need to exercise consistency in progress monitoring regarding curriculum source (Wayman et al., 2007).

The third theme to emerge was the need to match instructional and progress monitoring material (Wayman et al., 2007) l. Most of the studies revealed that students who were taught using literature-based series did not grow differently when monitored

with either literature-based probes or basal-series probes or instructional material or generic probes (Hintze et al., 1994; Hintze & Shapiro, 1997; Powell-Smith & Bradley-King, 2001; Riley-Heller, Kelly-Vance, & Shriver, 2005; Tindal et al., 1992).

The Effects of the Difficulty Level of Progress Monitoring Probes

When choosing probes for progress monitoring, two questions regarding the difficulty level of material should be considered (Wayman et al., 2007). Should the students be monitored using instructional-level material, or should they be measured with material outside their instructional level? Fuchs and Deno (1992) compared the criterion-related validity and developmental growth rates from progress monitoring materials of various difficulties which revealed no differences related to material difficulty in the magnitude of correlations. They also found that developmental growth rates decreased as the difficulty of the passages increased. Other studies found similar results when examining the influence of difficulty of the material on intraindividual growth rates (Shinn, Gleason, & Tindal, 1989). Their study monitored the reading-aloud passages for mildly handicapped students in third to eighth grades randomly assigned to either a group reading material one grade level below and above instructional placement, or two and four grade levels above instructional placement. The results provided comparable slopes. The students in the first group with reading material one grade level below increased 4.3 words per week, and they increased 3.7 words per week when monitored one above. The slopes for the second group differed from 4.55 words per week to 2.35 words per week. Their study was similar to Fuchs and Deno (1992) because both supported an increase in the difficulty level of tests leads to a flatter slope.

Another issue related to the difficulty of the CBM passages is establishing equivalent passages or “parallel” passages used for repeated progress monitoring

(Wayman et al., 2007). An example was several studies found mean score differences on the reading-aloud measure for passages of different difficulty levels (Dunn & Eckert, 2002; Fuchs & Deno, 1992; Hintze, et al., 1998; Shinn et al., 1989). This finding was a positive finding because it demonstrated the validity and sensitivity of the CBM reading-aloud measures, but it was a problematic finding regarding ongoing progress monitoring (Wayman et al., 2007). This finding implied that scores on repeated progress-monitoring passages were affected by variations in passage difficulty; therefore, it is important to establish equivalence of passage difficulty. Poncy, Skinner, and Axtell (2005) studied the effects of the variability of passages on reading-aloud scores for third graders. The participants were given passages with readabilities ranging from 2.8 to 3.1 in random order over a four day period. Through analysis of the data, they found that about 82% of the variance in the students' scores was credited to student skill, about 10% was credited to passage difficulty, and about 9% was without reason. As the difficulty of the passage on the basis of students' average scores was controlled, variance due to student skill increased and variance to passage difficulty decreased. In another study, Hintze and Christ (2004) analyzed a comparison of randomly selected materials from graded readers for students with grade-level material controlled for difficulty in second to fifth grade. The students were given both passages over the course of eleven weeks. The results suggested that the standard error of measurement (SEM) and the standard error of the slope (SEb) were smaller when passages were controlled for difficulty than when they were not.

Both studies (Hintze & Christ, 2004; Poncy et al., 2005) demonstrated the importance of establishing passage equivalence. The problem is this is not an easy task. Usually the common method for establishing parallel passages has been to study the

readability levels of the passages by using readability formulas (Wayman et al., 2007); however, Ardoin, Suldo, Witt, Aldrich, and McDonald (2005) only found a modest relationship between the reading levels given to passages using readability formulas and the number of words read correctly in one minute from third grade passages. Ardoin and others (2005) found that two components were significantly related to words read correctly per minute, syllables per 100 words and words on the Dale-Chall List of 3,000 words. The results suggested inconsistencies between the levels given to passages using various readability formulas. Compton, Appleton, and Hosp (2004) found certain components of a passage were related to words read correctly per minute. These components included the number of high-frequency and decodable words, the number of multisyllabic words (negatively related), and sentence length.

In summary, even if CBM probes are developed from different curriculum sources, the measures work consistently (Wayman et al., 2007). The probes do not necessarily need to be developed from the student's instructional material, such as the grade level basal reader. Students can be monitored with material that is easier or more difficult than their instructional level without compromising the technical adequacy of the measure (Wayman et al., 2007). On the other hand, if the material is too hard, such as two to three levels above the student's instructional level, rates of growth may be affected. A greater concern is regarding the issue difficulty relating to intraindividual growth monitoring (Wayman et al., 2007). If CBM progress monitoring is used as a part of a decision-making process, there is a need to establish passage equivalence (Wayman et al., 2007). If CBM progress monitoring is used to monitor student progress and evaluate instructional programs, then establishing passage equivalence is not as important (Ardoin et al., 2005). Ardoin and others (2005) and Compton and others (2004) suggested to field

test the selected passages with a large number of students and only use the passages that fall within one standard deviation of the mean of words correct per minute for CBM progress monitoring. The effects of this approach on the stability of growth rates created by reading-aloud passages have not been examined (Wayman et al., 2007). Stecker and others (2005) studied the positive treatment validity results by using passages selected from controlled sources, such as basal-series passages, and the study suggested that it may be sufficient for classroom use. It is necessary to establish equivalence of passage difficulty for progress monitoring if it is to be used as part of an eligibility decision-making process, or if CBM is used as part of a school- or district-wide decision-making process (Ardoin et al., 2005).

Growth Rates and Progress Monitoring

Utilizing effective and efficient assessments to monitor student progress and guide instruction is vital to the educational process (Wayman et al., 2005). Ineffectively using assessments to guide instruction creates a bigger problem (Ardoin, 2006). If a student is not identified early as a dysfluent reader, the gaps become larger as they get older (Salinger, 2003). If an advanced reader is not identified and instructed at an appropriate level providing rigor to his/her studies, then society is losing a chance to develop a tremendous natural resource, possibly a potential leader or professional talent (Davis & Rimm, 2004). Therefore, educators must cultivate the gifted and talented student's abilities as well as the at-risk student's abilities through matching instruction with the students' needs using valid and reliable assessments (Davis & Rimm, 2004). Educators must not leave any of teaching and learning to chance. Curriculum-Based Measurement of reading provides educators with a tool that has the potential to eliminate chance and be used in "the development of a seamless and flexible system of progress

monitoring that could be used across students of various ages and performance levels…from kindergarten to Grade 12" (Wayman et al., 2007, p. 116).

To make a difference, teachers and administrators must use frequent assessments that are meaningful and are able to guide instruction (Cobb, 2003).

> The decisions made in the classroom each day are ones that will have a direct influence on student learning. Many of these decisions are made every few minutes in the classroom. These decisions provide opportunities for students to decide if academic success is feasible and then decide if they are willing to do the work required (by the teacher) to achieve academic success. Demands of the teacher must take student learning needs into consideration, and teachers must correctly diagnose the educational needs of the students. Lesson time allocations and instructional implementation decisions will be made that will have a direct impact on student achievement. (Brown, 2007, p. 9)

Ongoing collaboration must take place regarding student work, instructional methods, and specific uses of curriculum (Cobb, 2003). Assessments for progress monitoring must be used, such as the CBM of oral reading fluency which provides not only validity and reliability, but also an efficient, effective, and inexpensive way for a teacher to assess a student. The results will be evident through the improvements on the statewide accountability testing, demonstrating, according to the state, school-effectiveness (Silberglitt, 2006). While each school wants to be deemed effective, "one must never forget that each piece of data reflects a student. Each piece of data that demonstrates growth is a representation of one more child who has learned" which should be the aim of every educator (Jones, 2008).

What the Research Means to this Study

Knowing the valuable information that can be gathered from curriculum-based measurement (CBM), there are areas that need further investigation which will be important to this study. Oral reading fluency measurement for making essential decisions has been validated by many studies, but up-to-date oral reading fluency norms can provide a progress monitoring tool to help in determining effective instructional programs (Hasbrouck & Tindal, 2006). Even though current studies suggest the predictive power of CBM and state summative scores in the early grades, more research needs to be conducted within the later grades to see if there is predictive validity (Silberglitt, 2006). According to Rasinski and others (2005), reading fluency is significant in a secondary student's reading development as well, but future studies need to focus on the hypothesis that improvements "in fluency could account for significant and substantial gains in students' reading comprehension" (p. 25). Most of the research focuses on struggling readers (Wayman et al., 2007). Future research should consider differential growth rate across more generalizable samples (Silberglitt & Hintze, 2007). Little is known in regards to the teacher's knowledge and understanding of CBM progress monitoring and data analysis within the Response to Intervention model (Wayman et al., 2007). Utilizing full implementation throughout all middle school language arts classrooms will provide answers to vital questions regarding helping each student to be successful in reading, such as how will the implementation of the RTI model work within the regular educational setting and how will it work for the general population. As teachers learn how to progress monitor and analyze data, they hopefully will see how the data can be a vital tool in guiding day-to-day instruction for all student populations. The plethora of studies shows the value of progress monitoring, and maybe

this study will provide a demonstration of how to effectively and efficiently monitor reading growth from kindergarten to twelfth grade (Wayman et al., 2007).

CHAPTER 3: METHODOLOGY AND RESEARCH PROCEDURES

The purpose of this study was to find a valid and reliable assessment for progress monitoring having predictive power of a student's future reading performance on a standardized reading achievement evaluation. This chapter describes the research context, the study design, and the procedures used during the implementation of the study.

Research Design, Context, and Access

This quantitative study took place in a northeast Georgia middle school where the researcher taught. The school is in Hall County with a population of 875 students for the 2007-08 school year. According to the Adequate Yearly Progress Report (AYP) for 2007-08, the school's demographics were as follows: 83.8% While, 1.9% Multiracial, 1% Black, 12% Hispanic, 11.7% students with disabilities, 4.1% English Language Learners (ELL), and 28% economically disadvantaged (GADOE, 2008a). Out of the 875 students at this middle school, 791 (the 2007-08 population) took the Reading/Language Arts CRCT last school year. There were 52 students, 6.6%, who tested at a basic level which did not meet standards; 481 students, 60.8% tested at a proficient level or met standards; and 258 students, 32.6% tested at an Advanced Level (GADOE, 2008a). The teachers and the researcher had access to the data because the data were frequently analyzed for potential academic strengths and weaknesses of the students, as well as identifying students who were advanced or at-risk. The data was recorded in a computerized grade book school-wide called Infinite Campus.

Population

The population was the students in the sixth, seventh, and eighth grade language

arts classes who attended the entire 2008-09 school year and took the Georgia Criterion Referenced Competency Test (CRCT) in 2008 and 2009. The 8th grade students who took the 9th Grade Literature and Composition course also took the End of Course Test (EOCT). The language arts teachers were taught by the researcher about fluency in reading and how to monitor reading growth through two fluency assessments, the oral reading fluency assessment (ORF) and the Maze fluency assessment. Each teacher received a teacher fluency pack with standardized directions and procedures and different tables to help them to analyze the data (see Appendix C). To establish interrater reliability, the researcher and the teachers practiced listening scoring students as they read, marking errors, and counting the words correct per minute. To establish procedural integrity throughout the study, the researcher visited the teachers to ensure that the correct methodology was followed. Even though all students were monitored using the RTI model, only the data of the students who were monitored with correct methodology (the correct monitoring goal level) and had complete data were used in the data analysis.

Instrumentation

Four different measurement tools were used for this study. The two progress monitoring tools were the oral reading fluency and Maze. Two of the measurement tools were the state-mandated high-stakes tests in Georgia, CRCT and EOCT (GADOE, 2008c).

The Validity of ORF and Maze Assessments

Validity is when an instrument that is used to measure something actually measures what it claims to measure (Ary et al., 2006). Establishing the validity of oral reading fluency and Maze assessments is a continuous process (Messick, 1989). Validity is determined by evidence that is gathered over time not by one study or one correlation

(Wayman et al., 2007). Using a Curriculum-Based Measurement (CBM) approach, the validity of the oral reading fluency and Maze fluency is determined by the examination of the extent to which the measure serves as a vital indicator of an academic domain (Deno, 1985). The early studies of Marston (1989) set a high standard for reliability and validity of CBM of reading where many correlations reached from .70 to .90. Through the synthesis of literature from Wayman and others (2007) on CBM utilizing ORF and Maze assessments, strong relations were established at .70 and above; moderate relations were from .50 to .70; and weak relations were those that were below .50.

The Reliability of ORF

The reliability of an instrument is established when a measurement is consistent in its measurement (Ary et al., 2006). Research from Marston's review (1989) supported the use of word identification and oral reading as indicators of general reading proficiency. In his review, five studies yielded test-retest reliability coefficients ranging from .82 to .97 with the majority of the coefficients above, and the interrater reliability was .99. Marston (1989) reviewed 14 studies, and the criterion-related validity coefficients with published measures of reading were from .63 to .90. When he analyzed criterion-related validity coefficients with basal reading series criterion mastery tests, the range was from .57 to .86, and over one half of the studies were above .80.

Some studies have begun to focus on the use of oral reading fluency measures to predict a student's performance on a statewide reading test; these studies usually include third or fourth grade. Crawford, Tindal, and Stieber (2001) studied using oral reading fluency with second and third graders to predict the students' performances on statewide achievement reading tests. They found a correlation of .66 for second-grade oral reading fluency and .60 for third-grade oral reading fluency in predicting reading test scores.

Fuchs and others (2001) studied oral reading fluency and the Reading Comprehension portion of the Iowa Tests of Basic Skills and found a correlation of .80. Good and others (2001) conducted a study of third graders and found a correlation of .67 between the students' spring oral reading fluency and their spring third-grade performance on the Oregon Statewide Reading Assessment. McGlinchey and Hixson (2004) studied several cohorts of fourth graders and found correlations .49 to .81 in reading between oral reading fluency and the students' performances on the Michigan Educational Assessment Test in reading. Stage and Jacobsen (2001) found a correlation of .44 between oral reading fluency in May and the reading portion of the Washington Assessment of Student Learning in May. Hintze and Pelle Petitte (2001) measured the reliability of ORF in grades 3 and 4 and found that the individual differences accounted for 62% of the variance and the reading group accounted for 15% of the variance. Morgan and Bradley-Johnson (1995) used alternate forms of oral reading fluency in grades 3-7 and found the reliability to be .88 to 93.

The Reliability of the Maze Assessment

Several studies suggested that the Maze assessment was a reliable progress monitoring tool. Brown-Chidsey, Davis, and Maya (2003) implemented the Maze assessment in grades 5-8 in a sample of 476 students and found the grade level accounted for 68 to 71% of the variance on two passages. The individual differences in scores accounted for 84% of the variance on one passage. Brown-Chidsey, Johnson, & Fernstrom (2005) used a Maze selection with a controlled passage and found that the controlled and literature-based passages showed significant growth across fall, winter, and spring; the mean scores from the controlled passages consistently were higher. Espin, Deno, Maruyama, and Cohen (1989) reported correlations between the Basic Academic

Skills Samples (BASS) Maze and 1-minute oral reading fluency passages for Grades 3, 4, and 5 which were .77, .86, and .86 respectively. Fuchs and Fuchs (1990) found correlations of .83 between scores on Maze and oral reading, and Gardner, Rudman, Karlsen, & Merwin (1982) found correlations between scores on Maze and the Reading Comprehension subtest of the Stanford Achievement Test (SAT) to be .77.

Overview of the CRCT

The reading portion of the Georgia Criterion Referenced Competency Test (CRCT) is administered annually to all students from first grade to eighth grade. The CRCT Reading Test is created to measure reading comprehension and to evaluate whether students are meeting state standards at each grade level. The same content standards are measured every year. The test is comprised of multiple-choice questions, and the students' performance on the CRCT in reading based on the Georgia Performance Standards (GPS) is categorized according to three levels of reading: Level 3-exceeding standards, Level 2-meeting standards, or Level-1 not meeting standards. The student performance on the CRCT is reported in scaled scores for each grade ranging from 600 to 900.

The Validity of CRCT

The validity of the CRCT was established through a clear identification of the test's purpose by the state legislature which was to be "a measure of how well students have mastered the state's curriculum" (GADOE, 2009August). The CRCT Reading Test then began to be developed with the state's mandated curriculum, Georgia Performance Standards (GPS), relying heavily on including educators from across the state. Committees of educators were formed which reviewed the curriculum and established the reading concepts, knowledge, and skills that would be assessed and how they would be

assessed (GADOE, 2009August). These results became the test blueprint and test specifications. From these two documents, another document was created giving details for the item writing phase which identified the way specific standards of the curriculum would be categorized into domains or strands. The test item specifications document provided detail about the kinds of items to be written including the item format, content scope and limits, and cognitive complexity. All of these documents were created with meetings of the Georgia Department of Education with the assessment contractor as well as substantial involvement of curricular specialists and Georgia educators. The content domain specifications became the CRCT Content Descriptions which are posted on the website to inform those interested of the test's content and method of assessment. A Content Weight document was produced to show the relative proportion of the items by the domain. These documents along with using Georgia educators throughout the process were part of the establishment of the validity of the CRCT (GADOE, 2009August). The final steps in creating the CRCT were the writing of the test items, review of field-tested items with statistics, development of the test form, the standard setting, and the technical documentation of the reliability and validity of the tests (GADOE, 2009August).

The Reliability of the CRCT Scaled Score

The validity of the CRCT was established and along with the validity the reliability was also determined (GADOE, 2009August; 2009September). There were several reliability indices that were reported for the CRCT. The first two statistics to measure reliability were Cronbach's alpha reliability and standard error of measurement (SEM) (GADOE, 2009August). Table 4 identifies these two statistics for the 2008 and 2009 CRCT.

Table 4

Reliability Coefficients (Cronbach's Alpha) and the Raw Score for SEM for Middle School Students Who Took the 2008 and 2009 Georgia CRCT

Grade	2008 CRCT Reading		2009 CRCT Reading	
	Alpha	SEM	Alpha	SEM
6	.88	2.54	.88	2.49
7	.87	2.60	.86	2.62
8	.87	2.50	.87	2.42

The Reliability of the CRCT Lexiles

The reliability coefficient can be compared from test to test, and they range from 0 to 1 (GADOE, 2009a). The consistency of the reliabilities and SEMs between the CRCT 2008 and 2009 with previous administrations suggest that the CRCT assessments are "sufficiently reliable for their intended purpose" (GADOE, 2009a, p. 5; GADOE, 2009b). The reliability of the Lexiles when compared to the CRCT scores for 16,363 students for grades 1-8 is .72 to .88 (GADOE, 2009August).

The Reliability of the CRCT Level

The contrast to the standard errors of measurement (SEM) is the condition standard errors of measurement (CSEMs) which is the third statistic to demonstrate reliability of the CRCT Level. The Rasch-based CSEMs for the 2008 and 2009 CRCT Level suggested that they were consistent with previous administrations and give an accurate picture of a student's reading performance (GADOE, 2009a). Table 5 shows the CSEMs for the 2008 and 2009 CRCT Level.

Table 5

The CSEMs at the Cut Scale Score that Define the Performance Levels of the 2008 and 2009 CRCT

	2008 CRCT CSEMs		2009 CRCT CSEMs	
Grade	Meets	Exceeds	Meets	Exceeds
6	7	10	8	10
7	7	11	7	10
8	7	10	7	10

The Validity of the EOCT

The content validity of the EOCT was established with item alignment to the standards and content representation establishing the purpose of the test. Because the educators had extensive input into the construction of the test, the validity was ensured as well. The control of measurement error helped to establish the reliability of the test scores which in turn played a role in the validity of the EOCT. Construct validity was established through the consistent results that the reliability indices, model fit, and dimensionality studies yielded, which suggested that the EOCT was properly scored and the scores could be generalized to the universe score.

The Reliability of the EOCT

The reliability data for the 2009 EOCT has not been published as of the date of this publication, but the data for the previous years demonstrates that the scores for the EOCT remain consistent establishing validity.

The coefficient alpha, SEMs, and CSEM across several administrations were used to demonstrate the 9th Grade Literature and Composition EOCT's reliability through the consistency of the scores. These statistics are as follows:

Table 6

Summary of Coefficient Alpha and SEMs across Administrations for 2007-08 Ninth-Grade Lit EOCT

Coefficient Alpha			SEMs		
Summer 2007	Winter 2007 Form1/Form2	Spring 2008 Form1/Form2	Summer 2007	Winter 2007 Form1/Form2	Spring 2008 Form1/Form 2
.90	0.92/0.92	0.92/0.92	3.65	3.45/3.41	3.34/3.33

Table 7

Summary of CSEM for 2007-08 Ninth-Grade Lit EOCT

	Minimum CSEM	Maximum CSEM	Average CSEM	CSEM at Meets	CSEM at Exceeds
Summer 2007	9	97	14.07	9	11
Winter 2007 Form1	9	61	13.39	9	12
Winter 2007 Form2	9	61	13.41	9	12
Spring 2008 Form 1	9	61	13.39	9	12
Spring 2008 Form 2	9	61	13.42	9	12

Procedures of Curriculum-based Measurement

Using standardized observational procedures for repeatedly sampling performance on core reading has established the reliability and validity of CBM (Deno,

2003). The main characteristics of CBM are the psychometric concepts of reliability and validity (Good & Jefferson, 1998; Shinn, 1989). Hosp and Hosp (2003) provided directions regarding how to conduct CBM for reading which were used to train the teachers how to give the assessments. The procedures for the ORF assessment were taught first. The first step was to gather different but equivalent reading passages or probes, a timer, a writing utensil, a graph of equal intervals to plot the data, and directions for administering and scoring the passages. The three passages were about 200 words from the same grade level preferably passages that the student had not read (Wesson, 1992). Two copies of each passage were created, a passage for the teacher and for the student (Hosp & Hosp, 2003). At the end of each line on the teacher passage, a running word count was written to help the teacher in scoring (See Appendix A). Both copies were placed in protective sheet covers to promote durability and eliminate unnecessary expenses. The researcher prepared the assessments to ensure passage equivalency establishing validity and reliability (Wayman et al, 2007). The teachers chose passages randomly to establish passage equivalency and to eliminate bias (Wayman et al., 2007).

Preparation of Probes

Due to the nature of this study and to demonstrate that any teacher can utilize the procedures provided throughout research, the researcher created the probes based on the directions provided by Hosp and Hosp (2003). The equivalence of the probes was established through common Lexiles. The researcher used literature-based books and checked the book's Lexile level through www.lexile.com. On this website, one can click on *Find a Book* and follow the directions to either find the Lexile of a book or to create a list of books at a particular Lexile level or Lexile range. The Lexiles of many classroom literature books can also be found at www.scholastic.com. The researcher verified the

publisher when checking the Lexiles as well because different versions of a book can have different Lexiles. This verification established passage equivalency. Four probes from each Lexile range were created. The Lexile ranges used were 200-299L, 300-399L, 400-499L, 500-599L, 600-699L, 700-799L, 800-899L…..1600-1699L. The passages were typed from books or cut and pasted from excerpts of books found at www.lexile.com.

Two notebooks were created, one for the teacher and one for the student to use. The standardized directions were typed on a brightly-colored laminated sheet to create its permanence and to make the sheet easy to find. The researcher followed the directions given by Hosp and Hosp (2003) putting both copies in protective sleeves. The teacher copy had a running number of words on the side so that the words were easily counted, and the student had the same copy in a protective sleeve without the numbers. A dry erase marker was used allowing the teachers to mark errors on the plastic sleeve and easily erase. A timer was provided to provide an accurate timing of the student's reading.

