

2014

The Relationship between mClass Reading 3D Assessment and the North Carolina End-of-Grade Assessment of Reading Comprehension in an Elementary School

Amy S. Bowles
Gardner-Webb University

Follow this and additional works at: https://digitalcommons.gardner-webb.edu/education_etd



Part of the [Educational Assessment, Evaluation, and Research Commons](#), and the [Elementary Education and Teaching Commons](#)

Recommended Citation

Bowles, Amy S., "The Relationship between mClass Reading 3D Assessment and the North Carolina End-of-Grade Assessment of Reading Comprehension in an Elementary School" (2014). *Education Dissertations and Projects*. 6.
https://digitalcommons.gardner-webb.edu/education_etd/6

This Dissertation is brought to you for free and open access by the School of Education at Digital Commons @ Gardner-Webb University. It has been accepted for inclusion in Education Dissertations and Projects by an authorized administrator of Digital Commons @ Gardner-Webb University. For more information, please see [Copyright and Publishing Info](#).

The Relationship between mClass Reading 3D Assessment and the North Carolina End-of-Grade Assessment of Reading Comprehension in an Elementary School

By
Amy S. Bowles

A Dissertation Submitted to the
Gardner-Webb University School of Education
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Education

Gardner Webb University
2014

Approval Page

This dissertation was submitted by Amy S. Bowles under the direction of the persons listed below. It was submitted to the Gardner-Webb University School of Education and approved in partial fulfillment of the requirements for the degree of Doctor of Education at Gardner-Webb University.

David Shellman, Ed.D.
Committee Chair

Date

Sydney Brown, Ph.D.
Committee Member

Date

Lory Morrow, Ed.D.
Committee Member

Date

Jeffrey Rogers, Ph.D.
Dean of the Gayle Bolt Price School
of Graduate Studies

Date

Acknowledgements

There are many people I would like to acknowledge and thank for their support. This is not a process that one can do alone; it takes many supporters, encouragers, and leaders to guide you through. I am truly appreciative for all who have walked by my side throughout this process. I am grateful to be able to share this success with you.

First and foremost, I would not have been able to walk through this journey without the grace, unconditional love, and blessings of my Lord and Savior Jesus Christ, without whom none of this would have been possible. I would like to sincerely thank the many family, friends, colleagues, and professors who provided support in so many different ways. My husband and best friend always standing by me and encouraging me to reach my goals, I would be at a loss without you in my life. My son, who even though he is only little, has been patient and supportive in his own sweet way. My parents, my first and longest supporters, their continual love and support have brought me to where I am now and I am forever grateful. My other parent “in-laws” have supported and loved me as if I was their own. It is a blessing to be a part of their family. My brother and sister-in-law, your success and support have inspired me to push through to the end. My friends, who I hold dear, have encouraged and supported me not only in this process but in my life. My co-workers have inspired me and encouraged me to continue even when it was hard, and I cannot thank them enough for always standing by me. To my dissertation chair and committee, your guidance and encouragement throughout this process has allowed me to reach a goal that I have long been pursuing. Thank you for taking the time to help me reach success.

Abstract

The Relationship between mClass Reading 3D Assessment and the North Carolina End-of-Grade Assessment of Reading Comprehension in an Elementary School. Bowles, Amy S., 2014: Dissertation, Gardner-Webb University, Elementary Schools/Teacher Education/Early Reading/Curriculum-Based Assessment

The North Carolina Department of Public Instruction (NCDPI) instituted the Reading Diagnostic Initiative in 2009 in which select elementary schools across the state were piloting the reading diagnostic tool mClass Reading 3D. This study investigated the relationship between results from the North Carolina End-of-Grade (NCEOG) Assessment of Reading Comprehension and the results from the mClass Reading 3D assessment in a North Carolina elementary school's third, fourth, and fifth grades, especially examining the degree to which mClass Reading 3D predicts scores on the reading comprehension measures of the NCEOG.

The quantitative, correlational study utilized a predictive design to determine if the predictor variables, mClass Reading 3D assessment scores – Oral Reading Fluency (ORF) and Text Reading and Comprehension (TRC) – are accurate predictors of third-, fourth-, and fifth-grade students' scale scores on the NCEOG. This study was conducted in two parts to best address the research questions. Part one consists of descriptive, variance, and inferential statistics (frequency counts, measures of central tendency and variability, and correlations) calculated by grade level and demographic variables. Part one describes the relationship between the predictor variable (mClass Reading 3D) and the outcome measure (NCEOG).

Part two consists of calculating multiple regression analyses using the assessment scores by grade level. Part two describes the predictability of mClass Reading 3D to student scale scores on the reading comprehension portion of the NCEOG. Results of this study are consistent with previous research, indicating mClass Reading 3D ORF and TRC measures statistically, significantly predict student scale scores on the NCEOG. This is important for educators to be able to accurately base instructional decisions on the data.

Table of Contents

	Page
Chapter 1: Introduction	1
Introduction	1
Statement of the Problem	2
Demographics	3
Reading Initiatives	4
Early Intervention	6
Assessments	7
mClass Reading 3D Assessment	9
Purpose of the Study	10
Significance of the Study	10
Research Questions	10
Definition of Terms	11
Limitations and Delimitations	13
Summary	13
Chapter 2: Literature Review	15
Introduction	15
Brief History of Reading Research	15
Elements of Teaching Reading Today	20
Identifying Students At Risk of Reading Failure	25
Interventions	28
Reading Assessments	31
Summary	38
Chapter 3: Methodology	40
Introduction	40
Research Questions	40
Participants	40
Setting	41
Instruments	42
Design	48
Procedures	48
Summary	50
Chapter 4: Results	51
Introduction	51
Research Questions	51
Data Collection and Screening Procedures	51
Descriptive and Inferential Statistics for Third-Grade participants	52
Descriptive and Inferential Statistics for Fourth-Grade Participants	76
Descriptive and Inferential Statistics for Fifth-Grade Participants	95
Summary	114
Chapter 5: Discussion and Conclusions	116
Introduction	116
Study Summary	116
Discussion of Study Results and Connection to Literature Review	117
Limitations	125

Implications for Education.....	126
Recommendations for Future Research	128
Conclusion	129
References.....	132
Appendix	
Coding Key for TRC Levels	142
Tables	
1 Demographics of 2010-2011 Study Participants Compared to School and District.....	41
2 2011 ORF Score Benchmark Cut Point Ranges for Grades 3-5	44
3 Text Gradient for Fountas and Pinnell Leveling System.....	45
4 2011 TRC Cut Point Ranges for Grades 3-5	46
5 2011 NCEOG Reading Achievement Level Ranges for Grades 3-5	47
6 Frequency Distributions for Third-Grade Study Participants	54
7 Descriptive Statistics for Third-Grade NCEOG Scale Score, ORF Score, and TRC Score	55
8 Shapiro-Wilk Test for Third-Grade Participants' Assessment Scores	62
9 Pearson Correlation Matrix for Third-Grade Participant Assessment Scores	63
10 Pearson Correlation Matrix for Third-Grade Scores by Ethnicity.....	64
11 Pearson Correlation Matrix for Third-Grade Scores by Gender.....	65
12 Casewise Diagnostics for Outlier Residuals for Third-Grade Participants	66
13 Correlations Coefficients for Third-Grade Participant Data.....	70
14 Tolerance Values for Third-Grade Participant Data	70
15 Regression Model Summary for Third-Grade Participant Data	72
16 ANOVA Test Results for Third-Grade Participant Data	74
17 Summary of Standard Multiple Regression for Third-Grade Participants	75
18 Variables Enter/Removed from Analysis for Third-Grade Gender and Ethnicity	76
19 Excluded Variables from Analysis for Third-Grade Gender and Ethnicity	76
20 Frequency Distributions for Fourth-Grade Study Participants	77
21 Descriptive Statistics for Fourth-Grade NCEOG Scale Score, ORF Score, and TRC Score	78
22 Shapiro-Wilk Test for Fourth-Grade Participants' Assessment Scores.....	84
23 Pearson Correlation Matrix for Fourth-Grade Participant Assessment Scores	85
24 Pearson Correlation Matrix for Fourth-Grade Assessment by Ethnicity.....	86
25 Pearson Correlation Matrix for Fourth-Grade Assessment by Gender.....	87
26 Casewise Diagnostics for Outlier Residuals for Fourth-Grade Participants.....	88
27 Correlations Coefficients for Fourth-Grade Participant Data.....	91
28 Tolerance Values for Fourth-Grade Participant Data	91
29 Regression Model Summary for Fourth-Grade Participant Data	92
30 ANOVA Test Results for Fourth-Grade Participant Data	93
31 Summary of Standard Multiple Regression for Fourth-Grade Participants.....	94
32 Variables Enter/Removed from Analysis for Fourth-Grade Gender and Ethnicity	95
33 Excluded Variables from Analysis for Fourth-Grade Gender and Ethnicity.....	95
34 Frequency Distributions for Fifth-Grade Study Participants	96

35	Descriptive Statistics for Fifth-Grade NCEOG Scale Score, ORF Score, and TRC Score.....	97
36	Shapiro-Wilk Test for Fifth-Grade Participants' Assessment Scores.....	104
37	Pearson Correlation Matrix for Fifth-Grade Participant Assessment Scores	104
38	Pearson Correlation Matrix for Fifth-Grade Scores by Ethnicity	105
39	Pearson Correlation Matrix for Fifth-Grade Scores by Gender	106
40	Correlations Coefficients for Fifth-Grade Participant Data.....	109
41	Tolerance Values for Fifth-Grade Participant Data	110
42	Regression Model Summary for Fifth-Grade Participant Data	111
43	ANOVA Test Results for Fifth-Grade Participant Data	112
44	Summary of Standard Multiple Regression for Fifth-Grade Participants	113
45	Variables Enter/Removed from Analysis for Fifth-Grade Gender and Ethnicity	114
46	Excluded Variables from Analysis for Fifth-Grade Gender and Ethnicity.....	114
Figures		
1	NCEOG, ORF, and TRC Mean Scores by Third-Grade Ethnicity	57
2	NCEOG, ORF, and TRC Mean Scores by Third-Grade Gender	58
3	NCEOG, ORF, and TRC Mean Scores by Third-Grade Students with Special Needs.....	59
4	Scatter Plot of NCEOG and ORF Scores for Third-Grade Participants	60
5	Scatter Plot of NCEOG and TRC Scores for Third-Grade Participants	61
6	Scatter Plot of ORF and TRC Scores for Third-Grade Participants	61
7	Scatter Plot of multiple regression residuals and predicted values for Third-Grade Participants.....	68
8	Partial Regression Plot of NCEOG and ORF Scores for Third-Grade Participants.....	68
9	Partial Regression Plot of NCEOG and TRC Scores for Third-Grade Participants.....	69
10	Normal P-P Plot of Third-Grade Regression Residuals (NCEOG-Dependent Variable)	72
11	NCEOG, ORF, and TRC Mean Scores by Fourth-Grade Ethnicity	80
12	NCEOG, ORF, and TRC Mean Scores by Fourth-Grade Gender	81
13	NCEOG, ORF, and TRC Mean Scores by Fourth-Grade Students with Special Needs.....	82
14	Scatter Plot of NCEOG and ORF Scores for Fourth-Grade Participants	83
15	Scatter Plot of NCEOG and TRC Scores for Fourth-Grade Participants	83
16	Scatter Plot of ORF and TRC Scores for Fourth-Grade Participants	84
17	Scatter Plot of Multiple Regression Residuals and Predicted Values for Fourth-Grade Participants	89
18	Partial Regression Plot of NCEOG and ORF Scores for Fourth-Grade Participants.....	89
19	Partial Regression Plot of NCEOG and TRC Scores for Fourth-Grade Participants.....	90
20	Normal P-P Plot of Fourth-Grade Regression Residuals (NCEOG-Dependent Variable)	92
21	NCEOG, ORF, and TRC Mean Scores by Fifth-Grade Ethnicity	99
22	NCEOG, ORF, and TRC Mean Scores by Fifth-Grade Gender	100

23	NCEOG, ORF, and TRC Mean Scores by Fifth-Grade Students with Special Needs.....	101
24	Scatter Plot of NCEOG and ORF Scores for Fifth-Grade Participants	102
25	Scatter Plot of NCEOG and TRC Scores for Fifth-Grade Participants	102
26	Scatter Plot of ORF and TRC Scores for Fifth-Grade Participants	103
27	Scatter Plot of Multiple Regression Residuals and Predicted Values for Fifth-Grade Participants.....	108
28	Partial Regression Plot of NCEOG and ORF Scores for Fifth-Grade Participants.....	108
29	Partial Regression Plot of NCEOG and TRC Scores for Fifth-Grade Participants.....	109
30	Normal P-P Plot of Fifth-Grade Regression Residuals (NCEOG-Dependent Variable)	110

Chapter 1: Introduction

Introduction

“Reading is the most important skill for success in school and society. Children who fail to learn to read will surely fail to reach their full potential” (Hall & Moats, 1999, p. 6). Based on the 2011 National Assessment of Educational Progress (NAEP) scores, our nation is in a reading crisis. Only 66% of fourth graders read at or above a basic level and only 32% of those students read above the proficient level of performance. These results show no significant increase from the 2009 NAEP assessment results (United States Department of Education [USDE], 2011).

Reading is the foundation of all school-based learning, yet reading failure is prevalent in the United States (Honig, Diamond, & Gutlohn, 2000). According to the National Assessment of Adult Literacy in 2003, there were 30 million people in the United States who were below basic in their reading ability level (National Center for Educational Statistics [NCES], 2003). The authors of *A Nation at Risk* discovered that some 23 million American adults are functionally illiterate by the simplest tests of everyday reading, writing, and comprehension (National Commission on Excellence in Education [NCEE], 1983). People in the United States who are illiterate represent 75% of the unemployed, 33% of mothers receiving Aid to Families with Dependent Children, 85% of juveniles who appear in court, and 60% of prison inmates (Hall & Moats, 1999).

The reading crisis is not just a national problem; it is also a state and local problem. North Carolina’s 2011 education accountability system, The ABC’s of Public Education, revealed only 71% of students in Grades 3-8 were proficient in reading across the state. The North Carolina county included in this study proved to be lower than the state average with only a 70% proficiency rate in reading (North Carolina Department of

Public Instruction [NCDPI], 2011). The county dropout rate of 4.46% in 2011 was higher than the 2011 state average of 3.43% (NCDPI, n.d.). The county unemployment rate of 10.8% in September 2011 (Gaston County Health Department, 2011) was higher than the September 2011 state average of 10.7% (Department of Numbers, n.d.).

Statement of the Problem

The school that served as the focus of this study is hereafter referred to as School A. According to a 3-year trend, School A's reading scores decreased from 40% proficient in 2008-2009 to 30% proficient in 2009-2010 and 34% proficient in 2010-2011. This signified that the reading crisis was not only a national, state, and county issue but a school issue as well.

In order to combat the problem of decreasing reading proficiency scores, NCDPI implemented the reading assessment tool, mClass Reading 3D, as a pilot program through the Reading Diagnostic Initiative. This is a feature of the Ready, Set, Go! initiative from the Budget Act of 2009-2010, Section 7.18(b) that stated,

The State Board of Education shall investigate and pilot a developmentally appropriate diagnostic assessment for students in elementary grades. This assessment will (i) enable teachers to determine student learning needs and individualize instruction, and (ii) ensure that students are adequately prepared for the next level of coursework as set out by the NC Standard Course of Study.

(North Carolina State Board of Education, 2010, p. 10)

This tool provides teachers with benchmark data that allow them to individualize and adjust their instruction on an ongoing and frequent basis.

This research study explored the mClass Reading 3D assessment to determine if it is an accurate predictor of student scores on the North Carolina End-of-Grade (NCEOG)

Assessment of Reading Comprehension. Teachers use mClass Reading 3D assessment results to drive their classroom instruction and provide early intervention; therefore, the assessments should align with the culminating NCEOG to be truly effective.

Demographics

School A is an inner-city school located in the southwestern piedmont area of North Carolina that houses kindergarten through fifth grade. In 2011, the school had 443 students. Its ethnic population was made up of 57% African-American, 21% White, 17% Hispanic, and 5% multi-ethnic.

In 2011, 82% of the students in School A qualified for free or reduced price lunch. This percentage qualified the school to receive Title I funds in the 2010-2011 school year. During this school year, Title I funds provided the school with a parent specialist, reading specialist, literacy facilitator, additional teacher assistants, class-size reduction, classroom teachers, technology tools, professional development, and classroom supplies.

During 2010-2011, School A had 20 classroom teachers with a 9% teacher turnover rate. Seventy percent of the teachers had less than 10 years of teaching experience. Thirteen percent of classroom teachers had advanced degrees. Three teachers had their National Board of Professional Teaching Standards Certification. All teachers and teacher assistants were highly qualified as defined by the No Child Left Behind Act of 2001 (NCLB).

The school was designated a priority school according to the 2010-2011 NCEOG test scores. This means that 50-60% or less of School A's students had reached grade-level proficiency. The overall reading score was 34.4% proficient, and the overall math score was 53% proficient. School A did not meet Adequate Yearly Progress (AYP) as

defined by NCLB, having only met eight of 13 performance targets for AYP.

Under NCLB, schools enter Title I school improvement status by not meeting target goals in the same subject for 2 years in a row. A school in Title I school improvement status must take certain measures to improve performance. Since School A did not meet AYP in 2009-2010, it entered year 3 of Title I school improvement status. This means that School A adhered to the following sanctions in 2010-2011: School Choice, Supplemental Educational Services, and Corrective Action.

Reading Initiatives

The passage of NCLB (2001) called for all students to reach towards the same academic standards. According to NCLB, by the year 2014, all students must be reading on grade level by the end of the third grade. In order to reach this target, many initiatives have been put into action across the nation.

The National Reading Panel (NRP) was established in 1997 by the United States Congress to assess the effectiveness of different approaches used to teach children to read. For over 2 years, the NRP reviewed research-based knowledge on reading instruction. In April 2000, the NRP concluded its work and submitted “The Report of the National Reading Panel: Teaching Children to Read” (National Institute of Child Health and Human Development [NICHD], 2000). This report created the foundation for teaching reading across the United States today. The report stated that there were five *big* ideas to teaching reading: (1) phonics (2) phonemic awareness, (3) fluency, (4) text comprehension, and (5) vocabulary. The findings also suggested that early intervention is critical for students experiencing early difficulties.

As a result of the NRP findings, the Reading First (RF) program was established in 2002 by the U.S. Department of Education in order to provide the early intervention

that was called for by the NRP report. This program focused on putting proven methods of early reading instruction in classrooms. Through RF, states and districts received support to apply scientifically based reading research (SBRR) – and the proven instructional and assessment tools consistent with this research – to ensure that all children learn to read on grade level by the end of third grade (USDE, 2009).

Most states put forth a state reading initiative to complement the national initiatives. The Florida initiative, enacted in 2001, was entitled “Just Read, Florida,” a comprehensive, coordinated reading initiative aimed at helping every student become a successful, independent reader (Florida Department of Education [FDE], 2001). “Just Read, Florida” had many components; the following are just a few: (1) early-reading instruction strategies and reading screenings or assessments for K-2 students, (2) reading intervention strategies for students who read below grade level, (3) reading activities in teacher preparation and professional development programs, and (4) increasing parental and family involvement in teaching and encouraging reading (FDE, 2001).

North Carolina’s reading initiative was based on the RF program. In 2003, the State of North Carolina was awarded a federally funded RF grant. Over the course of a 5-year period, from 2004 to 2009, this grant was used to train teachers in the RF schools in the principles and methodology of SBRR, with the ultimate goal of improving the reading skills of students in North Carolina’s lowest performing elementary schools. The overarching goal of North Carolina’s RF initiative was to ensure that all children learn to read on grade level by the end of the third grade through the systemic application of SBRR to reading instruction, including the following five components: phonemic awareness, phonics, fluency, vocabulary, and text comprehension (NCDPI, 2007).

Early Intervention

The overarching theme in the above reading initiatives was early intervention. Research says that students who are not reading at grade level by the end of the first grade have a high probability of being a poor reader by the end of the fourth grade (Juel, 1988). The Matthew Effect theory suggested that “the rich get richer and the poor get poorer,” meaning the literacy gap between students who learn to read early and those who struggle only widens as they get older (Stanovich, 1986, p. 382).

There is evidence to suggest that a significant number of reading difficulties are preventable (Snow, Burns, & Griffin, 1998). Torgesen (2000) estimated that as many as 50% of children who are most at risk for reading failure can be brought to normal levels of performance following effective early reading instruction and interventions (Reschly, 2010). Research also found that if those struggling readers are identified within the first few years of schooling and provided with targeted and intensive instruction, they are more likely to make the progress necessary to catch up with their peers who are reading at grade level (Torgesen, 2004).

The Chicago Longitudinal Study followed a cohort of 1,539 children from preschool through age 24 and determined the longitudinal effects of early school intervention. The study found associations between participating in an early school-based intervention and positive outcomes enduring through adulthood. The study participants who were in the early intervention programs had significantly higher rates of high school completion and 4-year college attendance, significantly lower rates of juvenile arrest for both violent and nonviolent offenses, lower rates of remedial services, higher percentages of full-time employment with higher income earnings, and lower percentages of receiving public aid assistance (Reynolds et al., 2007).

Karoly, Kilburn, and Cannon (2005) found similar results as the Chicago study. Karoly et al. found there were statistically significant benefits for participants in targeted early intervention programs. In some cases, the improved outcomes were demonstrated soon after the program ended, while other cases were observed through adolescence and adulthood. The long-lasting gains included areas such as educational progress and attainment, positive labor market outcomes, lowered dependency, pro-social behaviors, lowered special education placement, and grade retention.

Assessments

Without early intervention, a struggling reader is at risk of reading failure. In order to adequately serve all students, especially those deemed at risk for reading failure, teachers must be able to accurately assess student needs and subsequently plan and deliver instruction based on that assessment. Otherwise, it is difficult to ensure that all students will master the necessary skills to become proficient readers (Menziez, Mahdavi, & Lewis, 2008).

There are two major types of assessments in education, summative and formative, each possessing a different function. Summative assessments are those assessments designed to determine a student's academic development after a set unit of material (Stiggins, 2002). Formative assessments are assessments designed to monitor student progress during the learning process (Chappuis & Stiggins, 2002). Even though an assessment may be designed formative or summative, it is the methodology, data analysis, and use of results that determine whether an assessment is actually formative or summative (Dunn & Mulvenon, 2009).

Educators recognize that summative assessments, which occur at the end of units/terms/school year, provide too little information that arrives too late for planning

everyday instruction. These assessments do not provide teachers with information they can use for ongoing instruction (Heritage, 2007). Formative assessment practices, if implemented effectively, can provide teachers and their students with the data they need for appropriate interventions (Heritage, 2007).

Assessment should be seen as a moment of learning where students have an active role in their own assessment and an understanding of what it means to get better (Black & Wiliam, 1998). As expectations for reading instruction and the need for individualization of instruction increases, so does the expectation that teachers will regularly collect and make use of assessment data to inform their classroom instruction (Hupert, Heinze, Gunn, Stewart, & Honey, 2007). In order for assessment data to be useful to teachers, it must be (a) specific enough to show where students need help, (b) accessible in a timely manner so that teachers can act upon the information, and (c) comprehensible so that it can be translated into practice (Hupert et al., 2007).

Some examples of the more prominent formative reading assessments used today are *Dynamic Indicators of Basic Early Literacy Skills* (DIBELS) – a set of short fluency measures used to regularly monitor the development of early literacy and early reading skills from kindergarten through sixth grade (Good & Kaminski, n.d.); *Scholastic Reading Inventory* (SRI) – a computer-adaptive reading assessment program for students in Grades Kindergarten through 12 that measures reading comprehension on the Lexile Framework[®] for Reading (SRI, n.d.); *STAR Reading* – a computer-adaptive reading assessment program for students in Grades Kindergarten through 12 that provides an approximate measure of a student’s reading level; *Running Records* (RRs) – a reading assessment tool that allows a teacher to assess a student’s reading performance as she/he reads from a benchmark book (RRs and Benchmark Books, n.d.); *Text Comprehension*

and Reading – consisting of RRs that enable teachers to determine students' accuracy, fluency, reading strategies, and comprehension as they read authentic fiction and nonfiction texts (Reading 3D Brochure, 2009).

mClass Reading 3D

mClass Reading 3D is a formative assessment tool which combines the DIBELS assessment with the Text Reading and Comprehension assessment (TRC). The TRC component consists of RRs that enable teachers to determine students' accuracy, fluency, reading strategies, and comprehension as they read authentic fiction and nonfiction texts (Reading 3D Brochure, 2009). The DIBELS component consists of seven measures: Initial Sound Fluency (ISF), Letter Naming Fluency (LNF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Oral Reading Fluency (ORF) with Retell Fluency (RTF), and Word Use Fluency (WUF). These measures enable teachers to determine student phonetic, phonemic awareness and fluency abilities (mClass Reading 3D, 2010). They include benchmark assessments that are administered three times a year, as well as ongoing assessments for monitoring progress more frequently, focusing on students at risk (Reading 3D Brochure, 2009).

The mClass Reading 3D assessment is conducted using a computer. In the TRC component, the teacher follows along on the computer recording the student's performance while the student reads from a set of leveled reading books. In the DIBELS component, the teacher guides the student through a series of 1-minute probes and records their performance on the computer. This technology allows teachers to capture the data from the assessment in one central place for a full picture of a student's reading development. The instant access to data from the computer gives teachers critical knowledge about their students in real time. They can immediately turn that knowledge

into action by adjusting their teaching strategies to fit their students' needs (Montgomery County Public Schools, 2005).

Purpose of the Study

The purpose of this study was to investigate the relationship between results from the NCEOG Assessment of Reading Comprehension and results from the mClass Reading 3D assessment, especially examining the degree to which mClass Reading 3D predicts scores on the reading comprehension portion of the NCEOG.

Significance of the Study

Many studies exist examining the relationship between the DIBELS ORF component of mClass Reading 3D and state high-stakes tests; however, there is limited research related to using the mClass Reading 3D assessment as a whole (DIBELS and TRC) to predict achievement on high-stakes tests. Currently, no research has been found related to using the TRC assessment component as a predictor of performance on the reading comprehension portion of the NCEOG assessment.

NCLB (2001) mandates that each child progresses toward the same standards measured by a statewide system of accountability; therefore, the academic progress of each student should be monitored frequently through the use of effective formative assessment tools. Research on the formative assessment tools and their ability to predict performance on high-stakes tests is necessary for teachers to be able to accurately base instructional decisions on the data provided.

Research Questions

1. What is the relationship between the mClass Reading 3D assessment and the NCEOG Reading Comprehension assessment?
2. To what extent does the mClass Reading 3D assessment accurately predict

student scores on the NCEOG Reading Comprehension assessment?

Definition of Terms

Adequate yearly progress. AYP is the measure by which schools, districts, and states are held accountable for student performance under Title I of NCLB (AYP, 2004).

Benchmark assessment. Benchmark assessments are tests administered throughout the school year to give teachers immediate, formative feedback on how their students are performing (Fournier, North, & LaPointe, 2009).

Comprehension. Comprehension is intentional thinking during which meaning is constructed through interactions between text and reader (NICHD, 2000).

Corrective Action. Corrective Action is a sanction from the NCLB legislation that occurs when a school has failed to meet AYP goals for 4 consecutive years. The school must implement certain corrective actions, including one of the following areas: replace school staff relevant to the failure, institute and implement a new curriculum, significantly decrease management authority in the school, appoint outside experts to advise the school, extend the school year or school day, and restructure the internal organization of the school (Great Schools, n.d.).

Curriculum-based measurement (CBM). CBM is a methodology for indexing student proficiency in the curriculum (Madelaine & Wheldall, 1999).

Formative assessment. Formative assessments are assessments designed to monitor student progress during the learning process (Chappuis & Stiggins, 2002).

High-stakes assessment. High-stakes assessment is the evaluation of individual performance through assessment when the data have significant (high) consequences (stakes) (Braden & Schroeder, 2004).

Intervention. Interventions are instructional approaches and programs designed

to either prevent or remediate persistent academic difficulties (Tunmer, 2008).

mClass Reading 3D. mClass Reading 3D is assessment software that incorporates phonemic awareness, phonics, and fluency assessments with print concepts, reading behaviors, and comprehension measures (mClass:Reading 3D, 2010).

No Child Left Behind Act. NCLB is a federal legislation that enacts the theories of standards-based education reform. NCLB ensures that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging state academic achievement standards and state academic assessments (No Child Left Behind Act Law and Legal Definition, n.d.).

Oral reading fluency. ORF is a standardized, individually administered test of accuracy and fluency with connected text (University of Oregon Center on Teaching and Learning, n.d.).

Progress monitoring. Progress monitoring involves the frequent and repeated collection and analysis of student performance data. It provides a standardized and empirical method for evaluating the effectiveness of interventions (Florida Association of School Psychologists, n.d.).

Running Records. RRs is a reading assessment tool that allows you to assess a student's reading performance as she/he reads from a benchmark book (RRs and Benchmark Books, n.d.).

School Choice. School Choice is a sanction from the NCLB legislation that occurs when a school has failed to meet AYP goals for 2 consecutive years. Parents have the choice to transfer their children to schools which are (1) not identified for school improvement and (2) not identified by the state as a persistently dangerous school (Great Schools, n.d.).

Summative assessment. Summative assessments are those assessments designed to determine a students' academic development after a set unit of material (Stiggins, 2002).

Supplemental Educational Services. Supplemental Educational Services is a sanction from the NCLB legislation that occurs when a school has failed to meet AYP goals for 3 consecutive years. These services include tutoring or other extra education services that provide academic aid to students (Great Schools, n.d.).

Title I. Schools where at least 40% of the children in the school attendance area are from low-income families or at least 40% of the student enrollment are from low-income families are eligible to receive federal Title I funds. Title I funds are to be used for programs designed to improve the academic achievement of children from low-income homes (Great Schools, n.d.).

Limitations and Delimitations

The research was restricted by the use of data that was collected from one elementary school in the southwestern piedmont area of North Carolina, creating a small population sample. The school participated in a pilot program by the NCDPI to implement mClass Reading 3D. Data from only one school were analyzed in this study because it was the only elementary school in the district that implemented mClass Reading 3D to all third-, fourth-, and fifth-grade students in the 2010-2011 school year. The limitations are further explained in Chapter 5.

Summary

Due to the lack of prevalent research available on this study topic, research was needed to determine if there was a significant relationship between the mClass Reading 3D assessment and the NCEOG Reading Comprehension assessment. The purpose of the

study was to investigate if the mClass Reading 3D assessment is an accurate predictor of the scores on the reading comprehension portion of the NCEOG. The results of the study are significant for educators due to the widespread use of these assessments and its implications to daily instruction.

The remaining chapters include pertinent information for understanding this study. Chapter 2 provides a review of relevant literature. The study's methodology is described in Chapter 3. A summary of the results is presented in Chapter 4. Chapter 5 provides the implications of this study and areas of future research.

Chapter 2: Literature Review

Introduction

The ability to read is one of the most important skills to foster academic success; however, today's students have a prevailing weakness in this area. mClass Reading 3D has been implemented in School A to provide data on students' reading achievement levels. These data are used to drive the classroom reading instruction and provide interventions for students in need of support. This study explored mClass Reading 3D's ability to accurately predict students' scale scores on the reading comprehension portion of the NCEOG assessment to determine if mClass Reading 3D is an effective reading assessment tool for School A.

This literature review discusses the relevant theory and research on teaching reading and reading assessments. A brief history of reading research and the elements of teaching reading provide context for the study. The identifiers for students at risk of reading failure are explored. Interventions are investigated as a means to prevent reading failure. The remaining portions of the chapter focus on the foundations of the reading assessments involved in this study. The chapter concludes with a summary of the literature.

Brief History of Reading Research

There have been many attempts by researchers to determine the *best* way to teach reading. However, according to Snow et al. (1998), there has been no true consensus among groups of educators and researchers with regard to how to best teach children to read, causing "the reading wars" (p. v).

There are six influential U.S. studies that have shaped the field of reading research and contributed to the reading wars: (1) *The Cooperative Research Program in*

First-Grade Reading Instruction, (2) *Learning to Read: The Great Debate*, (3) *Becoming a Nation of Readers: The Report of the Commission on Reading*, (4) *Beginning to Read: Thinking and Learning about Print*, (5) *Preventing Reading Difficulties in Young Children*, and (6) *the Report of the National Reading Panel: Teaching Children to Read* (Cowen, 2003).

The 1955 publishing of the book *Why Johnny Can't Read* by Rudolf Flesch and the Russian launching of Sputnik I in 1957 prompted a committee from the National Conference on Research in English to investigate research that existed in reading and to launch one of the earliest comprehensive studies of how young children begin to learn to read (Cowen, 2003; Readance & Baron, 1998). Guy Bond and Robert Dykstra (1967) headed the study, coined as the First-Grade Studies. The First-Grade Studies examined beginning reading approaches, their effectiveness in relation to readiness, and environmental characteristics. Bond and Dykstra found, through 27 individual research projects, the importance of emergent literacy especially in the areas of systematic phonics and phonemic awareness instruction, and expressed a need for future research on the role of the teacher and the importance of teacher training through professional development (Bond & Dykstra, 1967, 1997; Cowen, 2003).

The other research that resulted from the National Conference on Research in English's meeting was the extensive study conducted by Jeanne Chall (1967) entitled *Learning to Read: The Great Debate*. Chall extensively reviewed previously published research, examined basal programs, visited classrooms, and interviewed many educators investigating the debate of whole language versus phonics approach in teaching children to read (Hunt, 1969). Her findings were similar to those found in the First Grade Studies even though they were conducted very differently. Chall found beginning reading should

have a code-emphasis approach, reading programs should be reexamined, reading grade levels should be reevaluated, better diagnostics and assessments should be developed, and more research should be conducted in the area of beginning reading instruction (Hunt, 1969). Her conclusions, coupled with the research of Bond and Dykstra (1967) set a course for future research in beginning reading (Chall, 1967; Cowen, 2003; Hunt, 1969).

Despite the research of Chall (1967), Bond and Dykstra (1967), and others, literacy problems still continued in the United States (Cowen, 2003). This was made evident to the public in a report published by NCEE (1983) entitled *A Nation at Risk: The Imperative for Educational Reform*. In response to this report, the National Academy of Education, National Institute of Education, and Center for the Study of Reading sponsored the most significant literacy research study since the 1960s known as *Becoming a Nation of Readers* (Anderson, Hiebert, Scott, Wilkinson, & The Commission on Reading, 1985) by the Commission on Reading (Cowen, 2003). This study concluded that skilled reading must be a constructive process, be fluent, be strategic, require motivation, and be a lifelong pursuit (Binkley, 1986). *Becoming a Nation of Readers* (Anderson et. al, 1985) stresses the importance of reading for meaning, text structure, a balanced reading approach, reading in context, reading authentic literature, and automatic word recognition more than other studies (Cowen, 2003; NCEE, 1983).

The Center for Reading chose Marilyn Jager Adams in 1986 to lead their research study on all aspects of phonics and early reading instruction in response to a call for research proposals set forth by the U.S. Department of Education's Reading Research and Education Center (Adams, 1990). Adams published *Beginning to Read: Thinking and Learning about Print* in 1990. This study was viewed almost as a sequel to Chall's

(1967) work. Adams had a theoretical advantage over Chall (1967) due to the amount of current and prior research on basic reading processes that existed in 1989 versus 1967 (Adams, 1990). Her book provides a complete review with the purpose of bringing balance and reason to the reading wars that still exist. Adams found that reading approaches that used both phonics and whole language components demonstrated higher results in reading achievement. She suggested that reading instruction should include phonemic awareness, phonics, independent reading, authentic literature, automaticity, and read alouds (Adams, 1990; Cowen, 2003). Her findings have shaped the studies that form today's foundations for teaching reading.

Adams (1990) provided an extensive review of reading research; however, her research did not provide any solutions to identify effective interventions for struggling readers. The U.S. Department of Education and National Academy of Sciences developed a report entitled *Preventing Reading Difficulties in Young Children* in 1998 to combat the growing demands for a higher literate and global society within the United States. With the demands of a literate, higher level thinking society increasing, the National Academy of Sciences conducted a study to determine effective interventions for young children at risk of learning how to read (Cowen, 2003; Snow et al., 1998). The report determined that there were no interventions which could take the place of a well-trained teacher (Snow et al., 1998). The authors of the study found that teachers of young children should use the following reading instructional methods: obtain meaning from print, frequent and intensive opportunities to read, exposure to regular spelling-sound relationships, determine the nature of the alphabetic writing system, and understand the structure of spoken words (Cowen, 2003; Snow et al., 1998). The report also provided grade-level recommendations and strategies for teachers of at-risk students (Snow et al.,

1998).

Building upon the National Research Council Committee findings published in the report *Preventing Reading Difficulties in Young Children* (Snow et al., 1998), the most recent large-scale research analysis on reading instruction was conducted by the NRP whose report was published in 2000. The NRP was developed through a congressional charge to the NICHD in consultation with the Secretary of Education. The panel was to assess the status of research-based knowledge, including the effectiveness of various approaches to teaching children to read. The report provided the panel's conclusions, an indication of the readiness for application in the classroom of the results of the research, and a strategy for rapidly disseminating the information to facilitate effective reading instruction in the schools (NICHD, 2000).

The NRP developed and applied an objective research review methodology to analyze research literature relevant to a set of selected topics judged to be the central importance in teaching children to read (NICHD, 2000). The major topics were determined by an extensive review of research studies and regional public hearings. The topics adopted were alphabetic (phonemic awareness and phonics instruction), fluency, comprehension (vocabulary, text comprehension instruction, teacher preparation, and comprehension strategies instruction), teacher education and reading instruction, and computer technology and reading instruction. NRP developed subgroups of panelists, each assigned to one or more of the major topic areas (NICHD, 2000).

Each topic was then thoroughly researched by the subgroup using a set of rigorous research standards established by the NRP and the subcommittee itself. The research studies were coded and those that met the criteria were analyzed. A statistical meta-analysis was conducted on the topics that had a sufficient number of studies. The

topics with too few studies to conduct a meta-analysis had a more subjective qualitative analysis conducted to provide the best possible information. The findings of the NRP research are important towards a complete reading program, as it forms the elements of teaching reading today (NICHD, 2000).

Elements of Teaching Reading Today

The NRP found that certain reading instructional methods are more effective than others. The five instructional areas they found to be most effective in teaching reading were phonemic awareness, phonics, fluency, comprehension, and vocabulary. The combination of these five areas creates a well-rounded reading program (NICHD, 2000).

Phonemic awareness involves teaching children to hear and manipulate sounds in spoken words. Correlational studies have identified phonemic awareness and letter knowledge as the two best school-entry predictors of how well children will learn to read during the first 2 years of instruction. The NRP's meta-analysis found that teaching children to manipulate phonemes in words was highly effective under a variety of teaching conditions with a variety of learners across a range of grade and age levels and that teaching phonemic awareness to children significantly improves their reading more than instruction that lacks any attention to phonemic awareness (NICHD, 2000).

The lack of phonemic awareness is the most powerful predictor of difficulty in learning to read (Honig et al., 2000). If students cannot hear and manipulate sounds in spoken words, they have an extremely difficult time learning how to map those sounds to letters and letter patterns (Adams, 1990). In fact, research clearly shows that phonemic awareness can be developed through instruction and, furthermore, that doing so significantly accelerates children's subsequent reading and writing achievement (Adams, Foorman, Lundberg, & Beeler, 1998).

Phonics instruction is the study of letter-sound correspondences to help identify written words. The primary focus of phonics instruction is to help beginning readers to link letters and sounds together in order to form letter-sound correspondence and spelling patterns (NICHD, 2000). A primary difference between good and poor readers is the ability to use sound/spelling correspondences to identify words (Juel, 1991). The NRP's meta-analysis revealed that systematic phonics instruction produces benefits for students in kindergarten through sixth grade and for children having difficulty learning to read. Systematic phonics instruction also had a positive and significant effect on disabled readers' reading skills. These children improved substantially in their ability to read words and showed significant gains in their ability to process text (NICHD, 2000).

Chall (1996) concluded that comprehensive, systematic phonics-first instruction was overwhelmingly supported by the vast majority of research. She found that the failure to acquire and use efficient decoding skills begins to take a toll on reading comprehension by third grade (Chall, 1996). Adams's (1990) research led to the conclusion that direct, explicit instruction of the alphabetic principle is necessary for some children. Beck and Juel (1995) found the early attainment of decoding skill is important because this accurately predicts later skill in reading comprehension.

Vocabulary is the body of words students must understand in order to read text with fluency and comprehension (Honig et al., 2000). There is a strong relationship between reading ability and vocabulary acquisition in that the amount of reading students do, both in and out of school, is an indicator of students' vocabulary size. It follows that students need to develop strong reading skills to be able to engage successfully in the volume of reading necessary for them to learn large numbers of words (Honig et al., 2000). The NRP's meta-analysis found that vocabulary instruction does lead to gains in

comprehension, but that methods must be appropriate to the age and ability of the reader.

Other findings were:

1. Vocabulary should be taught both directly and indirectly.
2. Repetition and multiple exposures to vocabulary items are important.
3. Learning in rich contexts enhances the acquisition of vocabulary.
4. Direct instruction should include task restructuring as necessary and actively engage the students.
5. Dependence on a single vocabulary instruction method will not result in optimal learning. (NICHD, 2000, pp. 4-27)

Vocabulary knowledge is also fundamental to reading comprehension. In fact, research has shown that the proportion of difficult words in text is the single most powerful predictor of text difficulty, and a reader's general vocabulary knowledge is the single best predictor of how well that reader can understand text (Honig et al., 2000). According to Nagy (1988), increasing the volume of student reading is the single most important thing teachers can do to promote large-scale vocabulary growth.

Comprehension is the process of constructing meaning from written texts (Honig et al., 2000). Reading comprehension is where the readers derive meaning from text when they engage in intentional, problem-solving thinking processes (NICHD, 2000). The NRP meta-analysis suggests the rationale for the explicit teaching of comprehension skills is that comprehension can be improved by teaching students to use specific cognitive strategies or to reason strategically when they encounter barriers to understanding what they are reading. Explicit instruction in the application of comprehension strategies has been shown to be highly effective in enhancing understanding. The evidence suggests that teaching a combination of reading

comprehension techniques is the most effective. When used in combination, these techniques can improve results in standardized comprehension tests (NICHD, 2000).

Fountas and Pinnell (2006) found in their research that comprehension is the vital, central core of the complex ability to reason. Comprehension is critical to students in the later elementary grades because it provides the foundation for further learning in secondary school (Sweet & Snow, 2003). Research has shown there are seven categories that appear to provide a scientifically based foundation for the improvement of comprehension: (1) comprehension monitoring, (2) cooperative learning, (3) graphic and semantic organizers, (4) question answering, (5) question generation, (6) story structure, and (7) summarization. These techniques can be effective in improving comprehension of other content areas and standardized comprehension tests (Butler, Urrutia, Buenger, & Hunt, 2010).

Fluent readers are able to read with speed, accuracy, and proper expression (NICHD, 2000). It significantly affects the reader's ability to comprehend and it is the mark of a proficient reader (Honig et al., 2000). The NRP's meta-analysis found there are two prominent instructional approaches to teaching fluency, guided repeated oral readings and independent silent reading. The panel found that guided, repeated oral readings had a significant and positive impact on word recognition, fluency, and comprehension across a range of grade levels, but they were unable to find a positive relationship between programs and instruction that encourages large amounts of independent reading and improvements in reading achievement, including fluency (NICHD, 2000).

Snow et al. (1998) emphasized in their research study the important role of fluency in a proficient reading process. Fountas and Pinnell (2006) found in their

research that fluency in itself is not the goal of reading: The concern is its connection with comprehension because when fluency is disrupted so is understanding. Allington (1983) agreed: “The most compelling reason to focus instruction on fluency is the strong correlation between reading fluency and reading comprehension” (p. 560). In 1995, a large scale descriptive study found high correlations between scores on a rubric measuring phrased and fluent oral reading and scores on tests of comprehension (Fountas & Pinnell, 2006).

Clay’s (2013) research on current reading instruction found that most often it focuses on items of knowledge – words, letters, and sounds. Students search for links between the items and relate new discoveries to the old knowledge. They are searching for relationships within the complexity of text in order to simplify it. The problem with past literacy instruction is students reach an end point when they are taught only within specific text and letter/sound patterns.

Clay has found that reading instruction needs to have a self-extending system of literacy behaviors, meaning they learn more about reading every time they read independent of instruction. She believed that the emphasis in literacy instruction should be based on strategic activities, not a set of instructions. This causes the students’ processing to be more progressive and accumulative. Strategic activities involve the reader using their foundational knowledge of letters, sounds, and text features to form strategies of how to work on print. The strategies allow the reader to apply what he/she knows about one text to another, even one more difficult than their current skill set. Using what they know now about how print works can lead them to connect to new items of knowledge on other text later (Clay, 2013).

Strategic activities produce a reader who gradually constructs a network of

different processes for working on print and is able to carry over those skills when shifting to reading silently. This type of reading instruction ensures that the student can continue to learn to read by reading (Clay, 2013).

Identifying Students at Risk of Reading Failure

Many Americans' educational careers are imperiled because they do not read well enough to ensure understanding and to meet the demands of an increasingly competitive society (Snow et al., 1998). Research says that students who are not reading at grade level by the end of the first grade have a high probability of being a poor reader by the end of the fourth grade (Juel, 1988). The Matthew Effect theory suggests that "the rich get richer and the poor get poorer," meaning the literacy gap between students who learn to read early and those who struggle, only widens as they get older (Stanovich, 1986, p. 382).

The majority of reading problems faced by today's adolescents and adults are the result of problems that might have been avoided or resolved in their early childhood years (Snow et al., 1998). Snow et al. (1998) conducted a study and determined the three categories of predictors with potential influence on reading achievement: child-based risk factors, family-based risk factors, and neighborhood/community/school-based risk factors.

Child-based risk factors are due to general learning problems as a result of cognitive or sensory limitations. The primary conditions include cognitive deficiencies, hearing impairments, early language impairment, attention deficits, and other conditions such as visual impairments (Snow et al., 1998).

Family-risk factors consist of biological and environmental conditions.

Demographic data suggest that a majority of reading problems tend to occur in children

from poor families with little education, but not exclusively. Also, being a member of a family in which reading difficulties have previously occurred can be a risk factor (Snow et al., 1998).

Neighborhood/community/school-based risk factors are closely tied in with family-based factors. Where the family lives, the cultural and economic community in which the family is a part, and the school where the child attends all play a role in the literacy environment of the child (Snow et al., 1998).

Clay (2013) found three major areas that students may struggle in with reading. A child may have the necessary abilities but may not have learned how to use those abilities in reading. They do not make moves to solve their own problems. Another child may have made insufficient development in one ability area to carry out some new operation without special help. A third area is when a child may have knowledge about literacy but they are unable to connect the pieces together in order to gain meaning from the text.

Once a child's risk factors for reading are established then their area of struggle must be identified. In order to determine where a child is failing in reading, it must first be determined what a good reader does to read successfully. "It is important to know what effective reading processing is like because that gives us a vision for what we want to help struggling readers learn how to do" (Fountas & Pinnell, 2009, p. 16).

Pearson, Dole, Duffy, and Roehler (1992) summarized the strategies that active readers use when constructing meaning from text. Pearson et al. found proficient readers

1. search for connections between what they know and the new information they encounter in the texts they read;
2. ask questions of themselves, the authors, and the texts they read;

3. draw inferences during and after reading;
4. distinguish important from less important ideas in text;
5. adept at synthesizing information within and across texts and reading experiences; and
6. monitor the adequacy of their understanding and repair faulty comprehension.

Pressley (1976) and Keene and Zimmerman (1997) added sensory imaging to the list.

Fountas and Pinnell (2009) found that a proficient reader thinks within the text, beyond the text, and about the text. Thinking within the text consists of actions such as solving words, monitoring and correcting, searching for and using information, summarizing, maintaining fluency, and adjusting. Thinking beyond the text includes predicting, making connections, synthesizing, and inferring. Thinking about the text includes analyzing and critiquing the text. The key factor for these actions is the reader's ability to initiate the strategic activities to gain independent inner control of these complex behaviors (Fountas & Pinnell, 2009).

Some students struggle to become proficient readers, readers who have the knowledge and skills that are fundamental for reading at their grade-level standards (Peterson, Caverly, Nicholson, O'Neal & Cusenbary, 2000). Even though there are risk factors that are genetic and environmentally based, the majority of learning problems exist not within the child but in the inadequacy of the system to find a way to teach him/her (Fountas & Pinnell, 2009). Without intervention, children fall into behaviors that are often diagnosed as reading disabilities. What was an early weakness that would have responded to instruction becomes a long-term deficit (Fountas & Pinnell, 2009).

Fountas and Pinnell (2009) listed some common areas of reading difficulty that occur in children having trouble learning to read: (a) language processing – the ability to

use the systems of language while reading, (b) phonological processing – the understanding that speech is made up of sounds, (c) visual processing – the ability to notice and use visual features of letters and words, (d) use of background knowledge to construct meaning – the ability to bring information that exists in the reader’s head to the processing of print, (e) connecting reading and writing – the ability to acquire information in one area and use it to support learning in the other area, (f) reading fluency – the ability to use print and language to convey the meaning of the text in oral reading, (g) attention – the ability to sustain one’s attention while reading and direct that attention to the most helpful and useful information, (h) memory – the ability to remember and access information while reading, (i) processing actions/cognitive actions – the ability to initiate *in the head* activities while reading, and (j) emotion and motivation – the affective factors that have an impact on all areas of reading. Learners are very diverse and will not exhibit all of these signs. They may know some aspects of reading but are unable to connect it to others.

Interventions

As many as one in five children have difficulty learning to read (Foorman, Francis, Fletcher, Schatschneider, & Mehta, 1998). Richardson and DiBenedetto (1996) estimated 10-20% of all first graders require intervention for literacy-related skills. There will always be a percentage of children who are at risk of reading failure for a variety of factors. There is no single approach to teaching reading that will meet the needs of all children; however, it is imperative that students at risk of reading failure are identified and interventions are put into place to catch them up to grade-level standards.

Early identification and treatment is the most effective course of action for the prevention of learning disabilities in reading (Menzies et al., 2008). Taylor, Pearson,

Clark, and Walpole (2000) determined that reading achievement in the primary grades is associated with instruction that includes the use of early interventions. Early identification and intervention in specific areas of deficit can improve children's skill levels immediately and prevent later difficulties (NICHD, 2000; Snow et al., 1998). Velluntino and Scanlon (2001) found that school-based, small group interventions are successful at improving student achievement as determined by performance on reading outcome measures. This is particularly true if the interventions are tailored towards specific areas of need for the student (Juel, 1988).

Research has proven that early intervention in reading increases student literacy achievement, but some students continue to struggle after third grade. This is, in part, due to the emphasis on reading instruction changing from learning to read to reading to learn beginning in third grade. This means that students who do not read proficiently by the end of third grade may face serious consequences in their academic achievement (Wanzek et al., 2013).

Chall and Jacobs (1983) found that many low-income third graders reading at grade level experienced a drop in reading scores by fourth grade. Scammacca et al. (2007) conducted a meta-analysis to address the reading interventions for struggling readers in Grades 4-12. The findings indicated the largest effects were found in those interventions that focused on multiple components of literacy instruction and comprehension. Wanzek et al. (2013) extended Scammacca's meta-analysis and found that extensive interventions can have a positive impact on student learning across a variety of reading outcomes for students in Grades 4-12. This verifies the value of continued reading intervention for struggling readers beyond third grade.

Menzies et al. (2008) conducted a study that evaluated the extent to which

students at risk for reading difficulties attained grade-level reading expectations when given instruction targeted to their needs. The researchers used ongoing assessment, a lower student-teacher ratio, and differentiated instruction as interventions to assist students in reading at grade-level proficiency. They found through their study that the students demonstrated significant growth over time. The intensive instruction resulted in reading gains for all study participants, with 90% reaching grade-level proficiency at the end of the year. Eight of 16 children who were identified as at risk for reading problems at the beginning of the year demonstrated advanced or above grade-level reading ability in the spring.

Another study that revealed the effectiveness of reading interventions was conducted by Fuchs, Fuchs, Mathes, and Simmons (1996). The researchers evaluated the effectiveness of an intervention that used a class-wide peer tutoring program in reading. It was evaluated for its effectiveness with three learner types: low achievers with disabilities, low achievers without disabilities, and learners of average achievement. They found that regardless of the type of measure and type of learner, students in peer tutoring classrooms demonstrated greater reading progress than the control students. The results also indicated that the lowest achieving students consistently make significantly better reading gains on multiple measures of reading achievement than contrasting low-achieving students receiving typical reading instruction (Mathis, Grek, Howard, Babyak, & Allen, 1999).

Coyne, McCoach, and Kapp (2007) conducted a study examining the efficacy of targeted reading supplemental instruction of children at risk for reading difficulty. Even though the target of the study was to compare two reading intervention programs, the conclusion was one that impacts the use of reading intervention overall. The study

suggested that children most at risk of reading failure benefit substantially from supplemental reading interventions, make greater gains in understanding words targeted in the intervention as compared to words included only in classroom-based instruction, and the supplemental instruction helped reduce the word knowledge gap between the at-risk children and their peers (Coyne et al.).

The reading intervention strategies researched by Snow et al. (1998) in their study report entitled *Preventing Reading Difficulties in Young Children* were broken down into before kindergarten, after kindergarten, classroom strategies, and strategies for children with persistent reading difficulties. Some of the strategies discussed in the report were (a) teaching parents to teach children, (b) preschools with high literacy environments, (c) develop phonological awareness, (d) reading aloud, (e) predictable books, (f) language experience, (g) play-based instruction, (h) alphabetic instruction, (i) explicit instruction in phonemic awareness and phonics, (j) independent reading, (k) promote comprehension, (l) reading strategies, (m) small groups, (n) small class size, (o) reading tutors, (p) book buddies, (q) reading one-on-one, and (r) computer support.

Reading Assessments

While interventions are imperative to the prevention of reading disabilities, it is assessment that should drive the planning of the interventions for each individual student. Menzies et al. (2008) stated,

To adequately serve all children, especially those deemed at risk for reading failure, instruction must be both focused and comprehensive, which requires that teachers be able to accurately assess student needs and subsequently plan and deliver instruction based on that assessment. Otherwise it is difficult to ensure that all children will master the necessary skills to become proficient readers. (p.

67)

Clay (2013) determined through her research that reading assessments should be child specific, consisting of recording what the student does when processing texts of specified difficulty. The assessment should refer to the student's skill strengths and weaknesses and literacy moves made while processing the text. The results should be compared with a model of similar behaviors used by children who make satisfactory progress in reading. The instruction that follows should be based on the results found in the individualized reading assessment.

Baily and Drummond (2006) conducted a study on teacher rationale for selecting students at risk of reading failure. Their findings suggested that teachers who did not use formal assessments as a source of evidence to determine an at-risk student were not able to sufficiently make accurate judgments of a student's full understanding of early literacy skills. The researchers supported teacher identification of reading difficulties to include proven literacy assessments in order to capture the intricacies of literacy-related skills that teachers are unable to determine by informal observation alone.

Coyne et al. (2013) conducted an experimental study which evaluated the effects of a supplemental beginning reading intervention and its effects on student performance. They found the experimental group, which had received the intervention with adjustments made based on student performance, outperformed the comparison group, who did not receive adjustments based on performance. The experimental group outperformed the comparison group on all posttest measures at the end of kindergarten and had a continued advantage with follow-up analyses at the end of first grade. Their findings suggested that adjusting intervention support in response to student performance was more advantageous for students.

This study evaluates the mClass Reading 3D assessment. This assessment is a combination of two smaller assessment measures: the DIBELS and the TRC. DIBELS is based on the approach to assessing known as CBM and TRC is based on the assessment known as RRs.

CBM. CBM is an alternative to commercially developed standardized assessments. CBMs began in the 1970s as a 6-year empirical research project out of the Institute for Research on Learning Disabilities in the University of Minnesota by Stanley Deno and associates. In an effort to make data on student achievement more a part of daily teacher decision making, Deno and associates developed a program of research. The primary goal of the research was to develop measurement and evaluation procedures that teachers could use routinely to make decisions about whether and when to modify a student's instructional program (Deno, 1985).

Deno (1985) stated the measures in CBMs would have to be

1. Reliable and valid if the results of their use were to be accepted as evidence regarding student achievement and the basis for making instructional decisions.
2. Simple and efficient if teachers were going to use them, or teach others to use them, to frequently monitor student achievement.
3. Easily understood so that the results could be clearly and correctly communicated to parents, teachers, and students.
4. Inexpensive since multiple forms were to be required for repeated measurement. (p. 221)

Reading CBM (R-CBM) probes are 1-minute measures based on grade-level curriculum materials that are sensitive to student growth over time. The administration

and scores are standardized and the probes are designed to assess fluency and accuracy. The probes consist of cloze, word identification, and read aloud tasks.

In 1989, Marston reviewed the existing research on CBM. Research in reading focused on two measures: word identification and reading aloud. The results of Marston's review provided support for the use of these two measures as indicators of general reading proficiency. In terms of reliability, results of five studies revealed test/retest reliability coefficients ranging from .82 to .97, with most coefficients above .90. In terms of validity, 14 studies were reviewed. Criterion-related validity coefficients ranged from .63 to .90, with most above .80 (Marston, 1989; Wayman, Wallace, Wiley, Ticha, & Espin, 2007).

Studies, as reported by Deno (2003), have shown high correlations (.65-.85) between CBM scores and performance on high-stakes tests. A meta-analysis study was conducted by Yeo in 2010. The purpose of the study was to examine the relationship between CBMs and statewide achievement tests in reading. His results found an overall correlation coefficient for 27 studies to be .689. Good, Simmons, and Kameenui (2001) found a significant relationship between R-CBMs and the Oregon Statewide Assessment, with 96% of students who met Grade 3 ORF benchmarks met or exceeded expectations on the OSA. A study conducted by Shapiro, Keller, Lutz, Santoro, and Hintze (2006) showed in Pennsylvania that CBMs have a moderate to strong correlation with midyear assessments in reading and standardized tests across school districts. The results suggest that CBMs can be one source of data that could identify those students likely to be successful or fail the statewide assessment measure.