Procedures

Procedures for ORF Assessment

The first assessment, ORF, was given as follows: The teacher collected the student's 2007-08 CRCT reading score and Lexile score which were reported together. Since the probes were from Lexile-based passages, the teacher used the student's 2007-08 Lexile score to get an approximate instructional level giving a beginning point to choose the first three passages for establishing a baseline. A teacher chose a passage near the CRCT Lexile, and two directly below. If the student read more than 125 words correctly, the teacher gave a higher Lexile-based probe. For example, if the student's Lexile was 750, then the teacher could give three passages in the 700 range since Lexiles

are presented as ranges. The teacher administered three probes one after the other or over three consecutive days determining the student's instructional level using the following ORF scripted standardized directions:

> Today, I will be listening to you read orally and checking your fluency. Fluency will be measured by the number of words you read correctly in one minute. When I say begin, you will read the passage out loud, and I will time you. You will have one minute to read. Make sure you focus not only on how fast you are reading but also on how many words you are reading correctly. When I say, 'begin,' start reading aloud at the top of this page. Read across the page [demonstrate by pointing]. Try to read each word. If you come to a word you don't know, I'll tell it to you. Are you ready? You may begin. (Start timer as you say begin) (Adapted from Hosp & Hosp, 2003; Wright, 1992).

Data Collection for ORF

The teacher listened to the student read for one minute and marked errors as the student read orally (see Appendix A). After the student read, the words correctly read per minute (WCPM) were recorded. Students who scored on the reading portion of the CRCT a scale score below 800 or close to 800 or those who read below 100 words correct per minute were considered at-risk and were monitored closely with weekly fluency checks as teachers and administrators utilized the Response to Intervention model. The Lexile score was an indicator of students who were at-risk based on Table 8. Table 8 also showed which Lexile-based passages to use to monitor.

After the initial three ORF assessments, the teacher administered one ORF assessment every other week to the at-risk readers and once every nine weeks to all students in regular language arts classrooms which was a workable scenario in

classrooms with as many as thirty students. The teachers recorded this data in the computerized grade book called Infinite Campus. In addition to the students who were identified as being at-risk in reading, five percent of the population were randomly selected using stratified random sampling (6^{th}, 7^{th}, and 8^{th} grade; Level 1, Level 2, and Level 3) to monitor with ORF assessments every other week. The five percent were based on the percentages of students within each level from the previous year's reading CRCT scores. The stratified random sampling was implemented to be able to gather ORF data to establish if there was a relationship with the ORF and the Criterion Referenced Competency Test (CRCT) or End of Course Test (EOCT) scores and to establish reliability and validity providing a representation of the population at the middle school.

Table 8

Correlation of Lexiles to Proficient and Exemplary Based on 2007-08 CRCT Scores

	6^{th}	7^{th}	8^{th}	6^{th} Monitor With	7^{th} and 8^{th} Monitor With
Passing Lexile Score (Below 800 CRCT Score. Students below this point did not meet standards.)	685 Lexile	800 Lexile	805 Lexile	700-900 Lexile Probes	800-1000 Lexile Probes
Grade Level Lexile Score (75% accuracy) 30 out of 40 CRCT Score from 800-849 Students met standards.	831 CRCT 955 Lexile	826 CRCT 1015 Lexile	828 CRCT 1080 Lexile	900-1200 Lexile Probes	1000-1200 Lexile Probes
Exemplary Lexile Score (90% accuracy) CRCT Score from 850 and above Students exceeded standards.	1210 Lexile	1210 Lexile	1265 Lexile	1200-1400 Lexile Probes	1200-1400 Lexile Probes

The median score of the three passages given at the beginning of the ORF assessments was recorded and charted as the beginning point on a scatter plot. The data were compared to Table 2 as shown by Hosp and Hosp (2003). This analysis was to identify if the student's instructional level met the criteria in Table 2 noting the number of errors and words read correctly per minute which denoted a frustrational level (Fuchs & Deno, 1982). Goal level Lexile-based probes (see Table 8) were used to progress monitor the student's growth throughout the year based on each student's performance on the 2007-08 CRCT. The goal level was based on what skills and level of curriculum the student should reach in one year, generally one year above instructional level (Shinn, Gleason, & Tindal, 1989). Hintze, Daly, and Shapiro (1998) recommended for grades 1 and 2 using 30 passages from the student's instructional level. For grades 3 and above, they recommended using goal level material or passages one year above instructional level because the passages could be more sensitive to growth and allow for frequent changes in the early grades but provide other diagnostic information for grades 3 and above. Since this study utilized middle school students, the goal level material was used and followed the criteria from Table 8. Table 8 provided the Lexile level that the student must read fluently to score at each CRCT level. According to Table 8, if students in 6th grade received a Level 1, they did not meet standards and received a scaled score below 800 on the Reading CRCT established by the state of Georgia, and their Lexile level was below 685 (GADOE, 2008c). The Lexile of 685 was the Lexile level of reading passages that the students needed to be able to read fluently to make a score of 800, or pass; therefore, the teacher monitored any 6th grader who scored below 800 (Level 1s) on the previous year's CRCT with the Lexile-based probes ranging from 700 to 900 because this level provided the goal level Lexile that the students needed to read to pass the CRCT in

their current school year. If a student received a score of 800-849 the previous school year and were in 6th grade, they met standards receiving Level 2 (GADOE, 2008c). The Level 2 students in 6th grade were monitored with 900 to 1200 Lexile-based passages because their goal level was to exceed standards or receive a Level 3. To be recognized as exceeding standards in 6th grade in 2007-08, a student needed to be able to read fluently 1210 Lexile probes. Students who scored 850 and above, or Level 3, were monitored using 1200-1400 Lexile-based probes. To summarize for 6th, 7th, and 8th grade, Level 1 students' goals were to reach Level 2, Level 2s to reach Level 3s, and Level 3s to go beyond what was even measured by the CRCT. This way *all* students were challenged and able to demonstrate growth.

The scoring and directions for all levels were the same (Hosp & Hosp, 2003). Errors were recorded as words that were mispronounced, omitted, substituted, inserted, or reversed. If the student hesitated on a word more than 3 seconds, the word was given, and an error marked. Repetitions were ignored, and self corrections within 3 seconds were counted correct (Hosp & Hosp, 2003).

Not only were the students' goal levels established at the beginning, but the expected growth was determined as well. In Table 3 (see Appendix C), Hosp and Hosp (2003) adapted and compiled data from two studies: Fuchs, Fuchs, Hamlett, Walz, and Germann (1993) showed realistic and ambitious growth rates for typically developing students, and Deno and others (2001) identified growth rates. A row was added and color-coded gray to show that it was not part of the original table but was added based on the number correlations in the chart from other grade levels and research regarding reading fluency to give the teachers an idea of where 7th and 8th graders should be (see Appendix C). The passages used to monitor growth should correspond for the expected

growth for that grade (see Table 3, Hosp & Hosp, 2003). For example, a sixth grade student reading at seventh grade instructional level would have expected growth corresponding to eighth grade. If the student's median score on eighth grade material was 95 words read correctly, and the student was monitored for 23 weeks, the expected ambitious growth was .66 x 23 weeks = 15.18, which was added to the 95 words read correctly or [15.18 + 95 = 110.18]. Thus, the determined goal was 110.18 or 110 words read correctly (Hosp & Hosp, 2003). These scores were plotted on the interval graph in which the vertical axis was the number of words read correctly, and the horizontal axis was the number of weeks monitored, allowing for data to be plotted one to two times a week. A copy of a scatter plot that can be graphed by hand or links to graph by a computer can be found at http://interventioncentral.org/htmdocs/interventions/cbmwarehouse.php.

For this study, the graphing by hand scatter plot was used. A goal line or regression line was drawn on the graph connecting the beginning median score to the expected ambitious growth, WCPM + 15 (15.18 was rounded to the nearest whole number), creating a regression line or goal line. When the ORF assessments were given, the data points were plotted every other week on the chart. If four consecutive scores fell below the goal line, instruction was reevaluated utilizing the RTI model. If the student was consistent in scoring above the line, the goal was raised (Fuchs, Fuchs, & Hamlett, 1989). Analyzing data effectively enabled progress monitoring to meet all students' needs and demonstrated the effectiveness of the Response to Intervention model.

Procedures for the Maze Fluency Assessment

The teachers followed standardized written procedures for the Maze which were given in training sessions (See Appendix C). In the beginning, the student was given

three Maze assessments to establish a baseline. The Maze assessments were probes that were Lexile-based having about 400 words. The Lexile-based probes were collected by the researcher just as the ORF probes. Every seventh word was replaced with two distractors and the correct answer. Some researchers found that the distractor should be within one letter of the correct choice, such as if the word has four letters, then the distractors should have three, four, or five letters. These Maze assessments will not focus on the length by letters because the researcher used a free template found at http://www.lextutor.ca/cloze/n_word/ which replaced the seventh word with three choices without regards to word length. The researcher cut and pasted the probes onto the template that were gathered from literature excerpts. Then the researcher pasted the first sentence in the space, telling the template to leave the first sentence in tact. Next, the researcher clicked on the feature to leave out every seventh word and have three answer choices, which the template did by putting a drop down box with three answer choices. Even though it did not allow for utilizing specific word length, it provided an efficient, teacher-friendly (The assessment was graded automatically on-line and off-line when the student clicked, *Check.*), cost effective (There was no cost for paper copies and template was free.) computerized template that was excellent for classroom progress monitoring. The template was valuable because it was easily adapted and used for all ages. The Maze Lexile-based probes were placed on the I-Drive of the school computers making the probes available to all language arts teachers from any school computer and also on the P-drive making the probes available to all students enabling the teacher to give the Maze to the class at once, to small groups, or to individual students.

The teacher followed the scripted directions (See Appendix C). The teacher kept the time for three minutes. At the end of three minutes, the teacher stopped the students'

assessments. The students clicked C*heck* on the computer. The computer template was programmed to grade, and it provided the number of correct responses which the students recorded and plotted on an individual scatter plot allowing for self-monitoring and assessing. The teacher recorded the correct responses for each student on her chart and in *Infinite Campus* as well. This procedure was repeated for three different probes.

Data Collection of the Maze

The median score from the three baseline scores was used as the beginning point, and the goal line point or end point was based on the following formula: 23 X .4 = 9.2 The 9.2 was rounded to 9 correct responses as expected growth for the 23 week period.

Number of weeks X .4 = _____ + median score = Maze Fluency Goal.

A regression line was drawn establishing the goal line. The number of correct responses was used to monitor progress (see Appendix C Maze Goal Setting Sheet). The graph was used to follow the Response to Intervention model.

The Georgia Criterion Referenced Competency Test (CRCT) and the End of Course Test (EOCT) were given in April 2009 and were Georgia's measurement of reading achievement according to the Georgia Performance Standards. Scores were available May 2009. This data were the final data collected for the study.

Data Analysis

An analysis of the scores from the CRCT and the ongoing probes was completed using a multiple linear regression of the means of each assessment according to the subgroups (Level 1, Level 2, and Level 3). Even though all students in the language arts classrooms were monitored with the Maze assessment, only students with complete data, (previous year's CRCT score, previous year's CRCT level and 12 Maze assessments), were used for the analysis. Some students had incomplete data due to absences and

moving. Some teachers monitored with incorrect goal level passages according to the directions provided; therefore, these students' data were not used for the analysis to eliminate bias.

Ary and others (2006) defined multiple regression as a statistical procedure that looks at the relationship between variables using the following equation:

$$\grave{Y} = a + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + b_5X_5 + b_6X_6$$

$\grave{Y}$ was the value of the dependent variable, the predicted CRCT score; a, b_1, b_2, and b_3 were the constants that were provided by the regression analysis; X_1 , X_2 , X_3 was the independent or predictor variables which was the mean ORF scores or mean Maze scores, the 2007-08 CRCT scaled score and 2007-08 CRCT level. The regression analysis yielded an *R* which was the coefficient of multiple correlation, indicating the relationship between the predictor variables in combination and the criterion (Ary et al., 2006). EOCT scores (dependent variable from the students in 8th grade taking 9th grade literature for high school credit) were analyzed with multiple regression to see if there was a significant relationship between the predictor variables and the criterion variable.

Potential Findings

1. Will using an oral reading fluency (ORF) or Maze assessment be a valid and reliable progress monitoring tool in predicting reading achievement for a middle school student?

2. Will the ORF or the Maze predict the reading performance of a middle school student on the Georgia Criterion Referenced Competency Test (CRCT) or the Georgia Literature End of Course Test (EOCT)?

3. Will the ORF and/or the Maze provide valid and reliable data for students scoring below standard, meeting standard, or exceeding standard on the summative evaluations?

CHAPTER IV

RESEARCH FINDINGS

This chapter summarizes the research findings that were obtained through data gathered from two Response to Intervention (RTI) tools, Oral Reading Fluency (ORF) and Maze Assessments that were given to the students by the Language Arts teachers at North Hall Middle School (NHMS) in Gainesville, Georgia during the school year 2008-2009. Chapter IV includes demographic data for the study participants and findings from the data of the randomly selected students who were given the Oral Reading Fluency Assessments, the students' Maze assessments, the students' Lexile and Criterion Referenced Competency Test (CRCT) scores from the previous year, 2008, and from the current year, 2009, and the students' scores from the reading portion of the 2009 End of Course Test (EOCT). Finally, the data were analyzed by multiple regressions, and the results were analyzed by the research questions to identify the effects of the two assessments, ORF and Maze, on the reading achievement of the students at NHMS as measured by the CRCT and the EOCT.

Demographic Data

Table 9 includes demographic data for all participants who received the Maze assessments. The Maze assessment was given every other week from September to March that the students were in attendance. The Maze assessment was a timed passage of three minutes and was completed individually in the Computer Lab, a whole class at one time. As Table 9 shows, there were a total of 833 students involved in the study, 290 of the participating students were in 6^{th} grade, 271 were in 7^{th} grade, and 272 were in 8^{th}.

Because of reasons, such as no previous year's CRCT score, moving, sickness, and absenteeism, some of the participant had incomplete data; therefore, any student with incomplete data was subtracted from the beginning number as shown in Table 9. The data were further reduced because incorrect methodology was used. The final total of student participants represents this deduction, 226 in 6th, 179 in 7th, and 151 in 8th grade.

Demographics of the Maze Assessment

Table 9

Number of Students Assessed with Maze Assessment

Grade	Total tested beginning in August 2008	Students with Incomplete Data (ID)	Total minus ID	Students monitored at Wrong Level (WL)	Total minus ID minus WL
6th	290	18	272	46	226
7th	271	21	250	71	179
8th	272	37	235	84	151

Table 9 compares the total number of participants who were given the Maze assessment in each grade level from the beginning of the year to the end of the year using the 2008 CRCT Level to determine a monitoring level to begin the progress monitoring process and the 2009 CRCT Level to measure general reading growth. For participants who transferred from other schools or who did not have a CRCT score in 2008, the teacher gave the Oral Reading Fluency Assessment (ORF) and assessed a beginning level through Lexile-based passages. The passages ranged from 200 Lexile (early 2nd grade) to 1600 Lexile (college level). Even though these students were monitored with progress monitoring throughout the school year, they were considered to have incomplete data because the two independent variables, 2008 CRCT score and level, were not available;

therefore, these students' data could not be used in the final analysis. Table 10 identifies the population of the students used for this study because some teachers monitored at levels that were not based on the standardized methodology provided, such as monitoring at a level too high or too low; therefore, the students who were monitored at the wrong level were pulled from the data so that there would not be a bias. Most of the participants that were monitored at the wrong level were at Level 2 and 3.

Table 10

Students with Complete Data Monitored with Correct Methodology

Grade	Year	1s (Below Standards)	2s (Meets Standards)	3s (Exceeds Standards)	No Score Available	Total
6th	2008	13	157	56	9	235
6th	2009	8	118	100	0	226
7th	2008	7	102	72	9	190
7th	2009	12	115	52	0	179
8th	2008	12	124	15	17	168
8th	2009	6	96	49	0	151

Demographics of the ORF Assessment

Unlike the Maze, the ORF assessment was given to each student individually. The student read orally to the teacher from Lexile-based passages for one minute, and the teachers recorded the number of words read for one minute accurately. Because of the large class sizes, random selection of [5%] of the student population was used. This decision was made to make it more teacher-friendly for teachers with large class sizes and yet still provide a valid and reliable representation of the target population. At the beginning of the school year, 2008-2009, the principal of North Hall Middle School randomly selected [5%] of the total number of students who took the 2008 CRCT for

each subgroup of Level 1s, 2s, and 3s from an Excel file by choosing every *nth* student. Table 11 indicates the number of students randomly selected to be given the Oral Reading Fluency Assessment every other week across the grade levels based on [5%] of the population. Grade 8 had one too many Level 2s and was one short for Level 3s.

Table 11

Totals for the 2008-09 School Year and Those Randomly Selected for ORF

Grade	Total Tested Who were Monitored All Year	Below Standards	Meets Standards	Exceeds Above Standards	Total Without CRCT Score in 2008
6th 2008	290	14	191	76	9
5%	14.5	.7	9.55	3.8	n/a
Randomly Selected	15	1	10	4	
7th 2008	271	7	127	125	12
5%	13.55	.35	6.35	6.25	n/a
Randomly Selected	15	1	7	7	
8th 2008	272	15	194	38	25
5%	13.6	.75	9.7	1.9	n/a
Randomly Selected	13	1	11	1	

Overall Explanation of Data

The data results will be organized as follows: the means of each assessment by grade level, the summary output of each grade level with charts, and two kinds of plots for each grade level, line fit and residual plots. The line fit plots identify the independent variable against the dependent variable and show the general linearity of the relationship. The residual plots show the difference between the observed value and the predicted value. The residual plots for all grade levels showed no patterns in the residuals which meant the models were unbiased. The high values of the multiple linear correlation coefficient (*R*) supported the models as being relativity strong, unbiased, and predictive.

Table 12

Averages of Assessments, Words Correct Per Minute (WCPM), Correct Responses (CR)

6th Grade Randomly Assigned CRCT and ORF Averages

Level	CRCT 2008	ORF	CRCT 2009	Growth Per Week WCPM
1	791.0	98.5	800	0.04
2	818.3	116.7	824.6	0.23
3	879.5	145.3	880.5	0.21

6th Grade CRCT and Maze Averages

CRCT Level	CRCT 2008	MAZE	CRCT 2009	Growth Per Week CR
1	790.2	15.9	804.6	0.1
2	828.6	21.8	835.5	0.3
3	863.1	27.7	869.3	0.4

7th Grade Randomly Assigned CRCT and ORF Averages

CRCT Level	CRCT 2008	ORF	CRCT 2009	Growth Per Week WCPM
1	794.0	83.8	816.0	1.13
2	817.9	116.2	821.3	0.44
3	865.1	156.6	846.3	1.78

7th Grade CRCT and Maze Averages

CRCT Level	CRCT 2008	MAZE	CRCT 2009	Growth Per Week CR
1	793.6	17.4	811.6	0.1
2	824.6	22.9	821	0.4
3	865.6	30.4	847.6	0.4

8th Grade Randomly Assigned CRCT and ORF Averages

CRCT Level	CRCT 2008	ORF	CRCT 2009	Growth Per Week WCPM
1	792.0	98.2	817	(0.2)
2	822.5	135.6	840.6	0.2
3	866.0	209.7	920	1.0

8th Grade CRCT and Maze Averages

Level	CRCT 2008	MAZE	CRCT 2009	Growth Per Week CR
1	790.9	20.3	809.2	0.4
2	824.5	28.3	836.3	0.4
3	856.6	37.0	851	0.5

8th Grade Students' CRCT, EOCT, and ORF Averages Who Took 9th Grade Literature

Level	CRCT 2008	ORF	CRCT 2009	EOCT 2009	Growth Per Week WCPM
1					
2	835.5	151.6	859.5	456	0.1
3	866.0	209.7	920.0	490	1.0

8th Grade Students' CRCT, EOCT, and Maze Averages Who Took 9th Grade Literature

Level	CRCT 2008	MAZE	CRCT 2009	EOCT 2009	Growth Per Week CR
1					
2	836.2	31.9	851.4	452.6	0.3
3	857.6	42.4	851.9	485.6	0.3

These charts show the average growth from the CRCT 2008 to CRCT 2009 for each level. It indicates that as a student increases in the number of words read correctly per minute orally or increases in the number of correct responses on the Maze in three minutes, he or she will also increase in the CRCT scaled scores. The differences between the mean of the CRCT scores reflects the differences within the population because the Maze represents the mean of all students who received correct methodology and the ORF represents the mean of the [5%] who were randomly selected.

Results of 6th Grade

Table 13

6th Grade ORF Summary Output

Regression Statistics	
Multiple R	0.9548
R Square	0.9117
Adjusted R Square	0.8876
Standard Error	10.5618
Observations	15.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3.0000	12,662.6673	4,220.8891	37.8380	0.0000
Residual	11.0000	1,227.0660	111.5515		
Total	14.0000	13,889.7333			

	Coefficients	*Standard Error*	*t Stat*
Intercept	168.0022	124.4925	1.3495
Level	5.2644	9.9867	0.5271
CRCT08	0.7820	0.1831	4.2709
AVG ORF	0.0571	0.2662	0.2147

	P-value	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.2043	(106.0039)	442.0083	(106.0039)	442.0083
Level	0.6086	(16.7162)	27.2450	(16.7162)	27.2450
CRCT08	0.0013	0.3790	1.1850	0.3790	1.1850
AVG ORF	0.8339	(0.5287)	0.6430	(0.5287)	0.6430

The overall regression was significant with the multiple correlation coefficient, *R*, being 0.9548 and the squared correlation coefficient (R^2) being 0.9117. In other words, 91% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together. The overall regression is significant because *F* = 37.8380 with and associated probability of .000; therefore, using all predictors the correlation is significantly greater than 0.

Table 14

6th Grade Maze Summary Output

Regression Statistics	
Multiple R	0.7153
R Square	0.5117
Adjusted R Square	0.5051
Standard Error	19.4216
Observations	226.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3.0000	87,733.9609	29,244.6536	77.5310	0.0000
Residual	222.0000	83,738.2913	377.1995		
Total	225.0000	171,472.2522			

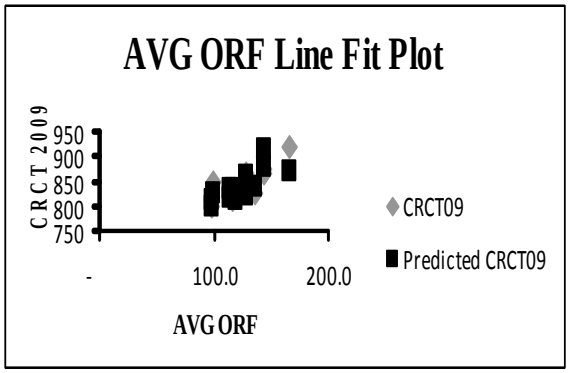
	Coefficients	*Standard Error*	*t Stat*
Intercept	302.4077	82.4351	3.6684
CRCT Level	7.4559	4.4139	1.6892
CRCT 07-08	0.6082	0.1099	5.5332
AVG Maze	0.6756	0.2833	2.3846

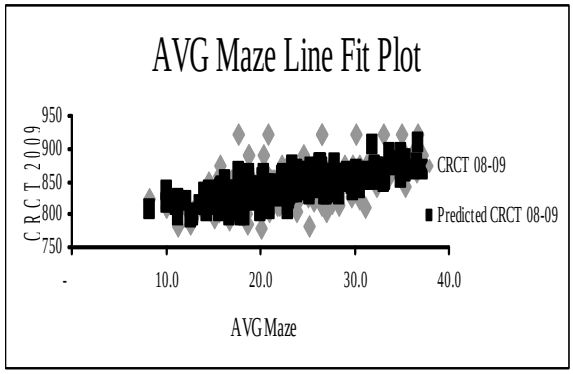
	P-value	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.0003	139.9522	464.8632	139.9522	464.8632
CRCT Level	0.0926	(1.2427)	16.1545	(1.2427)	16.1545
CRCT 07-08	0.0000	0.3916	0.8249	0.3916	0.8249
AVG MAZE	0.0179	0.1173	1.2340	0.1173	1.2340

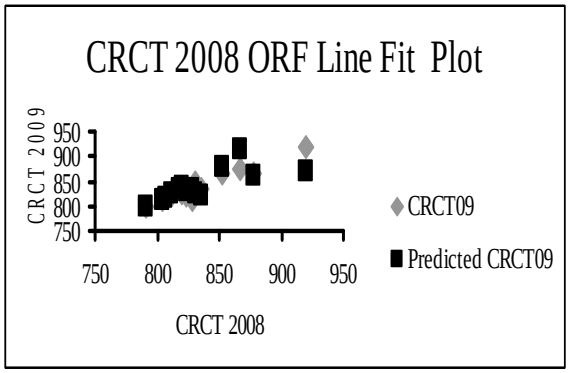
The overall regression was significant with the multiple correlation coefficient, R, being 0.7153 and the squared correlation coefficient (R^2) being 0.5117. In other words, 51% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together. The overall regression is significant because *F* = 77.5310 with an associated probability of .000; therefore, using all predictors the correlation is significantly greater than 0.

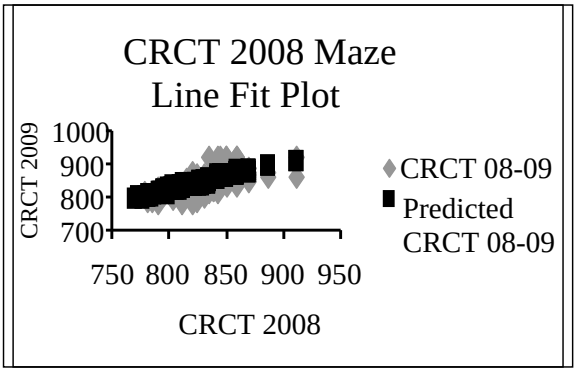
Figure 3

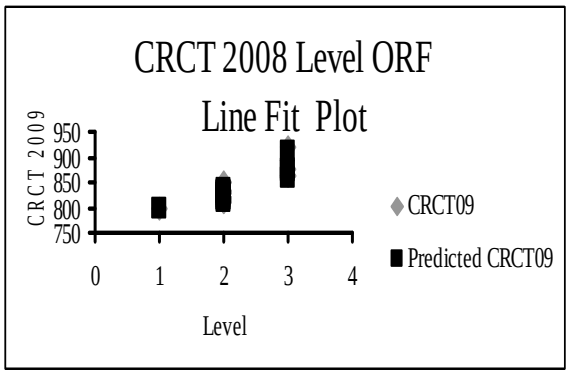
6th Grade ORF and Maze Line Fit Plots











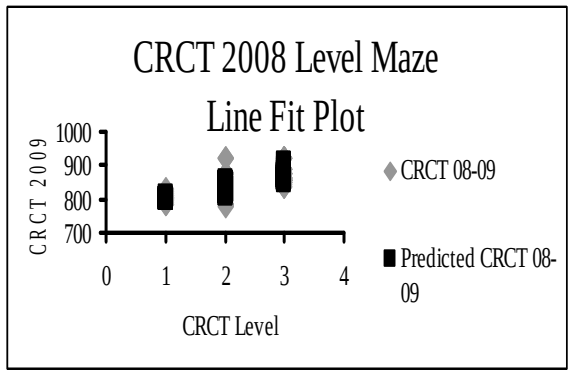
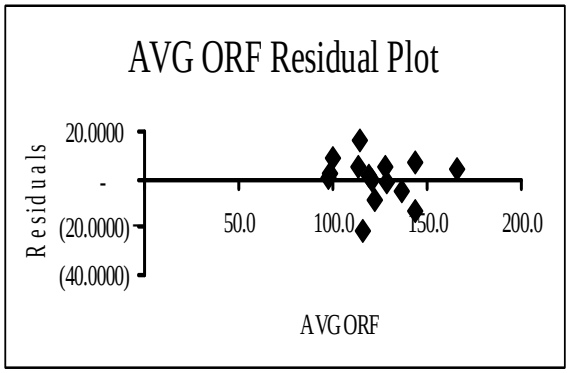
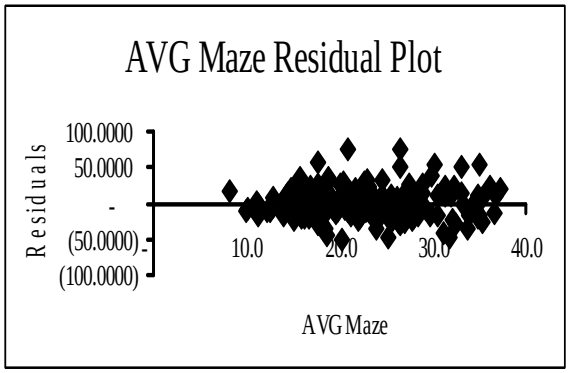
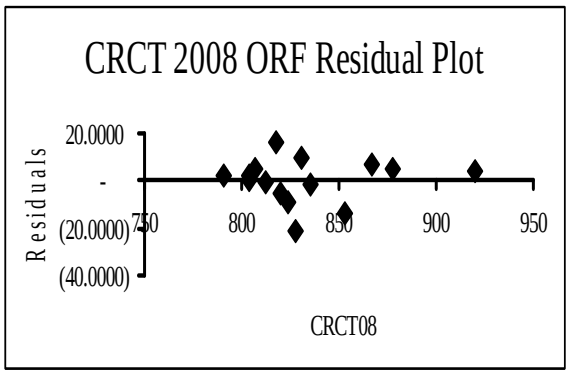


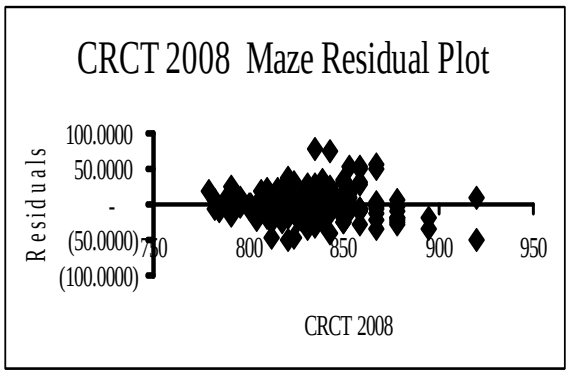
Figure 4

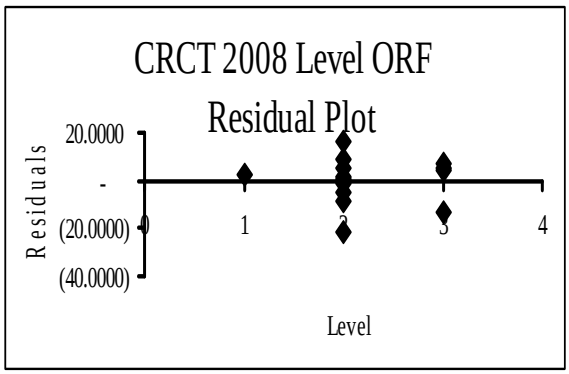
6th Grade ORF and Maze Residual Plots

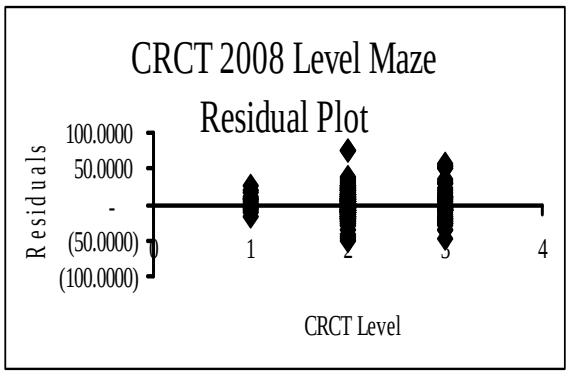












Results of 7th Grade

Table 15

7th Grade ORF Summary Output

Regression Statistics	
Multiple R	0.8036
R Square	0.6458
Adjusted R Square	0.5492
Standard Error	11.2520
Observations	15.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3.0000	2,538.9137	846.3046	6.6845	0.0078
Residual	11.0000	1,392.6863	126.6078		
Total	14.0000	3,931.6000			

	Coefficients	*Standard Error*	*t Stat*
Intercept	524.4125	175.3879	2.9900
Level	5.5961	11.1706	0.5010
CRCT 2008	0.3521	0.2377	1.4813
AVG ORF	(0.0029)	0.1099	(0.0260)

	P-value	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.0123	138.3862	910.4387	138.3862	910.4387
Level	0.6263	(18.9903)	30.1825	(18.9903)	30.1825
CRCT 08	0.1666	(0.1711)	0.8752	(0.1711)	0.8752
Avg ORF	0.9797	(0.2447)	0.2390	(0.2447)	0.2390

The overall regression was significant with the multiple correlation coefficient, R, being 0.8036 and the squared correlation coefficient (R^2) being 0.6458. In other words, 64.6% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together. The overall regression is significant because *F* = 6.6845 with an associated probability of .0078; therefore, using all predictors the correlation is significantly greater than 0.

Table 16

7th Grade Maze Summary Output

Regression Statistics	
Multiple R	0.7376
R Square	0.5440
Adjusted R Square	0.5362
Standard Error	13.5433
Observations	179.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3.0000	38,296.9392	12,765.6464	69.5975	0.0000
Residual	175.0000	32,098.7032	183.4212		
Total	178.0000	70,395.6425			

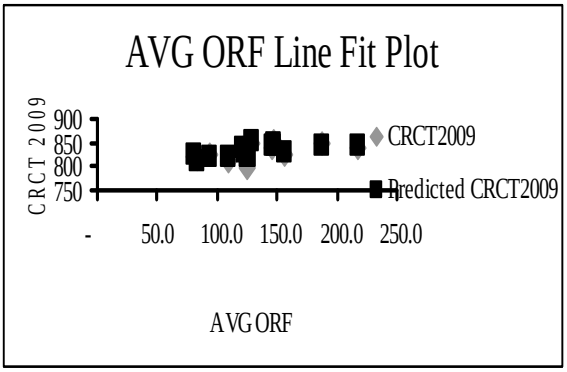
	Coefficients	*Standard Error*	*t Stat*
Intercept	566.0440	55.4328	10.2114
CRCT Level	8.2151	3.2028	2.5650
CRCT 07-08	0.2723	0.0751	3.6267
AVG MAZE	0.6674	0.2168	3.0787

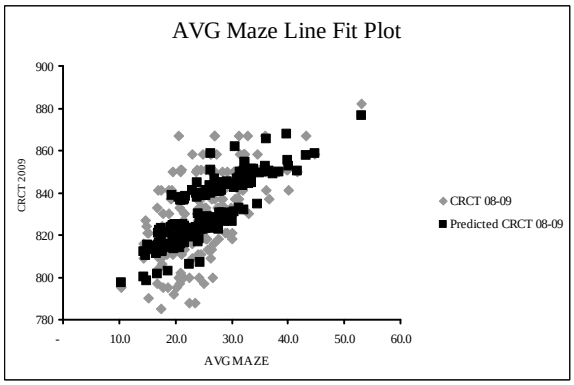
	P-value	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.0000	456.6412	675.4469	456.6412	675.4469
CRCT Level	0.0112	1.8940	14.5363	1.8940	14.5363
CRCT 07-08	0.0004	0.1241	0.4205	0.1241	0.4205
AVG Maze	0.0024	0.2395	1.0952	0.2395	1.0952

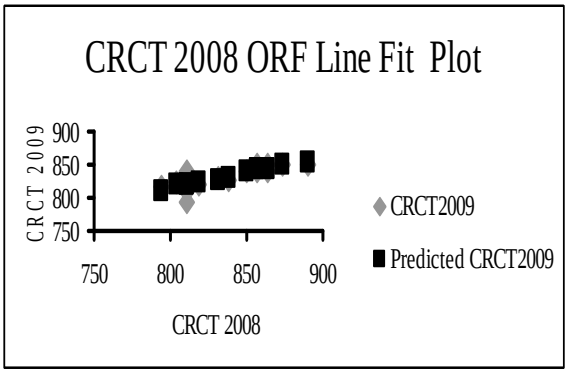
The overall regression was significant with the multiple correlation coefficient, R, being 0.7376 and the squared correlation coefficient (R^2) being 0.5440. In other words, 54.4% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together. The overall regression is significant because *F* = 69.5975 with an associated probability of .000; therefore, using all predictors the correlation is significantly greater than 0.

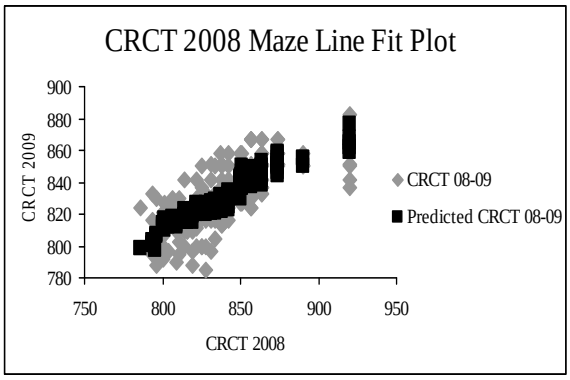
Figure 5

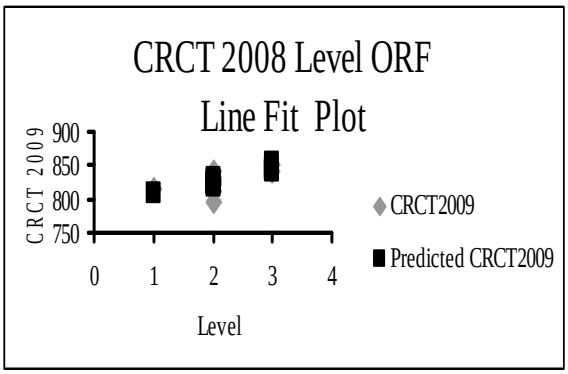
7th Grade ORF and Maze Line Fit Plots











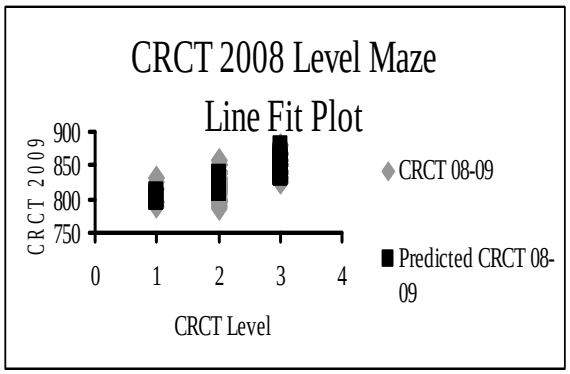
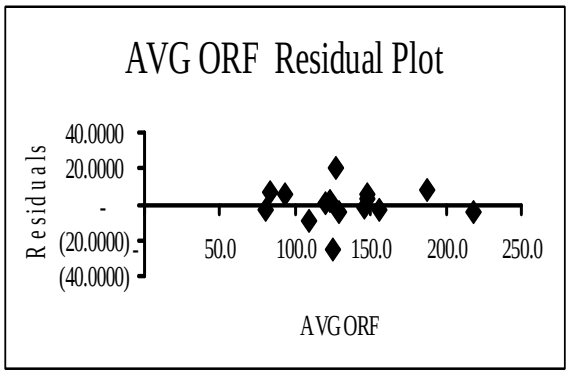
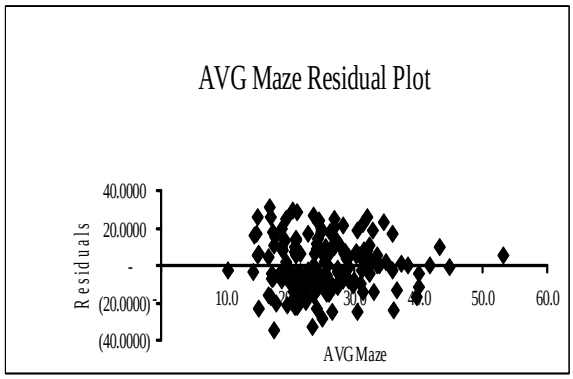
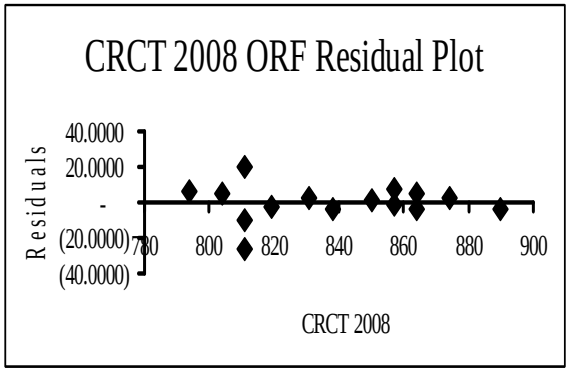


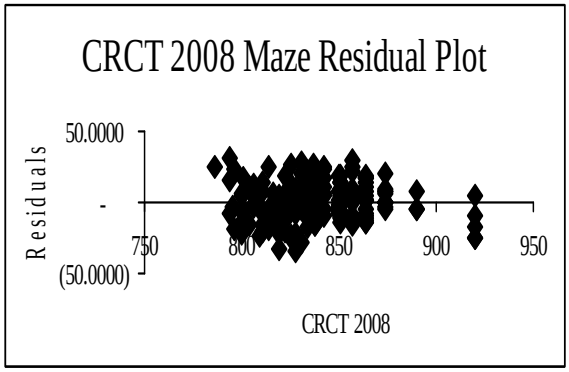
Figure 6

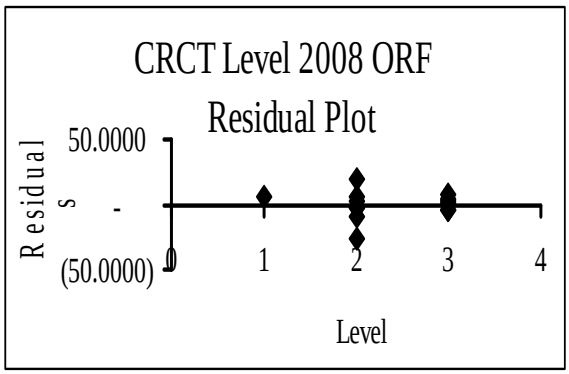
7th Grade ORF and Maze Residual Plots

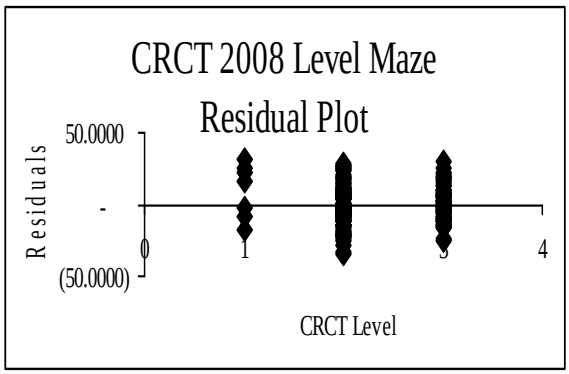












Results of 8th Grade

Table 17

8th Grade ORF Summary Output

Regression Statistics	
Multiple R	0.9022
R Square	0.8140
Adjusted R Square	0.7520
Standard Error	19.4437
Observations	13.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3.0000	14,892.4008	4,964.1336	13.1306	0.0012
Residual	9.0000	3,402.5223	378.0580		
Total	12.0000	18,294.9231			

	Coefficients	*Standard Error*	*t Stat*
Intercept	234.8755	474.7432	0.4947
Level	(20.1812)	21.3896	(0.9435)
CRCT08	0.6457	0.6405	1.0080
AVG ORF	0.8572	0.3200	2.6793

	P-value	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.6326	(839.0681)	1,308.8191	(839.0681)	1,308.8191
Level	0.3701	(68.5679)	28.2054	(68.5679)	28.2054
CRCT08	0.3398	(0.8033)	2.0947	(0.8033)	2.0947
Avg ORF	0.0252	0.1335	1.5810	0.1335	1.5810

The overall regression was significant with the multiple correlation coefficient, R, being 0.9022 and the squared correlation coefficient (R^2) being 0.8140. In other words, 81% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together. The overall regression is significant because *F* = 13.1306 with an associated probability of .0012; therefore, using all predictors the correlation is significantly greater than 0.

Table 18

8th Grade Maze Summary Output

Regression Statistics	
Multiple R	0.7243
R Square	0.5246
Adjusted R Square	0.5149
Standard Error	14.6747
Observations	151.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3.0000	34,935.1794	11,645.0598	54.0759	0.0000
Residual	147.0000	31,655.9729	215.3468		
Total	150.0000	66,591.1523			

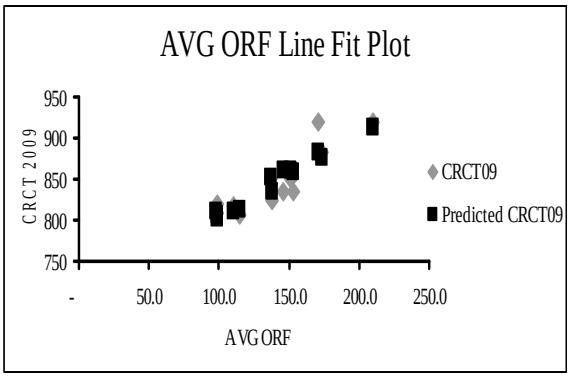
	Coefficients	*Standard Error*	*t Stat*
Intercept	136.2641	81.7608	1.6666
CRCT Level	(12.5332)	4.2681	(2.9365)
CRCT 07-08	0.8587	0.1089	7.8887
AVG MAZE	0.5654	0.2099	2.6934

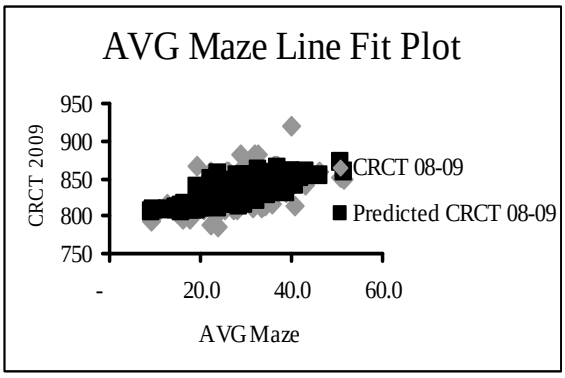
P-value	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
0.0977	(25.3143)	297.8426	(25.3143)	297.8426
0.0039	(20.9680)	(4.0984)	(20.9680)	(4.0984)
0.0000	0.6436	1.0739	0.6436	1.0739
0.0079	0.1505	0.9802	0.1505	0.9802

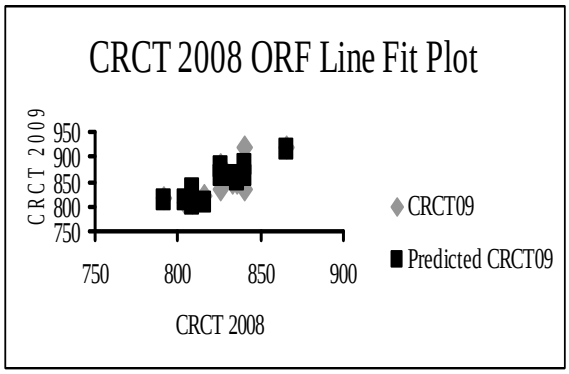
The overall regression was significant with the multiple correlation coefficient, R, being 0.7243 and the squared correlation coefficient (R^2) being 0.5246. In other words, 52% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together. The overall regression is significant because *F* = 54.0759 with an associated probability of .000; therefore, using all predictors the correlation is significantly greater than 0.