CBM and high-stakes assessment. There have been studies conducted on CBM's ability to predict performance on high-stakes assessments. A study by Barger

(2003) correlated the CBM of ORF scores to the NCEOG assessment. He correlated the spring ORF score of 38 third graders in North Carolina to their NCOEG Reading assessment. Barger found the correlation between the two assessments was high ($r=.73$). The study shows the DIBELS ORF measure may be an accurate predictor of whether or not a student will achieve a proficient score on the NCEOG Reading assessment. One hundred percent of the 26 students who scored 100 correct words per minute (cwpm) or better achieved a passing score on the NCEOG assessment.

Buck and Torgesen (2003) conducted a study of 13 schools in Florida to determine the predictability of ORF scores on the Florida Comprehensive Assessment test (FCAT). They found there was a significant correlation between the scores of the two assessments ($r=.70$). Ninety-one percent of the students who read at or above the 110 words per minute benchmark achieved adequate performance on the reading section of the FCAT.

Wood (2006) studied the relationship between ORF and the Colorado Student Assessment Program (CSAP). The study included 281 students in third through fifth grade in Colorado. He found a significant relationship between ORF and the CSAP. Results indicated ORF predicted performance equally well for CSAP in third ($r=.70$), fourth ($r=.67$), and fifth grades ($r=.75$).

A study by Wilson (2005) compared ORF median scores of 241 third-grade students from three RF schools with the Arizona Instrument to Measure Standards (AIMS). The purpose was to establish the ability of the ORF to determine a correlation with the AIMS. The results showed the correlation between AIMS and ORF to be positive and moderately large ($r=.741$). Students with higher levels of fluency tended to score higher on AIMS. Students who were deemed at-risk on the ORF measure did not

meet proficiency on AIMS.

Rowell (2009) conducted a study to investigate the relationship between the first grade scores of the DIBELS ORF and the subsequent third-grade scores of the Alaska Standards-Based Assessments (SBAs) in reading. A Pearson's r statistical test was performed on the data from both scores ($r=.723$). The results indicated that there was a positive correlation between ORF and comprehension on the Alaska SBAs in reading. The results validate that the majority of the students who scored proficient or higher on the ORF measure also scored proficient or higher on the Alaska SBAs in reading.

Running Records (RRs). The TRC portion of the mClass Reading 3D assessment is based on RRs. RRs is a formative reading assessment developed by Dr. Marie Clay in the 1960s. "If RRs are taken in a systematic way they provide evidence of how well children are learning to direct their knowledge of letters, sounds, and words to understanding the messages in the text" (Clay, 2005, p. 49). RRs are taken to guide teaching, assess text difficulty, and capture student reading progress (Clay, 2005). The prime purpose is to understand how children are using what they know to get to the message of the text or what reading processes they are using (Clay, 2013).

Typically a student's progress in learning to read is measured by testing the number of letters, sounds, or words they know. Most of the time, however, the students are asked to read continuous text; putting together the message transmitted by the letters, sounds, and words (Clay, 2013). Although the skills portions of a literacy assessment are important indicators, they would not show a complete picture of a student's literacy ability without combining that with an RRs-type assessment. RRs provide real time data that translates immediately into a teaching/intervention decision.

An RRs is conducted one-on-one with a student. The student reads aloud a

teacher-selected leveled text. The teacher records on a piece of paper what the child reads using standard conventional marks; this captures accuracy and fluency rate. At the conclusion of the reading, the child retells what he/she read to display comprehension. The teacher analyzes the accuracy (percentage of words read correctly), self-correction ratio, categories of errors (meaning, visual, or structure), fluency, and comprehension to determine the student's accurate reading levels and areas of needed intervention (Clay, 2005).

There is little research published to date on the reliability and validity of RRs. John Ross (2004) stated,

Although running records are frequently used in research, there is little psychometric data available about the procedure. Evidence about reliability is mixed. The consequential validity of running records has not been addressed because the effects of the assessment have not been disentangled from the instructional treatments in which the assessment is embedded. (p. 5)

The only evidence of reliability found is produced from Marie Clay's dissertation in 1966 where she found the scoring of the error rates to be correlated at $r=.98$ and for self-correction rates, $r=.68$. A chi square test found no significant differences at the .01 level for the raters' recording and scoring of error and self-correction behaviors (Clay, 2013).

Ross (2004) conducted a controlled experiment in which a sample of schools implemented RRs as a strategy for aligning literacy instruction with students' needs. School RRs scores were compared to schools that implemented an alternative school improvement strategy. Ross found schools assigned to RRs treatment outperformed schools assigned to an alternative strategy. After controlling for prior school

achievement and collective teacher efficacy, the RRs intervention accounted for 12% of the variance in reading and 7% in writing.

There is little to no research published to date on the correlation and/or predictability of RRs and high-stakes testing. This study provides current research on the topic.

Summary

This review of literature focused on the need for effective reading instruction in order to increase student achievement. The advancements in reading research have shown part of effective instruction is providing interventions for students at risk of reading failure. The research has shown us that assessments like mClass Reading 3D need to provide data for teachers to effectively administer interventions in the classroom. The purpose of the interventions is for students to grow in their reading achievement and be successful on high-stakes tests like the NCEOG; therefore, it needs to be determined if mClass Reading 3D is an accurate predictor of student success on the NCEOG Reading Comprehension assessment.

It has been determined that a gap exists in the research. There are numerous studies that have been conducted on the DIBELS ORF portion of the mClass Reading 3D assessment and its predictability to high-stakes testing like the NCEOG. However, little to no research has been found to date on the TRC portion and its predictability to high-stakes testing and the NCEOG assessment; therefore, this study fills the gap by researching the mClass Reading 3D assessment as a whole (ORF and TRC) and its predictability to the reading comprehension portion of the NCEOG assessment. The results of this study support teachers, ensuring they are using data that adequately prepare their students to be academically successful on grade-level literacy standards and in the

area of reading.

Chapter 3: Methodology

Introduction

The purpose of this study was to investigate the relationship between results from the NCEOG Assessment of Reading Comprehension and the results from the mClass Reading 3D assessment, especially examining the degree to which mClass Reading 3D predicts scores on the reading comprehension portion of the NCEOG. The following two questions were investigated.

Research Questions

1. What is the relationship between the mClass Reading 3D assessment and the NCEOG Reading Comprehension assessment?
2. To what extent does the mClass Reading 3D assessment accurately predict student scores on the NCEOG Reading Comprehension assessment?

Participants

The potential participants in this study were the 225 students enrolled in third, fourth, and fifth grades in School A during the 2010-2011 school year. The school had a total enrollment of 443 students. Its ethnic population was made up of 57% African-American, 21% White, 17% Hispanic, and 5% multi-ethnic. School A's special populations consisted of 22% Exceptional Children (EC), 1% Academically and Intellectually Gifted (AIG), and 11% Limited English Proficient (LEP). The free and reduced-priced lunch recipients made up 82% of the school population.

Students were eligible for participation in the study if they met the following criteria: (a) enrolled in Grades 3-5 at School A during 2010-2011 school year, (b) obtained an ORF score and TRC score from mClass Reading 3D from the EOY benchmark assessment in May 2011, and (c) obtained a score from the reading

comprehension portion of the North Carolina EOG assessment in May 2011. Students identified as EC and LEP were included in the study as long as they were not tested using the NCEXTEND 1 or 2 for reading.

The study participants consisted of 143 third-, fourth-, and fifth-grade students in School A meeting the study eligibility requirements. Table 1 shows the demographics of the selected participants for the study as compared to School A and Grades K-5 in School A's school district. The data were retrieved through NCWISE and North Carolina TetraData online databases.

Table 1

Demographics of 2010-2011 Study Participants Compared to School and District

Category	Study participants	School A (Grades 3-5)	School district (Grades 3-5)
# of Students	143	225	15442
# in Third Grade	60	87	2644
# in Fourth Grade	46	73	2609
# in Fifth Grade	37	65	2681
# Black	89	135	2937
# White	26	48	10060
# Hispanic	19	31	1643
# Multi-Racial	9	11	527
# Male	70	114	7978
# Female	73	111	7464
# of Transfers (Mobility)	17	30	1269
# Academically Gifted	4	4	1145
# Exceptional Children	23	42	1918
# Limited English Proficiency	4	17	1007

Setting

The study took place in one public elementary school in the southwestern piedmont area of North Carolina (School A). The school was selected to participate based on the number of students being assessed with mClass Reading 3D in third, fourth,

and fifth grades, permission from the principal, and the professional interest of the researcher as an educator in the school district where the school is located. School A was the only school in the district screening all kindergarten through fifth-grade students using mClass Reading 3D in the 2010-2011 school year. The six other schools in the district using mClass Reading 3D in 2010-2011 were screening kindergarten through third-grade students and only the lowest 20% of their fourth- and fifth-grade students.

During the 2010-2011 school year, participants from 10 third-, fourth-, and fifth-grade classrooms were individually assessed by a trained classroom teacher using the mClass Reading 3D assessment. All mClass Reading 3D EOY measures were administered one-on-one in the back of the student's classroom or outside of the classroom in the hallway. All reading comprehension measures on the NCEOG took place in the student's classroom with a trained test administrator. Students who required testing accommodations were provided those according to their Individualized Education Plan (IEP).

Instruments

mClass Reading 3D is a formative assessment tool which combines the DIBELS assessment with the TRC assessment. The measures include benchmark assessments that are administered three times a year as well as ongoing assessments for monitoring progress more frequently, focusing on students at risk (Reading 3D Brochure, 2009).

Dynamic Indicators of Basic Early Literacy Skills. The DIBELS assessment portion of mClass Reading 3D is based on CBMs.

DIBELS are a set of procedures and measures for assessing the acquisition of early literacy skills from kindergarten through sixth grade. They are designed to be short (1 minute) fluency measures used to regularly monitor the development

of early literacy and early reading skills. (General Information about DIBELS Measures, 2008, p. 1)

DIBELS were designed for use in identifying children experiencing difficulty in basic early literacy skills in order to provide support early and prevent the occurrence of later reading difficulties. DIBELS were designed to evaluate the effectiveness of interventions for those children receiving support to maximize learning growth (Dynamic Measurement Group, 2008).

The DIBELS component consists of seven measures: ISF, LNF, PSF, NWF, ORF, RTF, and WUF. These measures enable teachers to determine student phonetic, phonemic awareness and fluency abilities (mClass:Reading 3D, 2010). The measures are based on the five big ideas of reading determined by the NRP research – phonemic awareness, phonics, fluency, comprehension, and vocabulary (NICHD, 2000).

mClass Reading 3D requires third-, fourth-, and fifth-grade students to be benchmark assessed only on the ORF measure of DIBELS. The other measures are completed in kindergarten through second grade. Students who are considered performing below grade-level proficiency in Grades 3-5 may take the kindergarten through second-grade assessments as needed to determine areas of intervention through progress monitoring. This study focused on the ORF measure of DIBELS since the study's participants were third-, fourth-, and fifth-grade students and all were benchmark assessed on the ORF measure in May 2011.

The ORF measure targets fluency and comprehension. Students accurately and fluently read three grade-level passages in three 1-minute probes. mClass Reading 3D takes the median of all three probes to determine an overall score. The goal is that by the end of the year, third graders will read 110 words per minute, fourth graders will read 118

words per minute, and fifth graders will read 124 words per minute. Table 2 indicates the 2011 ORF benchmark cut point ranges for Grades 3, 4, and 5 as determined by the University of Oregon Center on Teaching and Learning.

Table 2

2011 ORF Score Benchmark Cut Point Ranges for Grades 3-5

Grade Level	Proficiency Level	BOY	MOY	EOY
3	Way Below	0-52	0-66	0-79
	Just Below	53-72	67-91	80-109
	At or Above	≥ 77	≥ 92	≥ 110
4	Way Below	0-70	0-82	0-95
	Just Below	71-92	83-104	96-117
	At or Above	≥ 93	≥ 105	≥ 118
5	Way Below	0-80	0-93	0-102
	Just Below	81-103	94-114	103-123
	At or Above	≥ 104	≥ 115	≥ 124

ORF is a reliable and valid assessment. According to the Dynamic Measurement Group (2008), the alternate form reliability for ORF is .90 for one probe, and the criterion related validity for ORF is .70-.80.

This study used the data from the DIBELS ORF EOY benchmark administration which occurred in May 2011. Each student was individually assessed by a trained classroom teacher to determine an overall score. The overall score was compared to the cut point goal for EOY benchmark in third, fourth, or fifth grade depending on the grade level of the student.

Text reading and comprehension assessment. The TRC assessment is a digital form of RRs. During the TRC, students are asked to read a book and complete one to two comprehension tasks. The teacher observes and records the student's oral reading

behaviors through the administration of RRs to determine reading accuracy percentage. The comprehension components help teachers determine whether the student understands the meaning of the text. There are three comprehension tasks within the TRC assessment: retelling, oral comprehension, and written comprehension. Retelling is required for text levels E and below and asks students to retell the beginning, middle, and end of the story they just read. Oral comprehension is required for text levels D and above and asks students to answer five text-specific questions about the text they just read. Written comprehension is required for text levels F and above and asks students to answer two comprehension questions in written form about the text they just read.

The accuracy percentage and comprehension component(s) together determine the student's overall instructional reading level (Text Reading and Comprehension, 2010). The instructional reading level is comprised of the following criteria: the accuracy percentage is 90-94% and proficiency on the assigned comprehension task(s) – a minimum of two on retell and/or four on oral comprehension, and/or two on written comprehension (Wireless Generation, 2010). It is represented by a letter (A-Z) from the Fountas and Pinnell (2010) leveling system. Table 3 represents the text gradient for the Fountas and Pinnell (2010) leveling system.

Table 3

Text Gradient for the Fountas and Pinnell Leveling System

Reading Level	Grade-Level Equivalency
A-C	Kindergarten
B-I	Grade 1
H-M	Grade 2
L-P	Grade 3
O-T	Grade 4
S-W	Grade 5

The mClass Reading 3D software provides the teacher with directions and prompts to maintain assessment fidelity. As each task is completed, the computer automatically calculates the student's score and provides a risk evaluation. The score provided is the instructional reading level of the student. The cut points for each proficiency level are reported as at or above grade-level proficiency, just below grade-level proficiency, or far below grade-level proficiency (mClass Reading 3D, n.d.). The goal is by the end of third grade to read on a level P, by the end of fourth grade to read on a level S, and by the end of fifth grade to read on a level U (Wireless Generation, 2010). Table 4 indicates the 2011 TRC cut point ranges for Grades 3-5 as determined by mClass Reading 3D through Wireless Generation.

Table 4

2011 TRC Cut Point Ranges for Grades 3-5

Grade Level	Proficiency Level	BOY	MOY	EOY
3	Far Below	≤ J	≤ K	≤ L
	Just Below	K-M	L-N	M-O
	At or Above	≥ N	≥ O	≥ P
4	Far Below	≤ L	≤ O	≤ P
	Just Below	M-O	P,Q	Q,R
	At or Above	≥ P	≥ R	≥ S
5	Far Below	≤ P	≤ R	≤ S
	Just Below	Q,R	S	T
	At or Above	≥ S	≥ T	≥ U

TRC is a reliable and valid assessment. According to the National Center on Response to Intervention, TRC's marginal reliability is 0.86, and the inter-rater reliability is 0.73. The predictive validity is 0.76, and the concurrent validity is 0.72 (mClass Reading 3D, n.d.).

This study used the data from the TRC EOY benchmark administration which occurred in May 2011. Each student was individually assessed by a trained classroom teacher to determine their instructional reading level. The instructional reading level was compared to the cut point score for the EOY benchmark in third, fourth, or fifth grade, depending on the grade level of the student.

NCEOG Assessment of Reading Comprehension Edition 3. The NCEOG Assessment of Reading Comprehension is administered each year to students in Grades 3-8 in the month of May. The reading comprehension measures of the NCEOG are designed to measure student performance on grade-level goals and objectives based on the North Carolina English Language Arts Standard Course of Study (NCDPI, 2011). The test is comprised of eight reading selections with corresponding questions for each selection. The reading selections vary from literary to informational text.

NCEOG scores are reported in achievement levels ranging from Level I to Level IV. Students must achieve at least a Level III to show grade-level reading comprehension skills and to be considered proficient. Table 5 indicates the 2011 achievement level cut scores for Grades 3-5 on the reading comprehension portion of the NCEOG as determined by NCDPI.

Table 5

2011 NCEOG Reading Achievement Level Ranges for Grades 3-5

Grade Level	Level I	Level II	Level III	Level IV
3	≤330	331-337	338-349	≥350
4	≤334	335-342	343-353	≥354
5	≤340	341-348	349-360	≥361

The NCEOG assessment is reliable and valid. The reliability is third grade–0.925, fourth grade–0.912, and fifth grade–0.900. The criterion-related validity is 0.66 for third grade, 0.63 for fourth grade, and 0.61 for fifth grade (NCDPI, 2011).

This study used the data from the reading comprehension measure of the NCEOG which occurred in May 2011. Each student was individually assessed by a trained test administrator to determine a scale score. The student scale score and proficiency level were compared to the achievement level cut point scores for third, fourth, or fifth grade, depending on the grade level of the student.

Design

This quantitative, correlational study utilized a predictive design to examine the relationship between scores on the mClass Reading 3D assessment and the reading comprehension portion of the NCEOG assessment. The predictive design was chosen for this study because it allows for calculation of the value of one variable based on the values of another variable (Madjidi, n.d.). In this case, the predictor variables are the scores from the mClass Reading 3D assessments, ORF and TRC; and the outcome measure is the EOG scale scores and proficiency levels. Using this design determined if the assessments from mClass Reading 3D were accurate predictors of student success on the NCEOG.

Procedures

The study was conducted in two parts in order to best address the two research questions. The sections below describe the procedure used to collect and analyze the data.

Data collection. The researcher collected archived data about the 143 study participants. The data collected included 2010-2011 demographic and NCEOG

assessment data from NCWISE and TetraData, the state student data collection and data analysis systems, along with ORF and TRC assessment scores from mClass Reading 3D. These data were coded according to category in preparation for data analysis.

The researcher collected and analyzed the quantitative data and answered the research questions based on the results. The purpose was to determine the relationship between the two assessments and whether mClass Reading 3D was a predictor of student success on the reading comprehension portion of the NCEOG.

Data analysis. Part one of the study focused on determining the relationship that exists between mClass Reading 3D and the reading comprehension portion of the NCEOG. The data collected and categorized were entered into the Statistical Package for the Social Sciences (SPSS) for analysis. Descriptive statistics (frequency counts and measures of central tendencies), measures of variability (standard deviations), and Pearson correlations (by gender and ethnicity) were calculated to determine any associations/relationships between the predictor variable (mClass Reading 3D) and the outcome measure (NCEOG).

Part two of the study focused on mClass Reading 3D's predictive ability for student success on the reading comprehension portion of the NCEOG. The assessment data collected and analyzed in part one from the ORF and TRC components of mClass Reading 3D and NCEOG were used in SPSS to calculate multiple regression analyses. A multiple regression analysis is a method of data analysis that may be appropriate whenever a quantitative variable is to be examined in relationship to any other factors (Berger, 2003). The analyses determined to what extent mClass Reading 3D predicted student scale scores on the NCEOG.

Summary

The purpose of this study was to investigate the relationship between results from the reading comprehension portion of the NCEOG and the results from the mClass Reading 3D assessment, especially examining the degree to which mClass Reading 3D predicted scores on the reading comprehension measures of the NCEOG. This chapter presented a discussion of the participants, setting, instruments, design, and procedure used in order to answer the research questions and fulfill the purpose of this study. Chapter 4 explains the analysis of data collected and answers to the research questions. Chapter 5 discusses interpretations, conclusions, and recommendations.

Chapter 4: Results

Introduction

The purpose of this study was to investigate the relationship between results on the NCEOG Assessment of Reading Comprehension and results on the mClass Reading 3D assessment, especially examining the degree to which mClass Reading 3D predicts scores on the reading comprehension portion of the NCEOG. This chapter describes how the data were collected and screened, analyzes the descriptive and inferential statistics that were utilized in order to address both of the research questions, and ends with a summary of the study results.

Research Questions

1. What is the relationship between the mClass Reading 3D assessment and the NCEOG Reading Comprehension assessment?
2. To what extent does the mClass Reading 3D assessment accurately predict student scores on the NCEOG Reading Comprehension assessment?

Data Collection and Screening Procedures

Prior to conducting statistical analyses, demographic and assessment data were gathered and entered into an Excel spreadsheet from NCWISE and TetraData (n.d.), the state student data collection and data analysis systems. Once the data entry was complete, it was copied into the SPSS statistical program database and appropriately coded with the assistance of a second viewer to maintain accuracy of coding and data entry.

There were a total of 225 students enrolled in third, fourth, and fifth grades in School A during the 2010-2011 school year. Students were eligible for participation in the study if they met the following criteria: (a) enrolled in Grades 3-5 at School A during

2010-2011 school year, (b) obtained an ORF score and TRC score from mClass Reading 3D from the EOY benchmark assessment in May 2011, and (c) obtained a score from the reading comprehension portion of the North Carolina EOG assessment in May 2011. Students identified as EC and LEP were included in the study as long as they were not tested using the NCEXTEND 1 or 2 for reading. After eliminating eight students not meeting the eligibility requirements, the resulting sample size consisted of 143 participants.

Descriptive and inferential statistics were both used to analyze the data collected from the study participants to determine the answers to the study's research questions. These analyses were calculated by grade level due to both mClass Reading 3D ORF scores and NCEOG scale scores changing achievement ranges at each grade level. The rest of the chapter discusses these analyses in terms of process and results by grade level, beginning with third grade (fourth and fifth grades follow).

Descriptive and Inferential Statistics for Third-Grade Participants

Descriptive statistics. Descriptive statistics is the analysis of data that helps describe, show, or summarize data in a meaningful way (Lund Research Group, 2013). Descriptive analyses were calculated for the third-grade study participants in order to determine what relationships existed between mClass Reading 3D and NCEOG.

Frequency distributions. Table 6 shows the results of the calculated frequency distributions on the data entered into SPSS to determine overall demographics for the third-grade study participants. The table includes frequencies for assessment proficiency levels as part of the demographics descriptors. Although the ORF and NCEOG scores change achievement ranges at each grade level, the proficiency levels have consistent descriptors across grade levels.

Table 2 in Chapter 3 reveals the specific ORF cut point ranges and Table 5 in Chapter 3 reveals the NCEOG achievement level ranges; both define the range of scores and their coordinating proficiency level. For example, according to Table 2, a third-grade student would be just below grade-level proficiency (yellow) if they scored an ORF score of 80-109 at the EOY benchmark; whereas, a fourth-grade student would be just below grade-level proficiency (yellow) if they scored an ORF score of 96-117 at the EOY benchmark. According to Table 5, a third-grade student would have an achievement level of II (limited understanding of grade-level standards) on the NCEOG if they scored a scale score of 331-337; whereas, a fourth-grade student would have an achievement level of II (limited understanding of grade-level standards) on the NCEOG if they scored a scale score of 335-342.

Table 6

Frequency Distributions for Third-Grade Study Participants (N=60)

Demographic Variable	n	Percentage of Third-Grade Participants
Male	30	50%
Female	30	50%
Black	38	63%
White	10	17%
Hispanic	9	15%
Multi-Racial	3	5%
AIG	2	3%
EC	10	17%
LEP	1	2%
NCEOG Level I	24	40%
NCEOG Level II	15	25%
NCEOG Level III	17	28%
NCEOG Level IV	4	7%
ORF Red	24	40%
ORF Yellow	24	40%
ORF Green	12	20%
TRC Red	20	33%
TRC Yellow	18	30%
TRC Green	22	37%

Note. AIG–Academically and Intellectually Gifted; EC–Exceptional Children; LEP–Limited English Proficient; NCEOG I–minimal understanding of grade-level standards; NCEOG II–limited understanding of grade-level standards; NCEOG III–grade-level proficient; NCEOG IV–above grade-level proficient; ORF/TRC Red–way below grade-level proficiency; ORF/TRC Yellow–just below grade-level proficiency; ORF/TRC Green–at or above grade-level proficiency.

Measures of central tendency and variability. Along with the frequency distributions, the measures of central tendency and variability were calculated to determine the mean, median, and standard deviations of the third-grade participants' assessment scores. The results are found in Table 7. For the purposes of statistical analyses, the researcher recoded the TRC levels from the assigned Fountas and Pinnell (2010) system letters to researcher-selected number codes. See the Appendix for the

coding key.

Table 7

Descriptive Statistics for Third-Grade NCEOG Scale Score, ORF Score, and TRC Score

Measure	Category	n	Mean	Median	Range	Standard Deviation
NCEOG	All Students	60	331.78	334.00	116	15.17
	Black	38	329.71	332.00	109	16.87
	White	10	336.00	338.00	47	14.18
	Hispanic	9	333.56	335.00	20	7.20
	Multi-Racial	3	338.67	341.00	25	12.66
	Male	30	334.40	335.50	47	11.62
	Female	30	329.17	333.50	104	17.87
	EC	10	318.00	322.00	97	26.64
	AIG	2	352.00	352.00	20	14.14
ORF	All Students	60	88.72	83.50	138	30.91
	Black	38	82.79	79.50	134	79.50
	White	10	106.90	103.50	105	34.04
	Hispanic	9	100.67	99.00	75	23.03
	Multi-Racial	3	67.33	74.00	58	29.57
	Male	30	90.20	84.50	138	32.64
	Female	30	87.23	81.50	125	29.57
	EC	10	74.70	79.00	103	35.39
	AIG	2	153.00	153.00	22	15.56
TRC	All Students	60	309.35	310.00	15	3.65
	Black	38	308.74	309.50	13	3.49
	White	10	311.90	312.00	12	3.60
	Hispanic	9	308.33	308.00	8	2.45
	Multi-Racial	3	311.67	314.00	11	5.86
	Male	30	309.97	310.00	13	3.69
	Female	30	308.73	308.50	15	3.57
	EC	10	307.60	307.00	13	4.90
	AIG	2	316.00	316.00	0	.00

Note. EC–Exceptional Children; AIG–Academically and Intellectually Gifted. Limited English Proficient (LEP) students were not included because of a low number of participants (n=1). TRC levels have been recoded from letters to numbers, see the Appendix.

Observations were made based on the analysis of the measures of central tendencies and variability. The mean NCEOG scale score for all third-grade participants ($M=331.78$) was equivalent to a proficiency level of II, limited understanding of grade-

level standards. The mean ORF score ($M=88.72$) and mean TRC score ($M=309.35$) for all third-grade participants were equivalent to just below grade-level proficiency (yellow).

With respect to ethnicity (Figure 1), White ($M=336.00$), Hispanic ($M=333.56$), and Multi-Racial ($M=338.67$) students' mean NCEOG scale scores were equivalent to a proficiency level of II, limited understanding of grade-level standards; the mean scale score of Black ($M=329.71$) students fell to the proficiency level of I, minimal understanding of grade-level standards. Black ($M=82.79$), White ($M=106.90$), and Hispanic ($M=100.67$) students' mean ORF scores were equivalent to just below grade-level proficiency (yellow); the mean ORF score for the Multi-Racial ($M=67.33$) students fell to way below grade-level proficiency (red). White ($M=311.90$) and Multi-Racial ($M=311.67$) students' mean TRC scores were equivalent to at or above grade-level proficiency (green); the mean TRC scores for Black ($M=308.74$) and Hispanic ($M=308.33$) students fell to just below grade-level proficiency (yellow).

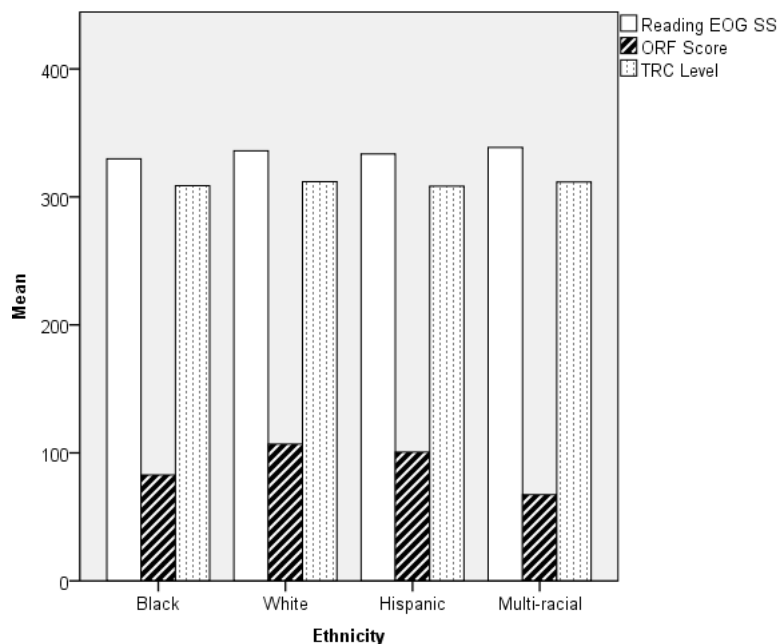


Figure 1. NCEOG, ORF, and TRC Mean Scores by Third-Grade Ethnicity.

With respect to gender (Figure 2), Male ($M=334.40$) participants scored a mean NCEOG scale score equivalent to a proficiency level of II, limited understanding of grade-level standards; the mean scale score for Female ($M=329.17$) students fell to a proficiency level of I, minimal understanding of grade-level standards. Male participants' mean ORF score ($M=90.20$) and a mean TRC score ($M=309.97$), and Female participants mean ORF score ($M=87.23$) and mean TRC score ($M=308.73$) were equivalent to just below grade-level proficiency (yellow).

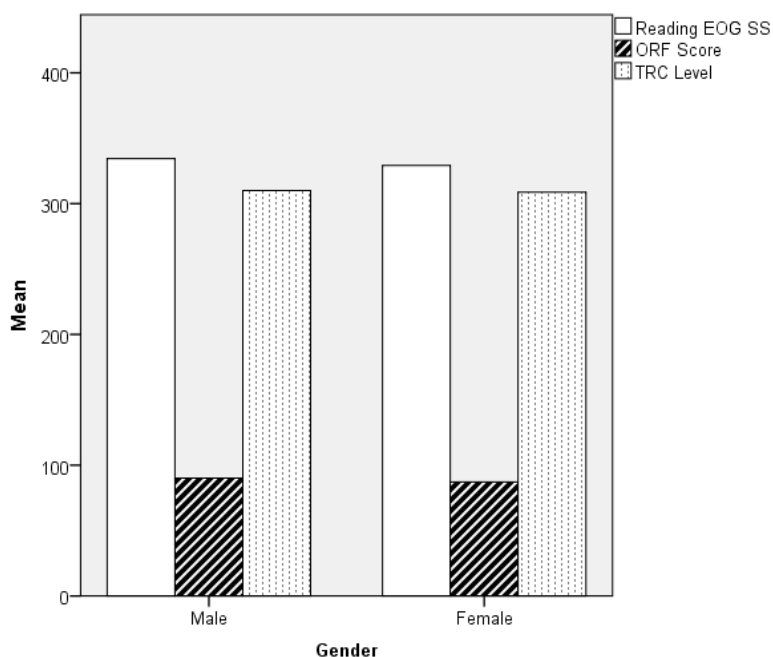


Figure 2. NCEOG, ORF, and TRC Mean Scores by Third-Grade Gender.

With respect to students identified with special needs (Figure 3), AIG students ($M=352.00$) scored a mean NCEOG scale score equivalent to a proficiency level IV, above proficiency of grade-level standards; EC students ($M=318.00$) fell to a proficiency level of I, minimal understanding of grade-level standards. AIG students' mean ORF score ($M=153.00$) and mean TRC score ($M=316.00$) were equivalent to at or above grade-level proficiency (green); EC students' mean ORF score ($M=74.70$) and mean TRC score ($M=307.60$) fell to way below grade-level proficiency (red). LEP students were not included due to the low number of participants identified ($n=1$).

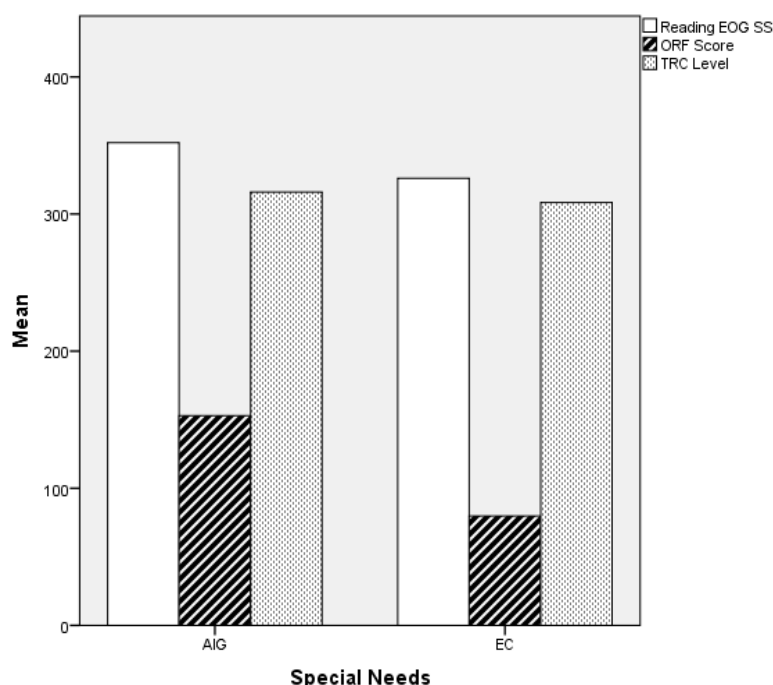


Figure 3. NCEOG, ORF, and TRC Mean Scores by Third-Grade Students with Special Needs.

Inferential statistics. Descriptive statistics do not allow researchers to make conclusions beyond the data that has been analyzed. However, inferential statistics can utilize the same data and make generalizations about a larger population (Lund Research Group, 2013). This study used methods of inferential statistics to further analyze the data collected and determine answers to the research questions. The criterion for all inferential tests of significance will utilize $\alpha=.05$.

Pearson product-moment correlation. The Pearson product-moment correlation was used to describe the relationship between the NCEOG and mClass Reading 3D assessments by determining the strength of the linear association between them (Lund Research Group, 2013). This, along with the calculated descriptive statistics, provided the answer to the research question “What is the relationship between the mClass Reading 3D assessment and the NCEOG Reading Comprehension assessment?”

The Pearson correlation was calculated by grade level due to the NCEOG scale score and ORF score achievement ranges changing by grade level. Preliminary analyses were conducted to allow for the study's findings to be easily generalized to a larger population and to support the study's validity. Scatter plots showed the relationships between the assessments to be positively linear (Figures 4, 5, and 6). One outlier was removed, case number 32, due to its unusually low NCEOG scale score (246), ORF score (29), and TRC score (301).

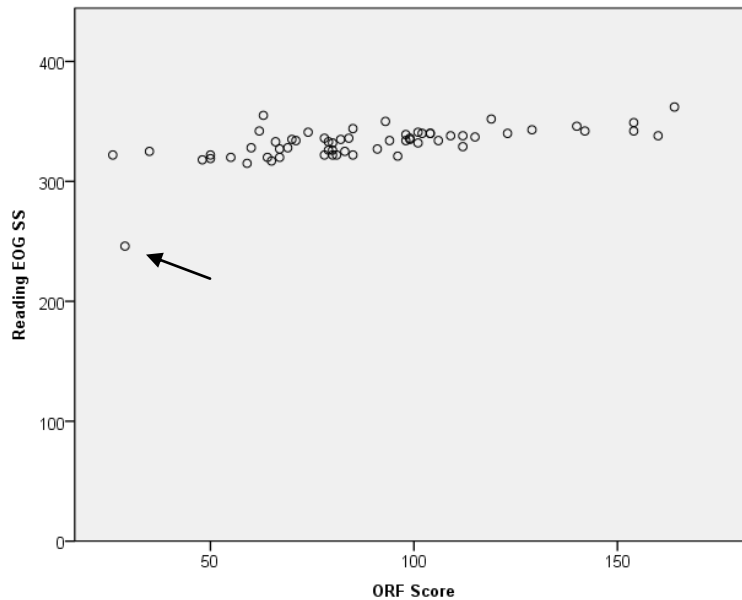


Figure 4. Scatter Plot of NCEOG and ORF Scores for Third-Grade Participants. The arrow indicates the outlier that was removed.

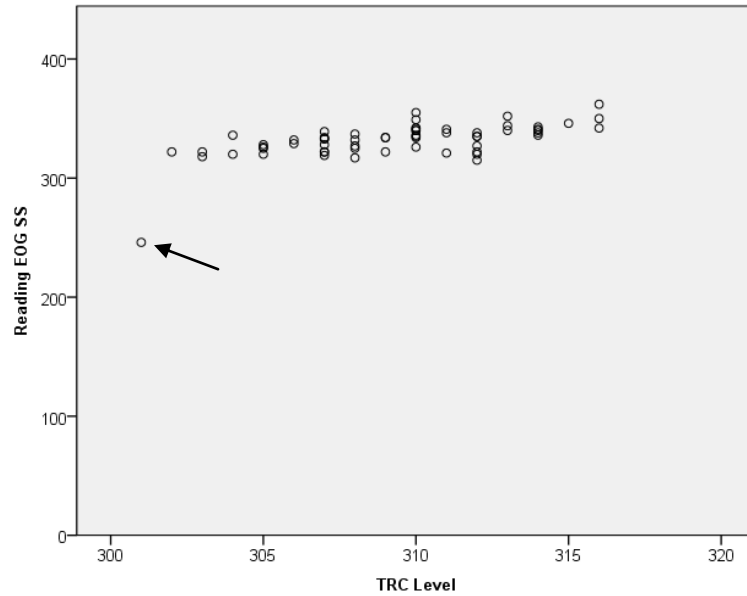


Figure 5. Scatter Plot of NCEOG and TRC Scores for Third-Grade Participants. The arrow indicates the outlier that was removed.

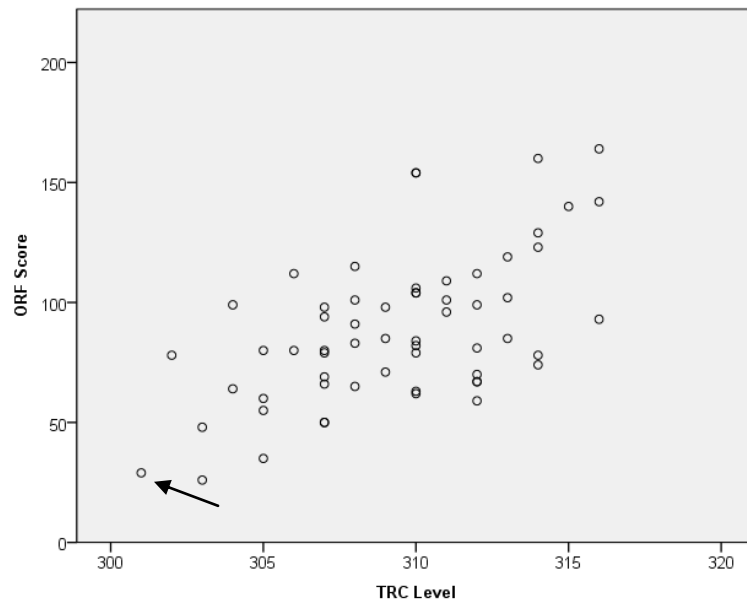


Figure 6. Scatter Plot of ORF Scores and TRC Scores for Third-Grade Participants. The arrow indicates the outlier that was removed.

All variables were normally distributed as assessed by the Shapiro-Wilk test.

This test is used with small sample sizes to determine if the data distribution is equal to a normal distribution (Lund Research Group, 2013). Table 8 shows the significance (Sig.)

($p > .05$) for each variable; therefore, the variables are normally distributed which is important in the generalizability of the study's results.

Table 8

Shapiro-Wilk Test for Third-Grade Participants' Assessment Scores

Assessment	Statistic	df	Sig.
NCEOG	.971	59	.163
ORF	.966	59	.097
TRC	.976	59	.302

Table 9 provides the Pearson correlations for third-grade NCEOG scale scores, ORF scores, and TRC scores. There was a positive correlation and statistically significant relationship ($\alpha = .05$) between NCEOG scale scores and ORF scores for third-grade participants, $r(59) = .654$, $p < .05$. This was determined by the positive correlation result of .654 ($r = .654$) and the correlation coefficient being greater than .5 (Lund Research Group, 2013).

There was a positive correlation and statistically significant relationship between NCEOG scale scores and TRC scores for third-grade students, $r(59) = .597$, $p < .05$. This was determined by the positive correlation result of .597 ($r = .597$) and the correlation coefficient being greater than .5 (Lund Research Group, 2013).

There was a positive correlation and statistically significant relationship between ORF scores and TRC scores for third-grade students, $r(59) = .556$, $p < .05$. This was determined by the positive correlation result of .556 ($r = .556$) and the correlation coefficient being greater than .5 (Lund Research Group, 2013).

Table 9

Pearson Correlation Matrix for Third-Grade Participant Assessment Scores

		NCEOG	ORF	TRC
NCEOG	Pearson Correlation			
	Sig. (2-tailed)	1		
	N			
ORF	Pearson Correlation	.654*	1	
	Sig. (2-tailed)	.000		
	N	59	59	
TRC	Pearson Correlation	.597*	.556*	
	Sig. (2-tailed)	.000	.000	1
	N	59	59	

Note. *Correlation is significant at the 0.01 level (2-tailed).

The researcher disaggregated the data from the third-grade NCEOG, ORF, and TRC scores for ethnicity and gender by calculating the correlation in the previous paragraphs separately for each ethnicity and gender variable. The Hispanic and Multi-Racial ethnic variables were combined and identified as Other ethnicity variable due to the low number of cases. Tables 10 and 11 show the correlations.

Table 10

Pearson Correlation Matrix for Third-Grade Scores by Ethnicity

Variable	Assessment	NCEOG	ORF	TRC
Black (N=37)	NCEOG	1	.595**	.554**
	ORF	.595**	1	.543**
	TRC	.554**	.543**	1
White (N=10)	NCEOG	1	.861**	.680*
	ORF	.861**	1	.747*
	TRC	.680**	.747*	1
Other (N=12)	NCEOG	1	.497	.654*
	ORF	.497	1	.213
	TRC	.654*	.213	1

Note. * $p < .05$, ** $p < .01$.

There was a positive correlation and statistically significant relationship for the Black and White third-grade participants between NCEOG scale scores and ORF scores (Black, $r(37) = .595$, $p < .05$; White, $r(10) = .861$, $p < .05$), between NCEOG scale scores and TRC scores (Black, $r(37) = .554$, $p < .05$; White, $r(10) = .680$, $p < .05$), and between ORF and TRC scores (Black, $r(37) = .543$, $p < .05$; White, $r(10) = .747$, $p < .05$). The third-grade Other ethnicity participants had a positive correlation between the NCEOG scale scores and ORF scores ($r(12) = .497$, $p > .05$), NCEOG scale scores and TRC scores ($r(12) = .654$, $p < .05$), and ORF and TRC scores ($r(12) = .213$, $p > .05$); however, the only correlation that was statistically significant was between the NCEOG scale scores and TRC scores.

Table 11

Pearson Correlation Matrix for Third-Grade Scores by Gender

Variable	Assessment	NCEOG	ORF	TRC
Male (N=30)	NCEOG	1	.624**	.599**
	ORF	.624**	1	.663**
	RC	.599**	.663**	1
Female (N=29)	NCEOG	1	.713**	.584**
	ORF	.713**	1	.419*
	RC	.584**	.419*	1

Note. * $p < .05$, ** $p < .01$.

There was a positive correlation and statistically significant relationship for the Male and Female third-grade participants between NCEOG scale scores and ORF scores (Male, $r(30) = .624$, $p < .05$; and Female $r(29) = .713$, $p < .05$), between NCEOG scale scores and TRC scores (Male, $r(30) = .599$, $p < .05$; and Female $r(29) = .584$, $p < .05$), and between ORF and TRC scores (Male, $r(30) = .663$, $p < .05$; and Female $r(29) = .419$, $p < .05$).

Overall, the data clarify that all three assessments have positive correlations and statistically significant relationships among the third-grade participants as a whole. Although close in correlation coefficients, the strongest correlation for all third-grade participants was between NCEOG and ORF scores where $r = .654$.

When the researcher broke down the data by ethnicity and gender, there were some noticeable observations. The Female and White variables held the highest correlation statistics for the NCEOG and ORF scores. The Male and White variables held the highest correlation for NCEOG and TRC scores and for the ORF and TRC scores, of those variables that were statistically significant.

Standard multiple regression analysis. Once the relationships had been

established, a standard multiple regression analysis was calculated to determine the answer to the research question “To what extent does the mClass Reading 3D assessment accurately predict student scores on the NCEOG Reading Comprehension assessment?”

Each standard multiple regression was calculated by grade level due to the NCEOG scale score and ORF score achievement ranges changing from one grade level to the next. A standard multiple regression was calculated to determine if ORF scores and TRC scores predicted NCEOG scale scores for third graders. Two outliers were removed after the casewise diagnostics for outliers was completed (see Table 12). Case 32, same case that was removed in the correlations statistics analysis, was removed due to its unusually low residual (as compared to the other variables) of -61.340 produced by its NCEOG scale score of 246. Case 58 was removed due to its unusually high residual (as compared to the other variables) of 25.496 produced by its NCEOG scale score of 355. Their removal supported the ability to generalize the results to larger populations and the validity of the study’s findings.

Table 12

Casewise Diagnostics for Outlier Results^a for Third-Grade Participants

Case Number	Std. Residual	NCEOG SS	Predicated Value	Residual
32	-5.436	246	307.34	-61.340
58	3.478	355	329.50	25.496

Note. a. Dependent Variable: NCEOG SS.

For a multiple regression to be a valid test to use, there are six assumptions that must be tested and held true: independence of errors/residuals, linearity between the independent and dependent variables, homoscedasticity of residuals (equal error

variances), no multicollinearity, no significant outliers or influential points, and the errors/residuals are normally distributed. These assumptions allowed the researcher to provide information on the accuracy of their predictions, test how well the regression model fits the study's data, determine the variation in the dependent variable explained by the independent variables, and test hypotheses on the regression equation (Lund Research Group, 2013). The researcher tested all six assumptions prior to calculating the multiple regression analysis.

There was an independence of errors as indicated by a Durbin-Watson statistic of 2.060. The Durbin-Watson result can range from zero to four, but a value of approximately two indicates that there is no correlation between residuals, meaning there was an independence of errors (Lund Research Group, 2013).

There was a linear relationship between the NCEOG and the ORF and TRC scores according to the scatter plot and partial regression plots completed as part of the analysis (Figures 7, 8, 9). The scatter plot (Figure 7) between third-grade residuals and predicated values formed a horizontal band to show the relationship between the dependent and independent variables was likely to be linear (Lund Research Group, 2013). The partial regression plots (Figures 8 and 9) between NCEOG (dependent variable), ORF and TRC (independent variables) also showed a linear relationship between the variables by forming horizontal bands.

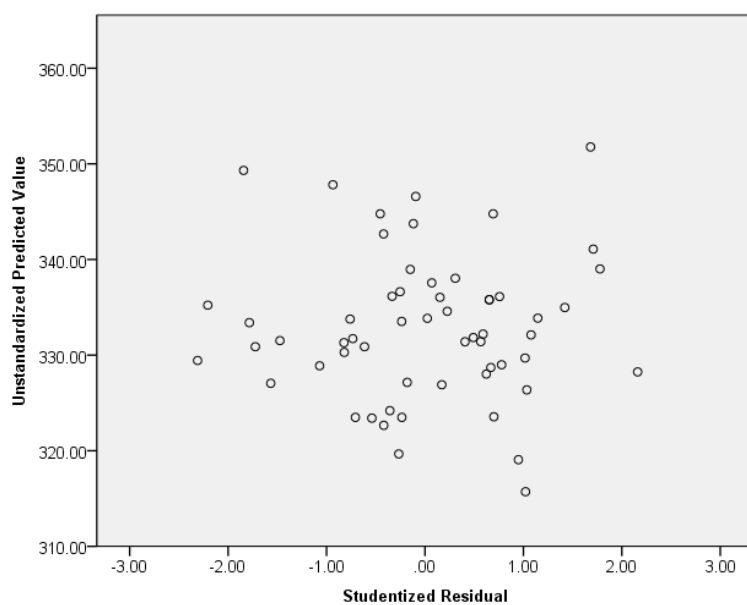


Figure 7. Scatter Plot of Multiple Regression Residuals and Predicted Values from Third-Grade Participants.

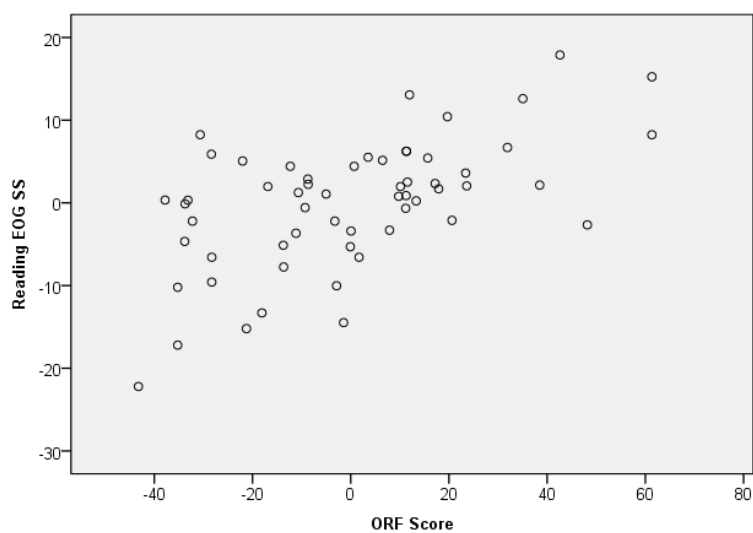


Figure 8. Partial Regression Plot of NCEOG and ORF Scores for Third-Grade Participants.

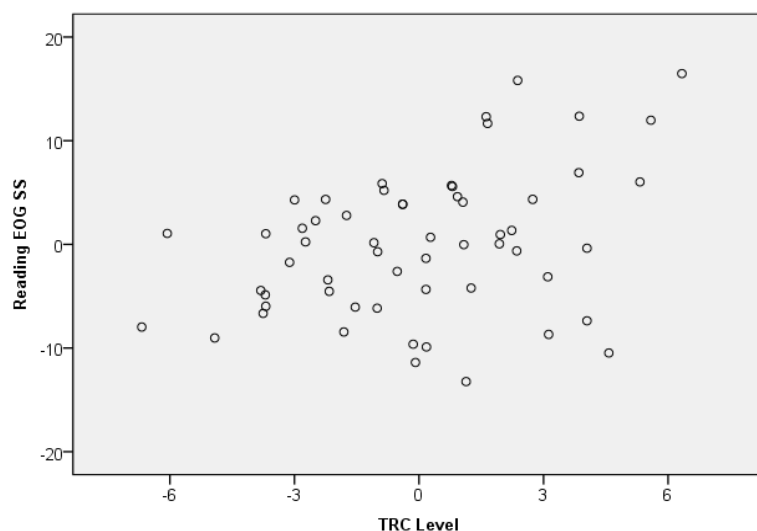


Figure 9. Partial Regression Plot of NCEOG and TRC Scores for Third-Grade Participants.

Homoscedasticity was tested to determine if the residuals were equal for all values of the predicted dependent variable. If there was homoscedasticity, the spread of the residuals would not increase or decrease as they moved across the predicted values (Lund Research Group, 2013). The scatter plot used to analyze linearity between the residuals and predicted values (Figure 7) was also used to test for homoscedasticity and determined that residuals were evenly spread across the predicated values.

Multicollinearity occurs when you have two or more independent variables that are highly correlated with each other. This leads to problems with understanding which variable contributes to the dependent variable. There are two stages to identifying multicollinearity: inspection of correlation coefficients and Tolerance values (Lund Research Group, 2013). Table 13 shows correlation coefficients for the dependent and independent variables. It was checked to determine that no independent variables had correlations greater than 0.7 (Lund Research Group, 2013). Since none of the correlation coefficients were larger than 0.7, the second stage, Tolerance, was tested. The Tolerance

values, seen in Table 14, were greater than 0.1, so the researcher could be confident that there was not a problem with multicollinearity within the third-grade participant data.

Table 13

Correlations Coefficients for Third-Grade Participant Data (N=58)

		NCEOG	ORF	TRC
Pearson Correlation	NCEOG	1		
	ORF	.721*	1	
	TRC	.617*	.562*	1

Note. * $p < .01$.

Table 14

Tolerance Values for Third-Grade Participant Data^a

Tolerance Value	
(Constant)	
ORF	.684
TRC	.684

Note. a. Dependent Variable: NCEOG.

No further outliers were detected through the casewise diagnostics and no residuals existed that had ± 3 standard deviations. After determining no existing outliers, the researcher tested for leverage. Testing for leverage identified those data points that were far away from the predictor values. If a data point has high leverage, it has a high potential to seriously alter the regression results if not removed (Simon, 2003). To determine whether any cases exhibited high leverage, the researcher considered leverage values less than 0.2 as safe, 0.2 to less than 0.5 as risky, and values above 0.5 as

dangerous (Lund Research Group, 2013). In this study, the highest leverage value was .11029 which allowed the researcher to determine there were no high leverage points.

The test for influential points was analyzed by utilizing Cook's distance values to measure for influence. Cook's distance determined data points with large residuals that may distort the outcome of the regression and affect its validity. Cook's distance values higher than one should be investigated and considered for removal (Lund Research Group, 2013). In this study the largest Cook value was .15171, which was below one, so the researcher determined there were no highly influential points.

A normal P-P plot (Figure 10) was used to test for normality. When analyzing inferential statistics, it is imperative that the residuals be normally distributed in order to support validity of study results. In a P-P plot, the residuals are normally distributed if the points align along the diagonal line; however, the residuals do not have to be perfectly aligned, only approximately (Lund Research Group, 2013). According to Figure 10, the researcher determined that the residuals were normally distributed.

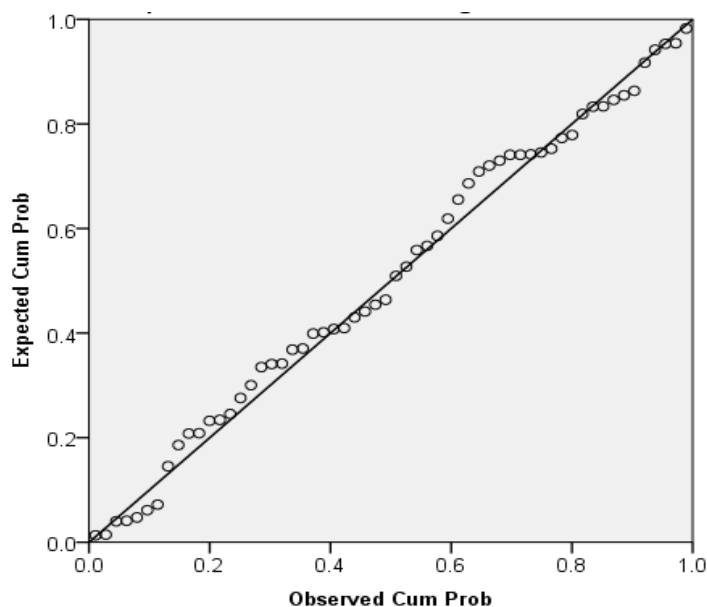


Figure 10. Normal P-P Plot of Third-Grade Regression Residuals (NCEOG-Dependent Variable).

All six assumptions were tested and held true for third-grade participant data, which led to the multiple regression test analysis. There were four measures that were used to determine how well the regression model fit the data: r , r^2 , adjusted r^2 , and statistical significance. Table 15 shows the results to the first three tests and Table 16 shows the results of statistical significance.

Table 15

Regression Model Summary^b for Third-Grade Participant Data

r	r^2	Adjusted r^2
.765 ^a	.585	.570

Note. a. Predictors: (Constant), TRC, ORF; b. Dependent Variable: NCEOG.

The r represents the multiple correlation coefficient which generalized the correlation coefficient r (found in Pearson correlation). R is considered one measure of

the quality of the prediction of the dependent variable. R ranges in value from zero to one. The higher the value of r the better the independent variable is at predicting the dependent variable (Lund Research Group, 2013). In this study the r value was .765 which indicates a high quality of prediction of the dependent variable.

The r^2 represents the proportion of variance in the dependent variable (NCEOG) that could be explained by the independent variables (ORF and TRC) (Lund Research Group, 2013). In this study ORF and TRC scores explained 58.5% ($r^2=.585$) of the variability of NCEOG scores. However, r^2 is based on all independent variables and assumes that all explain the variation, which can be considered a biased estimate, meaning it can be larger than it should be when generalizing to a larger population (Goodness of Fit, n.d.). The adjusted (adj.) r^2 attempts to correct for the bias by only including the independent variables that truly affect the dependent variable, therefore providing smaller values that are indicative of a larger effect size (Lund Research Group, 2013). In this study the adj. r^2 is 57% (adj. $r^2=.570$) which explained 57% of the variability ORF and TRC had on the NCEOG. The closer the value is to 100% (1) the better fit it is to the model (Goodness of Fit, n.d.). This study indicates that at 57% it is a moderate fit to the regression model.

Another way to test the model is the F-ratio. The F-ratio is configured through an ANOVA (Analysis of Variance) test that determines a ratio of the variance between groups to the variance within groups, ultimately whether or not the difference between the variables is statistically significant. Table 16 shows that the ORF and TRC statistically significantly predict the NCEOG, $F(2,55)=38.728$, $p<.05$, meaning that the regression model is a good fit for the data.