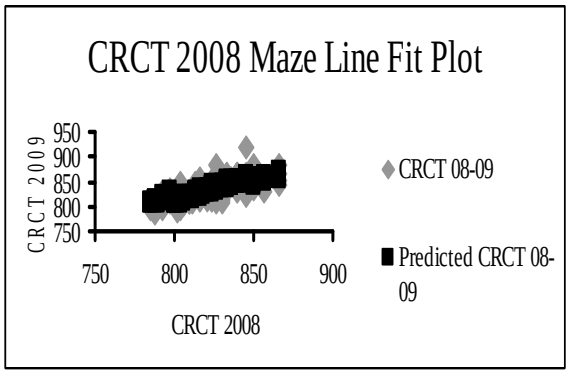
Figure 7

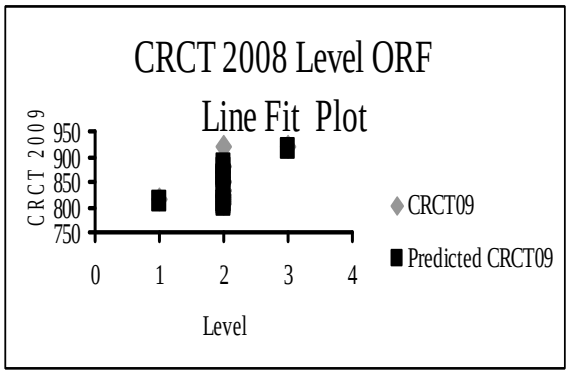
8th Grade ORF and Maze Line Fit Plots











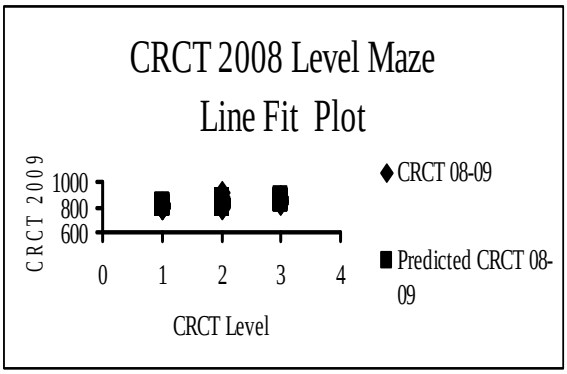
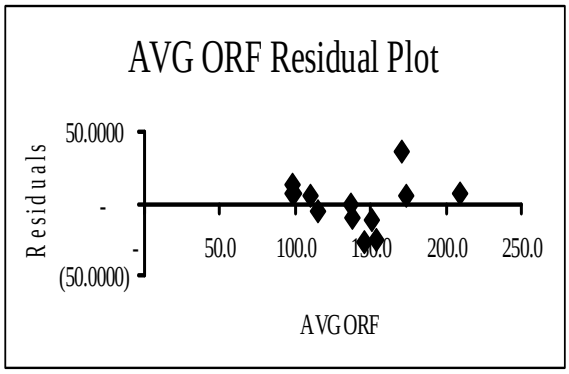
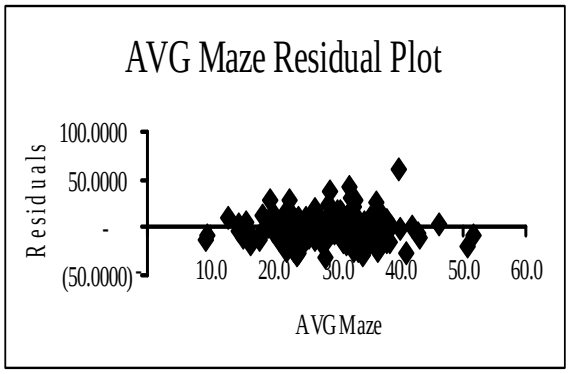
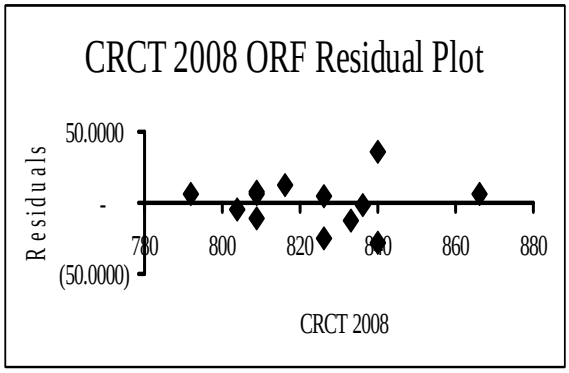


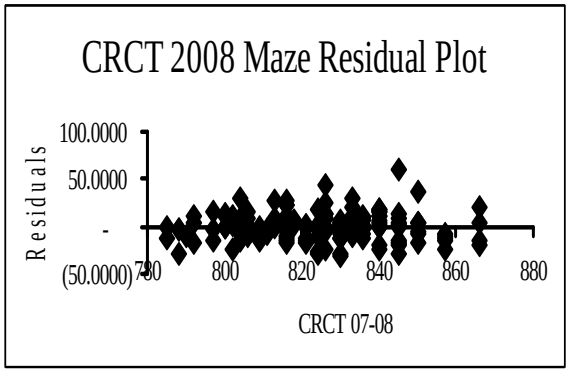
Figure 8

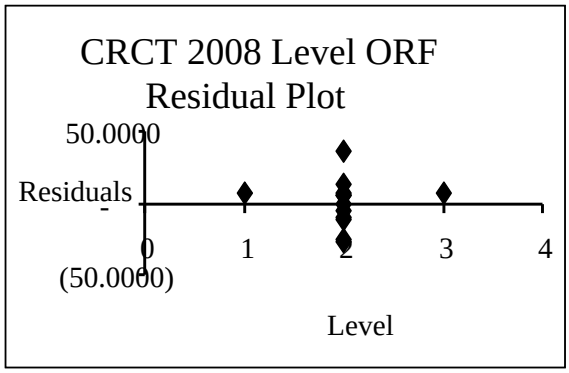
8th Grade ORF and Maze Residual Plots











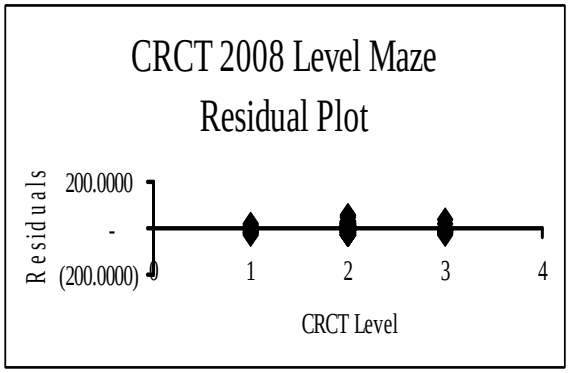


Table 19

8th Grade ORF Summary Output For Students Taking EOCT

Regression Statistics	
Multiple R	0.9936
R Square	0.9872
Adjusted R Square	0.9488
Standard Error	7.0184
Observations	5.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3.0000	3,797.5415	1,265.8472	25.6981	0.1438
Residual	1.0000	49.2585	49.2585		
Total	4.0000	3,846.8000			

	Coefficients	*Standard Error*	*t Stat*
Intercept	33.2800	477.7676	0.0697
Level	(99.0766)	24.6367	(4.0215)
CRCT08	0.3619	0.6179	0.5857
AVG ORF	2.1011	0.2807	7.4842

	P-value	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.9557	(6,037.3324)	6,103.8923	(6,037.3324)	6,103.8923
Level	0.1552	(412.1153)	213.9620	(412.1153)	213.9620
CRCT08	0.6627	(7.4893)	8.2131	(7.4893)	8.2131
AVG ORF	0.0846	(1.4660)	5.6682	(1.4660)	5.6682

The overall regression was significant with the multiple correlation coefficient, R, being 0.9935 and the squared correlation coefficient (R^2) being 0.9872. In other words, 99% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together. The limitation to this statistic is the small sample even though it represents more than [5%] of the student population (37 students) who were randomly selected and who took the EOCT.

Table 20

8th Grade Maze Summary Output for Students Taking EOCT

Regression Statistics	
Multiple R	0.5598
R Square	0.3133
Adjusted R Square	0.2509
Standard Error	26.8145
Observations	37.0000

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3.0000	10,827.4702	3,609.1567	5.0196	0.0056
Residual	33.0000	23,727.5568	719.0169		
Total	36.0000	34,555.0270			

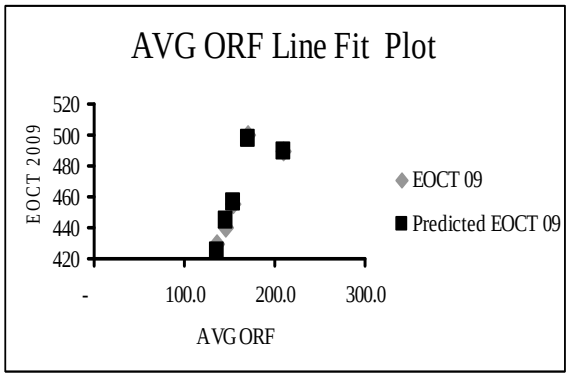
	Coefficients	*Standard Error*	*t Stat*
Intercept	325.2097	576.7151	0.5639
CRCT Level	10.4046	21.2893	0.4887
CRCT 07-08	0.0489	0.7254	0.0674
AVG MAZE	2.0598	0.8097	2.5440

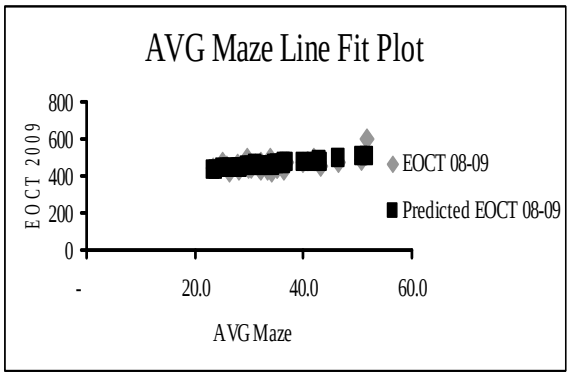
	P-value	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	0.5766	(848.1261)	1,498.5454	(848.1261)	1,498.5454
CRCT Level	0.6283	(32.9087)	53.7179	(32.9087)	53.7179
CRCT 07-08	0.9467	(1.4271)	1.5248	(1.4271)	1.5248
AVG MAZE	0.0158	0.4125	3.7071	0.4125	3.7071

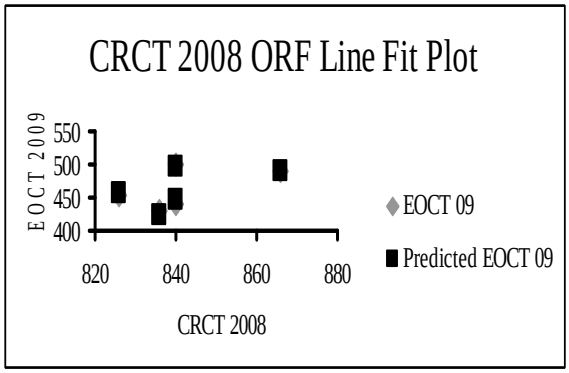
The overall regression was not as significant with the multiple correlation coefficient, R, being 0.5598 and the squared correlation coefficient (R^2) being 0.3133. In other words, 31.3 % of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together.

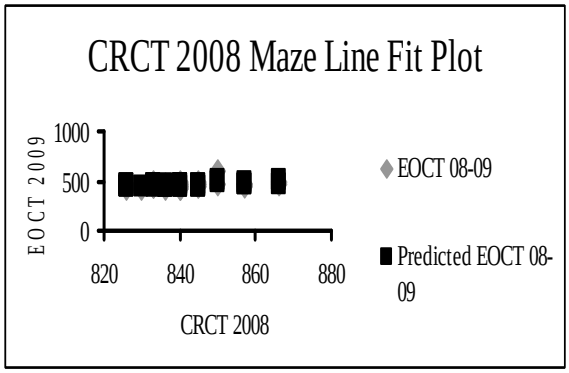
Figure 9

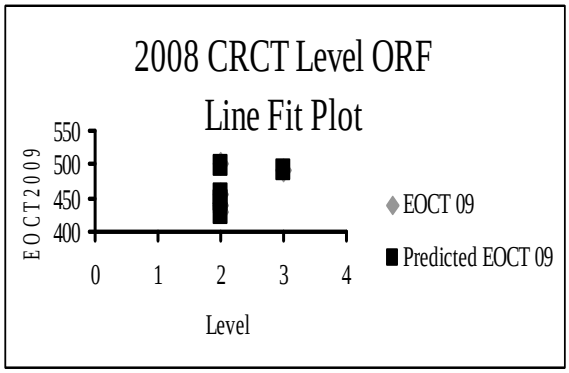
8th Grade ORF and Maze Line Fit Plots from Students Taking EOCT











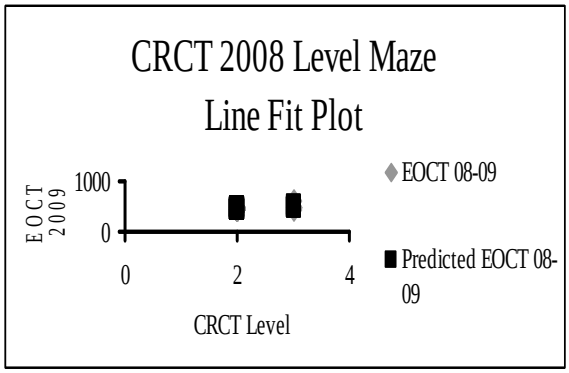
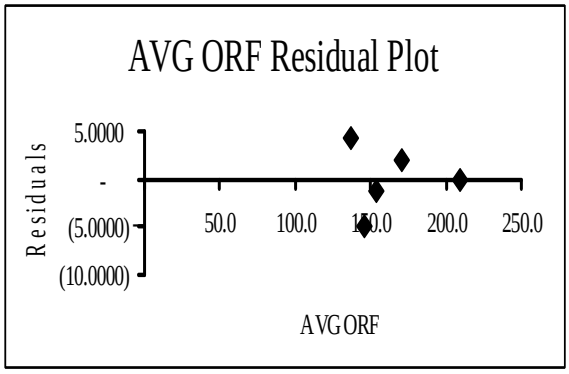
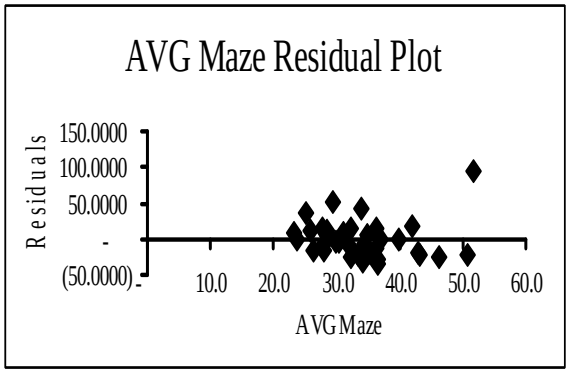
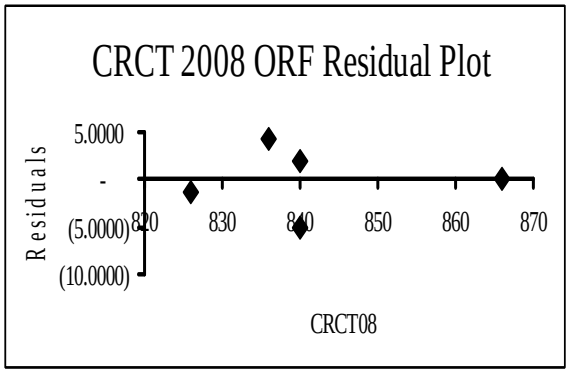


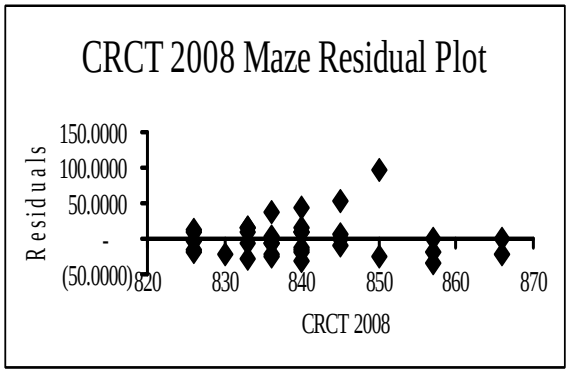
Figure 10

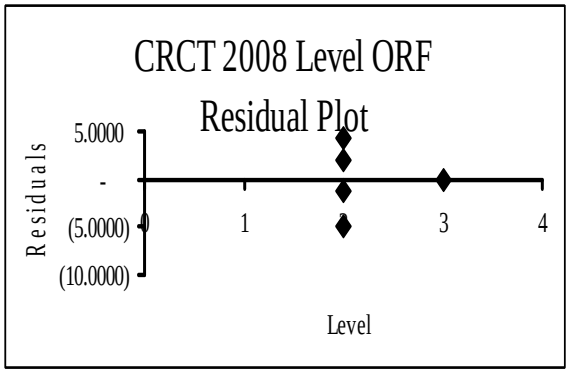
8th Grade ORF and Maze Residual Plots from Students Taking EOCT

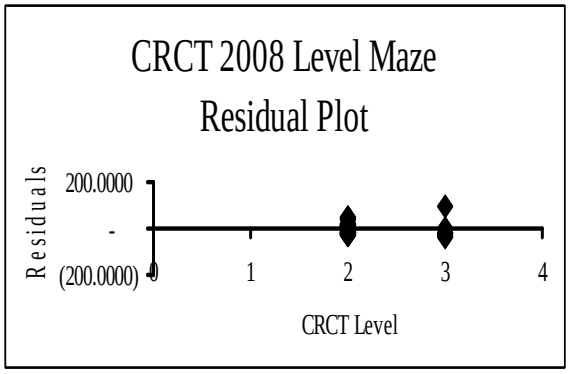












Research Questions

Three questions guided the framework for the primary focus of the study to determine if a relationship existed between two different independent variables, ORF and Maze assessments, and the dependent variables, the CRCT or EOCT score.

Research Question 1

Will using an oral reading fluency (ORF) or Maze assessment be a valid and reliable progress monitoring tool in predicting reading achievement for a middle school student?

Hypothesis (H_1): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for 6^{th} grade students.

H_1 can be accepted because the overall regression was significant with the multiple correlation coefficient *R*, being 0.9548 and the squared correlation coefficient (R^2) being 0.9117.The overall regression is significant because $F = 37.8380$ with an associated probability of .000; therefore, using all predictors the correlation is significantly greater than 0.

Hypothesis (H_2): There is a statistically significant relationship between the scores from the Maze assessment and the reading scores from the Georgia Criterion-Referenced Competency Test for 6^{th} grade students.

H_2 can be accepted because the overall regression was significant with the multiple correlation coefficient, R, being 0.7153 and the squared correlation coefficient (R^2) being 0.5117. The overall regression is significant because $F = 77.5310$ with an associated probability of .000; therefore, using all predictors the correlation is significantly greater than 0.

Hypothesis (H_3): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for 7th grade students.

H_3 can be accepted because the overall regression was significant with the multiple correlation coefficient, R, being 0.8036 and the squared correlation coefficient (R^2) being 0.6458. The overall regression is significant because $F = 6.6845$ with an associated probability of .0078; therefore, using all predictors the correlation is significantly greater than 0.

Hypothesis (H_4): There is a statistically significant relationship between the scores from the Maze assessment and the reading scores from the Georgia Criterion-Referenced Competency Test for 7th grade students.

H_4 can be accepted because the overall regression was significant with the multiple correlation coefficient, R, being 0.7376 and the squared correlation coefficient (R^2) being 0.5440. The overall regression is significant because $F = 69.5975$ with an associated probability of .000; therefore, using all predictors the correlation is significantly greater than 0.

Hypothesis (H_5): There is a statistically significant relationship between the mean of the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for 8th grade students.

H_5 can be accepted because the overall regression was significant with the multiple correlation coefficient, R, being 0.9022 and the squared correlation coefficient (R^2) being 0.8140. The overall regression is significant because $F = 13.1306$ with and

associated probability of .0012; therefore, using all predictors the correlation is significantly greater than 0.

Hypothesis (H_6): There is a statistically significant relationship between the mean of the scores from the Maze assessment and the reading scores from the Georgia Criterion-Referenced Competency Test for 8th grade students.

H_6 can be accepted because the overall regression was significant with the multiple correlation coefficient, R, being 0.7243 and the squared correlation coefficient (R^2) being 0.5246. The overall regression is significant because $F = 54.0759$ with an associated probability of .000; therefore, using all predictors the correlation is significantly greater than 0.

Hypothesis (H_7): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the scores from the End of Course Test for 8th grade students taking 9th Grade Literature and Composition.

H_7 can be accepted because the overall regression was significant with the multiple correlation coefficient, R, being 0.9935 and the squared correlation coefficient (R^2) being 0.9872. The limitation to this statistic is the small sample even though it represents more than [5%] of the student population (37 students) who were randomly selected and who took the EOCT.

Hypothesis (H_8): There is a statistically significant relationship between the scores from the Maze assessments and the scores from the End of Course Test for 8th grade students taking 9th Grade Literature and Composition.

H_8 can be accepted but the correlation was not as significant with the multiple correlation coefficient, R, being 0.5598 and the squared correlation coefficient (R^2) being 0.3133.

The data results support ORF and Maze as valid and reliable tools to use for progress monitoring in 6^{th}, 7^{th}, and 8^{th} grade with the coefficient of correlation, *R*, using ORF assessment being .95, .80, and .90 respectively, and *R* equal to .72, .74, and .72 using the Maze assessment. The ORF had the strongest relationship; therefore it was a stronger predictor along with the previous year's CRCT score and level than the Maze.

Research Question 2

Will the ORF or the Maze be able to predict a middle school student's reading performance on the Georgia Criterion Referenced Competency Test (CRCT) or the Georgia Literature End of Course Test (EOCT)?

Hypothesis (H_9): The oral reading fluency assessments is a predictor variable for reading performance of 6^{th} grade students on the Georgia Criterion-Referenced Competency Test.