Table 16

ANOVA^a Test Results for Third-Grade Participant Data

	Sum of Squares	df	Mean Square	F	Sig.
Regression	3284.573	2	1642.286	38.728	.000 ^b
Residual	2332.324	55	42.406		
Total	5616.897	57			

Note. a. Dependent Variable: NCEOG; b. Predictors: (Constant), TRC, ORF.

Unstandardized coefficients (B_1) indicate how much the dependent variable varies with an independent variable when all other independent variables are held constant (Lund Research Group, 2013). Table 17 shows the unstandardized coefficients for this study. The B_1 for ORF is equal to .180. This means that for each point increase in ORF, there is an increase in the NCEOG scale score of .180. The B_1 for TRC is equal to .866. This means that for each level increase in TRC, there is an increase in the NCEOG scale score of .866.

Table 17 also shows the results of the standard multiple regression test to determine the statistical significance of each of the independent variables (ORF and TRC) on the dependent variable (NCEOG) to show predictability. If $p < .05$, the researcher can conclude that the coefficients are statistically significant. In this study, the TRC coefficients and the ORF coefficients for third-grade participants are statistically significant ($p < .05$).

Table 17

Summary of Standard Multiple Regression for Third-Grade Participants

Variables	B	SE _B	β	t	Sig.
(Constant)	48.583	89.303		.544	.589
ORF Score	.180	.035	.547	5.206	.000
TRC Score	.866	.294	.309	2.946	.005

Note. B=unstandardized regression coefficient; SE_B=standard error of the coefficient; β =standardized coefficient (beta).

After the assumptions of linearity, independence of errors, homoscedasticity, normality, and significance were met, the multiple regression test revealed that the ORF and TRC scores statistically significantly predicted NCEOG scale scores, $F(2,55)=38.728, p<.05, \text{adj. } r^2=.570$. Both ORF and TRC scores added statistically significantly to the prediction of NCEOG scale scores, $p<.05$.

A standard multiple regression analysis was computed including ethnicity and gender. The Black and Male variables were not included in the analysis because they were perfectly collinear (see Tables 18 and 19, Tolerance=.000), which, if included in the analysis, would result in the unstable estimation of model parameters, potentially greatly reducing the model's statistical power (Denis, 2011). Therefore, because the researcher could not accurately determine the statistical significance and predictability of full ethnicity and gender, this component was not included in the study.

Table 18

Variables Enter/Removed for Third-Grade Gender and Ethnicity^a

Model	Variables Entered	Variables Removed	Method
1	Female, White, Other Ethnicity ^b		Enter

Note. a. Dependent Variable: NCEOG; b. Tolerance=.000 limits reached.

Table 19

Excluded Variables from Analysis for Third-Grade Gender and Ethnicity^a

Variables	Beta In	t	Sig.	Tolerance	VIF	Min. Tolerance
Black	.b			.000		.000
Male	.b			.000		.000

Note. a. Dependent Variable: NCEOG; b. Predictors in the Model: (Constant), Female, White, Other Ethnicity.

The descriptive and inferential statistical analyses based on the data collected from the third-grade study participants revealed several consistent findings. A positive correlation between NCEOG and mCLASS Reading 3D revealed a relationship between the two assessments. The strongest relationship existed between NCEOG and ORF.

Descriptive and Inferential Statistics for Fourth-Grade Participants

Descriptive statistics. Descriptive analyses were calculated for the fourth-grade study participants in order to determine what relationships exist between mClass Reading 3D and NCEOG.

Frequency distributions. Table 20 shows the results of the calculated frequency distributions on the data entered into SPSS to determine overall demographics for the fourth-grade study participants. The table includes frequencies for assessment

proficiency levels as part of the demographics descriptors. Although the ORF and NCEOG scores change achievement ranges at each grade level, the proficiency levels have consistent descriptors across grade levels. Table 2 in Chapter 3 reveals the specific ORF cut point ranges and Table 5 in Chapter 3 reveals the NCEOG achievement level ranges; both define the range of scores and their coordinating proficiency level.

Table 20

Frequency Distributions for Fourth-Grade Study Participants (N=46)

Demographic Variable	n	Percentage of Fourth-Grade Participants
Male	23	50%
Female	23	50%
Black	29	63%
White	9	20%
Hispanic	5	11%
Multi-Racial	3	6%
AIG	1	2%
EC	9	20%
LEP	2	4%
NCEOG Level I	11	24%
NCEOG Level II	22	48%
NCEOG Level III	11	24%
NCEOG Level IV	2	4%
ORF Red	18	39%
ORF Yellow	16	35%
ORF Green	12	26%
TRC Red	11	24%
TRC Yellow	4	9%
TRC Green	31	67%

Note. AIG—Academically and Intellectually Gifted; EC—Exceptional Children; LEP—Limited English Proficient; NCEOG I—minimal understanding of grade-level standards; NCEOG II—limited understanding of grade-level standards; NCEOG III—grade-level proficient; NCEOG IV—above grade-level proficient; ORF/TRC Red—way below grade-level proficiency; ORF/TRC Yellow—just below grade-level proficiency; NCEOG IV Green—at or above grade-level proficiency.

Measures of central tendency and variability. Along with the frequency distributions, the measures of central tendency and variability were calculated to

determine the mean, median, and standard deviations of the fourth-grade participants' assessment scores. The results are found in Table 21. For the purposes of statistical analyses, the researcher recoded the TRC levels from the assigned Fountas and Pinnell (2010) system letters to researcher-selected number codes. See the Appendix for the coding key.

Table 21

Descriptive Statistics for Fourth-Grade NCEOG Scale Score, ORF Score, and TRC Score

Measure	Category	N	Mean	Median	Range	SD
NCEOG	All Students	46	335.65	338.50	123	21.32
	Black	29	334.48	338.00	106	18.44
	White	9	332.67	344.00	113	34.64
	Hispanic	5	341.20	342.00	13	5.45
	Multi-Racial	3	346.67	339.00	31	16.86
	Male	23	329.83	338.00	104	247.78
	Female	23	341.48	339.00	40	9.32
	EC	9	315.00	333.00	104	41.26
	LEP	2	339.50	339.50	7	4.95
ORF	All Students	46	100.61	103.50	158	34.73
	Black	29	100.07	101.00	148	31.47
	White	9	98.22	94.00	152	49.21
	Hispanic	5	106.00	114.00	64	25.93
	Multi-Racial	3	104.00	104.00	92	46.00
	Male	23	87.70	90.00	148	36.12
	Female	23	113.52	105.00	122	28.52
	EC	9	64.56	58.00	119	40.74
	LEP	2	118.50	118.50	9	6.36
TRC	All Students	46	409.46	411.00	11	3.44
	Black	29	409.52	411.00	11	3.33
	White	9	408.89	412.00	11	4.40
	Hispanic	5	410.60	412.00	5	2.19
	Multi-Racial	3	408.67	410.00	8	4.16
	Male	23	408.83	411.00	11	3.89
	Female	23	410.09	411.00	9	2.86
	EC	9	405.78	404.00	11	4.68
	LEP	2	411.00	411.00	2	1.41

Note. EC—Exceptional Children; LEP—Limited English Proficient. Academically and Intellectually Gifted (AIG) students were not included because of a low number of participants (n=1). TRC levels have been recoded from letters to numbers, see the Appendix.

Observations were made based on the analysis of the measures of central

tendencies and variability. The mean NCEOG scale score for all fourth-grade participants ($M=335.65$) was equivalent to a proficiency level of II, limited understanding of grade-level standards. The mean ORF score ($M=100.61$) and mean TRC score ($M=409.46$) for all fourth-grade participants was equivalent to just below grade-level proficiency (yellow).

With respect to ethnicity (Figure 11), Multi-Racial ($M=346.67$) students' mean NCEOG scale scores were equivalent to a proficiency level of III, grade-level proficiency. Hispanic ($M=341.20$) students fell to a proficiency level of II, limited understanding of grade-level standards; Black ($M=334.48$) and White ($M=332.67$) students fell further to the proficiency level of I, minimal understanding of grade-level standards. Black ($M=100.07$), White ($M=98.22$), Hispanic ($M=106.00$), and Multi-Racial ($M=104.00$) students' mean ORF scores were equivalent to just below grade-level proficiency (yellow). Hispanic ($M=410.60$) students' mean TRC scores were equivalent to at or above grade-level proficiency (green); Black ($M=409.52$), White ($M=408.89$), and Multi-Racial ($M=408.67$) students fell to just below grade-level proficiency (yellow).

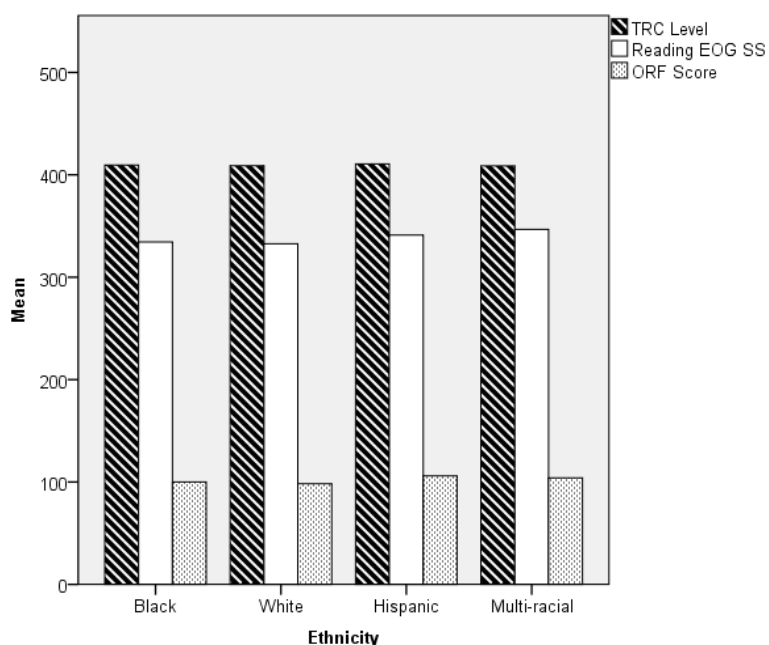


Figure 11. NCEOG, ORF, and TRC Mean Scores by Fourth-Grade Ethnicity.

With respect to gender (Figure 12), Female ($M=341.48$) participants scored a mean NCEOG scale score equivalent to a proficiency level of II, limited understanding of grade-level standards; the mean scale score for Male ($M=329.83$) students fell to a proficiency level of I, minimal understanding of grade-level standards. Female participants' mean ORF score ($M=113.52$) was equivalent to just below grade-level proficiency (yellow); Male ($M=87.70$) students fell to way below grade-level proficiency (red). Female participants' mean TRC score ($M=410.09$) was equivalent to at or above grade-level proficiency (green); Male ($M=408.83$) students fell to just below grade-level proficiency (yellow).

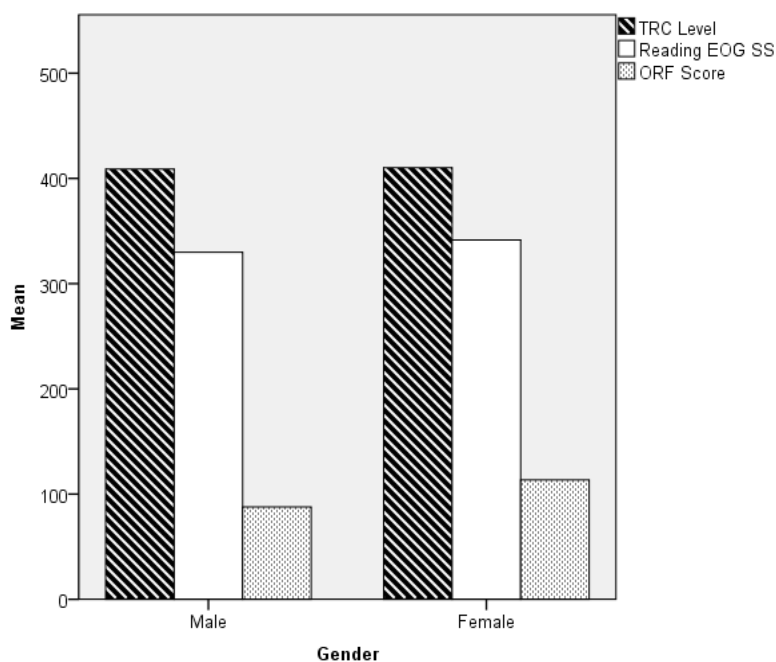


Figure 12. NCEOG, ORF, and TRC Mean Scores by Fourth-Grade Gender.

With respect to students identified with special needs (Figure 13), LEP students ($M=339.50$) scored a mean NCEOG scale score equivalent to a proficiency level II, limited understanding of grade-level standards; EC students ($M=315.00$) fell to a proficiency level of I, minimal understanding of grade-level standards. LEP students' mean ORF score ($M=118.50$) and mean TRC score ($M=411.00$) were equivalent to at or above grade-level proficiency (green); EC students' mean ORF score ($M=64.56$) and mean TRC score ($M=405.78$) fell to way below grade-level proficiency (red). AIG students were not included due to the low number of participants identified ($n=1$).

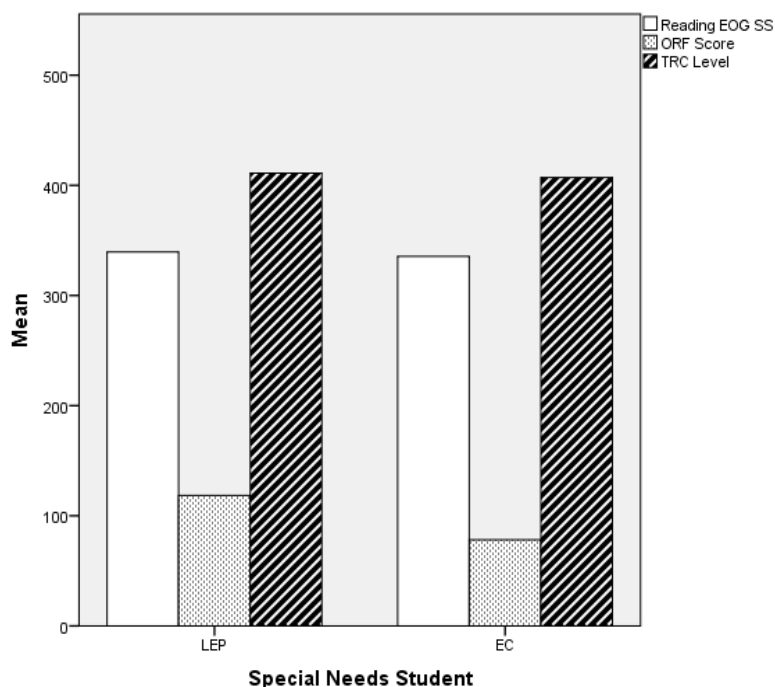


Figure 13. NCEOG, ORF, and TRC Mean Scores by Fourth-Grade Students with Special Needs.

Pearson product-moment correlation. The Pearson correlation was calculated and preliminary scatter plots showed the relationships between the assessments to be positively linear (Figures 14, 15, and 16). Three outliers were removed, case numbers 9 and 11, due to their unusually low NCEOG scale score (243), ORF score (14,20), and TRC score (401); and case number 32, due to its unusually high NCEOG scale score (366), ORF score (150), and TRC score (412) as observed on the scatter plots.

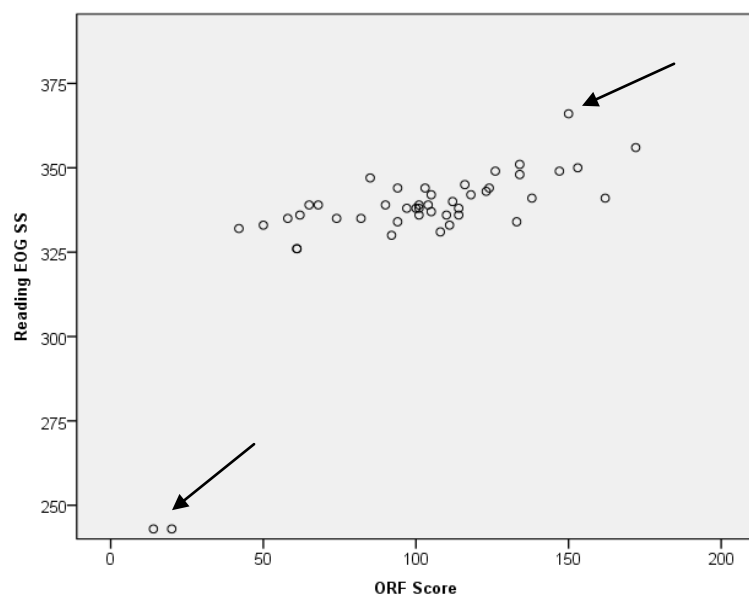


Figure 14. Scatter Plot of NCEOG and ORF Scores for Fourth-Grade Participants. The arrow indicates the outliers that were removed.

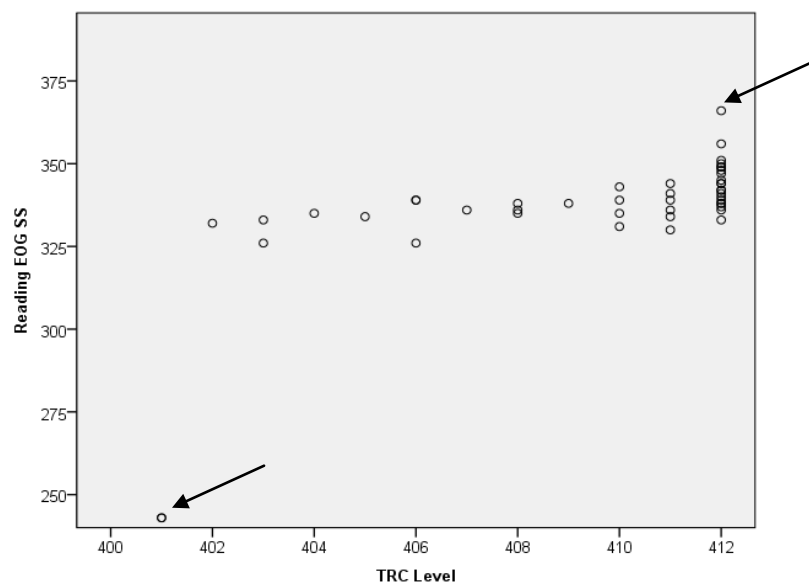


Figure 15. Scatter Plot of NCEOG and TRC Scores for Fourth-Grade Participants. The arrow indicates the outlier that was removed.

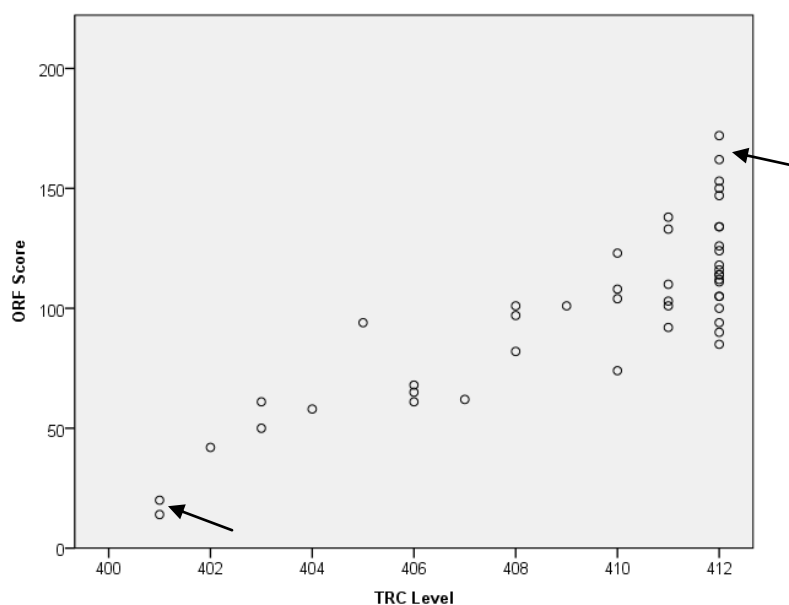


Figure 16. Scatter Plot of ORF and TRC Scores for Fourth-Grade Participants. The arrow indicates the outlier that was removed.

The Shapiro-Wilk test determined not all variables were normally distributed. Table 22 shows the Sig. ($p > .05$) for each variable. NCEOG ($p = .690$) and ORF ($p = .765$) variables were normally distributed ($p > .05$); however, the TRC ($p = .001$) was not normally distributed ($p < .05$). The Shapiro-Wilk test is somewhat robust to deviations from normality, so the researcher decided that with the robustness of the test and since two of the three variables were normally distributed to continue forward with the Pearson correlation test (Lund Research Group, 2013).

Table 22

Shapiro-Wilk Test for Fourth-Grade Participants' Assessment Scores

Assessment	Statistic	df	Sig.
NCEOG	.981	43	.690
ORF	.983	43	.765
TRC	.752	43	.000

Table 23 provides the Pearson correlations for fourth-grade NCEOG scale scores, ORF scores, and TRC scores. There was a positive correlation and statistically significant relationship ($\alpha=.05$) between NCEOG scale scores and ORF scores for all fourth-grade participants, $r(43)=.676, p<.05$. There was also a positive correlation and statistically significant relationship between NCEOG scale scores and TRC scores for fourth-grade participants, $r(43)=.584, p<.05$, along with a positive correlation and statistically significant relationship between ORF scores and TRC scores for fourth-grade participants, $r(43)=.766, p<.05$.

Table 23

Pearson Correlation Matrix for Fourth-Grade Participant Assessment Scores

		NCEOG	ORF	TRC
NCEOG	Pearson Correlation	1		
	Sig. (2-tailed)			
	N			
ORF	Pearson Correlation	.676*	1	
	Sig. (2-tailed)	.000		
	N	43		
TRC	Pearson Correlation	.584*	.766*	1
	Sig. (2-tailed)	.000	.000	
	N	43	43	

Note. * $p<.01$.

The researcher disaggregated the data from the fourth-grade NCEOG, ORF, and TRC scores for ethnicity and gender by calculating the correlation in the previous paragraphs separately for each ethnicity and gender variable. The Hispanic and Multi-Racial ethnic variables were combined and identified as Other ethnicity variable due to

low case numbers. Tables 24 and 25 show the correlations.

Table 24

Pearson Correlation Matrix for Fourth-Grade Scores by Ethnicity

Variable	Assessment	NCEOG	ORF	TRC
Black (N=28)	NCEOG	1	.674**	.608**
	ORF	.674**	1	.762**
	TRC	.608**	.762**	1
White (N=8)	NCEOG	1	.757*	.706
	ORF	.757*	1	.740*
	TRC	.706	.740*	1
Other (N=7)	NCEOG	1	.732	.609
	ORF	.732	1	.890**
	TRC	.609	.890**	1

Note. * $p < .05$, ** $p < .01$.

There was a positive correlation and statistically significant relationship for Black and White fourth-grade participants between NCEOG scale scores and ORF scores (Black, $r(28) = .674$, $p < .05$; White, $r(8) = .757$, $p < .05$), NCEOG scale scores and TRC scores (Black, $r(28) = .608$, $p < .05$; White, $r(8) = .706$, $p < .05$), and ORF and TRC scores (Black, $r(28) = .762$, $p < .05$; White, $r(8) = .740$, $p < .05$). Other ethnicity fourth-grade participants had a positive correlation between NCEOG scale scores and ORF scores ($r(7) = .732$, $p > .05$), NCEOG scales scores and TRC scores ($r(7) = .609$, $p > .05$), and ORF and TRC scores ($r(7) = .890$, $p < .05$); however, there was not a statistically significant relationship between the assessments, except between the ORF and TRC scores (which may be due to its low participant number [$n = 7$]).

Table 25

Pearson Correlation Matrix for Fourth-Grade Scores by Gender

Variable	Assessment	NCEOG	ORF	TRC
Male (N=21)	NCEOG	1	.434*	.560**
	ORF	.434*	1	.796**
	TRC	.560**	.796**	1
Female (N=22)	NCEOG	1	.845**	.638**
	ORF	.845**	1	.770*
	TRC	.638**	.770**	1

Note. * $p < .05$, ** $p < .01$.

There was a positive correlation and statistically significant relationship for Male and Female fourth-grade participants between NCEOG scale scores and ORF scores (Female, $r(22) = .845$, $p < .05$; Male, $r(21) = .434$, $p < .05$), between NCEOG scale scores and TRC scores (Female $r(22) = .638$, $p < .05$; Male, $r(21) = .560$, $p < .05$), and between ORF scores and TRC scores (Female $r(29) = .770$, $p < .05$; Male, $r(.796) = .663$, $p < .05$).

Overall the data clarify, that all three assessments have positive correlations and statistically significant relationships among the fourth-grade participants as a whole. Although close in correlation coefficients, the strongest correlation for all fourth-grade participants with the NCEOG was ORF scores where $r = .676$.

When the researcher disaggregated the data by ethnicity and gender, there were some noticeable observations. The Female and White variables held the highest correlation statistics between the NCEOG and ORF scores and the NCEOG and TRC scores. The Male and Black variables held the highest correlation for ORF and TRC scores of those variables that were statistically significant.

Standard multiple regression analysis. Once the relationships had been

established, a standard multiple regression analysis was calculated to determine if ORF scores and TRC scores predicted NCEOG scale scores for fourth-grade participants. Each standard multiple regression was calculated by grade level due to the NCEOG scale score and ORF score achievement ranges changing from one grade level to the next.

Three outliers were removed after the casewise diagnostics for outliers was computed (see Table 26). Cases 9, 11, and 32 were the same cases removed in the correlations computations. They were removed due to unusually low/high residuals (as compared to the other variables). Case 9 had a low residual of -53.972 produced by its NCEOG scale score of 243. Case 11 had a low residual of -53.677 produced by its NCEOG scale score of 296.68. Case 32 had a high residual of 18.601 produced by its NCEOG scale score of 366. Their removal will support the ability to generalize the results to larger populations and the validity of the study's findings.

Table 26

Casewise Diagnostics for Outlier Results^a for Fourth-Grade Participants

Case Number	Std. Residual	NCEOG SS	Predicated Value	Residual
9	-3.527	243	296.97	-53.972
11	-3.508	243	296.68	-53.677
32	3.260	366	347.40	18.601

Note. a. Dependent Variable: NCEOG SS.

The researcher tested for the six assumptions that must be held true prior to analyzing a multiple regression. There was an independence of errors as indicated by a Durbin-Watson statistic of 2.042. There were linear relationships between the NCEOG and the ORF and TRC scores according to the scatter plot and partial regression plots calculated as part of the analysis (Figures 17, 18, 19). The scatter plot used to analyze

linearity (Figure 17) was also used to test for homoscedasticity. The researcher determined that residuals were evenly spread across the predicated values.

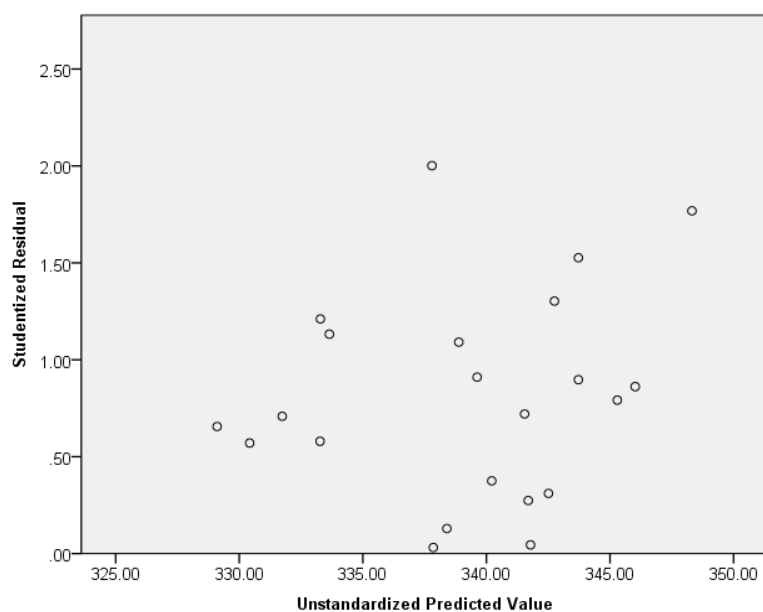


Figure 17. Scatter Plot of Multiple Regression Residuals and Predicted Values from Fourth-Grade Participants.

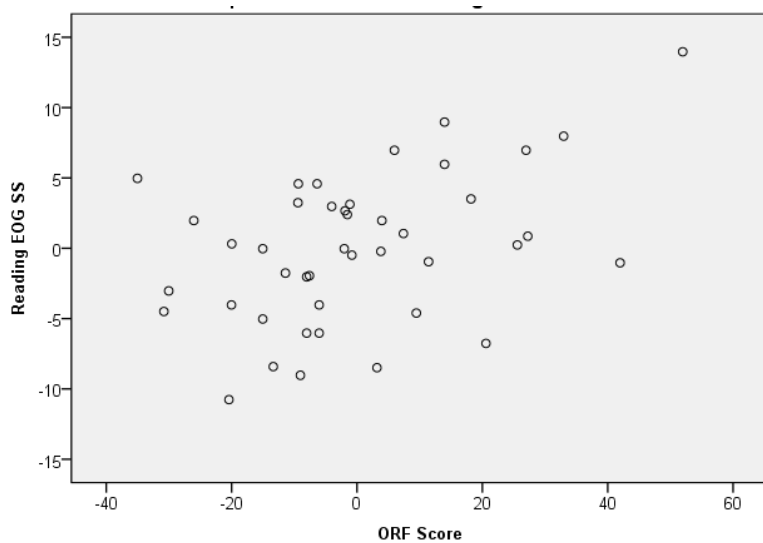


Figure 18. Partial Regression Plot of NCEOG and ORF Scores for Fourth-Grade Participants.