H_9 can be accepted because R^2 is 0.9117 meaning 91% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together.

Hypothesis (H_{10}): The Maze assessments is a predictor variable for reading performance of 6^{th} grade students on the Georgia Criterion-Referenced Competency Test.

H_{10} can be accepted because R^2 is 0.5117 meaning that 51% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together.

Hypothesis (H_{11}): The oral reading fluency assessments is a predictor variable for reading performance of 7^{th} grade students on the Georgia Criterion-Referenced Competency Test.

H_{11} can be accepted because R^2 is 0.6458 meaning that 64.6% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together.

Hypothesis (H_{12}): The Maze assessments is a predictor variable for reading performance of 7th grade students on the Georgia Criterion-Referenced Competency Test.

H_{12} can be accepted because R^2 is 0.5440 meaning that 54.4% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together.

Hypothesis (H_{13}): The oral reading fluency assessments is a predictor variable for reading performance of 8th grade students on the Georgia Criterion-Referenced Competency Test.

H_{13} can be accepted because R^2 is 0.8140 meaning that 81.40% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together.

Hypothesis (H_{14}): The Maze assessments is a predictor variable for reading performance of 8th grade students on the Georgia Criterion-Referenced Competency Test.

H_{14} can be accepted because R^2 is 0.5246 meaning that 52.46% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together.

Hypothesis (H_{15}): The oral reading fluency assessments is a predictor variable for reading performance of 8th grade students taking the Georgia End of Course Test for Ninth Grade Literature and Composition.

H_{15} can be accepted because R^2 is 0.9872 meaning that 98.72% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together.

Hypothesis (H_{16}): The Maze assessments is a predictor variable for reading performance of 8th grade students taking the Georgia End of Course Test for Ninth Grade Literature and Composition.

H_{16} can be accepted but is a weaker correlation because R^2 is 0.3133 meaning that 31.33% of the variability in the predicted CRCT scores can be accounted for by variability in the three predictors considered together.

The plethora of supporting research from chapter two helped to establish the ability of the ORF and the Maze assessments in a general classroom setting to predict a student's achievement on a summative evaluation, and the evidence collected and analyzed from this study demonstrated that both ORF and Maze were valid and reliable predictors of a student's reading success on the Georgia CRCT and EOCT. The future CRCT scaled score or future EOCT scaled score, which were the dependent variables, were found to be predicted in the following ways:

1) Using the following independent variables, previous year's CRCT score, previous year's CRCT level, and the mean of at least twelve ORF or Maze progress monitoring assessments along with the coefficients from the statistical data of this study, the future CRCT or EOCT score can be predicted with the regression equation (see Figure 11). The future CRCT scaled score can be predicted with confidence because the coefficient of multiple correlation *R* in each of the grade levels as well as each assessment indicated a strong relationship between the predictor, the independent variables, in combination with the

criterion, the dependent variable. The future EOCT scaled score can be predicted using the ORF assessment based on the statistical analysis, but the statistical data using the Maze assessments did not indicate a relationship that was as strong with the dependent variable. Future studies with a larger sampling are needed to see the predictability for the EOCT.

2) By using the ORF as a progress monitoring tool, a teacher can predict if student will pass, meet, or exceed standards by using the data from Table 12. In general, at-risk students read below 100 words correctly per minute (wcpm). Students who met standards read about 120 wcpm, and students who exceeded standards read above 150 wcpm.

3) By using the Maze as a progress monitoring tool, the teacher can use Table 12 to predict whether a student will be below, meet, or exceed standards. The researcher's findings were closely aligned to the research-based chart on the Maze Scoring Sheet (see Appendix C).

Figure 11

Regression Equations to Predict the Future CRCT Score:

For 6th Graders Using ORF

Y' (Predicted CRCT 2009) = (5.2644) (CRCT Level) + (0.7820) (CRCT 2008) + (0.0571) (Average ORF) + 168.0022

For 6th Graders Using Maze

Y' (Predicted CRCT 2009) = (7.4559) (CRCT Level) + (0.6082) (CRCT 2008) + (0.6756) (Average Maze) + 302.4077

For 7th Graders Using ORF

Y' (Predicted CRCT 2009) = (5.5961) (CRCT Level) + (0.3521) (CRCT 2008) +

(-0.0029) (Average ORF) + 524.4125

For 7th Graders Using Maze

Y' (Predicted CRCT 2009) = (8.2151) (CRCT Level) + (0.2723) (CRCT 2008) +

(0.6674) (Average Maze) + 566.0440

For 8th Graders Using ORF

Y' (Predicted CRCT 2009) = (20.1812) (CRCT Level) + (0.6457) (CRCT 2008) +

(0.8572) (Average ORF) + 234.8755

For 8th Graders Using Maze

Y' (Predicted CRCT 2009) = (-12.5332) (CRCT Level) + (0.8587) (CRCT 2008) +

(0.5654) (Average Maze) + 136.2641

Research Question 3

Will the ORF and/or the Maze provide valid and reliable data for all students, such as students scoring below standards, at standards, or above standards on the summative evaluations?

Hypothesis (H_{17}): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students scoring below standards.

H_{17} can be accepted because the student's previous years CRCT Level 1 (below standards), Level 2 (met standards), or Level 3 (exceeded standards) was found to be a valid predictor along with the student's previous year's CRCT score and the average of 12 oral reading fluency assessments for 6th, 7th, and 8th grade. For the oral reading fluency assessment for students below standards, the average words read correctly per minute was less than 100 words.

Hypothesis (H_{18}): There is a statistically significant relationship between the scores from the Maze assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students scoring below standards.

H_{18} can be accepted because the student's previous years CRCT Level 1 (below standards), Level 2 (met standards), or Level 3 (exceeded standards) was found to be a valid predictor along with the student's previous year's CRCT score and the average of 12 Maze assessments for 6th, 7th, and 8th grade. The mean of the Maze assessments for the students who scored below standards ranged from 16 (6th grade) to 20 (8th grade) correct responses. The average weekly growth of correct response was 0.1, 0.1, and .04 respectively for 6th, 7th, and 8th grade.

Hypothesis (H_{19}): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students meeting standards.

H_{17} can be accepted because the student's previous years CRCT Level 1 (below standards), Level 2 (met standards), or Level 3 (exceeded standards) was found to be a valid predictor along with the student's previous year's CRCT score and the average of 12 oral reading fluency assessments for 6th, 7th, and 8th grade. For the oral reading fluency assessment for students who met standards, the average words read correctly per minute ranged from 116 (6th grade to 135 (8th grade).

Hypothesis (H_{20}): There is a statistically significant relationship between the scores from the Maze assessments and the reading scores from the Georgia

Criterion-Referenced Competency Test for middle school students meeting standards.

H_{18} can be accepted because the student's previous years CRCT Level 1 (below standards), Level 2 (met standards), or Level 3 (exceeded standards) was found to be a valid predictor along with the student's previous year's CRCT score and the average of 12 Maze assessments for 6^{th}, 7^{th}, and 8^{th} grade. The mean of the Maze assessments for the students who met standards ranged from 22 (6^{th} grade) to 28 (8^{th} grade) correct responses. The average weekly growth for the Maze for those meeting standards was 0.3, 0.4, and 0.4 for 6^{th}, 7^{th}, and 8^{th} grade respectively.

Hypothesis (H_{21}): There is a statistically significant relationship between the scores from the oral reading fluency assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students exceeding standards.

H_{17} can be accepted because the student's previous years CRCT Level 1 (below standards), Level 2 (met standards), or Level 3 (exceeded standards) was found to be a valid predictor along with the student's previous year's CRCT score and the average of 12 oral reading fluency assessments for 6^{th}, 7^{th}, and 8^{th} grade. For the oral reading fluency assessment for students who exceeded standards, the average words read correctly per minute ranged from 145 (6^{th} grade) to 209 (8^{th} grade).

Hypothesis (H_{22}): There is a statistically significant relationship between the scores from the Maze assessments and the reading scores from the Georgia Criterion-Referenced Competency Test for middle school students exceeding standards.

H_{18} can be accepted because the student's previous years CRCT Level 1 (below

standards), Level 2 (met standards), or Level 3 (exceeded standards) was found to be a valid predictor along with the student's previous year's CRCT score and the average of 12 Maze assessments for 6th, 7th, and 8th grade. The mean of the Maze assessments for the students who exceeded standards ranged from 28 (6th grade) to 37 (8th grade) correct responses. The average weekly growth for the Maze for those exceeding standards was 0.4, 0.4, and 0.5 for 6th, 7th, and 8th grade respectively.

The ORF and Maze provided valid and reliable data for all students including the students who did not meet the standards, the students who met standards, and those who exceeded standards as measured by the CRCT. The line fit plots identified this validity and reliability through the linear relationships. The individual scatter plots of the data gathered on each ORF and Maze assessment by each teacher were used in the classroom to provide a visual representation to help the teacher to determine if the student was making adequate reading progress, if the interventions that were used for the at-risk were effective, and if the students were being challenged. Table 12 identified data that was not only valid and reliable data to use but also data that was practical and easy to understand. The data substantiated that for a student to pass the CRCT, he or she must consistently read above 100 words correctly per minute on at least 700L passages for 6th graders and 800L for 7th and 8th graders. For most students to exceed standards, they averaged reading above 150 words correctly per minute on 1200 Lexile-based passages. The data indicated the number of correct responses that a student should average on the Maze assessment in order to predict whether the student will not meet, meet, or exceed standards. In 6th, 7th, and 8th grade, the students who averaged 15.9, 17.4, and 20.3 when reading at least 700L for 6th grade and 800L for 7th and 8th grade respectively, did not meet expectations on the CRCT, and those who averaged 27.7, 30.4, and 37 respectively on at least 1200L

passages, exceeded expectations. Even though the sample was small for the group of students taking the EOCT, the ORF assessment was the stronger predictor variable for a student's performance on the EOCT when compared to the Maze.

Summary of Findings

This study examined data from middle school students who had previously taken the CRCT and who were given ongoing progress monitoring assessments using the ORF and Maze. The researcher analyzed the relationship between three independent variables, the previous year's CRCT score and level, the average of twelve progress monitoring assessments, either the ORF or Maze, and the dependent variable, the predicted CRCT score and/or the EOCT. After 23 weeks of progress monitoring, the data supported that using the three independent variables one can make a substantial prediction regarding the future CRCT scores. The ORF along with the previous year's CRCT score and level were good predictors for the EOCT. The overall regression was significant for each grade level for each assessment with the ORF being a stronger predictor than the Maze.

CHAPTER V

DISCUSSION, IMPLICATIONS, AND RECOMMENDATIONS

The purpose of this study was to analyze the relationship between data gathered from middle school students using the progress monitoring tools, ORF and Maze, and the Georgia Criterion Referenced Competency Test to find a valid and reliable assessment for progress monitoring and predictor of a student's future performance on the CRCT and the EOCT. The Response to Intervention Model was implemented to study the data and identify the effectiveness of instruction for all students, the at-risk, general, and advanced population. The purpose of this chapter is to discuss the findings, derive overall conclusions, identify practical implications, and create recommendations for future research.

Discussion of Findings

This quantitative study was guided by three research questions. The discussion of the results will focus on how the major findings are relevant to education today. The discussion of each of the research questions help to demonstrate the effectiveness of two Response to Intervention tools, ORF and Maze assessment, in the language arts classrooms of middle school students.

The first research question asked which assessment was a valid and reliable progress monitoring tool in predicting reading achievement for a middle school student. For grades 6, 7, and 8, the ORF assessment was found to be a stronger predictor than the Maze according to the multiple linear correlation coefficients (R) and coefficient of determination (R^2) for the student's future reading performance on the Georgia CRCT

using a multiple regression analysis of the previous year's CRCT score, previous year's CRCT level, and a mean of twelve ORF or Maze assessments given over a 23 week period. When using the Maze in the multiple regression analysis, the overall regression was also significant, but not as strong as the ORF for all grade levels and all achievement levels; below standards, meeting standards, and exceeding standards.

For practical classroom use, the Maze was a more efficient progress monitoring tool to use because of the time involved in assessing a student one-on-one with the ORF. In an average classroom size, completing the ORF with an entire class took from one to two 70 minute class periods. Most teachers did not find that they could give up one to two days of instruction every other week; therefore, in practicality and statistically, the Maze was the more efficient assessment for a progress monitoring tool for all students in the normal educational process along with the ORF assessment every nine weeks. The Maze assessment was more teacher-friendly because it was given as a whole group assessment on the computer. The Maze assessment took only about 15 minutes from start to finish as compared to at least one 70 minute class period for the ORF. The teachers were taught how to give the Maze assessment using the computers and how to record it like the ORF. A free template was found that took out every 7^{th} word and replaced them with three choices. As the student read the passage, they selected one response from the computer generated drop down box. When time was up, the students clicked the box that read, "Check." The computer scored the number of correct choices. Some of the teachers taught the students how to record their own data on a scatter plot, and the teachers checked the students' responses and recorded the number of responses that were correct on their chart as well. The teachers liked this assessment and found value in it, and some wanted to know if they could do more practice with it. If the computers were working

correctly, three classrooms could complete the Maze in one class period. Because the students were choosing the correct words as they read silently, some of the teachers stated that they felt this measured reading achievement better than the ORF because it measured their comprehension as well (Fuchs & Fuchs, 1992). This finding coincided with research that indicated the Maze task to be useful with upper elementary, middle, and high school students who were fluent in oral reading (Espin, Wallace, Lembke, & Campbell, 2004; Jenkins & Jewell, 1993). Having the Maze on the computer allowed for more efficiency because the assessment was scored immediately, eliminated paper costs, and was sensitive to performance over time as stated by Fuchs and Fuchs (1992). They also stated that teachers and students preferred the Maze which was found in this study.

The second major finding was the strong regression model that the ORF and Maze yielded demonstrating the validity and reliability of both tools in predicting reading success or failure on the CRCT. The ORF and Maze assessments provided strong predictor variables along with the previous year's CRCT scaled score and level. Using the scatter plot to record the student's progress for each assessment provided a means to visually represent whether the student was making normal reading progress and gave a clear model for those invested in the student's educational process to predict whether the student would pass or fail the future CRCT. This finding was also indicated in other studies (Espin et al., 2007; Espin & Deno, 1993a, 1993b; Espin & Deno, 1995; Espin & Foegen, 1996; Fewster & MacMillan, 2002; Muyskens & Marston, 2006; Tichá, et al., 2007; Yovanoff, Duesbery, Alonzo, & Tindal, 2005).

The third finding was the validity and the reliability of the data for all students, including those below standards, meeting standards, and exceeding standards. The multiple regression models were strong for each group of students, and the Response to

Intervention model was a valid and reliable instrument to use to validate a student's reading performance. The collected and charted data were used frequently to determine the effectiveness of an intervention or to determine weaknesses within the student's educational process which was a vital component for teachers to incorporate based on the findings of Ardoin (2006). Students who were found to be at-risk or needing extra practice in literacy to fill in gaps in reading based on analysis of data were placed in Tier 2 of the RTI model following the research of Barnes and Harlacher (2008) and Gresham and others (2005). The scatter plots were used in RTI meetings to identify if Tier 2 interventions were working. Generally, if a student was struggling in other subject areas, the reading data would visually indicate a problem as well, such as reading below 100 words per minute. If progress monitoring data did not follow normal progression in reading, some students were referred for Special Education services according to Fuchs and Fuchs (2005). The data also validated students who needed testing accommodations.

Because of the frequent monitoring of reading progress, the data were used in meetings with parents, teachers, and administrators to make decisions to implement effective interventions (Ardoin, 2006), and one of the main goals was to establish "exemplary, scientifically based literacy practices within each classroom by focusing on environmental quality…a comprehensive curriculum and research-based literacy strategies" (Gettinger & Stoiber, 2007, p. 202). The data provided the general and advanced population information that they were making normal reading progress and gave a visual representation of the goal to reach based on valid and reliable research. Because the ORF assessments were given at the beginning of the school year, teachers were able to identify at-risk readers through the data provided by Fuchs and Deno (1982), Fuchs and others (1993), and Deno and others (2001). The data collected on the scatter

plots visualized the students's weaknesses and strengths, the students who were making normal reading progress, and a prediction of whether a student was going to pass or fail the CRCT as the research suggested (Christ & Silberglitt, 2007; Graney & Shinn, 2005; Silberglitt, 2006; Silberglitt & Hintze; 2007; Stecker, et al., 2008; Wiley & Deno, 2005).

Another major finding was the expertise that the teachers learned throughout the process by creating individualized plans for the students that demonstrated to all involved in the educational process that the student was progressing in reading. Beginning in August of 2008, the language arts teachers were trained in using the RTI progress monitoring tools, ORF and Maze assessments. Because of the depth of information, the teachers were given an overview, and then specific instructions were taught for the ORF assessment first. They were taught how to give the assessment and how to record it. In general, the teachers found the information gathered from the ORF assessment valuable in assessing the student's ability to read at grade level based on the Lexile leveled passages that correlated with the Lexiles of the CRCT. The ORF also allowed for quick identification of reading weaknesses and strengths and provided the teacher information regarding the instructional Lexile level of the students.

As the teachers gave three ORF assessments to each individual student to get a baseline score, many of them found the ORF assessments to be frustrating because of the time involvement, especially for those in the larger classrooms. All of the students were assessed by ORF each quarter. Students who were found to be at-risk students or those who were randomly selected were given the ORF assessment every other week by the teacher. Even though some of the teachers expressed frustrations regarding the amount of time and the problems associated with the students at their seats while the individual student was reading to the teacher, the teachers found value in the information. The

teachers learned to implement new ideas and to make instructional decisions based on the action research and the data that they gathered. The progress monitoring assessments became a critical component of teaching and learning as they connected curriculum, assessment, and instruction (Cobb, 2003; Hintze, Daly, & Shapiro, 1998; Stecker, Lembke, & Foegen, 2008; Taylor, 2004). The teachers began to look at the assessments as Guskey (2003) suggested. The assessments were useful to teachers and students in setting goals to reach. Secondly, the assessments were not an end but a checkpoint of learning because as the teacher made adjustments in instruction through feedback or interventions, the students were able to demonstrate growth again (Guskey, 2003).

The teachers found the Lexile-based passages to be a good measurement tool because most of the class sets of books in the library were Lexile-based. The teachers had previous knowledge of the Lexiles because of the information given with the CRCT scores. The teachers used the Lexiles to guide their instruction. Their questions and dialogue began to focus on the Lexile of the book that they were going to read in class and whether it was too difficult or too easy, and they used the information gathered to help the students pick a book to read independently. The teachers learned to match the Lexile reader measure to the Lexile text measure (MetaMetrics, 2008). The teachers gained confidence as they made educational decisions regarding students with reading weaknesses, and they worked together with other teachers and administrators to correctly place the students with a proper intervention using the RTI model. Several students at the beginning of the year were placed in Tier 2 based on the ORF baseline assessments because of poor fluency (less than 100 wcpm). Overall, the teachers saw value in the ORF assessment using the Lexile-based passages but found it time consuming, and as stated in the research, their preferred assessment was the Maze (Fuchs & Fuchs, 1992).

Using the Lexile-based assessments, the teachers related the ORF and Maze assessments to other data that were Lexile-based assessments, such as the CRCT and Lexile-based books. The teachers had discussions on the Lexile-based passages and the range of difficulty of the passages. They began to notice how the length of words, sentences, and paragraphs worked together to create the reading level of the passage. Because many of the passages used were from the Classics, such as *Little Women, Don Quixote, The Strange Case of Dr. Jekyll and Mr. Hyde,* and *Gone With the Wind,* they found them interesting, and some were motivated to try the Maze assessments themselves. Overall, the teachers saw the parallel between the Lexiles of the assessments, and the Lexiles provided by the CRCT resulting in more teacher confidence in using data analysis to guide instruction and to predict reading achievement.

The teachers gained confidence in their ability to teach because they could visually see evidence of effective instruction and could modify their instruction if the data demonstrated a regression away from the regression line. The teachers saw value in the students having a goal to reach because it provided intrinsic motivation. The students were measured according to their own personal growth, and they also learned to choose books based on the information from the assessments. Some students would actually cheer when they went above the regression line, and when they fell below, they knew they had to work harder. This self evaluation made the students assume responsibility in the learning process which was found important in the study of Palinscar and Brown (1984). They found that students need to be active participants in their learning and be goal-oriented and intentional as they read. The common Lexile-based ORF and Maze measurement coupled with the validity and the reliability of the assessments and the RTI model created a way for teachers to take ownership of the educational process of each

student because there was a definite individual reading goal for each student to reach and a way to visually follow the progress along the way.

The most promising finding was whether there was a tool that would be effective in providing a continuum to measure a student's reading progress (Wayman et al., 2007). Throughout the implementation of this study, the focus was on the reading progress for all students and finding a way to visually and effectively represent it. This goal was accomplished because the study focused not on labels and capabilities, but on the premise that if expectations are clear and high, all students will make progress in reading. The RTI model took the focus from the achievement and intelligent measurements to the relevant day-to-day classroom instruction being a more proactive and preventative approach (National Association of State Directors of Special Education (NASDSE), 2006). Using Lexile-based assessments that were teacher-friendly, cost effective, and valid and reliable, a continual (next year the school will continue from this past year) and flexible (the data were analyzed and guided instruction) reading profile of each student at North Hall Middle School has been created. Because of the predictive power, the cost effectiveness, and the commonality of the Lexile measurements in multiple areas, such as books and assessments, schools can implement these tools from elementary to high school and have an uninterrupted reading progression profile of each student.

Utilizing the ORF and Maze created an effective and efficient progress monitoring tool that could connect all grade levels. Because the research supported both assessments especially in the elementary years and the ongoing research, along with this study, validated the reliability and the validity in the middle grades, these two progress monitoring tools hold a wealth of information to determine a student's reading ability (Wayman et al., 2007). Also, connecting the grade levels through ORF and Maze

progress monitoring assessments that were Lexile-based to the Lexile-based standardized assessments, such as the CRCT, EOCT, ITBS, SAT, and ACT provided a common measurement tool where educators gained confidence as they implemented the RTI model to make instructional decisions to create a seamless and effective instructional process for each student. The multiple regression analysis of the EOCT scores gave insight into future possibilities of study for predictor variables of the EOCT. Research continues to substantiate that fluency is the strongest predictor (Wayman et al., 2007).

Implications

Practical implications were derived from this study that could be beneficial for those who are involved in the educational process. The results of this study demonstrated the importance of reading out loud and improving fluency. Teachers need to find ways to practice fluency in the classroom on a regular basis. Parents need to have their children read out loud on a regular basis. Students can improve their reading achievement by practicing fluency while reading (Salinger, 2003).

Another practical implication was the necessity to have a goal for each student that was visual, attainable, and yet rigorous. Students need expectations Palinscar & Brown, 1984). Table 21 demonstrates an unintended finding in this study. As stated earlier, some of the data could not be used because of incorrect monitoring levels. Some students were monitored at levels that were too low (unchallenging), and some were monitored too high (rigorous). The data that were collected and analyzed for this study only used the students who were monitored with the correct methodology. Even though all data were not used for data analysis in this study, the data were still used to monitor reading progress across all grade levels. Table 21 represents a comparison the percentages of all students' data and those monitored with the correct level. The student's

CRCT scores from 2008 to 2009 were compared and an increase was indicated by a (+) and a decrease by a (-). Theoretically, the data should decrease in Level 1 and 2 and increase in Level 3 from the beginning to the end of the year. This goal was achieved in grade 6 and grade 8. Sixth grade had the fewest students' data that were dropped due to using incorrect methodology. Most of the students who were monitored incorrectly were monitored with passages on their current reading level, not their goal level except for some of the 8th grade students who were monitored with passages that were too high. Could the big increase in students scoring a Level 3 in 8th grade reflect that students achieve what is expected of them?