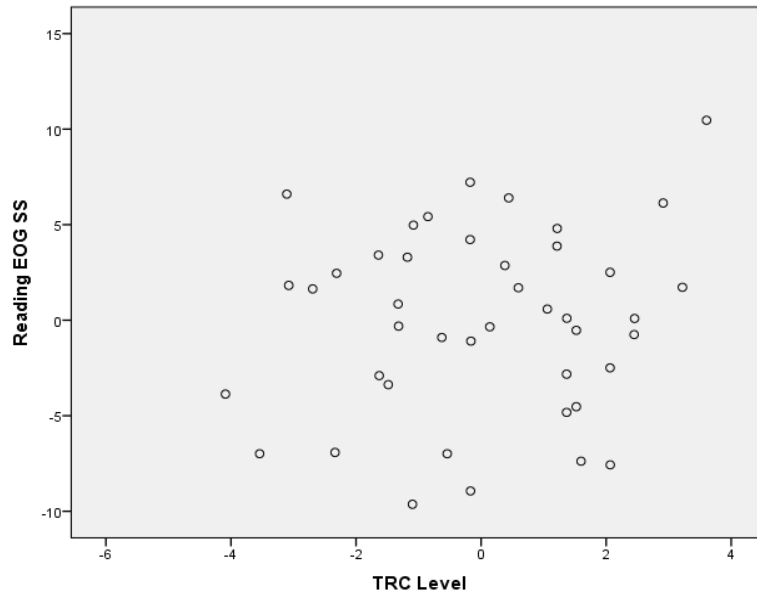


Figure 19. Partial Regression Plot of NCEOG and TRC Scores for Fourth-Grade Participants.

The researcher found no problems with multicollinearity within the fourth-grade participant data through the examination of the correlation coefficients, finding none greater than 0.7 (Table 27) and all Tolerance values were greater than 0.1 (Table 28). No further outliers were detected through the casewise diagnostics and no residuals existed that had ± 3 standard deviations. In this study, the highest leverage value was .18691, which was below two, which allowed the researcher to determine there were no high leverage points.

Table 27

Correlations Coefficients for Fourth-Grade Participant Data (N=43)

		NCEOG	ORF	TRC
Pearson Correlation	NCEOG	1		
	ORF	.676*	1	
	TRC	.584*	.766*	1

Note. * $p < .05$.

Table 28

Tolerance Values for Fourth- Grade Participant Data^a

	Tolerance Value
(Constant)	
ORF	.413
TRC	.413

Note. a. Dependent Variable: NCEOG.

The test for influential points was analyzed by utilizing Cook's distance values to measure for influence. The largest Cook value was .27735 which was below one, so the researcher determined there were no highly influential points. A normal P-P plot (Figure 20) was used to test for normality and the researcher determined that the residuals were normally distributed because they formed along the diagonal line.

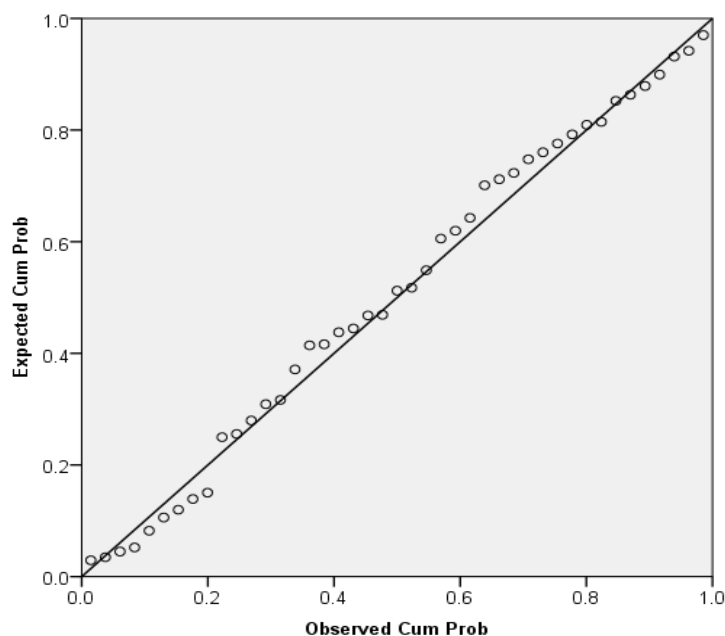


Figure 20. Normal P-P Plot of Regression Residuals, NCEOG-Dependent Variable.

The multiple regression analysis began with four measures that were used to determine how well the regression model fit the data: r , r^2 , adjusted r^2 , and statistical significance. Table 29 shows the results to the first three tests and Table 30 shows the results of statistical significance.

Table 29

Regression Model Summary^b for Fourth-Grade Participant Data

r	r^2	Adjusted r^2
.684 ^a	.468	.441

Note. a. Predictors: (Constant), TRC, ORF; b. Dependent Variable: NCEOG.

R is considered one measure of the quality of the prediction of the dependent variable and its values range from zero to one. In this study the r value was .684 which indicates a high quality of prediction of the dependent variable by the independent variables. The r^2 value showed that ORF and TRC scores explained 46.8% ($r^2=.468$) of

the variability of NCEOG scores. The adjusted (adj.) r^2 value of .441 explained 44.1% of the variability ORF and TRC had on the NCEOG, which indicated that it was a moderate fit to the regression model.

Another way to test the model is the F-ratio through an ANOVA (Analysis of Variance) test. Table 30 shows that the ORF and TRC statistically significantly predict the NCEOG, $F(2,40)=17.559$, $p<.01$, meaning that the regression model is a good fit for the data.

Table 30

ANOVA^a Test Results for Fourth-Grade Participant Data

	Sum of Squares	df	Mean Square	F	Sig.
Regression	839.721	2	419.861	17.559	.000 ^b
Residual	956.465	40	23.912		
Total	1796.186	42			

Note. a. Dependent Variable: NCEOG; b. Predictors: (Constant), TRC, ORF.

Unstandardized coefficients (B_1) indicate how much the dependent variable varies with an independent variable when all other independent variables are held constant (Lund Research Group, 2013). Table 31 shows the unstandardized coefficients for this study. The B_1 for ORF is equal to .121. This means that for each point increase in ORF, there is an increase in the NCEOG scale score of .121. The B_1 for TRC is equal to .349. This means that for each level increase in TRC, there is an increase in the NCEOG scale score of .349.

Table 31 also shows the results of the standard multiple regression test to determine the statistical significance of each of the independent variables (ORF and TRC) on the dependent variable (NCEOG) to show predictability. In this study, the ORF

coefficients (p value=.004, $p<.05$) for fourth-grade participants are statistically significantly different from zero. However, the TRC coefficients ($p=.376$, $p>.05$) for fourth-grade participants are not statistically significantly different from zero.

Table 31

Summary of Standard Multiple Regression for Fourth-Grade Participants

Variables	B	SE _B	β	t	Sig.
(Constant)	183.711	156.829		1.171	.248
ORF Score	.121	.039	.553	3.078	.004
TRC Score	.349	.390	.161	.894	.376

Note. B=unstandardized regression coefficient; SE_B=standard error of the coefficient; β =standardized coefficient (beta).

After the assumptions of linearity, independence of errors, homoscedasticity, normality, and significance were met, the multiple regression test revealed that the ORF and TRC scores statistically significantly predicted NCEOG scale scores, $F(2,40)=17.559$, $p<.05$, adj. $r^2=.441$; however, only the ORF score variable added statistically significantly to the prediction of NCEOG scale scores ($p=.004$, $p<.05$) because the TRC score variable p value was greater than .05 ($p=.376$).

A standard multiple regression analysis was conducted including ethnicity and gender. The White and Male variables were not included in the analysis because they were perfectly collinear (see Tables 32 and 33, Tolerance=.000), which, if included in the analysis, would result in the unstable estimation of model parameters, potentially greatly reducing the model's statistical power (Denis, 2011). Therefore, because the researcher could not accurately determine the statistical significance and predictability of full ethnicity and gender, this component was not included in the study.

Table 32

Variables Enter/Removed from Analysis for Fourth-Grade Gender and Ethnicity^a

Model	Variables Entered	Variables Removed	Method
1	Female, Black, Other Ethnicity ^b		Enter

Note. a. Dependent Variable: NCEOG; b. Tolerance=.000 limits reached.

Table 33

Excluded Variables from Analysis for Fourth-Grade Gender and Ethnicity^a

Variables	Beta In	t	Sig.	Tolerance	VIF	Min. Tolerance
Black	.b			.000		.000
Male	.b			.000		.000

Note. a. Dependent Variable: NCEOG; b. Predictors in the Model: (Constant), Female, Black, Other Ethnicity.

The descriptive and inferential statistical analyses based on the data collected from the fourth-grade study participants revealed several consistent findings. A positive correlation between NCEOG and mCLASS Reading 3D revealed a relationship between the two assessments. The strongest relationship with the dependent variable existed between NCEOG and ORF. The analyses also revealed that both the ORF scores and TRC scores statistically significantly predicted the student scale scores on the reading comprehension portion of the NCEOG.

Descriptive and Inferential Statistics for Fifth-Grade Participants

Descriptive statistics. Descriptive analyses were calculated for the fifth-grade study participants in order to determine what relationships exist between mClass Reading

3D and NCEOG.

Frequency distributions. Table 34 shows the results of the calculated frequency distributions on the data entered into SPSS to determine overall demographics for the fifth-grade study participants. The table includes frequencies for assessment proficiency levels as part of the demographics descriptors. Although the ORF and NCEOG scores change achievement ranges at each grade level, the proficiency levels have consistent descriptors across grade levels. Table 2 in Chapter 3 reveals the specific ORF cut point ranges and Table 5 in Chapter 3 reveals the NCEOG achievement level ranges; both define the range of scores and their coordinating proficiency level.

Table 34

Frequency Distributions for Fifth-Grade Study Participants (N=37)

Demographic Variable	n	Percentage of Fifth-Grade Participants
Male	17	46%
Female	20	54%
Black	22	60%
White	7	18%
Hispanic	5	14%
Multi-Racial	3	8%
AIG	1	3%
EC	4	11%
LEP	1	3%
NCEOG Level I	5	14%
NCEOG Level II	13	35%
NCEOG Level III	19	51%
NCEOG Level IV	0	0%
ORF Red	12	33%
ORF Yellow	9	24%
ORF Green	16	43%
TRC Red	13	35%
TRC Yellow	7	19%
TRC Green	17	46%

Note. AIG—Academically and Intellectually Gifted; EC—Exceptional Children; LEP—Limited English Proficient; NCEOG I—minimal understanding of grade-level standards; NCEOG II—limited understanding of grade-level standards; NCEOG III—grade-level proficient; NCEOG IV—above grade-level proficient; ORF/TRC Red—way below grade-level proficiency; ORF/TRC Yellow—just below grade-level proficiency; ORF/TRC Green—at or above grade-level proficiency.

Measures of central tendency and variability. Along with the frequency distributions, the measures of central tendency and variability were calculated to determine the mean, median, and standard deviations of the fifth-grade participant assessment scores. The results are found in Table 35. For the purposes of statistical analyses, the researcher recoded the TRC levels from the assigned Fountas and Pinnell (2010) system letters to researcher-selected number codes. See the Appendix for the coding key.

Table 35

Descriptive Statistics for Fifth-Grade NCEOG Scale Score, ORF Score, and TRC Score

Measure	Category	N	Mean	Median	Range	Standard Deviation
NCEOG	All Students	37	347.48	349.00	26	6.63
	Black	22	347.64	347.00	26	6.99
	White	7	347.29	348.00	20	7.32
	Hispanic	5	349.00	351.00	16	6.44
	Multi-Racial	3	348.00	349.00	11	5.57
	Male	17	345.88	345.00	26	8.15
	Female	20	349.40	349.00	16	4.64
	EC	4	342.75	344.50	14	6.08
ORF	All Students	37	116.14	120.00	114	33.96
	Black	22	115.68	121.50	107	31.79
	White	7	111.43	111.00	98	31.95
	Hispanic	5	137.40	150.00	114	45.88
	Multi-Racial	3	95.00	109.00	58	31.43
	Male	17	109.59	107.00	109	35.44
	Female	20	121.70	123.50	112	32.50
	EC	4	112.75	103.50	88	37.59
TRC	All Students	37	504.76	505.00	5	1.44
	Black	22	504.73	505.00	4	1.35
	White	7	504.86	505.00	3	1.35
	Hispanic	5	505.00	506.00	5	2.24
	Multi-Racial	3	504.33	504.00	3	1.53
	Male	17	504.59	505.00	5	1.42
	Female	20	504.90	506.00	4	1.48
	EC	4	505.25	505.00	1	.50

Note. EC—Exceptional Children. Limited English Proficient (LEP) and Academically and Intellectually Gifted (AIG) students were not included because of a low number of participants (n=1). TRC levels have been recoded from letters to numbers, see the Appendix.

Observations were made based on the analysis of the measures of central tendencies and variability. The mean NCEOG scale score for all fifth-grade participants ($M=347.48$) was equivalent to a proficiency level of II, limited understanding of grade-level standards. The mean ORF score ($M=116.14$) for all fourth-grade participants was equivalent to just below grade-level proficiency (yellow). The mean TRC score ($M=504.76$) for all fourth-grade participants was equivalent to way below grade-level proficiency (red).

With respect to ethnicity (Figure 21), Hispanic ($M=349.00$) students' mean NCEOG scale scores were equivalent to a proficiency level of III, grade-level proficiency. Black ($M=347.64$), White ($M=347.29$), and Multi-Racial ($M=348.00$) students fell to a proficiency level of II, limited understanding of grade-level standards. Hispanic ($M=137.40$) students' mean ORF scores were equivalent to grade-level proficiency (green). Black ($M=115.68$) and White ($M=111.43$) students fell to just below grade-level proficiency (yellow); Multi-Racial ($M=95.00$) students fell further to way below grade-level proficiency (red). Hispanic ($M=505.00$) students' mean TRC scores were equivalent to just below grade-level proficiency (yellow); Black ($M=504.73$), White ($M=504.86$), and Multi-Racial ($M=504.33$) students fell to way below grade-level proficiency (red).

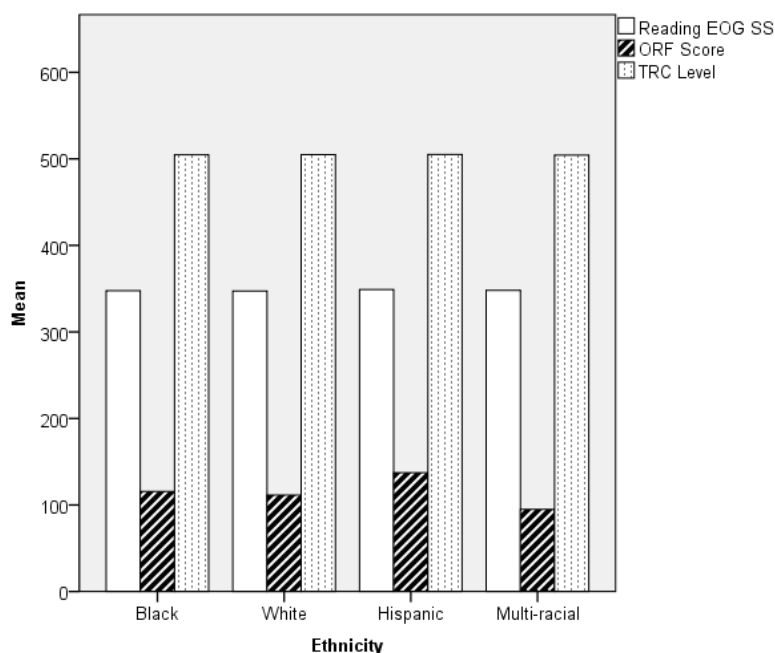


Figure 21. NCEOG, ORF, and TRC Mean Scores by Fifth-Grade Ethnicity.

With respect to gender (Figure 22), Female ($M=349.40$) participants scored a mean NCEOG scale score equivalent to a proficiency level of III, at or above grade-level proficiency; the mean scale score for Male ($M=345.88$) students fell to a proficiency level of II, limited understanding of grade-level standards. Female ($M=121.70$) and Male ($M=109.59$) participants' mean ORF scores were equivalent to just below grade-level proficiency (yellow). Female ($M=504.90$) and Male ($M=504.59$) participants' mean TRC scores were equivalent to way below grade-level proficiency (red).

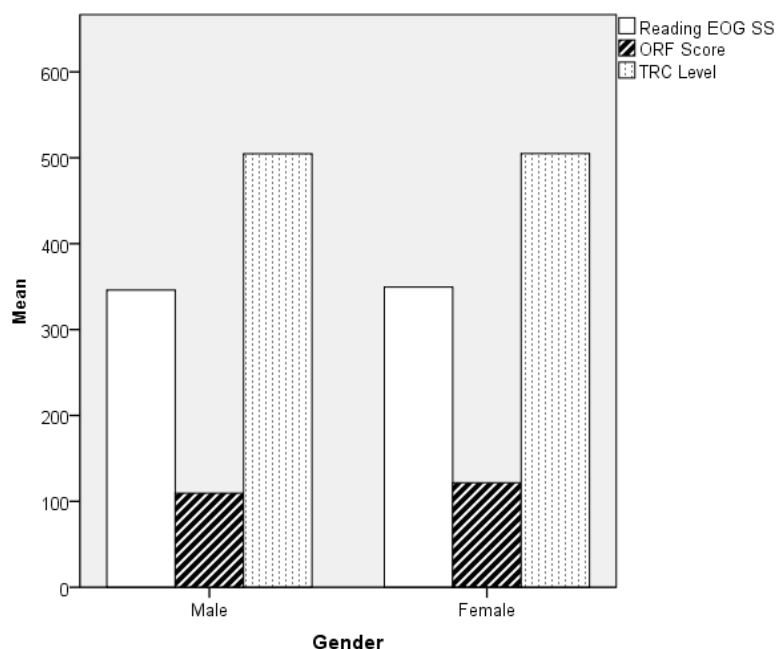


Figure 22. NCEOG, ORF, and TRC Mean Scores by Fifth-Grade Gender.

With respect to students identified with special needs (Figure 23), EC students ($M=342.75$) scored a mean NCEOG scale score equivalent to a proficiency level II, limited understanding of grade-level standards. EC students' mean ORF score ($M=112.75$) and mean TRC score ($M=505.25$) were equivalent to just below grade-level proficiency (yellow). AIG and LEP students were not included due to the low number of participants identified ($n=1$).

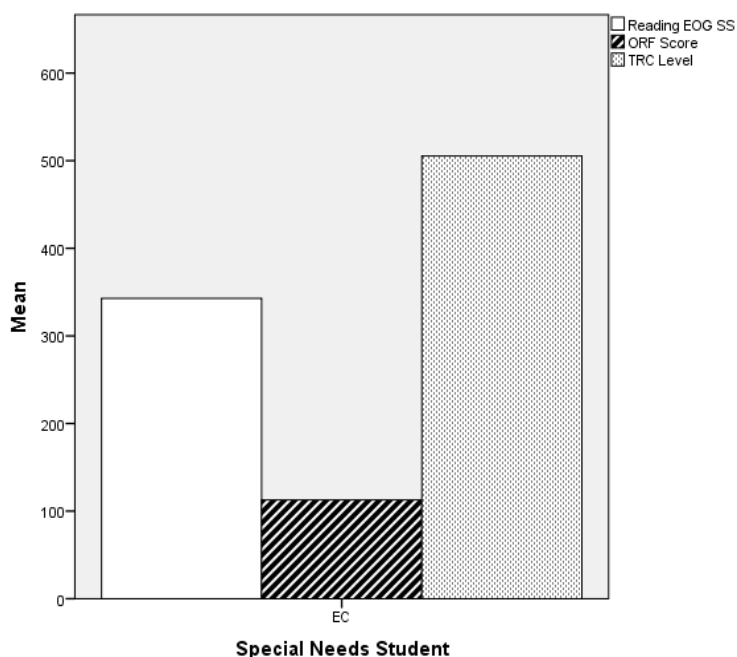


Figure 23. NCEOG, ORF, and TRC Mean Scores by Fifth-Grade Students with Special Needs.

Pearson product-moment correlation. The Pearson correlation was calculated and preliminary scatter plots showed the relationships between the assessments to be positively linear (Figures 24, 25, 26). One outlier was removed, case number 7, due to its unusually low NCEOG scale score (338), ORF score (57), and TRC score (501).

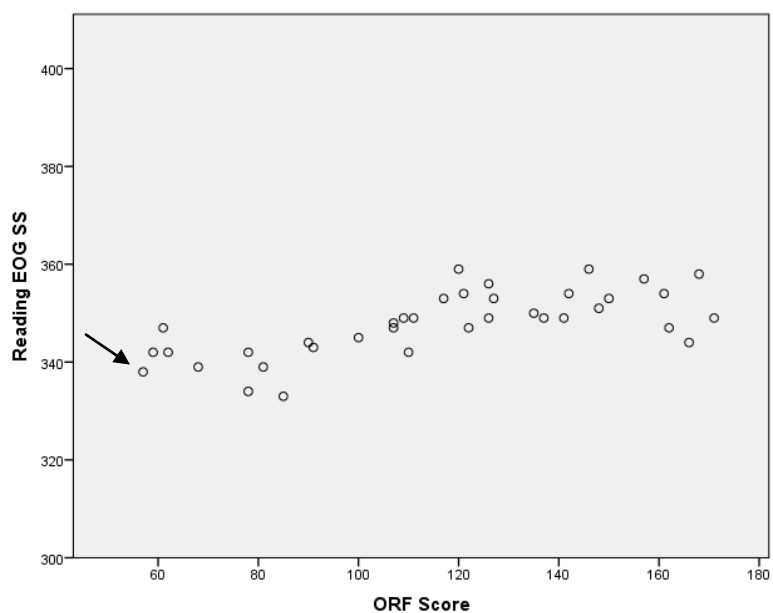


Figure 24. Scatter Plot of NCEOG and ORF Scores for Fifth-Grade Participants. The arrow indicates the outliers that were removed.

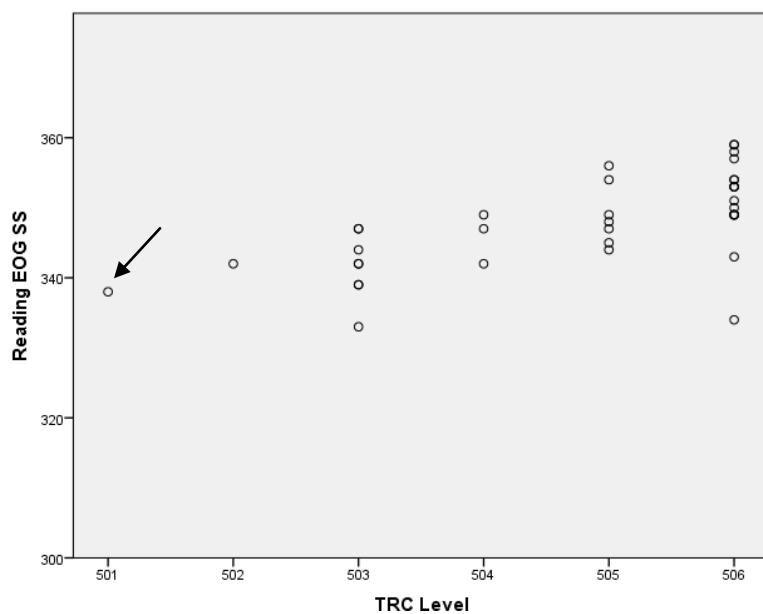


Figure 25. Scatter Plot of NCEOG and TRC Scores for Fifth-Grade Participants. The arrow indicates the outlier that was removed.

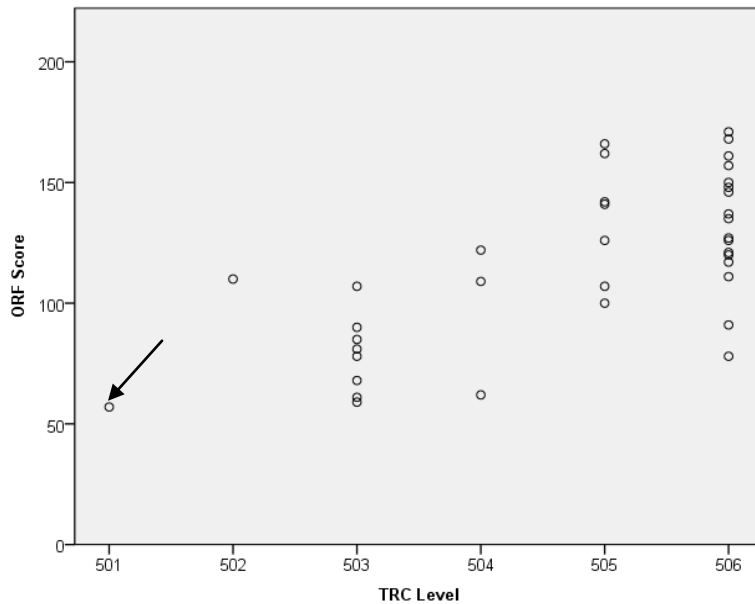


Figure 26. Scatter Plot of ORF and TRC Scores for Fifth-Grade Participants. The arrow indicates the outlier that was removed.

The Shapiro-Wilk test determined not all variables were normally distributed. Table 36 shows the significance (Sig.) ($p > .05$) for each variable. NCEOG ($p = .468$) and ORF ($p = .237$) p -values were greater than .05, determining those variables to be normally distributed. The TRC's p -value ($p < .001$) was less than .05, determining it was not normally distributed. However, the Shapiro-Wilk test is somewhat robust to deviations from normality, so the researcher decided that with the robustness of the test and since two of the three variables were normally distributed to continue forward with the Pearson correlation test (Lund Research Group, 2013).

Table 36

Shapiro-Wilk Test for Fifth-Grade Participants' Assessment Scores

Assessment	Statistic	df	Sig.
NCEOG	.972	36	.468
ORF	.961	36	.237
TRC	.784	36	.000

Table 37 provides the Pearson correlations for fifth-grade NCEOG scale scores, ORF scores, and TRC scores. There was a positive correlation and statistically significant relationship ($\alpha=.05$) between NCEOG scale scores and ORF scores for fifth-grade students, $r(36)=.669, p<.05$. There was also a positive correlation and statistically significant relationship between NCEOG scale scores and TRC scores for fifth-grade students, $r(36)=.616, p<.05$, along with a positive correlation and statistically significant relationship between ORF scores and TRC scores for fifth-grade students, $r(36)=.643, p<.05$.

Table 37

Pearson Correlation Matrix for Fifth-Grade Student Assessment Scores (N=36)

		NCEOG	ORF	TRC
NCEOG	Pearson Correlation	1		
	Sig. (2-tailed)			
	N			
ORF	Pearson Correlation	.669*	1	
	Sig. (2-tailed)	.000		
	N	36		
TRC	Pearson Correlation	.616*	.643*	1
	Sig. (2-tailed)	.000	.000	
	N	36	36	

Note. *Correlation is significant at the 0.01 level (2-tailed).

The researcher disaggregated the data from the fifth-grade NCEOG, ORF, and TRC scores for ethnicity and gender by running the correlation in the previous paragraphs separately for each ethnicity and gender variable. The Hispanic and Multi-Racial ethnic groups were combined together to create an Other ethnicity variable due to the low number of cases. Tables 38 and 39 show the correlations.

Table 38

Pearson Correlation Matrix for Fifth-Grade Scores by Ethnicity

Variable	Assessment	NCEOG	ORF	TRC
Black (N=22)	NCEOG	1	.740**	.493*
	ORF	.740**	1	.551**
	TRC	.493*	.551**	1
White (N=7)	NCEOG	1	.441	.885**
	ORF	.441	1	.665
	TRC	.885**	.665	1
Other (N=7)	NCEOG	1	.745	.866*
	ORF	.745	1	.887**
	TRC	.866*	.887**	1

Note. * $p < .05$, ** $p < .01$.

There was a positive correlation and statistically significant relationship for Black and Other fifth-grade participants between NCEOG scale scores and ORF scores (Black, $r(22) = .740$, $p < .05$; Other, $r(7) = .745$, $p < .05$), NCEOG scale scores and TRC scores (Black, $r(22) = .493$, $p < .05$; Other, $r(7) = .866$, $p < .05$), and ORF and TRC scores (Black, $r(22) = .551$, $p < .05$; Other, $r(7) = .887$, $p < .05$). White fifth-grade participants had a positive correlation between NCEOG scale scores and ORF scores ($r(7) = .441$, $p > .05$), NCEOG scale scores and TRC scores ($r(7) = .885$, $p < .05$), and ORF and TRC scores ($r(7) = .665$, $p < .05$); however, there was not a statistically significant relationship between the

assessments, except between the NCEOG and TRC scores (which may be due to its low participant number [n=7]).

Table 39

Pearson Correlation Matrix for Fifth-Grade Scores by Gender

Variable	Assessment	NCEOG	ORF	TRC
Male (N=16)	NCEOG	1	.650**	.562*
	ORF	.650**	1	.483
	TRC	.562*	.483	1
Female (N=20)	NCEOG	1	.732**	.809**
	ORF	.732**	1	.757**
	TRC	.809**	.757**	1

Note. * $p < .05$, ** $p < .01$.

There was a positive correlation and statistically significant relationship for Male and Female fifth-grade participants between NCEOG scale scores and ORF scores (Female, $r(20) = .732$, $p < .05$; Male, $r(16) = .650$, $p < .05$), between NCEOG scale scores and TRC scores (Female $r(20) = .809$, $p < .05$; Male, $r(16) = .562$, $p < .05$), and between ORF scores and TRC scores (Female $r(20) = .757$, $p < .05$; Male, $r(16) = .483$, $p < .05$).

Overall, the data clarify that all three assessments have positive correlations and statistically significant relationships among the fifth-grade participants as a whole. Although close in correlation coefficients, the strongest correlation for all fifth-grade participants with the NCEOG was ORF scores where $r = .669$.

When the researcher disaggregated the data by ethnicity and gender, there were some noticeable observations. The Female gender and Other ethnicity variables held the highest correlation statistics between the NCEOG and ORF scores and ORF and TRC scores of the variables that were statistically significant. The Female gender and White

ethnic variables held the highest correlation for NCEOG and TRC scores.

Standard multiple regression analysis. Once the relationships had been established, a standard multiple regression analysis was calculated to determine if ORF scores and TRC scores predicted NCEOG scale scores for fifth-grade participants. Each standard multiple regression was calculated by grade level due to the NCEOG scale score and ORF score achievement ranges changing from one grade level to the next.

One outlier was removed after the residuals were sorted to determine any data points that had ± 3 standard deviations that could result in skewing the multiple regression analysis. Case 16 had a residual of -3.30901. The case's low (as compared to other scores) NCEOG scale score of 334 and ORF score of 78 were included in the researcher's decision to remove the case as an outlier. Its removal supports the ability to generalize the results to larger populations and the validity of the study's findings.

The researcher tested for the six assumptions that must be held true prior to analyzing a multiple regression. There was an independence of errors as indicated by a Durbin-Watson statistic of 2.020. There were slight linear relationships between the NCEOG and the ORF and TRC scores according to the scatter plot and partial regression plots run as part of the analysis (Figures 27, 28, and 29). The scatter plot used to analyze linearity (Figure 27) was also used to test for homoscedasticity. The researcher determined that residuals were evenly spread across the predicated values.

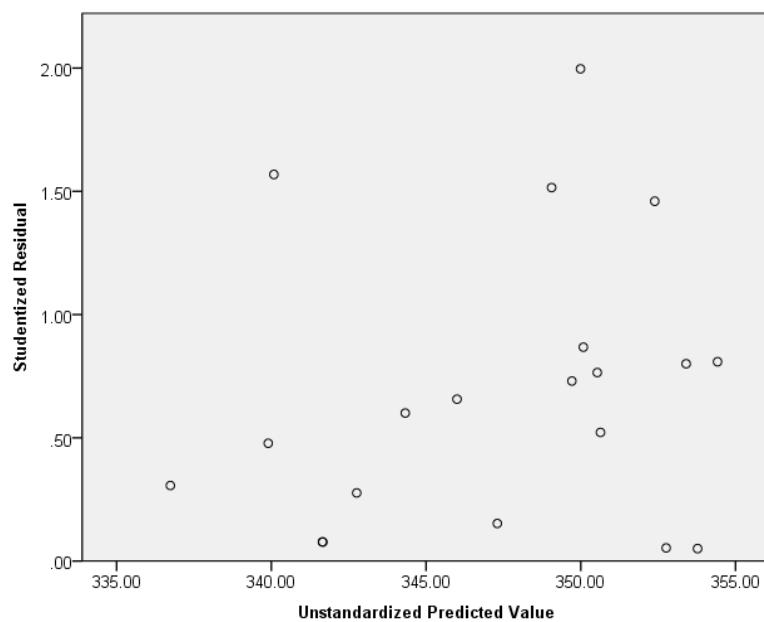


Figure 27. Scatter Plot of Multiple Regression Residuals and Predicted Values for Fifth-Grade Participants.

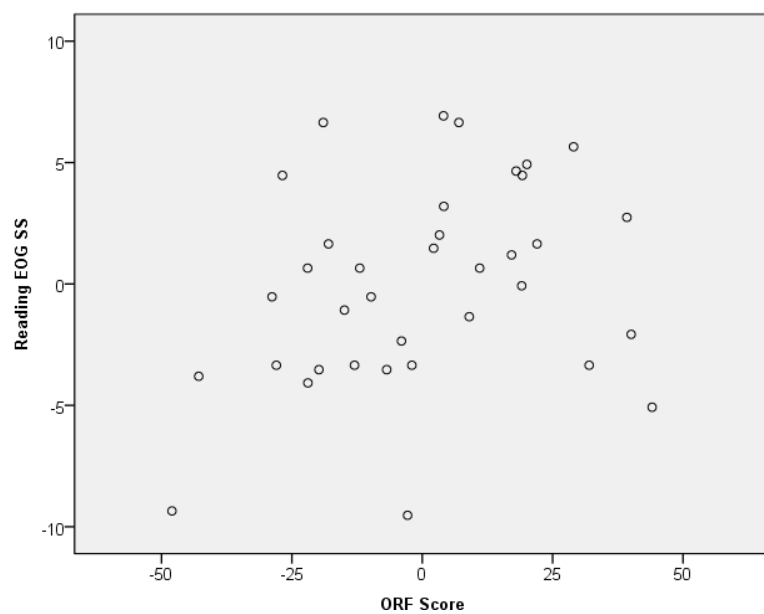


Figure 28. Partial Regression Plot of NCEOG and ORF Scores for Fifth-Grade Participants.

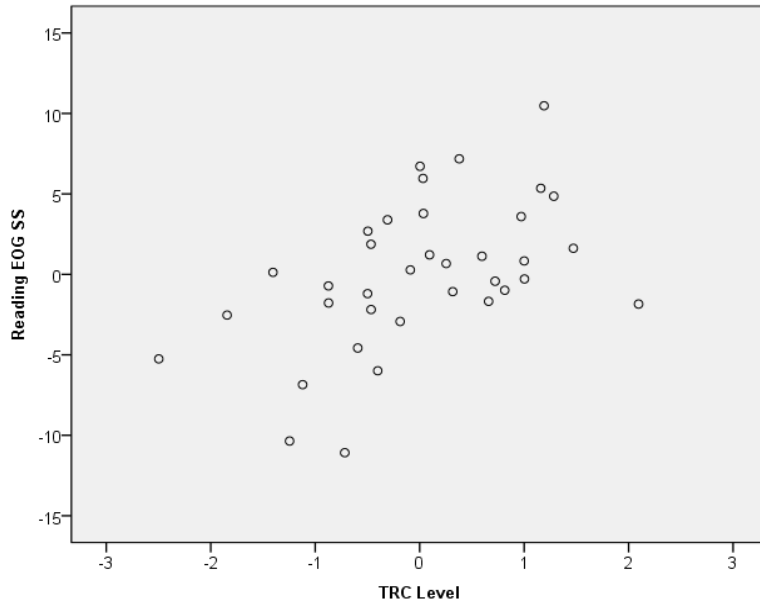


Figure 29. Partial Regression Plot of NCEOG and TRC for Fifth-Grade Participants.

The researcher found no problems with multicollinearity within the fifth-grade participant data through the examination of the correlation coefficients, finding none greater than 0.7 (Table 40), and all Tolerance values were greater than 0.1 (Table 41). No further outliers were detected through the casewise diagnostics and no residuals existed that had ± 3 standard deviations. In this study, the highest leverage value was .18871, which was below two, and allowed the researcher to determine there were no high leverage points.

Table 40

Correlations Coefficients for Fifth-Grade Participant Data (N=36)

		NCEOG	ORF	TRC
Pearson Correlation	NCEOG	1		
	ORF	.681*	1	
	TRC	.752*	.730*	1

Note. *p<.05.

Table 41

Tolerance Values for Fifth-Grade Participant Data^a

Tolerance Value	
(Constant)	
ORF	.468
TRC	.468

Note. a. Dependent Variable: NCEOG.

The test for influential points was analyzed by utilizing Cook's distance values to measure for influence. The largest Cook value was .18799, which was below one, so the researcher determined there were no highly influential points. A normal P-P plot (figure 30) was used to test for normality, and the researcher determined that the residuals were normally distributed because they formed along the diagonal line.

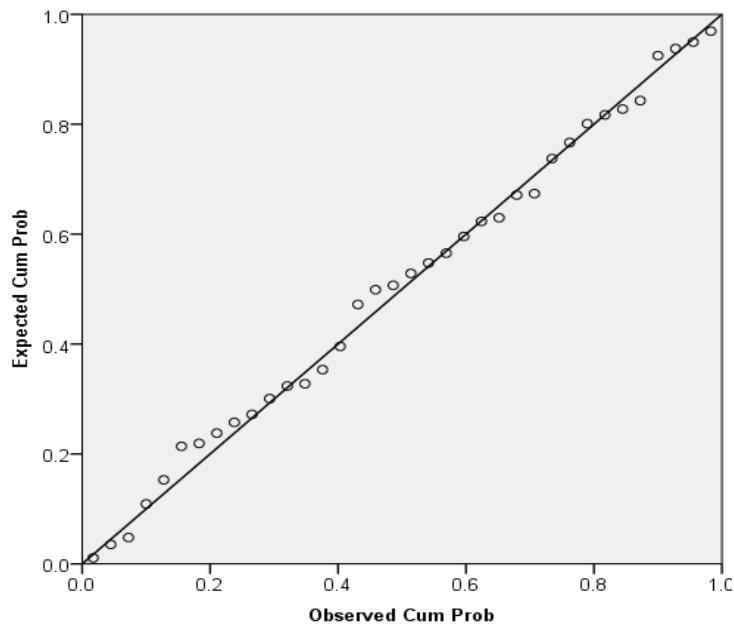


Figure 30. Normal P-P Plot of Fifth-Grade Regression Residuals (NCEOG-dependent variable).

The multiple regression analysis began with four measures that were used to determine how well the regression model fit the data: r , r^2 , adjusted r^2 , and statistical significance. Table 42 shows the results to the first three tests and Table 43 shows the results of statistical significance.

Table 42

Regression Model Summary^b for Fifth-Grade Participant Data

r	r^2	Adjusted r^2
.776 ^a	.602	.578

Note. a. Predictors: (Constant), TRC, ORF; b. Dependent Variable: NCEOG.

R is considered one measure of the quality of the prediction of the dependent variable, and its values range from zero to one. In this study, the r value was .776 which indicates a high quality of prediction of the dependent variable by the independent variables. The r^2 value showed that ORF and TRC scores explained 60.2% ($r^2=.602$) of the variability of NCEOG scores. The adjusted (adj.) r^2 value of .578 explained 57.8% of the variability ORF and TRC had on the NCEOG, which indicated that it was a moderate fit to the regression model.

Another way to test the model is the F-ratio through an ANOVA (Analysis of Variance) test. Table 43 shows that the ORF and TRC statistically significantly predict the NCEOG, $F(2,33)=24.990$, $p<.001$, meaning that the regression model is a good fit for the data.

Table 43

ANOVA^a Test Results for Fifth-Grade Participant Data

	Sum of Squares	df	Mean Square	F	Sig.
Regression	836.614	2	418.307	24.990	.000 ^b
Residual	552.386	33	16.739		
Total	1389.000	35			

Note. a. Dependent Variable: NCEOG; b. Predictors: (Constant), TRC, ORF.

Unstandardized coefficients (B_1) indicate how much the dependent variable varies with an independent variable when all other independent variables are held constant (Lund Research Group, 2013). Table 44 shows the unstandardized coefficients for this study. The B_1 for ORF is equal to .121. This means that for each point increase in ORF, there is an increase in the NCEOG scale score of .121. The B_1 for TRC is equal to .349. This means that for each level increase in TRC, there is an increase in the NCEOG scale score of .349.

Table 44 also shows the results of the standard multiple regression test to determine the statistical significance of each of the independent variables (ORF and TRC) on the dependent variable (NCEOG) to show predictability. In this study, the TRC coefficient (p value=.002, $p<.05$) for fifth-grade participants is statistically significantly different from zero. However, the ORF coefficient ($p=.087$, $p>.05$) for fifth-grade participants is not statistically significantly different from zero.

Table 44

Summary of Standard Multiple Regression for Fifth-Grade Participants

Variables	B	SE _B	β	t	Sig.
(Constant)	-855.565	350.345		-2.442	.020
ORF Score	.053	.030	.283	1.765	.087
TRC Score	2.373	.699	.545	3.393	.002

Note. B=unstandardized regression coefficient; SE_B=standard error of the coefficient; β =standardized coefficient (beta).

After the assumptions of linearity, independence of errors, homoscedasticity, normality, and significance were met, the multiple regression test revealed that the ORF and TRC scores statistically significantly predicted NCEOG scale scores, $F(2,33)=24.990$, $p<.05$, adj. $r^2=.578$; however, only the TRC score variable added statistically significantly to the prediction of NCEOG scale scores ($p=.002$, $p<.05$) because the ORF score variable p value was greater than .05 ($p=.087$).

A standard multiple regression analysis was conducted including ethnicity and gender. The Other ethnicity and Male variables were not included in the analysis because they were perfectly collinear (see Tables 45 and 46, Tolerance=.000), which, if included in the analysis, would result in the unstable estimation of model parameters, potentially greatly reducing the model's statistical power (Denis, 2011). Therefore, because the researcher could not accurately determine the statistical significance and predictability of full ethnicity and gender, this component was not included in the study.

Table 45

Variables Enter/Removed from Analysis for Fifth-Grade Gender and Ethnicity^a

Model	Variables Entered	Variables Removed	Method
1	Female, Black, White ^b		Enter

Note. a. Dependent Variable: NCEOG; b. Tolerance=.000 limits reached.

Table 46

Excluded Variables from Analysis for Fifth-Grade Gender and Ethnicity^a

Variables	Beta In	t	Sig.	Tolerance	VIF	Min. Tolerance
Other Ethnicity	. ^b			.000		.000
Male	. ^b			.000		.000

Note. a. Dependent Variable: NCEOG; b. Predictors in the Model: (Constant), Female, Black, White.

The descriptive and inferential statistical analyses based on the data collected from the fifth-grade study participants revealed several consistent findings. A positive correlation between NCEOG and mCLASS Reading 3D revealed a relationship between the two assessments. The strongest relationship with the dependent variable existed between NCEOG and ORF. The analyses also revealed that both the ORF scores and TRC scores statistically significantly predicted the student scale scores on the reading comprehension portion of the NCEOG.

Summary

All three grade-level participants, based on the descriptive and inferential statistics, had several consistent findings. Overall, there was a positive correlation between NCEOG and mCLASS Reading 3D, revealing a relationship between the two

assessments. Fourth grade had the strongest correlation between NCEOG and ORF with $r=.676$, and fifth grade had the strongest correlation between NCEOG and TRC with $r=.616$. This provides the information necessary to answer the research question “What is the relationship between the mCLASS Reading 3D assessment and the NCEOG Reading Comprehension assessment?”

The analyses also revealed that in all three grade levels both the ORF and TRC scores statistically significantly predicted the student scale scores on the reading comprehension portion of the NCEOG. This determines that mCLASS Reading 3D statistically significantly predicted the student scale scores on the reading comprehension portion of the NCEOG. Grade 3 had the strongest predictability by revealing that the ORF ($p<.05$) and TRC ($p=.005, p<.05$) scores from mCLASS Reading 3D added to the statistical significance of the prediction for NCEOG scale scores in their grade level; whereas, Grade 4 revealed only the ORF ($p=.004, p<.05$) scores from mCLASS Reading 3D added to the statistical significance of the predication of the NCEOG scale scores for their grade level. Also, Grade 5 revealed only the TRC ($p=.002, p<.05$) scores from mCLASS Reading 3D added to the statistical significance of the predication of the NCEOG scale scores for their grade level. This provides the information necessary to answer the research question “To what extent does the mClass Reading 3D assessment accurately predict student scores on the NCEOG Reading Comprehension assessment?” Chapter 5 will further discuss these analyses and findings in terms of purpose, connection to past and current literature, and connection to future research.

Chapter 5: Discussion and Conclusions

Introduction

The ability to read is one of the most important skills to foster academic success; however, today's students have a prevailing weakness in this area. It is imperative that students at risk of reading failure are identified and interventions are put into place to catch them up to grade-level standards. The planning of the interventions should be driven by assessment results. Assessments like mClass Reading 3D provide data for teachers to effectively administer interventions in the classroom. The purpose of intervention is for students to grow in their reading achievement and be successful on high-stakes tests like the NCEOG; therefore, it is essential to determine if mClass Reading 3D is an accurate predictor of student success on the NCEOG Reading Comprehension assessment. The remainder of Chapter 5 provides a summary of this study, a discussion of the findings, implications for education, the study's limitations, as well as recommendations for future research.

Study Summary

The purpose of this study was to investigate the relationship between results on the NCEOG Assessment of Reading Comprehension and results on the mClass Reading 3D assessment, especially examining the degree to which mClass Reading 3D predicts scores on the reading comprehension portion of the NCEOG. The study explored the following research questions.

1. What is the relationship between the mClass Reading 3D assessment and the NCEOG Reading Comprehension assessment?
2. To what extent does the mClass Reading 3D assessment accurately predict student scores on the NCEOG Reading Comprehension assessment?

In order to answer these research questions, quantitative data were collected from the study participants' May 2011 NCEOG Reading Comprehension assessment scores and 2011 EOY benchmark mClass Reading 3D scores. Descriptive statistics were used to describe the overall demographics and assessment data for the study participants. Measures of central tendency and variability were calculated to determine the mean, median, and standard deviations of the participants' assessment scores. This portion of the overall analysis supported the determination of the relationship between NCEOG and mClass Reading 3D.

The Pearson product-moment correlation was also used to further describe the relationship between the NCEOG and mClass Reading 3D assessments. The correlations were based on assessment scores by grade level, gender, and ethnicity. There was a positive correlation and a statistically significant relationship between the two assessments at all three grade levels (3, 4, and 5); however, there were differing results within gender and ethnicity.

Once the relationships were established, a standard multiple regression analysis was calculated and analyzed to reveal that mClass Reading 3D did statistically significantly predict NCEOG scores for Grades 3, 4, and 5. However, when the analysis examined the specific mClass assessment variables, the researcher found that the only grade level where both the ORF and TRC variables added statistical significance was third grade.

Discussion of Study Results and Connection to Literature Review

Pellegrino (2004) stated,

If social and public goals regarding academic achievement are to be attained, then we must make more effort to improve assessment, especially assessment practices

that can directly support enhanced outcomes for students. Thus assessment can become part of the solution rather than be part of the problem. (p. 5)

As increased accountability equates to increased assessment in schools, it is imperative to ensure that the assessments align and directly relate to instruction and intervention.

Educational researchers, for many years now, have investigated how instruction, intervention, and student scores on formative assessments relate to and predict student results on high-stakes assessments.

In Chapter 2, the review of literature revealed phonemic awareness, phonics, fluency, comprehension, and vocabulary as the most effective elements of teaching reading (NICHD, 2000). Clay (2013) added the importance of utilizing reading strategies to increase understanding of text. Pearson et al. (1992) found that proficient readers make connections, ask questions, draw inferences, distinguish important ideas, synthesize information, and monitor their comprehension of the text. mClass Reading 3D provides data that help determine a student's ability to process and utilize these skills and elements according to their grade-level expectations. This study provided further research on utilizing mClass Reading 3D's data to effectively instruct and intervene in the areas above so students can become more proficient readers. Researchers have found that there will always be a percentage of children who are at risk of reading failure for a variety of reasons (Richardson & DiBenedetto, 1996), and the early identification and intervention in specific areas of deficit can improve children's skill levels immediately and prevent later difficulties (NICHD, 2000). Menzies et al. (2008) found that teachers must be able to accurately assess student needs and subsequently plan and deliver instruction and interventions based on that assessment in order to identify student deficits and help students to become proficient readers. Baily and Drummond (2006) determined that

teacher identification of reading difficulties needed to include proven literacy assessments in order to capture the intricacies of literacy-related skills that teachers were unable to determine by observation alone. This study provides data and research on the mClass Reading 3D assessment and its relationship and predictability of the NCEOG that would provide confidence to educators in utilizing the data to plan and deliver instruction and interventions that would identify student deficits and build proficient readers.

The literature review in this study showed a glimpse into the history of reading research and its connection to the current emphasis on literacy assessments. However, despite the amount of research that exists, there are new assessments and interventions being utilized regularly in schools each year. These new assessments should have a supporting research base that determines if they are accurate predictors of student success on the high-stakes assessments like the NCEOG. By analyzing mClass Reading 3D and its relationship and predictability to the NCEOG, this study is adding to the research base of a new assessment that is expanding across North Carolina. The results of this study provide information that increases student success on the NCEOG by providing accurate data for instruction and interventions.

The literature review revealed numerous studies that have been conducted on the DIBELS ORF portion of the mClass Reading 3D assessment and its relationship with and predictability of high-stakes assessments like the NCEOG. This study found similar results to those past studies. This study found a positive correlation and statistically significant relationship between NCEOG and the ORF portion of the mClass Reading 3D assessment in Grades 3 ($r=.654, p<.05$), 4 ($r=.676, p<.05$), and 5 ($r=.669, p<.05$). It also found the ORF portion of NCEOG statistically significantly predicted student scale scores on NCEOG in Grades 3 ($F(2,55)=38.728, p<.05$), 4 ($F(2,40)=17.559, p<.05$), and

5 ($F(2,33)=24.990, p<.05$). Past studies such as the one conducted by Barger (2003), who found the correlation between ORF scores to the NCEOG assessment was high ($r=.73$) resulting in the ORF measure as an accurate predictor of proficient scores on the NCEOG, reveal similar results as this current study. Buck and Torgesen (2003) found there was a significant correlation between the scores of ORF and the Florida state assessment ($r=.70$). Wood (2006) found a significant relationship between ORF and the CSAP. Results indicated ORF predicted performance for CSAP in third ($r=.70$), fourth ($r=.67$), and fifth grades ($r=.75$).

The part of this study that makes it different from previous ones is its inclusion of the TRC portion of mClass Reading 3D in its analysis. According to Ross (2004), there is little psychometric data available for RRs, which are the foundation of the TRC assessment. Also, there is very little research to date on the correlation and/or predictability of RRs/TRC and high-stakes testing that has been published. The results of this study add research to the field on the TRC assessment's correlation and predictability to high-stakes testing (NCEOG).

This study found a high positive correlation and statistically significant relationship between NCEOG and the TRC portion of the mClass Reading 3D assessment in Grades 3 ($r=.597, p<.05$), 4 ($r=.584, p<.05$), and 5 ($r=.616, p<.05$). It also found the TRC portion of NCEOG statistically significantly predicted student scale scores on NCEOG in Grades 3 ($F(2,55)=38.728, p<.05$), 4 ($F(2,40)=17.559, p<.05$), and 5 ($F(2,33)=24.990, p<.05$).

Descriptive statistics. Analysis of the descriptive statistics in this study brought to light several observations. These statistics were used to determine what relationships existed between mClass Reading 3D and NCEOG. The analyses were calculated by

grade level due to both the ORF and NCEOG scores changing achievement ranges at each grade level.

All three grade-level mean NCEOG scale scores were equivalent to a proficiency level of II, limited understanding of grade-level standards; mean ORF scores were equivalent to just below grade-level proficiency (yellow); and mean TRC scores for third and fourth grades were equivalent to just below grade-level proficiency (yellow), while fifth grade was equivalent to way below grade-level proficiency (red). This revealed that, on average, the grade levels had equivalent proficiency levels on the NCEOG, ORF, and TRC, with the exception of fifth-grade TRC scores.

The researcher utilized the frequency counts of the demographic variables of gender and ethnicity to determine mean scores for each, helping to further analyze the relationship that existed between the two assessments. In this study, female participants scored higher means and proficiency levels than male participants on the NCEOG, ORF, and TRC assessments in Grades 4 and 5, while males scored higher on all assessments in the third grade.

In this study, in terms of ethnicity, Hispanic participants scored higher means and proficiency levels than other ethnicities on the fourth-grade ORF and TRC and fifth-grade NCEOG, ORF, and TRC. Multi-Racial participants scored higher means and proficiency levels on the third- and fourth-grade NCEOG, while White participants scored higher means and proficiency levels on the third-grade ORF and TRC. One noted observation was the Black participants, who had the highest number of participants in all three grade levels, never had the highest mean score at any grade level on any of the assessments examined in this study, which could be an area of future study.

Overall, based on the descriptive statistics, it could be stated that an observed

relationship does exist between NCEOG and mClass Reading 3D assessments. This relationship is revealed in several commonalities: (a) the participants' proficiency levels are parallel between the two assessments in Grades 3, 4, and 5; (b) consistent growth data for males and females in all three grade levels on both assessments; and (c) consistent proficiency levels and mean scores for Black participants in all three grade levels on both assessments. The inferential statistics, discussed in the next section, reveal the disaggregated data by gender and ethnicity, confirming that a relationship does exist between NCEOG and mClass Reading 3D assessments and to what extent it exists.

Inferential statistics. Analysis of the inferential statistics in this study brought to light several observations. These statistics were used to further determine the relationships and predictability that existed between NCEOG and mClass Reading 3D. The analyses were calculated by grade level due to both the ORF and NCEOG scores changing achievement ranges at each grade level.

The initial observations based on the descriptive statistics appeared to show a relationship existed between the two assessments; but the rest of the analysis determined to what extent the relationship existed in terms of correlation, statistical significance, and predictability. The researcher utilized the gender and ethnicity frequency counts and descriptive statistics to determine the correlation coefficients for each; but due to the low number of cases for the Hispanic and Multi-Racial ethnicities, the researcher combined them to form the Other ethnicity variable when running the Pearson correlation statistics.

The Pearson correlation determined that there was statistical evidence of a positive correlation and statistically significant relationships between NCEOG and mClass Reading 3D scores at all three grade levels involved in the study. As the correlation analysis was examined more closely, it revealed that the strongest relationship

for each grade level was between NCEOG and ORF scores.

Third-, fourth-, and fifth-grade male and female participants showed a positive correlation and statistically significant relationship between the NCEOG and mClass Reading 3D assessments, which is consistent with the observed data in the descriptive statistics. All ethnicities, except the following variables, also revealed a positive correlation and statistically significant relationship between the two assessments: third- and fourth-grade Other ethnicity (not statistically significant with ORF in third and neither ORF nor TRC in fourth), and fifth-grade White ethnicity (not statistically significant with ORF). This was an interesting find in relation to the descriptive statistics data which revealed Hispanic (which was part of the Other ethnicity) participants had the highest proficiency levels in fourth-grade ORF and TRC assessments. The Pearson correlation test does show a positive correlation exists between Hispanic participants' assessment scores; it was just not found to be statistically significant, meaning it has higher probability that the results happened by chance (StatPac, 2013). This could be due to the low number of participants who are included in this variable.

In all three grade levels, the female participants had the highest correlation between NCEOG, ORF, and TRC, with the exception of third-grade males who had a higher correlation between NCEOG and TRC than females. The third- and fourth-grade White participants had the highest correlation between NCEOG, ORF, and TRC, and fifth-grade White participants between NCEOG and TRC. The fifth-grade Other ethnicity participants had the highest correlation between NCEOG and ORF.

A relationship has been established between NCEOG and mClass Reading 3D, which answers the first research question of this study. The second research question refers to the extent that the mClass Reading 3D assessment predicts student scores on the

NCEOG. To answer this question, the researcher analyzed results from a multiple regression test. This test revealed that in all three grade levels, mClass Reading 3D statistically significantly predicted the student scores on the NCEOG. Third grade had the strongest predictability by determining that both portions of mClass Reading 3D, ORF and TRC, added to the statistical significance of the prediction for student scores on the NCEOG; whereas, in fourth grade only the ORF portion, and in fifth grade only the TRC portion of mClass Reading 3D, added statistical significance of the predication of student scores on the NCEOG. Due to the exclusion of some gender and ethnicity variables at all three grade levels, the researcher was unable to determine the predictability of mClass Reading 3D to NCEOG according to the gender and ethnicity variables.

The multiple regression analysis revealed findings that correlate to the descriptive statistics and Pearson correlation analysis, where in all three grade levels the proficiency levels were consistent between NCEOG and mClass Reading 3D, the Pearson correlation determined consistent positive correlations and statistical significance between the two assessments, and the multiple regression revealed mClass Reading 3D did statistically significantly predict scores on the NCEOG.