Table 21

Comparison of All Participants and Those Monitored with Correct Methodology (CM)

	Grade	1s (Below Standard)	2s (Meets Standard)	3s (Exceeds Standard)
All	6th	-21.4%	-22.5%	+61.9%
CM	6th	-38.4%	-24.8%	+78.6%
All	7th	+71.4%	+7.3 %	-43.5%
CM	7th	+71.4%	+12.7%	-27.8%
All	8th	-50%	-19.9%	+116.2%
CM	8th	-50%	-22.6%	+226.7%

The overall implication was the need to provide progress monitoring to all students. The teachers used the data to identify students who were at-risk in reading and to use the data to guide instruction making sure that the students were taught on an instructional level. The teachers and administrators used the data within the RTI model to raise students to Tier 2 or Tier 3 depending on the data that were gathered and to write Individual Educational Plans (IEP). Utilizing progress monitoring within the RTI model

provided intrinsic motivation to the teachers showing them that their instruction was effective and to the students providing them a visual goal to reach. The scatter plots were visuals for teachers, parents, and students that reflected either a problem with reading progress, normal reading progress, or advanced reading progress. Focusing on every student allowed the teachers to have data to create differentiated instruction.

Limitations

Throughout this study, several limitations surfaced. The findings of this study were not generalizable to other schools or other subjects but were specific to the school setting and population of North Hall Middle School.

One limitation was the effect of teacher instruction on the students as well as the way in which a teacher implements CBM. Whether the progress monitoring is completed as part of the classroom instruction or whether a teacher takes the student apart from the class to listen to them read should be taken into consideration.

Another limitation was the student's history. Some outliers within the data were Level 1s who scored too high or Level 3s scoring too low. These areas need further study.

The instrumentation was a threat to validity. There is limited research on the comparison of Lexile-based probes and the readabilities of other probes. Also, sometimes there was a problem with the server for the school computers, and the teachers had to reschedule their Maze assessment. Even though these were normal school circumstances, these circumstances do affect the attitudes of the teachers as well as the time involvement and could possibly affect outcomes.

The random sampling was a threat to validity associated with the oral reading assessments. A larger sampling may produce stronger findings and yield more precise expectations for growth rates per week; therefore, larger samplings should be studied.

As stated earlier, even though the teachers were trained together and the teachers were checked periodically to ensure there were no problems, this study had many concepts from reading to mathematical understandings of the regression lines and scatter plots. The way the different teachers understood and implemented the multiple steps throughout the process could have created some biases to the data, such as some teachers completed the ORFs in the class, and some took the students in the hallway. Another difference that may have caused a threat to validity was the particular time that the Maze was given. Even though every class took the Maze the same week during the language arts period, some students received the test the first period of the day; whereas, some took the assessments the last period of the day. This time factor may have played a role in the ways that the students performed. Not only did the time factor pose a threat to the validity, but also the particular computer lab that was used for the assessments. There were three computer labs used to give the Maze assessments. The three labs were on the particular halls of each grade level, 6^{th}, 7^{th}, and 8^{th} grade. The computers in the 7^{th} grade lab posed a little more problem than the other labs because they were sometimes slow. The extra time that they caused created some frustrations. Also, the school's county log in procedures for the computers changed halfway through the year which created some issues when students did not know their passwords or usernames which added to the time factor. Other educational concerns that may be a threat to the validity were making up the assessments when students missed them due to absences. Sometimes these had to be made up during class on the classroom computer. Other educational concerns were the subject content of the probes that were used and whether it was interesting to the students or not. To overcome this concern, multiple passages were used with different subject content, but there was still the risk of a student finding interest in one over another. Also,

some of the probes had more characters' names in them than others. This threat to validity was due to the issue of reading unknown words because sometimes the names followed no phonetic rules and caused some students to pause more than necessary. These educational concerns are potential threats that teachers must address on a day-to-day basis in classroom situations. Overcoming these obstacles in the classroom plays a part on the impact of student learning.

Other threats to validity were those dealing with human error. Every other week the teachers collected the data for the Maze and put the data in the school's computerized gradebook, *Infinite Campus.* First of all, the teachers collected the data by hand on a chart and then recorded the data in the computer. Each teacher had four classes and taught from 75 to 100 students; therefore, each teacher recorded approximately 100 pieces of data every other week on a chart and in the computer creating the possibility of entering a wrong piece of data which was not seen in the data collection process. Human error was also a possibility in the administration of the ORF, such as hearing a word incorrectly, marking a word incorrectly, or counting incorrectly the number of words read. At the beginning of this study, some of the teachers found some typo errors in some of the probes which had to be corrected. In education, human error will occur which poses a threat to validity. Hopefully, as a teacher recognizes the potential of human error, he or she will be cautiously reflective of the educational process and will make changes to address and possibly eliminate the shortcomings.

Recommendations

This study found substantial information, but limitations in some areas suggested a need for future studies. Recommendations for future studies would be as follows:

1. An analysis of variance in the data to identify specific factors that may also have an influence on the study, such as teacher instruction, student background, and specific subgroups (students with disabilities, students who speak English as their second language, and gifted students) may suggest factors that influence reading progress. Some of the outliers in the regression models were of students who were at-risk but were making progress above the norms, and some of the outliers of advanced level were not making as much progress. A more in depth study of possible factors would be beneficial to see if the variability is due to student behavior, teacher procedures or instruction, student history, or type of instrument. Also, it should be noted the at-risk population that was monitored every other week became comfortable with fluency checks, but students who were at the average or above average levels demonstrated more nervousness when reading their quarterly fluency checks, such as wringing hands and more mistakes. A study would be helpful to see the different reactions of students in classrooms where fluency was taught on a regular basis.
2. Future research needs to study students at the same level who are monitored using different levels to see if the monitoring has an impact on the results.
3. Further studies would be helpful that compares Lexile-based passages to other leveled passages as the progress monitoring tools to see if reading in the Lexile ranges are more effective measurements.
4. A longitudinal study that follows students throughout several years would be helpful to see if the reading profile is consistent. Will a high school student

still be able to follow a regression model? Will an elementary student have the same success?

5. A more focused study of predictor variables of students taking high school credit in middle school would help to see the particular factors involving a student's success with higher level material.
6. Another study that would hold a wealth of information is one of high school students that analyzes how fluency affects the EOCT scores and other subject areas as well as the SAT and ACT scores.
7. This study provided valuable material that the researcher's school is continuing to use to monitor reading progress. Some teachers find oral reading fluency checks too time consuming, but the results suggested the ORF assessments hold valuable information for each student at any level. A further study would be helpful to analyze the effects of oral reading fluency using a larger sampling and whether the time factor involved is time well spent.

Summary

Although education is consistently changing, fluency remains a constant that should be assessed and taught to ensure that a student's fluency does not interfere with his or her ability to comprehend because fluency is a reliable predictor of reading ability (Stanovich, 1991). A way for an educator to identify problems in reading is to use progress monitoring with the ORF or the Maze assessment. Espin and Foegen (1996), Espin and others (2007), Muyskens and Marston (2006), and Tichá, and others (2007) found that oral reading fluency and the Maze were valid and reliable in predicting reading success. Rasinski and others (2005) stated that future studies were needed with secondary students to determine the significance of fluency on reading comprehension which this

study answered through identifying the overall significance of using ORF and Maze through multiple regression. Silberglitt and Hitze (2007) suggested that future studies should consider differential growth rates across generalizable samples which this study accomplished by focusing on the groups, students below standards, students who met standards, and students who exceeded standards in the regular language arts classroom. Silberglitt (2006) also suggested that research was needed in later grades to test the predictive validity of CBM. This study found the CBM tools, ORF and Maze assessment, were significant predictors of a student's future performance on the Georgia's Criterion Referenced Competency Test and even though not as strong, they have some validity and reliability in being predictors for the 9th Grade Literature and Composition End of Course Tests. Wayman and others (2007) summarized multiple studies and suggested that a CBM of reading has the potential to develop "a seamless and flexible system of progress monitoring that could be used….from kindergarten to Grade 12" (p. 116). This study of middle school students provided a model that could be a seamless and flexible system. Eliot and Fuch's study (1997) identified important steps that were consistent with this study. These steps were 1) progress monitoring should be quick to administer, 2) have adequate reliability and validity, 3) demonstrate what a student is learning, 4) help in developing interventions, and 5) be sensitive to improvements in academic performance to identify the effectiveness of interventions.

The RTI model along with the progress monitoring tools, ORF and Maze, provided a model that educators can make decisions based on valid and reliable data for all students. The assessments were not an end, but the beginning of new goals to be reached which guided the instruction to new and refined levels of knowledge based on the curriculum or standards (Guskey, 2003). Throughout history, educators have

influenced how curriculum, assessment, and instruction are implemented in the teaching and learning cycle, and effective educators demonstrate the importance of integrating and balancing the three components (Jones, 2008).

References

Allington, R. L. (1983). Fluency: The neglected reading goal in reading instruction. *The Reading Teacher, 36,* 556-561.

Anderson, R., Hiebert, E., Scott, J., & Wilkinson, I. (1985). *Becoming a nation of readers: the report on the commission on reading.* Champaign, IL: University of Illinois, Center for the Study of Reading.

Ardoin, Scott P. (2006). The response in response to intervention: Evaluating the utility of assessing maintenance of intervention effects. *Psychology in the Schools, 43*(6), 713-725. Retrieved from www.interscience.wiley.com.

Ardoin, S. P., Suldo, S. M., Witt, J. C., Aldrich, S., & McDonald, E.. (2005). Accuracy of readability estimates' predictions of CBM performance. *School Psychology Review, 20,* 1-22.

Ary, D., Jacobs, L. C., Razavieh, A., & Sorensen, C. (2006). *Introduction to research in education* (7th ed.). Belmont, CA: Thomson Wadsworth.

Baker, S. K. & Good, R. (1995). Curriculum-based measurement of English reading with bilingual Hispanic students: A validation study with second-grade students. *School Psychology Review, 24,* 561-578.

Barrington, Jackie. (2003). Oral reading fluency presentation. Vancouver: Simon Fraser University. Retrieved from http://www.cenmi.org/msdbLIO/downloads/Literacy/ReadingWritingBraille/ReadingFluency.doc.

Barnes, A. C. & Harlacher, J. (2008). Clearing the confusion: Response-to Intervention as a set of principles. *Education and treatment of children, 31*(3), 417-431.

Biancarosa, G., & Snow, C. E. (2004). Reading next-A vision for action and research in middle and high school literacy: A report from Carnegie Corporation of New York. Washington, DC: Alliance for Excellent Education.

Bell, S.M., Zielger, M., & McCallum, S. (2004). What adult educators know compared with what they say that they know about providing research-based reading instruction. *Journal of Adolescent and Adult Literacy, 47*(7), 542-563.

Bush, President G. W. (2007, January). President Bush Delivers State of the Union Address [Full Manuscript]. Washington, DC: The White House. Retrieved from http://www.whitehouse.gov/news/releases/2007/01/20070123-2.html

Bradley-Klug, K. L., Shapiro, E. S., Lutz, K. G. & DuPaul, G. J. (1998). Evaluation of oral reading rate as a curriculum-based measure within literature-based curriculum. *Journal of School Psychology, 36,* 183-197.

Brown, R. B. (2007). An examination of the relationship between participation in school-sponsored extracurricular activities and student achievement. Ed.D. dissertation, The University of Alabama, United States -- Alabama. Retrieved June 24, 2009, from Dissertations & Theses: Full Text database. (Publication No. AAT 3313703).

Brown-Chidsey, R., Davis, L., & Maya, C. (2003). Sources of variation in curriculum-based measure of silent reading. *Psychology in the Schools, 40,* 363-377.

Brown-Chidsey, R., Johnson, P. Jr., & Fernstrom, R. (2005). Comparison of grade-level controlled and literature-based maze CBM reading passages. *School Psychology Review, 34,* 387-394.

Brynildssen, S. (2001, December). *Highlights of reading research in the labs and centers of the U.S. Department of Education.* (ERIC Document Reproduction Service No. ED459425).

Brynildssen, S. (2002). *Recent reading initiatives: Examples of national state and professional organizations' efforts.* (ERIC Document Reproduction Service No. ED469927)

Byrnes, M. (2008). *Taking sides: Clashing views in special education,* 3rd ed. Dubuque: McGraw-Hill Contemporary Learning Series.

Christ, T. J. (2006, Winter). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35*(1), pp. 128-134.

Christ, T. J., & Silberglitt, B. (2007, March). Estimates of the standard error of measurement for curriculum-based measures of oral reading fluency. *School Psychology Review, 36*(1), pp. 130-147.

Cobb, C. (2003). Effective instruction begins with purposeful assessments. *The Reading Teacher, 57*(4), 386-389.

Compton, D. L., Appleton, A. C., & Hosp, M. K. (2004). Exploring the relationship between text-leveling systems and reading accuracy and fluency in second-grade students who are average and poor decoders. *Learning Disabilities Research & Practice, 19,* 176-184.

Coyne, M. D. & Harn, B. (2006). Promoting beginning reading success through meaningful assessment of early literacy skills. *Psychology in the Schools, 43,* 33-43.

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

Davis, G. & Rimm, S. (2004). Education *of the Gifted and Talented* (5th ed.). Boston, MA: Pearson Education, Inc.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 57,* 219-232.

Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30,* 507-524.

Dooley, C. (1993, April). The challenge: Meeting the needs of gifted readers. (cover story). *Reading Teacher, 46*(7), 546.

Dudley-Marling, C. & Paugh, P. (2004). Tapping the power of student voice through whole language practices. *Reading & Writing Quarterly,* (20), 385-399.

Dunn, E. K., & Eckert, T. L. (2002). Curriculum-based measurement in reading: A Comparison of similar versus challenging material. *School Psychology Quarterly, 17,* 36-45.

Elliott, S.N., & Fuchs, L.S. (1997). The utility of curriculum-based measurement and performance assessments as alternatives to traditional intelligence and achievement tests. *School Psychology Review,* (26), 224-233.

Espin, C. A., & Deno, S. L. (1993a). Content-specific and general reading disabilities of secondary students: Identification and educational relevance. *The Journal of Special Education, 27,* 321-337.

Espin, C. A., & Deno, S. L. (1993b). Performance in reading from content area text as an

indicator of achievement. *Remedial and Special Education, 14,* 47-59.

Espin, C.A. & Deno, S. L. (1994-1995). Curriculum-based measures for secondary students: Utility and task specificity of text-based reading and vocabulary measures for predicting performance on content-area tasks. *Diagnostique, 29*, 121-142.

Espin, C. A., Deno, S. L., Maruyama, G., & Cohen, C. (1989, March). *The Basic Academic Skills Samples(BASS): An instrument for the screening and identification of children at risk for failure in regular education classrooms.* Paper presented at the National Convention of the American Educational Research Association, San Francisco, CA.

Espin, C. A., & Foegen, A. (1996). Validity of general outcome measures for predicting secondary students' performance on content-area tasks. *Exceptional Children, 62,* 497-514.

Espin, C. S., Wallace, T., Lembke, E. S., & Campbel, H. (2004). *Preparing secondary students for state standards tests: Monitoring reading with curriculum-based measures.* Paper presented at the annual meeting of the Council for Exceptional Children, New Orleans, LA.

Espin, C. A., Wallace, T., Lembke, E., Campbell, H., & Long, J. (2007). *Creating a progress measurement system in reading for secondary students: Monitoring progress towards meeting high stakes standards.* Manuscript submitted for publication.

Everatt, J., Bradshaw, M. F., & Hibbard, P.B. (1998). Individual differences in reading and eye movement control. In G. Underwood (Ed.), *Eye Guidance in Reading and Scene Perception* (pp. 223-242). Oxford, England: Elsevier.

Everatt, J., & Underwood, G. (1994). Individual differences in reading subprocesses: Relationships between reading ability, lexical access, and eye movement control. *Language and Speech, (37),* 283-297.

Feldhusen, J. F., & Wyman-Robinson, A. J. (1980). Super Saturday: Design and implementation of Purdue's special program for gifted children. *Gifted Child Quarterly. 24,* 15-21.

Fewster, S., & MacMillan, P. D., (2002). School-based evidence for the validity of curriculum-based measurement of reading and writing. *Remedial and Special Education, 23,* 149-156.

Foertsch, M. (2003). *A study of reading practices, instruction, and achievement in District 31 schools.* Retrieved March 2, 2008, from the North Central Regional Educational Laboratory Web site: http://www.ncrel.org.

Foorman, B., Francis, D., Beeler, T., Winikates, D., & Fletcher J. (in press). Early interventions for children with reading problems: Study designs and preliminary findings. *Learning Disabilities: A Multi-disciplinary Journal.*

Fry, E., Kress, J. & Fountoudidis, D. (1993). *The reading teacher's book of lists,* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Fuchs, L. S. (1986). Monitoring progress among mildly handicapped pupils: Review of current practice and research. *Remedial and Special Education, 7*(5), 5-12.

Fuchs, L. S., & Deno, S. L. (1982). *Developing goals and objectives for educational programs.* Washington, DC: American Association of Colleges for Teacher Education.

Fuchs, L. S., & Deno, S. L. (1992). Effects of curriculum within curriculum-based measurement. *Exceptional Children, 58,* 232-243.

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

Fuchs, L. S. & Fuchs, L. S. (1990). [Relations among the Reading Comprehension subtest of the Stanford Achievement Test and maze, recall, question answering, and fluency measures]. Unpublished data.

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

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

Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for Reconceptualizing the identification of learning disabilities. *Learning Disabilities Research Practice, 13*(3), 204-219.

Fuchs, L. S. & Fuchs D. (2005). Responsiveness-to-intervention: A blueprint for practitioners, policymakers, and parents. *Teaching Exceptional Children, Sept/Oct 2005,* 57-61.

Fuchs, L. S., Fuchs, D., & Hamlett, C.L. (1989). Effects of instrumental use of curriculum-based measurement to enhance instructional programs. *Remedial and Special Education, 10,* 43-52.

Fuchs, L. S., Fuchs, D., & Hamlett, C.L., & Ferguson, C. (1992). Effects of expert system consultation within curriculum-based measurement using a reading maze task. *Exceptional Children, 58,* 436-451.

Fuchs, L. S., Fuchs, D., Hamlett, C.L., Waltz, L., & Germann, G. (1993). How much growth can we expect? *School Psychology Review, 22,* 27-48.

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

Fulton, O. (2004). *The Framework for Reading* [PowerPoint]. MetaMetrics, Inc., Retrieved from www.cesa7.k12.wi.us/content/literacy/**Lexile**WIpresentation1104.pdf

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

Georgia Department of Education. (2008a). *2008 Adequate yearly progress rate report.* Retrieved from http://public.doe.k12.ga.us/ayp2008 .

Georgia Department of Education. (2008b). *Language Arts standards.* Retrieved from http://www.georgiastandards.org .

Georgia Department of Education. (2008c). *Standards, instruction and assessment (Testing).* Retrieved from http://gadoe.org/ci_testing.aspx.

Georgia Department of Education. (2008d). *What Georgia educators need to know about Georgia's testing program.* Retrieved from http://public.doe.k12.ga.us/DMGetDocument.aspx/Final%20-%20testing%20newspaper.pdf?p=4BE1EECF99CD364EA5554055463F1FBBF5D074D5FB1F2CAEB3B63B3ECB220CDD26C2114F3C57D8D2040FD88C56B817B3&Type=D

Georgia Department of Education. (2009, August). *Validity and reliability for the 2008 Criterion-Referenced Competency Test.* Assessment Research and Development Division of the Georgia Department of Education.

Georgia Department of Education. (2009, September). *Validity and reliability for the 2009 Criterion-Referenced Competency Test.* Assessment Research and Development Division of the Georgia Department of Education.

Gettinger, M., & Stoiber, K. (2007). Applying a Response-to-Intervention Model for Early Literacy Development in Low-Income Children. *Topics in Early Childhood Special Education, 27*(4), 198-213.

Golden, A. (1999). *Memoirs of a Geisha.* New York: Random House Publishers.

Good, R. H., III, Simmons, D. C., & Kameenui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills f or third-grade high stakes outcomes. *Scientific Studies of Reading, 5*, 257-288.

Goodman, Y.M., Watson, D.J. & Burke, C.L. (1987). *Reading miscue inventory.* Katonah, NY: Richard C. Owen.

Graney, S., & Shinn, M. (2005, Spring). Effects of Reading Curriculum-Based Measurement (R-CBM) teacher feedback in general education classrooms. *School Psychology Review*, *34*,(2), 184-202.

Graves, A. W., Plasencia-Peinado, J., Deno, S. L., & Johnson, J. R. (2005). Formatively evaluating the reading progress of first-grade English Learners in multiple-language classrooms. *Remedial and Special Education, 26*, 215-225.

Groff, P. (2001). *Decodable words versus predictable text.* Strasburg, VA: The National Right To Read Foundation. http://www.nrrf.org.

Grossen, B. (1997). 30 *Years of research: What we now know about how children learn to read.* Santa Cruz, CA: Center for the Future of Teaching and Learning.

Gunning, T.G. (1995). Word Building: A strategic approach to the teaching of phonics.

The Reading Teacher, 48, 484-488.

Guskey, T.R. (2003). How classroom assessments improve learning. *Educational Leadership, 60*(5), 6-11.

Gutek, G.L. (2005). *Historical and philosophical foundations of education: A biographical Introduction,* (4th ed.). Upper Saddle River, NJ: Pearson Education.

Harris, T. L., & Hodges, R. E. (1995). *The Literacy Dictionary.* Newark, DE: International Reading Association.

Hartman, J. M. & Fuller, M. L. (1997). The development of curriculum-based measurement norms in literature-based classrooms. *Journal of School Psychology, 35,* 377-389.

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

Hintze, J. M., Callahan, J. E., Matthews, W. J., Williams, S. A. S. & 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., & Christ, T. J. (2004). An examination of variability as a function of passage variance in CBM progress monitoring. *School Psychology Review, 33,* 204-217.

Hintze J., Daly III E., & Shapiro, E. (1998). An investigation of the effects of passage difficulty level on outcomes of oral reading fluency progress monitoring. *School Psychology Review* [serial online] *27*(3):433.

Hintze, J. M., Owen, S. V., Shapiro, E. S. & Daly, E. J. (2000). Generalizability of oral

reading fluency measures: Application of G theory to curriculum-based measurement. *School Psychology Quarterly, 15,* 52-68.

Hintze, J. M. & Pelle Petitte, H. A. (2001). The generalizability of CBM oral reading fluency measures across general and special education. *Jouranl of Psychoeducational Assessment, 19,* 158-170.

Hintze, J. M., & Shapiro, E. (1997). Curriculum-based measurement and literature-based readings: Is curriculum-based measurement meeting the needs of changing reading curricula? *Journal of School Psychology,* (35), 351-375.

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

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

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

Hixson, M. D., & McGlinchey, M. T. (2004). The relationship between race, income, and oral reading fluency and performance on two reading comprehension measures. *Journal of Psychoeducational Assessment, 22,* 351-364.

Horry County Schools. (2008). *High school SAT/ACT coaches roles & responsibilities.* Conway, SC. Retrieved from http://www3.hcs.k12.sc.us/.

Hosp, M., & Hosp, J. (2003). Curriculum-based measurement for reading, spelling, and math: How to do it and why. *Preventing School Failure, 48*(1), 10-17.

Hudson, R., Lane, H., & Pullen, P. (2005, May). Reading fluency assessment and instruction: What, why, and how? *The Reading Teacher, 58*(8), 702-714.

Hudson, R. F., Mercer, C. D., & Lane, H. B. (2000). *Exploring reading fluency: A paradigmatic overview.* Unpublished manuscript, University of Florida, Gainesville.

Ikeda, M. J., Grimes, J., Tilly III, W. D., Allison, R., Kurns, S., & Stumme, J. (2002). Implementing an intervention-based approach to service delivery: A case example. In M.R. Shinn, H. M. Walker & G. Stoner (Eds.), *Interventions for academic and behavior problems II: Preventive and remedial approaches* (pp. 53-69). Bethesda, MD: National Association of School Psychologists.

Individual With Disabilities Education Improvement Act of 2004 (IDEA). Retrieved from http://idea.ed.gov.

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

Johns, J.L. (1993). *Informal reading inventories.* DeKalb, IL: Communitech.

Johnston, P.H., & Winograd, P.N. (1985). Passive failure in reading. *Journal of Reading Behavior, 4,* 279-301.

Jones, Donna L. (2008) "The Teaching and Learning Cycle: Integrating Curriculum, Instruction, and Assessment," *Christian Perspectives in Education*: Vol. 2: Iss. 1, Article 2. Retrieved from http://digitalcommons.liberty.edu/cpe/vol2/iss1/2.

Klein, J. R., & Jimerson, S. R. (2005). Examining ethnic, gender, language, and socioeconomic bias in oral reading fluency scores among Caucasian and Hispanic students. *School Psychology Quarterly, 20,* 23-50.

Kranzler, J. H., Miller, M. D. & Jordan, L. (1999). An examination of racial/ethnic and gender bias on curriculum-based measurement of reading. *School Psychology Quarterly, 14*, 327-342.

Lyons, G.R. (1994). *Frames of reference for the assessment of learning disabilities: New views on measurement methods.* Baltimore, MD: Paul H. Brookes.

Lyons, G.R. (1995). Toward a definition of dyslexia. *Annals of Dyslexia, 45*, 3-27.

Lyon, G. R. (2001). *Measuring success: Using assessments and accountability to raise student achievement.* Subcommittee on Education Reform Committee on Education and the Workforce,U. S. House of Representatives. Washington, D.C. from http://www.nrrf.org/lyon_statement3-01.htm.

Marchand-Martella, N., Martella, R., & Przychodzin-Havis. (2007). *The research base and validation of SRA's Corrective Reading Program.* Columbus, OH: SRA/McGraw Hill.http://www.sraonline.com

Marcon, R.A. (1995). *Moving up the Grades: Relationship between Preschool Model and Later School Success.* University of Florida: Early Childhood Research & Practice (EDRP). Retrieved from http://ecrp.uiuc.edu/v4n1/marcon.html.

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

Marston, D.B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it? In M.R. Shinn (Ed.), *Curriculum-based Measurement: Assessing Special Children*, 18-78. New York: Gilford Press.

Marston, D. M., Lau, P., & Carter, A. (2003). Problem-solving model for decision

making with high-incidence disabilities : The Minneapolis experience. *Learning Disabilities Research & Practice, 18,* 187-200.

McGlinchey, M. T. & Hixon, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review, 33,* 192-203.

MetaMetrics, Inc. (2006). Linking assessment with reading instruction. *The Lexile Framework for Reading.* Retrieved from http://www.lexile.com/PDF/TheLexileFrameworkforReading.pdf.

MetaMetrics, Inc. (2007). The Lexile framework for reading technical report. Retrieved from www.ade.state.az.us/azreads/lexile/Research/LFforReadingTechnical Report_042007.pdf.

MetaMetrics, Inc. (2008). FAQ. *The Lexile Framework for Reading.* Retrieved from http://www.lexile.com/DesktopDefault.aspx?view=ed&tabindex=6&tabid=18.

Miller, Congressman George. (2001). *Full history of the No Child Left Behind Bill.* Retrieved from http://georgemiller.house.gov/eseainfo.html.

Mitchell, J.V. (1985). *The ninth mental measurements yearbook.* Lincoln, NC: University of Nebraska Press.

Morgan, S. K., & Bradley-Johnson, S. (1995). Technical adequacy of curriculum-based measurement for Braille readers. *School Psychology Review, 24,* 94-103.

Muyskens, P., & Marston, D. B. (2006). *The relationship between curriculum-based measurement and outcomes on high-stakes tests with secondary students.* Minneapolis Public Schools. Unpublished manuscript.

National Association of State Directors of Special Education. (2006). *Response to*

Intervention: Policy considerations and implementation. Alexandria, VA: NASDSE, Inc.

National Center for Education Statistics (NAEP). (2008). The *Nation's Report Card.* Washington, DC: U. S. Department of Education. Retrieved from http://nces.ed.gov/nationsreportcard/.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read*. Washington, D.C.: National Institute of Child Health and Human Development. http://www.nationalreadingpanel.org.

Olinghouse, N., Lambert, W. & Compton, D. (2006). Monitoring children with reading disabilities' response to intervention: Are there differences between intervention aligned and general skill progress monitoring assessments?. *Exceptional Children 73*(1), 90-107.

Olson, R. K., Kliegl, R., Davidson, B. J., & Foltz, G. (1985). Individual differences and developmental differences in reading disability. In G. MacKinnon & T. G. Waller (Eds.), *Reading Research: Advances in Theory and Practice* (pp. 1-64). New York: Academic Press.

Palinscar, A. S. & Brown, A. L. (1984, Spring). Reciprocal teaching of comprehension-fostering and Comprehension-monitoring activities. *Cognition & Instruction, 1*(2), 117-175.

Pikulski, J., (1994). Preventing reading failure: A review of five effective programs. *The Reading Teacher 48*(1), 30-39.

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). *Listening to Children Read Aloud.* Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. (1995). *Listening to Children Read Aloud.* Washington, DC: Office of Educational Research and Improvement, U.S. Department of Education.

Pittsburgh Public Schools. (2007-2008). *Data Analysis/Decision making and MAZE comprehension training* [PowerPoint]. Retrieved from http://www.pps.k12.pa.us/1431101271026276 7/lib/1431101271026276 7/Data_Analysis_and_MAZE.ppt.

Poncy, B. C., Skinner, C. H., & Axtell, P. K. (2005). An investigation of the reliability and standard error of measurement of words read correctly per minute using curriculum-based measurement. *Journal of Psychoeducational Assessment, 23,* 326-338.

Powell-Smith, K. A. & Bradley-Klug, K. L. (2001). Another look at the "C" in CBM: Does it really matter if curriculum-based measurement reading probes are curriculum-based? *Psychology in the Schools, 38,* 326-338.

Quatroche, D., (1999). *Helping the underachiever in reading.* Bloomington, IN: ERIC Clearinghouse on Reading English and Communication.

Rasinski, T., Padak, N., McKeon, C., Wilfong, L., Friedauer, J., & Heim, P. (2005, September 1). Is reading fluency a key for successful high school reading? *Journal of Adolescent and Adult Literacy, 49*(1), 22. (ERIC Document Reproduction Service No. EJ719844).

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological Bulletin, (85),* 618-660.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological bulletin, (124), 372-422.

Reschly, D. J. (2005). Learning disabilities identification: Primary intervention, secondary intervention, and then what? *Journal of Learning Disabilities, 38,* 510-515.

Riley-Heller, N., Kelly-Vance, L., & Shriver, M. (2005). Curriculum-based measurement: Generic Vs. curriculum-dependent probes. *Journal of Applied School Psychology, 21*(1), 141-162.

Salinger, T. (2003). Helping older, struggling readers. *Preventing school failure, 47*(2), 79-85.

Samuels, S. J. (1988). Decoding and automaticity: Helping poor readers become automatic at word recognition. *The Reading Teacher, 41,* 756-760.

Scarborough, H.S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97-110). New York: Guilford.

Schreiber, P. A. (1980). On the acquisition of reading fluency. *Journal of Reading Behavior*, 7, 177-186.

Schulte, A. C., Villwock, D. N. , Wichard, M., & Stallings, C. F. (2001). High stakes testing and expected progress standards for students with learning disabilities: A five-year study of one district. *School Psychology Review, 30,* 487-506.

Share, D.L., & Stanovich, K.E. (1995). Cognitive processes in early reading developments: Accommodating individual differences in the acquisition of learning to read. *Reading Research Quarterly, Issues in Education: Contributions from Educational Psychology, 1,* 1-57.

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

Shinn, M.R. (Ed.). (1998). *Advanced applications of curriculum-based management.* New York: Guilford.

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

Shin, M. R., Gleason, M. M., & Tindal, G. (1989). Varying the difficulty of testing Materials: Implications for curriculum-based measurement. *The Journal of Special Education, 23,* 223-233.

Shriver, E. K. (2007, August). *Overview: Establishment and history.* Washington, D.C.: National Institute of Child Health and Human Development. Retrieved from http://www.nichd.nih.gov/about/overview/history/.

Sibley, D., Biwer, D., & Hesch, A. (2001, April). Establishing curriculum-based measurement oral reading fluency performance standards to predict success on local and state tests of reading achievement. Paper presented at the annual meeting of the National Association of School Psychologists, Washington, D.C.

Silberglitt, B. (2006). Relationship of reading fluency assessment data with state accountability test scores: A longitudinal comparison of grade levels. *Psychology in the Schools, 43*(5).

Silberglitt, B., & Hintze, J. (2007). How much growth can we expect? A conditional analysis of R-CBM growth rates by level of performance. *Exceptional Children 74*(1). 71-85.

Smith, C. (2003). The use of balanced instruction in language arts. *Topical Bibliography and Commentary.* (ERIC Document Reproduction Service No.

ED480636).

Snow, C. E. Burns, M.S., & Griffin, P. (1998). *Preventing reading difficulties in young children.* Washington, DC: National Research Council.

Speece, D. (2005, Nov/Dec). Hitting the moving target known as reading development: Some thoughts on screening children for secondary interventions. *Journal of Learning Disabilities, 38*(6), pp. 487-494.

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

Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of the learning to read. *Research Reading Quarterly, 21*, 360-407.

Stecker, P., Lembke, E., & Foegen, A. (2008). Using progress-monitoring data to improve instructional decision making. *Preventing school failure, 52*(2), 48-59.

Stenner, J. (2006). *Lexile 101: Readers, Texts and Comprehension* [PowerPoint]. MetaMetrics, Inc., Retrieved from 0-www.**lexile**.com.iii.ocls.info/PDF/PREPS%20opening%20session%202.22.06.pdf.

Stenner, J., Sanford, E., Burdick, D., & Burdick, H. (2006). How accurate are Lexile text measures? *Journal of Applied Measurement,* 7(3), 307-322, MetaMetrics, Inc.

Stevenson, R. L. (1993). *The strange case of Dr. Jekyll and Mr. Hyde.* Bradford, England: Waternill Press.

Taylor, G. M. (2004). Linking *formative assessment and collaborative decision-making: A systemic approach to improving reading instruction*. Ed.D. dissertation,

Northern Arizona University, United States -- Arizona. Retrieved February 24, 2008, from ProQuest Digital Dissertations database. (Publication No. AAT 3138551).

Thurlow, R., & van den Broek, P. (1997). Automaticity and inference generation. *Reading and Writing Quarterly, (40),* 165-184.

Tichá, R., Espin, E. A., & Wayman, M. M. (2007). *Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading aloud and maze selection measures.* Manuscript submitted for publication.

Tindal, G., Flick, D., & Cole, C. (1992). The effect of curriculum on inferences of reading performance and improvement. *Diagnostique, 18,* 69-84.

Tindal, G., & Marston, D. (1990). *Classroom-based assessment: Testing for teachers.* Columbus, OH: Merrill.

Tindal, G., Marston, D., Deno, S. L., & Germann, G. (1982). Curriculum differences in direct repeated measures of reading. *Institute for Research on Learning Disabilities, 93.*

Torgesen, J. (1982). The learning-disabled child as an inactive learner: Educational Implications. *Topics in Learning and Learning Disabilities, 2,* 42-52.

Torgesen, J. (2005). *Reading K-12: The view from 10,000 feet above school level* [PowerPoint]. Florida State University: Florida Center for Reading Research. Retrieved from http://www.fcrr.org/science/powerpoint/torgesen/Leadershipconferencelunch2005.ppt#257,1,Slide1.

Torgesen, J.K., Rashotte, C., Alexander, A., & MacPhee, K. (2002, June). *Progress*

toward understanding the conditions that are required to remediate reading difficulties in older children. Paper presented at the meeting of the National Dyslexia Research Foundation, Kohala Coast, HI.

Treiman, R. (1992). The role of intrasyllabic units in learning to read and spell. In P.B. Gough, L.C, Ehri, & R. Treiman (Eds.), *Reading acquisition*, (pp.65-106). Hillsdale, NJ: Erlbaum.

Underwood, G., Hubbard, A., & Wilkinson, H. (1990). Eye fixations predict reading comprehension: The relationship between reading skill, reading speed and visual inspection. *Language and Speech, (33),* 69-81.

U. S. Department of Education. (2003). Fact Sheet on the Major Provisions of the Conference Report to H.R. 1, the No Child Left Behind Act. [Archived Information]. Retrieved on November 20, 2007 from: http://www.ed.gov/nclb/overview/intro/factsheet.html .

U. S. Department of Education. (2004). Public *law print of PL 107-110, the No Child Left Behind Act of 2001.* Retrieved from http://www.ed.gov/policy/elsec/leg/esea02/index.html.

Volker, M., Lopata, C., & Cook-Cottone, C. (2006). Assessment of children with intellectual giftedness and reading disabilities. *Psychology in the Schools, 43*(8), 855-868.

Wayman, M., Wallace, T., Wiley, H., Tichá, R., & Espin, C. (2007, Summer). Literature Synthesis on Curriculum-Based Measurement in Reading. *Journal of Special Education*, *41*(2), 85-120. Retrieved from Academic Search Premier database.

Wesson, C.L. (1992). Using curriculum-based measurement to create instructional groups. *Preventing School Failure*, *36*(2), 17-20.

Wiley, H. & Deno, S. (2005, July-August). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education 26*(4). 207-215.

Whitehurst, G. J. & Lonigan, C. J. (1998). Child development and emergent literacy, *Child Development, 69,* 848-872.

Whitehurst, G. J. & Lonigan, C. J. (2001). Emergent literacy: Development form prereaders to readers. In S. B. Neuman & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 11-29). New York: Guilford.

Woodcock, R. W. (1973). *Woodcock reading mastery tests manual.* Circle Pines, MN: American Guidance Service.

Wright, Jim. (1992). Curriculum-Based Measurement: Directions for Administering and Scoring CBM Probes in…ORAL READING FLUENCY, Excerpt from *Curriculum-Based Measurement: A Manual for Teachers, Jim Wright, School Psychologist.* New York: Syracuse City Schools. Retrieved from www.interventioncentral.org.

Yovanoff, P., Duesbery, L., Alonzo, J., & Tindal, G. (2005). Grade-level invariance of a theoretical causal structure predicting reading comprehension with vocabulary and oral reading fluency. *Educational Measurement: Issues and Practice, 24*(3), 4-12.

APPENDIXES

APPENDIX A

ORF PROBE

ORF Probe: Student Copy (1000 Lexile, ORF, 258 words)

Inside this tipsy house I lived something of a lopsided life. Because from my earliest years I was very much like my mother, and hardly at all like my father or older sister. My mother said it was because we were made just the same, she and I — and it was true — we both had the same peculiar eyes of a sort you almost never see in Japan. Instead of being dark brown like everyone else's, my mother's eyes were a translucent gray, and mine are just the same. When I was very young, I told my mother I thought someone had poked a hole in her eyes and all the ink had drained out, which she thought very funny. The fortune-tellers said her eyes were so pale because of too much water in her personality, so much that the other four elements were hardly present at all — and this, they explained, was why her features matched so poorly. People in the village often said she ought to have been extremely attractive, because her parents had been. Well, a peach has a lovely taste and so does a mushroom, but you can't put the two together; this was the terrible trick nature had played on her. She had her mother's pouty mouth but her father's angular jaw, which gave the impression of a delicate picture with much too heavy a frame. And her lovely gray eyes were surrounded by thick lashes that must have been striking on her father, but in her case only made her look startled.

Golden, Arthur. (1999). *Memoirs of a Geisha.* New York: Random House Publishers.

ORF Probe: Teacher Copy (1000 Lexile, ORF, 258 words)

Inside this tipsy house I lived something of a lopsided life. 11
Because from my earliest years I was very much like my mother, and 24
hardly at all like my father or older sister. My mother said it was because 39
we were made just the same, she and I — and it was true — we both had 55
the same peculiar eyes of a sort you almost never see in Japan. Instead of 70
being dark brown like everyone else's, my mother's eyes were a 81
translucent gray, and mine are just the same. When I was very young, I 95
told my mother I thought someone had poked a hole in her eyes and all 110
the ink had drained out, which she thought very funny. The fortune- 122
tellers said her eyes were so pale because of too much water in her 135
personality, so much that the other four elements were hardly present at 147
all — and this, they explained, was why her features matched so poorly. 159
People in the village often said she ought to have been extremely 171
attractive, because her parents had been. Well, a peach has a lovely taste 184
and so does a mushroom, but you can't put the two together; this was the 199
terrible trick nature had played on her. She had her mother's pouty mouth 212
but her father's angular jaw, which gave the impression of a delicate 224
picture with much too heavy a frame. And her lovely gray eyes were 237
surrounded by thick lashes that must have been striking on her father, but 250
in her case only made her look startled. 258

Golden, Arthur. (1999). *Memoirs of a Geisha.* New York: Random House Publishers.

APPENDIX B

MAZE PROBE

MR. UTTERSON the lawyer was a man of a rugged countenance that was never lighted by a smile; cold, scanty and embarrassed in discourse; backward in sentiment; lean, long, dusty, dreary and yet somehow lovable. At friendly meetings, and when the wine [1] to his taste, something eminently human [2] from his eye; something indeed which [3] found its way into his talk, [4] which spoke not only in these [5] symbols of the after-dinner face, but [6] often and loudly in the acts [7] his life. He was austere with [8] drank gin when he was alone, [9] mortify a taste for vintages; and [10] he enjoyed the theater, had not [11] the doors of one for twenty [12] But he had an approved tolerance [13] others; sometimes wondering, almost with envy, [14] the high pressure of spirits involved [15] their misdeeds; and in any extremity [16] to help rather than to reprove. [17] incline to Cain´s heresy,´ he used [18] say quaintly: ´I let my brother [19] to the devil in his own [20] In this character, it was frequently [21] fortune to be the last reputable [22] and the last good influence in [23] lives of down going men. And to [24] as these, so long as they [25] about his chambers, he never marked [26] shade of change in his demeanour.

No doubt the feat was easy [27] Mr. Utterson; for he was undemonstrative [28] the best, and even his friendship [29] to be founded in a similar [30] of good-nature. It is the mark [31] a modest man to accept his [32] circle ready-made from the hands of [33] and that was the lawyer´s way. [34] friends were those of his own [35] or those whom he had known [36] longest; his affections, like ivy, were [37] growth of time, they implied no aptness [38] the object. Hence, no doubt, the [39] that united him to Mr. Richard [40] his distant kinsman, the well-known man [41] town. It was a nut to [42] for many, what these two could [43] in each other, or what subject [44] could find in common. It was [45] by those who encountered them in [46] Sunday walks, that they said nothing, [47] singularly dull, and would hail with [48] relief the appearance of a friend. [49] all that, the two men put [50] greatest store by these excursions, counted [51] the chief jewel of each week, [52] not only set aside occasions of [53] but even resisted the calls of [54] that they might enjoy them uninterrupted.

Name : Student

Stevenson, R. L. (1993). *The strange case of Dr. Jekyll and Mr. Hyde.* Bradford, England: Waternill Press.

APPENDIX C

TEACHER FLUENCY PACKET

Teacher Fluency Packet

Oral Reading Fluency

Reading Rate
Grade 1 – 80 words per minute
Grade 2 – 90 words per minute
Grade 3 – 110 words per minute
Grade 4 – 140 words per minute
Grade 5 – 160 words per minute
Grade 6 – 180 words per minute
Adults read about 300 words per minute
If the rates are below these levels, the child may be having difficulty with fluency.
Formula for reading rate per minute: Total number of words read x60/number of seconds read.

Barrington, Jackie. (2003). Oral reading fluency presentation. Vancouver: Simon Fraser University. Retrieved from http://www.cenmi.org/msdbLIO/downloads/Literacy/ReadingWritingBraille/ReadingFluency.doc

A student at the 10th percentile reads about 60,000 words a year in 5th grade.
A student at the 50th percentile reads about 900,000words a year in 5th grade.
Average students receive about 15 times as much practice in a year (Anderson, R. C., 1992*)*

Retrieved from:

Torgesen, J. (2005). *Reading K-12: The view from 10,000 feet above school level* [PowerPoint]. Florida State University: Florida Center for Reading Research. Retrieved from http://www.fcrr.org/science/powerpoint/torgesen/Leadershipconferencelunch2005.ppt#257,1,Slide1

Tables for Determining Placement Level and Growth Rates

Table 1: Correlation of Lexiles to Proficient and Exemplary CRCT Scores

	6th	7th	8th
Passing Lexile Score (800 CRCT Score)	685 Lexile	800 Lexile	805 Lexile
Exemplary Lexile Score (850 CRCT Score)	1210 Lexile	1210 Lexile	1265 Lexile

Table 2: Determining Placement Level in Reading Material

Grade Level	Placement Level	Correct Words per Minute (CWPM)	Errors per Minute
1-2	Frustration	<40	>4
	Instructional	40-60	4 or less
	Mastery	>60	4 or less
3-6	Frustration	<70	>6
	Instructional	70-100	6 or less
	Mastery	>100	6 or less
**7-8 Possibility from combination of research	Frustration	<100	>8
	Instructional	100-125	8 or less
	Mastery	>125	8 or less

Note. Adapted from *Developing goals and objectives for educational programs by* Fuchs, L. S., & Deno, S. L. (1982). Washington, DC: American Association of Colleges for Teacher Education.

Table 3: Growth Rates for Reading

Grade	Realistic Growth Rates		Ambitious Growth Rates
1	2		3
2	1.5		2
3	1		1.5
4	.85		1.1
5	.5		.8
6	.3		.65
Grade		Growth Rates	
1		1.80	
2		1.66	
3		1.18	
4		1.01	
5		.58	
6		.66	

Note. Adapted from "Formulative evaluation of academic progress: How much growth can we expect? By Fuchs, L. S., Fuchs, D., Hamlett, C. L., Walz, L., & Germann, G. (1993), *School Psychology Review, 22,* 27-48, and "Using curriculum-based measurement to establish growth standards for students with learning disabilities" by Deno, S. L., Fuchs, L. S., Marston, D., & Shin, J. (2001). School *Psychology Review, 30,* 507-524.