Overall, based on the descriptive and inferential statistics, it could be stated that a relationship does exist between NCEOG and mClass Reading 3D assessments, and at all three grade levels included in the study, mClass Reading 3D student scores serve as a predictor of student success on the NCEOG Reading Comprehension assessment.

Contributions. This study has made several contributions to the field of education and to the research on predictability of formative assessments to high-stakes assessments. The available research on the formative assessment mClass Reading 3D as

a whole, especially in terms of the TRC portion, and its relationship and predictability of high-stakes assessments like the NCEOG, is lacking. This study adds significantly to the body of research that is needed as mClass Reading 3D is growing in its utilization across states like North Carolina.

The design of the study and the literature review provides a strong background and research base for educators. Even though it targets two specific assessments and uses historical data, it can be easily replicated to include other similar assessments across the nation. It provides research and data that reveal to educators why it is imperative to determine the relationship and predictability of the formative assessments to high-stakes assessments. The findings of this study may also help to support effective student interventions by basing the interventions on accurate data, ultimately leading to an increase of student success on the EOY summative assessments in reading.

Limitations

Even though this study's results were consistent with previous research, the findings do have limitations that should be considered. The school in the study was the only school in the district at the time that was using mClass Reading 3D for all students in Grades 3-5. Since the results were from only one elementary school in North Carolina, creating a small sample size, the ability to generalize results may be limited. If the study were to be replicated with a larger sample size, it could provide more generalizability of the results.

The data gathered for analysis only included the EOY benchmark scores from mClass Reading 3D. Since the EOY benchmarks for mClass Reading 3D occurred only a few weeks prior to the NCEOG assessment, there was little time in between to change the outcomes of NCEOG based on the results of mClass Reading 3D. This is a potential

limitation that could be overcome in future studies by using all three benchmark periods within the assessment year for prediction of student scores on the NCEOG.

This study was cross-sectional, examining one set of participants at one point in time. The data were collected from participants during the 2010-2011 school year. The observations and conclusions in this study are not longitudinal. This limitation could be overcome in future studies by replicating the study as a longitudinal study.

There were several extraneous factors out of the researcher's control that could have impacted student scores on mClass Reading 3D and/or NCEOG: (a) the teachers utilized the same standards, curriculum, and resources; however, they each had their own individual way of incorporating these into daily instruction; (b) the teachers used different instructional and management strategies, creating different classroom climates; and (c) the students, even though they were from the same neighborhoods, each had different home support systems and backgrounds that were reflected in their classroom environments.

Implications for Education

This study has several implications for education. As the nation continues to strive towards student accountability through formative and high-stakes testing, it is important that the assessments align and prepare students for success on statewide accountability targets. The results of this study reveal that mClass Reading 3D has a statistically significant relationship and is predictive of the high-stakes test in reading for North Carolina, NCEOG, in Grades 3-5. This shows that the ORF and TRC portion of the mClass Reading 3D assessment assesses skills necessary for third- through fifth-grade students to show proficiency on the reading comprehension portion of the NCEOG. It also determines that the ORF and TRC contain data that can provide accurate progress

monitoring and interventions towards student success on grade-level expectations.

The data in this study are historical, which helps to form a comparative baseline for current and future implications. Since this study's data were collected, several statewide initiatives have taken place to further emphasize the importance of its results. North Carolina has put into action an Elementary and Secondary Education Act (ESEA) flexibility waiver which provides flexibility on specific requirements of NCLB (NCDPI, 2012), which included a shift from federal accountability sanctions (Annual Yearly Progress) to state accountability designations by Annual Measureable Objectives (AMO). The school in this study moved from the AYP sanction of Corrective Action to the AMO designation of priority school. As a result, they received the School Improvement Grant (SIG) which created an administrative and staff turnover and implementation of new initiatives in order to raise student proficiency on grade-level standards as reflected on the NCEOG. The school continued to implement mClass Reading 3D school-wide, and their reading proficiencies on the NCEOG increased from 34% in 2010-2011 to 46% in 2011-2012, which, after this study, it can be determined that mClass Reading 3D was one of the contributing factors to the increase in reading proficiency scores.

North Carolina also expanded the Pilot program for mClass Reading 3D to kindergarten through third grade statewide in 2012-2013, as part of the Excellent Public Schools Act Read to Achieve House Bill 950/S.L. 2012-142 Section 7A. This study adds to the support of the program expansion by providing statistical evidence that mClass Reading 3D is a predictor of student success on the NCEOG and can be utilized as data to drive instruction and interventions.

Since, mClass Reading 3D is a valid predictor of student success on the NCEOG, there are some recommendations to be considered in light of recent initiatives and data

reflections for the study's targeted school and district. The first is to continue to provide professional development interpreting the assessment data, intervention strategies based on the data, and giving the assessment with fidelity. Guskey (2003) stated, "Assessments must be followed by high-quality, corrective instruction designed to remedy whatever learning errors the assessment identified" (p. 8). Knowing that mClass Reading 3D data are predictors to student scores on NCEOG, utilizing that data to guide corrective instruction will help to support increases in student proficiency on grade-level standards and, ultimately, scores on the NCEOG. The school and district administrators can also use the data provided by mClass Reading 3D as contributing factors in making educational decisions about instructional programs, resource allocations, staff distributions, and school scheduling in order to effectively run a data-driven school.

Recommendations for Future Research

This study on the relationship and predictability of mClass Reading 3D and NCEOG assessments was a smaller scale study with limited generalizability due to the sample size, but its results do impact daily instruction and interventions. Based on the results of the study and the literature review, several recommendations for future research have been made. The recommendations are listed below.

1. Future replications of this study across different schools in other districts to provide generalizability of the findings.
2. Future studies using all three benchmark periods in mClass Reading 3D within the assessment year for prediction of student scores on the NCEOG.
3. Future longitudinal study following a cohort of earlier grade levels to determine the impact of growth over time on the relationship/predictability of mClass Reading 3D and NCEOG assessments.

4. Future studies correlating the NCEOG based on Common Core standards and mClass Reading 3D.

5. Future studies analyzing the relationship between mClass Reading 3D and NCEOG based on other variable factors such as AIG, EC, LEP, students who have been retained, and students with high mobility rates.

6. Future studies analyzing the relationship of other DIBELS assessments that are part of the mClass Reading 3D assessment in lower grade levels such as NWF and PSF, with high-stakes assessments.

7. Future studies analyzing the impact of interventions based on mClass Reading 3D data and student proficiency levels on the NCEOG.

8. Future studies on the impact of progress monitoring through mClass Reading 3D and student proficiency level growth on the NCEOG.

These are just a beginning point for the areas that could be possibilities for future research and literature. Future research will continue to determine the relationship between mClass Reading 3D and high-stakes assessments as standards, assessments, and instructional strategies change over time. This is will ensure that as long as it is being utilized as a formative assessment in the classroom, it will be relevant to increasing student proficiency levels on the EOY high-stakes assessment.

Conclusion

In response to the federal and state expectations and initiatives in student accountability, there is a growing need for the use of formative assessments to inform instruction and best meet needs of all students. State- and district-wide initiatives promote district, school, and classroom data collection to guide data-driven instructional decision making through daily instruction and interventions. Therefore, the formative

assessment data should be predictive of the student success on the high-stakes assessments that directly correlate to state, district, and school accountability targets.

While this study was not void of limitations, it may hold significance to those stakeholders considering using or already using mClass Reading 3D in the State of North Carolina. Overall findings of this study have implications on current and future initiatives such as the Read to Achieve bill in North Carolina, school and classroom instructional decisions, and student success on the NCEOG. The findings of this study are consistent with previous research suggesting the ORF portion of mClass Reading 3D can be used to predict performance on high-stakes assessments of reading. The findings also added research to an area that was lacking, revealing that the TRC portion of mClass Reading 3D statistically significantly predicated student scores on the NCEOG.

Results of the study clearly supported the use of mClass Reading 3D in third, fourth, and fifth grades as a data source for determining data-driven instruction and interventions to use for prediction of reading proficiency on the NCEOG. This provides educators with the confidence to utilize the mClass Reading 3D data as an effective source of instructional decision making.

Results of this study should be of interest to all educators. The results, in part, reveal the importance of formative assessments like mClass Reading 3D being predictors of performance on high-stakes assessments like NCEOG in order to inform educational decisions. The study findings also provide opportunities for educators to adjust daily instruction and improve student outcomes by providing data-driven interventions. This study should add strength to the educational field and urge researchers to continue with the recommendations for future research, ensuring mClass Reading 3D continues to hold a statistically significant relationship and be a predictor of NCEOG scores, even as

initiatives and standards change over time.

References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.
- Adams, M. J., Foorman, B., Lundberg, I., & Beeler, T. (1998). The elusive phoneme: Why phonemic awareness is so important and how to help children develop it. *American Educator*, 22(1-2), 18-29.
- Adequate Yearly Progress. (2004). Retrieved May 19, 2011, from <http://www.edweek.org/ew/issues/adequate-yearly-progress/>
- Allington, R. (1983). Fluency: The neglected reading goal. *The Reading Teacher*, 36(6), 556-561.
- Anderson, R. C., Hiebert, E. H., Scott, J. A., Wilkinson, I. A., & The Commission on Reading. (1985). *Becoming a nation of readers: The report of the commission on reading*. Washington, DC: National Academy of Education.
- Baily, A. L., & Drummond, K. V. (2006). Who is at risk and why? Teachers' reasons for concern and their understanding and assessment of early literacy. *Educational Assessment*, 11(3&4), 149-178. doi:10.1207/s15326977ea1103&4_2
- Barger, J. (2003). *Comparing the DIBELS Oral Reading Fluency indicator and the North Carolina end of grade reading assessment* (Technical Report). Asheville, NC: North Carolina Teacher Academy.
- Beck, I., & Juel, C. (1995). The role of decoding in learning to read. *American Educator*, 19(2), 8-25.
- Berger, D. (2003). *Introduction to multiple regression*. Retrieved from wise.cgu.edu/downloads/Regression.doc
- Binkley, M. R. (1986). *Becoming a nation of readers: Implications for teachers*. Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED271741). Retrieved July 21, 2011, from EBSCOHost ERIC database
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice*, 5(1), 7-74. doi:10.1080/0969595980050102
- Bond, G. I., & Dykstra, R. (1967). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly*, 2, 5-142. doi:10.23071746948

- Bond, G. I. & Dykstra, R. (1997). The cooperative research program in first-grade reading instruction. *Reading Research Quarterly*, 32, 345-427.
doi:10.1598/RRQ.32.4.4
- Braden, J. P., & Schroeder, J. L. (2004). High stakes testing and No Child Left Behind: Information and strategies for educators. Retrieved from <http://www.nasponline.org/communications/spawareness/highstakes.pdf>
- Buck, J., & Torgeson, J. (2003). *The relationship between performance on a measure of oral reading fluency and performance on the Florida Comprehensive Assessment Test* (Technical Report). Tallahassee, FL: Florida Center for Reading Research.
- Butler, S., Urrutia, K., Buenger, A., & Hunt, M. (2010). *A review of the current research on comprehension instruction*. National Reading Technical Assistance Center. Retrieved from <http://www2.ed.gov/programs/readingfirst/support/compfinal.pdf>
- Chall, J. S. (1967). *Learning to read: The great debate*. New York: McGraw Hill.
- Chall, J. S. (1996). *Learning to read: The great debate (revised)*. New York: McGraw Hill.
- Chall, J. S., & Jacobs, V. A. (1983). Writing and reading in the elementary grades: Developmental trends among low SES children. *Language Arts*, 60, 617-626.
- Chappuis, S., & Stiggins, R. J. (2002). Classroom assessment for learning. *Educational Leadership*, 60(1), 40-43.
- Clay, M. M. (2005). *An observation survey of early literacy achievement*. Portsmouth, NH: Heinemann.
- Clay, M. M. (2013). *An observation survey of early literacy achievement* (3rd ed.). Portsmouth, NH: Heinemann. (Original work published, 1993)
- Cowen, J. E. (2003). *A balanced approach to beginning reading instruction: A synthesis of six major U.S. research studies*. Newark, DE: International Reading Association.
- Coyne, M. D., McCoach, B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disabilities Quarterly*, 30(2), 74-88.
doi:10.2307/30035543
- Coyne, M. D., Simmons, D. C., Hagan-Burke, S., Leslie, E. S., Kwok, O., Minjung, K., Rawlinson, D. M. (2013). Adjusting beginning reading intervention based on student performance: An experimental evaluation. *Exceptional Children*, 80(1), 25-44.

- Denis, D. J. (2011). *Multiple linear regression using SPSS, part I*. Retrieved from http://psychweb.psy.umt.edu/denis/datadecision/front/stat_II_2011/psyx_521_multiple_regression_part_I.pdf
- Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children*, 52(3), 219-232.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, 37(3), 184-192. doi:10.1177/00224669030370030801
- Department of Numbers. (n.d.). *North Carolina unemployment*. Retrieved from <http://www.deptofnumbers.com/unemployment/north-carolina/>
- Dunn, K., & Mulvenon, S. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research, & Evaluation*, 14(7), 1-11. Retrieved from <http://pareonline.net/getvn.asp?v=14&n=7>
- Dynamic Measurement Group. (2008). *Linking assessment to instruction: Using Dynamic Indicators of Basic Early Literacy Skills in an outcomes-driven model*. Oregon: Kaminski, R. & Cummings, K.
- Florida Association of School Psychologists. (n.d.). *Information paper 8 progress monitoring*. Retrieved from http://www.fasp.org/PDF_Files/New_&_Important/Iowa8ProgressMonitor.pdf
- Florida Department of Education. (2001). *Just Read Florida*. Retrieved from <http://www.justreadflorida.com/>
- Foorman, B. R., Francis, D. J., Fletcher, J. M., Schatschneider, C., & Mehta, P. (1998). The role of instruction in learning to read. *Journal of Educational Psychology*, 90(1), 37-55. doi:10.1037//0022-0663.90.1.37
- Fountas, I., & Pinnell, G. (2006). *Teaching for comprehending and fluency*. Portsmouth, NH: Heinemann.
- Fountas, I., & Pinnell, G. (2009). *When readers struggle: Teaching that works*. Portsmouth, NH: Heinemann.
- Fountas, I., & Pinnell, G. (2010). *The continuum of literacy learning, Grades Pre-K-8* (2d ed.). Portsmouth, NH: Heinemann.
- Fournier, R., North, C., & LaPointe, M. (2009, Jan/Feb). What are benchmark assessments and how do they work? *EdEvidence Newsletter*. Retrieved from <http://www.relnei.org/newsletters.php?nlid=19&nlapno=6>

- Fuchs, D., Fuchs, L., Mathes, P., & Simmons, D. (1996). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *The American Educational Research Journal*, 34, 174-206. doi:10.3102/00028312034001174
- Gaston County Health Department. (2011). *2011 state of the county health report Gaston County, NC*. Retrieved from <http://www.gastongov.com/docs/health-department/gaston-sotch-report-2011-final.pdf>
- General Information about DIBELS Measures. (2008). Retrieved January 23, 2012, from <https://dibels.uoregon.edu/docs/dibelsinfo.pdf>
- Good, R. H., & Kaminski, R. A. (n.d.) What are DIBELS? Dynamic Indicators of Basic Early Literacy Skills. *Dynamic Measurement Group*. Retrieved from <http://www.dibels.org/dibels.html>
- Good, R. H., Simmons, D. C., & Kameenui, E. J. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third grade high-stakes outcomes. *Scientific Studies of Reading*, 5, 257-288. doi:10.207/S1572799XSSR0503_4
- Goodness of Fit: R vs. Adjusted R. (n.d.). Retrieved September 11, 2013, from <http://www.bus.ucf.edu/faculty/rhofler/file.axd?file=2012%2F2%2FR2+vs+adj+R2.pdf>
- Great Schools. (n.d.). *NCLB definitions*. Retrieved May 20, 2012, from <http://www.greatschools.org/definitions/or/nclb.html>
- Guskey, T.R. (2003). How classroom assessments improve learning. *Educational Leadership*, 60(5), 6-11.
- Hall, S., & Moats, L. (1999). *Straight talk about reading*. New York, NY: McGraw-Hill.
- Heritage, M. (2007). Formative assessment: What do teachers need to know and do? *Phi Delta Kappan*, 89(2), 140-145.
- Honig, B., Diamond, L., & Gutlohn, L. (2000). *Teaching reading sourcebook*. Emeryville, CA: CORE.
- Hunt, L.C. (1969). Learning to read: The great debate EKNE research review. Washington, DC: Department of Elementary-Kindergarten-Nursery Education, National Education Association. (ERIC Document Reproduction Service No. ED033011). Retrieved July 21, 2011, from EBSCOHost ERIC database

- Hupert, N., Heinze, J., Gunn, G., Stewart, J., & Honey, M. (2007). An analysis of technology assisted progress monitoring to drive improved student outcomes. *Wireless Generation*. Retrieved from http://www.wirelessgeneration.com/pdf/white-papers/Progress_Monitoring_White_Paper.pdf
- Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grade. *Journal of Educational Psychology*, 80(4), 437-447.
- Juel, C. (1991). Beginning reading. In Barr, Kamil, Mosenthal, and Pearson (eds.), *Handbook of Reading Research*, (pp. 759-788). New York: Longman.
- Karoly, L. A., Kilburn, M. R., & Cannon, J. S. (2005). *Early childhood interventions: Proven results, future promise*. Santa Monica, CA: RAND Corporation.
- Keene, E., & Zimmerman, S. (1997). *Mosaic of thought: Teaching comprehension in a reader's workshop*. Portsmouth, NH: Heinemann.
- Lund Research Group. (2013). *Descriptive and inferential statistics*. Retrieved from <https://statistics.laerd.com/statistical-guides/descriptive-inferential-statistics.php>
- Madelaine, A., & Wheldall, K. (1999). Curriculum-based measurement of reading. *International Journal of Disability, Development, and Education*, 46(1), 71-85. doi:10.1080/103491299100731
- Madjidi, F. (n.d). *Non-experimental research design* [PowerPoint slides]. Retrieved from <http://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&ved=0CEAQFjAD&url=http%3A%2F%2Ffaculty.pepperdine.edu%2Ffmadjidi%2FStatistics%2Felap%2Fnonexp.ppt&ei=SSOHT7nlJqSD8AGm8dWjCA&usg=AFQjCNHlK-t76gDzKQK57SeuNQeeHdWunw>
- Marston, D. B. (1989). A curriculum-based measurement approach to assessing academic performance: What it is and why do it. *Curriculum-Based Measurement: Assessing Special Children*, 18-78.
- Mathis, P., Grek, M., Howard, J., Babyak, A., & Allen, S. (1999). Peer-assisted learning strategies for first grade readers: A tool for preventing early reading failure. *Learning Disabilities Research and Practice*, 14(1), 50-60. doi:10.1207/SLDRP1401_5
- mClass:Reading 3D-Text Reading & Comprehension. (n.d.). *In Rti4Success*. Retrieved from http://www.rti4success.org/tools_charts/popups_screening/area/mclass3d-reading.html
- mClass:Reading 3D. (2010). *In mClass Reading 3D*. Retrieved from https://www.mclasshome.com/wgenhelp/R3D2_Harcourt/Reading_3D/The_mCLASS_Reading_3D_System.htm

- Menzies, H., Mahdavi, J., & Lewis, J. (2008). Early intervention in reading: From research to practice. *Remedial and Special Education, 29*(2), 67-77. doi:10.1177/10741932508315844
- Montgomery County Public Schools. (2005, June 3). *Montgomery County Public Schools announces partnership with innovative technology company providing handheld and web-based assessment solutions for teachers* [Press release]. Retrieved from http://www.wirelessgeneration.com/pdf/press-releases/Reading_3D_Montgomery_County_Press_Release_01.pdf
- Nagy, W. E. (1988). *Teaching vocabulary to improve reading comprehension*. Newark, DE: International Reading Association.
- National Center for Education Statistics (NCES). (2003). *National assessment of adult literacy* [Data file]. Retrieved from http://nces.ed.gov/naal/kf_demographics.asp
- National Commission on Excellence in Education (NCEE). (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: United States Department of Education. doi:10.1086/461348
- National Institute of Child Health and Human Development (NICHD). (2000). *Report of the National Reading Panel. Teaching children to read: an evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. Retrieved April 25, 2011, from <http://www.nichd.nih.gov/publications/nrp/smallbook.htm>
- No Child Left Behind Act of 2001, PL 107-110, 115 Stat. 1425, 20 U.S.C. §§ 6301 et seq.
- No Child Left Behind Act Law and Legal Definition. (n.d.). Retrieved May 19, 2011, from <http://definitions.uslegal.com/n/no-child-left-behind-act/>
- North Carolina Department of Public Instruction. (n.d.). *Public schools of North Carolina annual dropout reports*. Retrieved from <http://www.ncpublicschools.org/research/dropout/reports/>
- North Carolina Department of Public Instruction. (2007). *North Carolina Reading First initiative third annual evaluation report*. Retrieved from www2.ed.gov/programs/readingfirst/evaluationreports/nc1.doc
- North Carolina Department of Public Instruction, Division of Accountability Services, Testing Section. (2011). Retrieved December 6, 2011, from www.ncpublicschools.org/accountability/testing
- North Carolina Department of Public Instruction. (2012). *Elementary Secondary Education Act*. Retrieved from <http://www.ncpublicschools.org/program-monitoring/esea/>

- North Carolina State Board of Education. (2010). *Minutes of the North Carolina State Board of Education Meeting*. Retrieved from <http://www.ncpublicschools.org/docs/stateboard/meetings/2010/minutes/final/11minutes.pdf>
- Pearson, P. D., Dole, J. A., Duffy, G. G., & Roehler, L. R. (1992). Developing expertise in reading comprehension. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about reading instruction* (2nd ed., pp. 145-199). Newark, DE: International Reading Association.
- Pellegrino, J. W. (2004). *The evolution of educational assessment: Considering the past and imagining the future*. Princeton, NJ: Educational Testing Service.
- Peterson, C., Caverly, D., Nicholson, S., O'Neal, S., & Cusenbary, S. (2000). *Building reading proficiency at the secondary level: A guide to resources*. Austin, TX: Southwest Educational Development Laboratory.
- Pressley, M. (1976). Mental imagery helps eight year olds remember what they read. *Journal of Educational Psychology*, 68, 355-359. doi:10.1037/110022-0663.68.3.355
- Readance, J., & Barone, D. (1998). Revisiting the first grade studies. *Reading Research Quarterly*, 32(4), 1-111. doi:10.1598/RRQ.32.4.2
- Reading 3D Brochure. (2009). Retrieved May 6, 2011, from http://www.wirelessgeneration.com/pdf/brochures/mCLASS_Reading_3D_Brochure.pdf
- Reschly, A. (2010). Reading and school completion: Critical connections and Matthew effects. *Reading and Writing Quarterly*, 26, 67-90. doi:10.1880/10573560903397023
- Reynolds, A. J., Temple, J. A., Ou, S., Robertson, D. L., Mersky, J. P., Topitzes, J. W., & Niles, M. D. (2007). Effects of a school-based, early childhood intervention on adult health and well-being: A 19 year follow-up of low-income families. *JAMA Pediatrics*, 161(8), 730-739. doi:10.100//archpedi.161.8.730
- Richardson, E., & DiBenedetto, B. (1996). Identifying dyslexic students. In L.R. Putnam (Ed.), *How to become a better reading teacher: Strategies for assessment and intervention* (pp. 53-63). Englewood Cliffs, NJ: Merrill.
- Ross, J. (2004). Effects of running records assessment on early literacy achievement. *The Journal of Educational Research*, 97, 186-194. doi:10.3200/JOER.97.4.196-195

- Rowell, P.S. (2009) *The relationship between the dynamic indicators of basic early literacy skills (DIBELS) oral reading fluency (ORF) and the Alaska standards based assessment (SBA) for proficiency in reading*. (Doctoral dissertation, Liberty University). Retrieved from: <http://digitalcommons.liberty.edu/cgi/viewcontent.cgi?article=1149&context=doctoral>
- Running Records and Benchmark Books. (n.d.). In *Reading A-Z*. Retrieved from <http://www.readinga-z.com/guided/runrecord.html>
- Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C. K., & Torgesen, J. K. (2007). *Reading interventions for adolescent struggling readers: A meta-analysis with implications for practice*. Portsmouth, NH: RMC Research, Center on Instruction.
- Scholastic Reading Inventory (SRI). (n.d.). Retrieved May 26, 2011, from http://teacher.scholastic.com/products/sri_reading_assessment/programoverview.htm
- Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment*, 24(1), 19-35.
- Simon, G. (2003). Multiple regression diagnostics. Retrieved September 10, 2013, from <http://people.stern.nyu.edu/wgreene/Statistics/MultipleRegressionDiagnosticsCollection.pdf>
- Snow, C. E., Burns, S., & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.
- Stanovich, K. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360-406. doi:10.1598/RRQ.21.4.1
- StatPac. (2013). Statistical significance. Retrieved September 22, 2013, from <http://www.statpac.com/surveys/statistical-significance.htm>
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758-765.
- Sweet, A., & Snow, C. (2003). *Rethinking reading comprehension: Solving problems in the teaching of literacy*. New York: Guilford Press.
- Taylor, B. M., Pearson, P. D., Clark, K., & Walpole, S. (2000). Effective schools and accomplished teachers: Lessons about primary-grade reading instruction in low-income schools. *Elementary School Journal*, 101, 121-165. doi:10.1086/499662

- TetraData. (n.d.). (*School District of School A*) demographic and assessment data for 2010-2011 school year [Data File]. Retrieved from <https://nc-gaston.tetradata.com/ANALYSIS/Default.aspx>
- Text Reading and Comprehension. (2010). In *mClass Reading 3D*. Retrieved from https://www.mclasshome.com/wgenhelp/R3D2_Harcourt/Reading_3D/Assessment_Scoring/TRC/Text_Reading_and_Comprehension.htm
- Torgeson, J. (2000). Individual differences in response to early interventions in reading: The lingering problem of treatment resisters. *Learning Disabilities Research and Practice, 15*(1), 55-64. doi:10.1207/SLDRP1501_6
- Torgesen, J. (2004). Avoiding the devastating downward spiral: The evidence that early intervention prevents reading failure. *American Educator, 28*(3). Retrieved from <http://www.aft.org/newspubs/periodicals/ae/fall2004/index.cfm>
- Tunmer, W. (2008). Recent developments in reading intervention research: Introduction to the special issue. *Reading and Writing, 21*(4), 299-316. doi:10.1007/S11145-007-9108-4
- United States Department of Education (USDE). (2009). *Reading first*. Retrieved from <http://www2.ed.gov/programs/readingfirst/index.html>
- United States Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2011 Reading Assessment. (2011). Retrieved December 6, 2011, from <http://nces.ed.gov/nationsreportcard/pdf/stt2011/2012454NC4.pdf>
- University of Oregon Center on Teaching and Learning. (n.d.). *DIBELS oral reading fluency and retell fluency*. Retrieved from <https://dibels.uoregon.edu/measures/orf.php>
- Vellutino, F. R., & Scanlon, D. M. (2001). Emergent literacy skills, early instruction, and individual differences as determinants of difficulties in learning to read: The case for early intervention. In S. B. Neuman & D. K. Dickinson (eds.), *Handbook of early literacy research* (pp. 295-321). New York: Guilford Press.
- Wanzek, J., Vaughn, S., Scamacca, N. K., Metz, K., Murray, C., Roberts, G., & Danielson, L. (2013). Extensive reading interventions for students with reading difficulties after grade 3. *Review of Educational Research, 83*(2), 163-195. doi:10.3102/0034654313477212
- Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum based measurement in reading. *The Journal of Special Education, 41*(2), 85-120.

- Wilson, J. (2005). The relationship of dynamic indicators of basic early literacy skills (DIBELS) oral reading fluency to performance on Arizona instrument to measure standards (AIMS). (Technical report). Retrieved July 2013, from <https://dibels.uoregon.edu/docs/techreports/arizona.pdf>
- Wireless Generation. (2010). *North Carolina diagnostic assessment for reading: Training of trainers manual*. Unpublished manual.
- Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11*, 85-104. doi:10.1207/515326977ea1102_1
- Yeo, S. (2010). Predicting performance on state achievement tests using curriculum-based measurement in reading: A multilevel meta-analysis. *Remedial and Special Education, 31*(6), 412-422. doi:10.1177/10741932508327463

Appendix

Coding Key for TRC Levels

Coding Key for TRC Levels

*Number	**F&P Level	*Number	**F&P Level	*Number	**F&P Level
301	B	314	S	411	T
302	F	315	T	412	U
303	H	316	U	501	I
304	I	401	E	502	K
305	J	402	I	503	R
306	K	403	L	504	S
307	L	404	M	505	T
308	M	405	N	506	U
309	N	406	O		
310	O	407	P		
311	P	408	Q		
312	Q	409	R		
313	R	410	S		

Note. *First number in each code indicates grade level; **F&P=Fountas and Pinnell. Observations were made based on the analysis of the measures of central tendency.