Table 4. Correlation of Lexiles to Proficient and Exemplary CRCT Scores

	6th	7th	8th	6th Monitor With	7th and 8th Monitor With
Passing Lexile Score (Below 800 CRCT Score. Students below this point do not meet standards.)	685 Lexile	800 Lexile	805 Lexile	700-900 Lexile Probes	800-1000 Lexile Probes
Grade Level Lexile Score (75% accuracy) 30 out of 40 CRCT Score from 800-849	831 CRCT 955 Lexile	826 CRCT 1015 Lexile	828 CRCT 1080 Lexile	900-1200 Lexile Probes	1000-1200 Lexile Probes
Exemplary Lexile Score CRCT Score from 850 and above	1210 Lexile	1210 Lexile	1265 Lexile	1200-1400 Lexile Probes	1200-1400 Lexile Probes

Oral Reading Fluency Directions

Acknowledgements:
Wright, Jim. (1992). Curriculum-Based Measurement: Directions for Administering and Scoring CBM Probes in…ORAL READING FLUENCY, Excerpt from *Curriculum-Based Measurement: A Manual for Teachers, Jim Wright, School Psychologist.* New York: Syracuse City Schools. Retrieved from www.interventioncentral.org.

Beginning ORF assessments

- Use the student Lexile from the CRCT and choose the Lexile passage that is closet to that number, one just below that number, and one that is just above that Lexile number.
- Give assessments of ORF, one passage at a time either on one day or three consecutive days.
- Use the median score as the beginning point for correct responses as well as the beginning Lexile passage to use unless it is under 100 wcpm. Then go to a lower Lexile until the WCPM is at least 100.
- Record the median Lexile score in Infinite Campus (see directions below).
- Record that as the student's beginning point on the scatter plot.
- The ending point, or goal is:

  the number of weeks x .66 words + WCPM = Goal or
  23 weeks X .66 + WCPM = Goal
  Example if student reads a 100 WCPM:
  23 x .66 + 100 WCPM= 15.8 + 100 = 115.8 WCPM Goal
- Plot the goal line point on the scatter plot.
- Connect the two points with a line (regression line).

Directions for monthly assessment, every other week.

- Monitor using the goal level passage based on Table 5.
- Count the number of Words Correct Per Minute (WCPM).
- Chart the number of correct responses on the scatter plot.
- Record WCPM in Infinite Campus under Oral Reading Fluency Task, Assignment: ORF-August-#1.
- When all the students are complete, do Gradebook Export procedure listed above.
- Email file to me donna.jones@hallco.org.
- After the 5th point that is charted, analyze the data using the 4-point rule.
- After the 9th point is charted, analyze using the Trend-line rule.

Table 5: Monitoring Levels

Grade	CRCT Score	CRCT Level	Monitoring Leve Based on Lexiles Passages
6th	Below 800	1	700L-900L
	800-849	2	900L-1200L
	850	3	1200L-1400L
7th/ 8th (the same	Below 800	1	800L-1000L
	800-849	2	1000L-1200L
	850	3	1200L-1400L

Administration of CBM, Oral Reading Fluency, reading probes

The examiner and the student sit across the table from each other. The examiner hands the student the unnumbered copy of the CBM reading passage. The examiner takes the numbered copy of the passage, shielding it from the student's view. The examiner can read the regular print to the whole group to save time and read the italics to each individual student before he or she begins reading.

The examiner says to the student:

Today, I will be listening to you read orally and checking your fluency. Fluency will be measured by the number of words you read correctly in one minute. When I say begin, you will read the passage out loud, and I will time you. You will have one minute to read. Make sure you focus not only or how fast you are reading but also on how many words you are reading correctly. *When I say, 'start,' begin reading aloud at the top of this page. Read across the page* [demonstrate by pointing]. *Try to read each word. If you come to a word you don't know, I'll tell it to you. Are you ready? When I say begin, start reading out loud. Ready. Begin*. (Start timer as you say begin). [Pause] *Start.*

The examiner begins the stopwatch when the student says the first word.
If the student does not say the initial word within 3 seconds, the examiner says the word and starts the stopwatch. As the student reads along in the text, the examiner records any errors by marking a slash (/) through the incorrectly read word. If the student hesitates for 3 seconds on any word, the examiner says the word and marks it as an error. At the end of 1 minute, the examiner says, *Stop* and marks the student's concluding place in the text with a bracket ( ]).

Putting Data in Infinite Campus

To begin:

- Create an Infinite Campus file. This is where you will save all of your ORF and Maze assessments. Follow the following steps. Click on each bullet:
    - Lesson Planner
    - New Task Group
    - Name (Oral Reading Fluency), take off weight, Check all boxes for all terms
    - Save
    - New Assignment
        - Name: ORF-August-1
        - Abbreviation: ORF1
        - Group: Scroll down to Oral Reading Fluency Term 1 Current Grade
        - Due date: Date that you give the ORF assessment
    - Reports
        - Gradebook Export
        - File Format: Scroll to Comma Separated (recommended)
        - Generate Report
        - Save
        - Click on Desktop on left side
        - Click on Yellow Folder at top (Create New Folder)
        - Name Folder: Oral Reading Fluency Assessments
        - Click on Open
        - File name: extract_ORF#1
        - Save

      When giving next assessment, always name the same way, such as extract_ORF #2, Extract ORF#3, etc.

Scoring

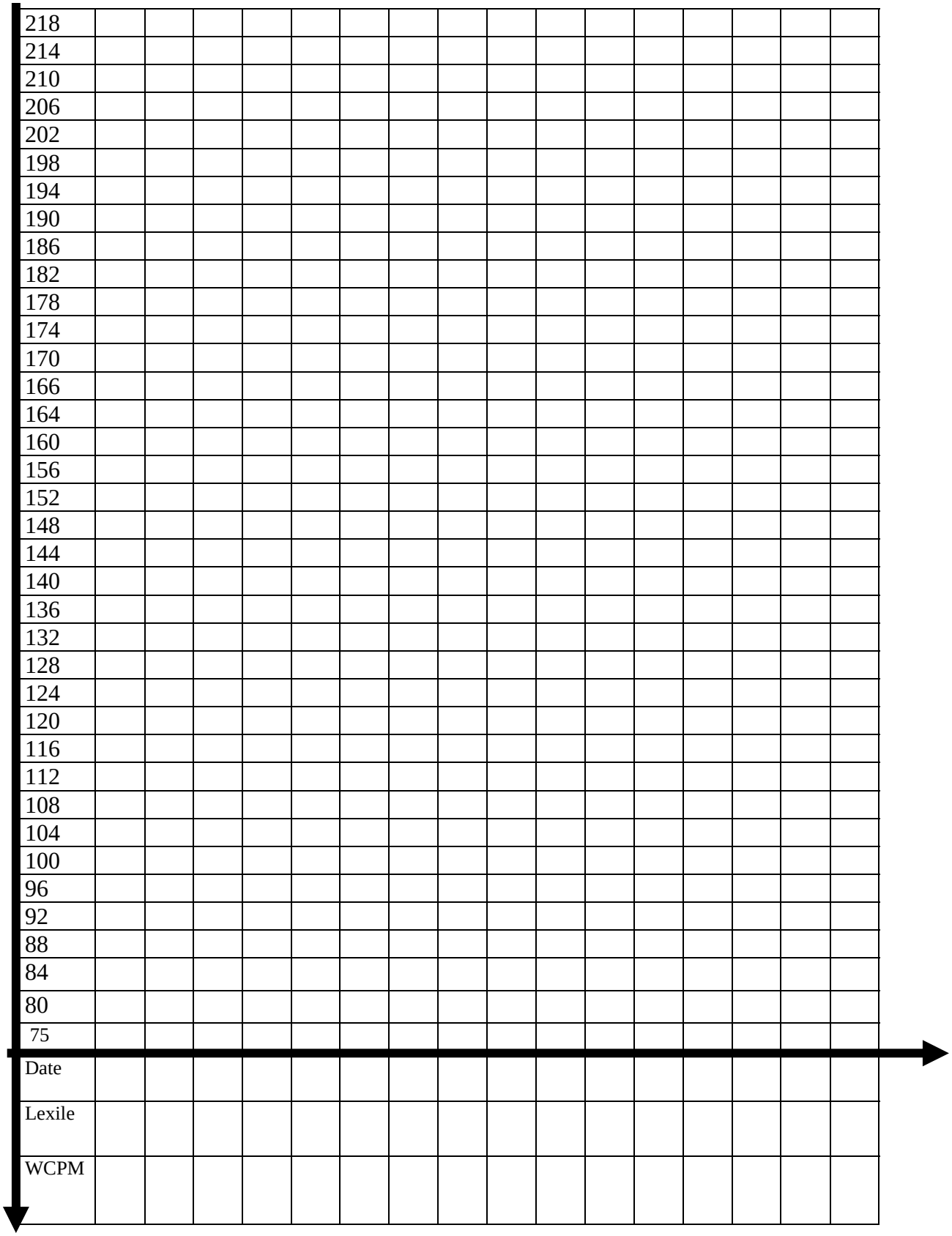
Reading fluency is calculated by first determining the total words attempted within the timed reading probe and then deducting from that total the number of incorrectly read words.

Errors

The following scoring rules will aid the instructor in marking the reading probe:

1. Words read correctly are scored as correct:
    --Self-corrected words are counted as correct.
    --Repetitions are counted as correct.
    --Examples of dialectical speech are counted as correct.
    --Inserted words are ignored.
2. Mispronunciations are counted as errors.
    Example:
        Text: The small gray fox ran to the cover of the trees.
        Student: "The smill gray fox ran to the cover of the trees."
3. Substitutions are counted as errors.
    Example:
        Text: When she returned to the house, Grandmother called for Franchesca.
        Student: "When she returned to the home, Grandmother called for Franchesca.
4. Omissions are counted as errors.
    Example
        Text: Anna could not compete in the last race.
        Student: "Anna could not in the last race."
5. Transpositions of word-pairs are counted as 1 error.
    Example
        Text: She looked at the bright, shining face of the sun.
        Student: "She looked at the shining bright face of the sun."
6. Words read to the student by the examiner after 3 seconds have gone by are counted as errors.

Words Correct Per Minute Growth (Bar Graph for Student)



Maze Assessment

What is it?

- MAZE is a multiple-choice cloze task that students complete while reading silently a passage of about 400 words.
- The first sentence is left as is.
- After the first sentence, every 7th word is replaced with three words (one correct word and two distractors).
-

Why use the Maze assessment?

- Provides a corroborative measure for reading progress
- Assists in making instructional decisions and improves accountability.

How is the Maze given?

- Students read silently for 3 minutes from Lexile MAZE passages
- Determine the number of correct answers
- Record the total number of correct answers.

Measure	Time	Test Arrangements	What Is Scored
Maze Reading	3 minutes	Individual, Small Group, Classroom Group	# of Correct Answers

How is the Maze administered?

- MAZE is a standardized test.
- Procedures and directions must be uniform.
- Once students are familiar with the test directions, the shortened "familiar" directions may be used.

What is important to know?

- Administer a simple practice test to familiarize the student with the procedure.
- *Have the students look at the teacher once the passage has been clicked on the computer* so that the student does not begin the test prematurely.
- Monitor student to ensure that he/she is choosing one answer.
- Stop the MAZE passage and administer another if there are any interruptions.

Where do I start when I give the Maze?
To begin:

- Create an Infinite Campus file. This is where you will save all of your ORF and Maze assessments. Follow the following steps. Click on each bullet:
    - Lesson Planner
    - New Task Group
    - Name (Maze), take off weight
    - Click each term's box
    - Save
    - New Assignments
        - Name: Maze-August-1
        - Abbreviation: Maze1
        - Group: Scroll down to Maze Fluency Term 1 Current Grade
        - Due date: Date that you give the Maze assessment
    - Reports
        - Gradebook Export
        - File Format: Scroll to Comma Separated (recommended)
        - Generate Report
        - Save
        - Click on Desktop on left side
        - Click on Yellow Folder at top (Create New Folder)
        - Name Folder: Maze Fluency Assessments
        - Click on Open
        - File name: extract_Maze#1
        - Save
        - Follow same procedure on next assessment except call it extract_Maze#2, extract_Maze#3, etc.

- Use the student Lexile from the CRCT and choose the Lexile passage that is closet to that number, one just below that number, and one that is just above that Lexile number.
- Give one passage at a time either on one day or three consecutive days.
- Use the median score as the beginning point for correct responses as well as the beginning Lexile passage to use.

Directions for monthly assessment, every other week.

- Give each student the next Lexile leveled passage based the goal level (See Table 5).
- Chart the number of correct responses on the scatter plot and worksheet (Maze Assessment Goal Setting Form). Students can do this as a self assessment with close teacher monitoring.
- Record Responses Correct in Infinite Campus and export file to Excel.
- Email file to me donna.jones@hallco.org.
- Chart the median as the beginning point, and the end point based on the Maze fluency goal.
- After the 5th point that is charted, analyze the data using the 4-point rule.
- After the 9th point is charted, analyze using the Trend-line rule.

Curriculum Based Measurement Reading-Maze (CBM R-MAZE) Standardized Directions (Front of Teacher's Copy)

Say: Today you will take a Maze assessment. Before you begin, we will practice using an example on the computer to show you how to complete the test.
Click on My Computer icon on the Desktop.
Next click on apps on 'nhms-apps' (R :)
Next click on 2-Online Maze Assessments.
At the top corner of the computer you see a -, a box, and an X. Click on the box so that your screen is maximized.
Next click on CLOZE_230_Lexile_Mummies. When the screen appears look at me so that I can see your eyes. You may not begin reading until I have everyone's eyes, and we all start at the same time so that it will be fair for everyone. [*This is important because some will read ahead before time starts.]*

When I say begin you will read the Maze passage for 3 minutes which is the passage in red. When you come to a box, scroll down and you will see a group of three choices. Highlight the one word that makes the most sense. Work as quickly as you can without making mistakes. Do not jump around the passage. Work in order because that will save you time. Continue working until I say stop. Do you have any questions?
[Set the timer for 3 minutes. Start the timer.] Begin.
[Monitor students to make sure they understand that they are to choose 1 word. At the end of 3 minutes say..]
***Stop. Do not click on anymore answers. Find the gray box that says "Check." There is one at the top of the passage and one at the bottom. Click "Check."
You now see the gray box on the screen. If you look at the top, it tells you what answer choices were correct and incorrect. The ones in ( ) are the ones that you did not attempt. Under the incorrect answers, you will see the total correct. For instance, it says, you have 14 out of 58 or 20 out of 56. The number correct is what we want. When I come around, I would like you to tell me the number you answered correctly. Under that section, you will see the percentage correct. We do not want the percentage because that is based on the fact that you finished the entire passage. We only want the total answered correctly.
[Walk around and collect the number of correct choices from each student and record.]
Click on the X in the top corner. Go back to the web page with E's

> *If the check box does not work, follow this procedure:*
> Everyone look at the gray bar across the top of the screen.
> Find the Pop-ups blocked icon. It may say, "14 blocked." It is a warning sign, a red circle with a red slash across it. Click on that so that it will say, "Pop-ups okay." This will allow you to check your answers. Once you see "Pop-ups okay." Click on "Check."
> Collect the R-MAZE scores total answers correct, not percentages.

Procedures to follow after the students are familiar with the Maze assessment:

You will do another Maze assessment. Click on _____ (Give the Lexile that you want them to do.). Remember look at me as soon as you click on the passage. I will watch when the passages are loaded and will tell you when to start so that we all start at the same time. Now I see everyone's eyes. When I say 'Begin' click on the computer screen and start reading silently. *[Set the timer for 3 minutes. Start the timer.]* Ready. Begin.
Follow same procedure as at the top when finished, beginning at the ***.

Back of Directions: Monitoring Levels

To get the Baseline score for the Maze, follow the following procedure:
**Give the same passages to all the students for the baseline.
**Give the first Lexile-leveled passage in each 100 level. For example, the first 500 Lexile level that we have a passage for is 520 Lexile. That is where I would start in the 500s, and then continue following that for each Lexile, such as 600 Lexile, 710 Lexile, 810 Lexile, 960 Lexile, 1020 Lexile, 1100 Lexile, etc. (These are the first levels in each 100s Lexile group.)

Give the following passages using the first Lexile level for each particular group:

6^{th} grade
At risk students or co- taught classes, use first 500 Lexile level, first 600 Lexile level, and 700, first Lexile level.
Regular classes, use the first one in the 800, 900, and 1000.
Advanced classes, use the first one in the 900, 1000, and 1100.

7^{th} and 8^{th} grade teachers
At risk or co-taught, use 800, 900, and 1000.
Regular classes use 900, 1000, and 1100.
Advanced students get a baseline with 1000, 1100, and 1200.

After 3 Maze passages are given, use the median score as the beginning point on the scatter plot. From then on, only give one Maze every other week following the chart listing dates. Plot each point. Monitor each student with the passage based on his or her CRCT score.
Table 5 provides the Lexile level that represents the goal level for the student to be monitored with the Maze assessment.

MAZE Assessment Goal Setting Form

Student Name:_____Grade: _____ Date: _____

CRCT Level 2007-2008: _____ Monitoring Lexile Level: _____

Baseline MAZE Scores: [ ] [ ] [ ] Median = [ ]

Length of time, in weeks, the progress will be monitored: [ ]

Lexile Level Reading Passages	Number of Responses Correct	Performance Level Goal		Spring Benchmarks Per Grade Level	Responses Correct (RC) 5th through 50th Percentiles
Sept				First	3 – 7
Sept				Second	9 – 14
Oct				Third	10 – 15
Oct				Fourth	13 – 19
Nov				Fifth	17 – 24
Nov				Sixth	18 – 25
Dec				Seventh	18 – 26
Jan				Eighth	19 – 25
Jan					
Feb					
Feb					
Mar					
Mar					

RC = Responses Correct E = Errors

Grade Level Text	Fuchs et al. Expected Growth Rate
All Grades	.4 words per week

Formula:

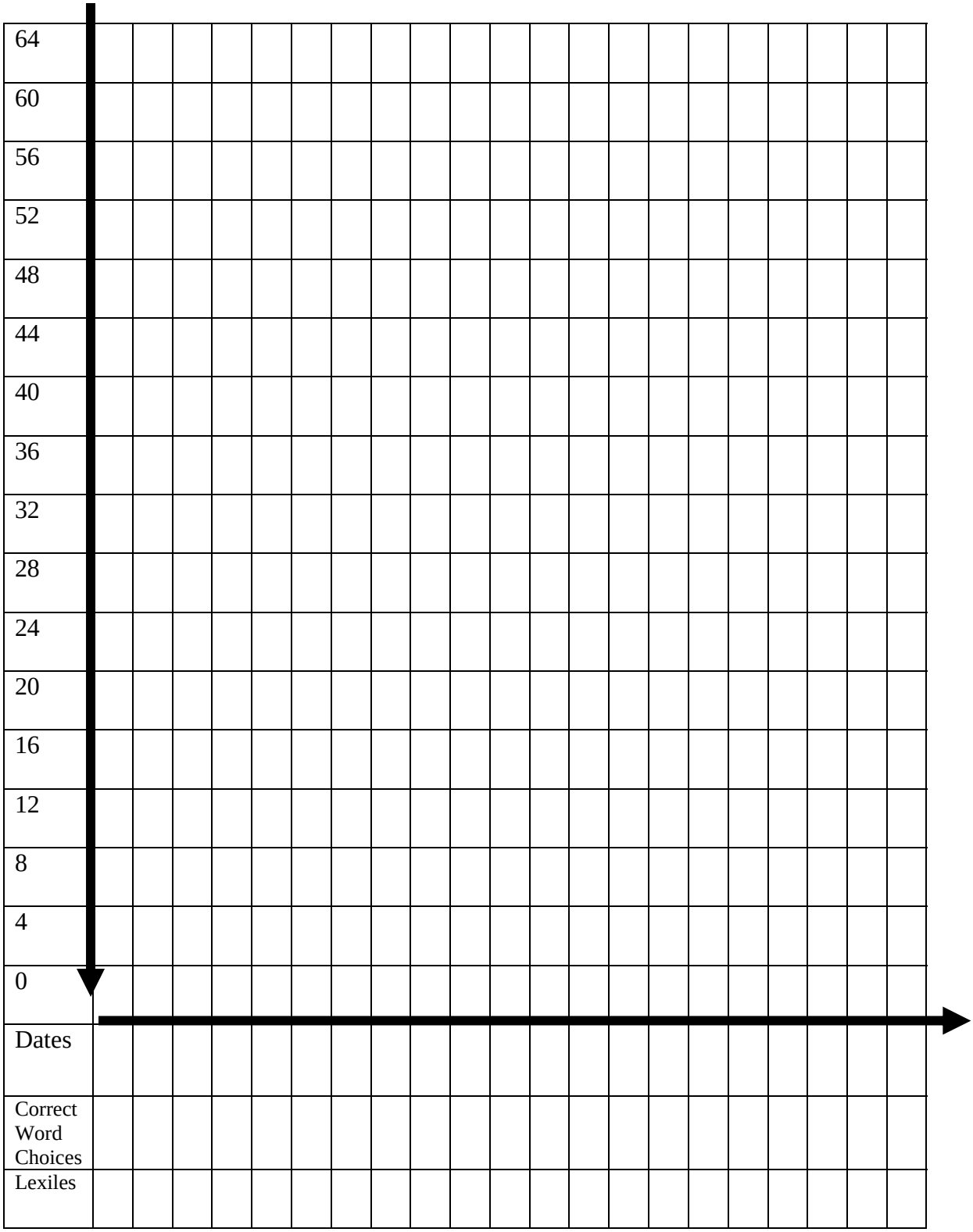
# of weeks [ ] X .4 = [ ] + median score = [ ]

MAZE Fluency Goal

Revised 8/08

Adapted from Pittsburgh Public Schools. (2007-2008). *Data Analysis/Decision making and MAZE comprehension training* [PowerPoint]. Retrieved from http://www.pps.k12.pa.us/1431101271026276 7/lib/1431101271026276 7/Data_Analysis_and_MAZE.ppt

Student Bar Graph for Recording MAZE Fluency Assessment



Response to Intervention (RTI)
Data Analysis and Decision Guidelines for ORF and Maze

- At least three characteristics of graphed data can be used to describe and summarize student performance:
  - Level of performance (WRC)
  - Variability of performance
  - Slope of performance

- Two methods for data-based decision making:
  - 4-Point Rule
  - Trend-Line Rule

4-Point Rule

When at least 6 points have been collected (not counting the beginning point) examine the 4 most recent points.

-If all 4 are above Goal Line, increase goal.

-If all 4 are below Goal Line, make an instructional change.

-If the data points are both above and below the goal line, continue collecting data until the 4-Point Rule or Trend-Line Rule can be applied.

Trend-Line Rule

When at least 8 data points have been collected, review trend of current performance and compare to Aim Line (Goal Line).

- If trend of student progress is steeper than Aim Line, raise goal.
- If trend of student progress is less steep than Aim Line, make an instructional change.

Documentation of Instructional Interventions

The following instructional elements may be altered to enhance student performance (Stecker & Lembke, 2005):

- Instructional strategies
- Size of instructional group
- Time allocated for instruction
- Materials
- Reinforcement