ABSTRACT


Standardized tests are an education reality and an important accountability consideration in most states and school systems. Most states require standardized assessments to meet requirements of the federal No Child Left Behind Act of 2001. Changes to curriculum and instruction and to the school culture frequently occur through a school improvement process, and standardized test data are often used to inform these decisions. The school improvement process and professional development should focus on student learning, but how administrators and teachers perceive standardized testing and the ensuing data analyses is an important consideration in understanding what needs improvement and the professional development that best supports student learning. Ultimately, standardized assessment results should inform changes to curriculum and instruction. However, who decides what change is needed or how to implement the change? In this study, the researcher seeks to understand the value of existing students’ testing in middle school as it relates to and, perhaps, predicts their high school achievement in ninth grade.

Keywords: accountability, criterion-referenced test, Georgia Criterion-Referenced Competency Test (CRCT), Georgia High School Graduation Test (GHSGT), Georgia End-of-Course Test (EOCT), Georgia Performance Standards (GPS), high-stakes testing, secondary education, standardized assessment, summative assessment
DEDICATION

Throughout this dissertation journey, my Bible study and my thoughts have often been of God as our rock. I have found Psalms 62:6 an especially powerful message from Him:

“He only is my rock and my salvation, My stronghold; I shall not be shaken” (NASB).

As much as I rely on Our Heavenly Father as my eternal, ever-present stronghold, my rock on this Earth is my husband of 33 years, Michael. He has always given me constant, unfailing encouragement and love. However much our sons and I swirled around in our many endeavors, Michael is our solid rock and keeps our family grounded. His support, encouragement, and unfailing patience, especially in light of becoming “empty nesters” 18 months ago, has been evidence of the wonderful, self-sacrificing man he is. Michael, this dissertation is dedicated to you, with my love.
ACKNOWLEDGMENTS

What a journey! First, to my husband, my sons, and their wives — Michael, Chris & Bobbie, Eric & Kristin — your love, prayers, and support have given me support and the confidence to complete this doctoral program. Wife and mother are my favorite and most important roles.

I want to commend Liberty University for remaining a steadfast Christian university in word and deed. Growing spiritually as well as professionally was a powerful aspect of my learning.

To Dr. Judy Shoemaker, you have been a blessing with your encouragement, availability, comments, and advice as a professor and as my dissertation committee chairperson. Dr. Puga and Dr. Curry, thank you for serving on my dissertation committee. Jim, I am especially grateful that you encouraged me toward a teaching certificate years ago and gave me my first administrative position a few years later. You taught me so much! Reinhardt University is fortunate to have you guiding their school of education.

I have been blessed with a wonderful editor, Dave Gorman, who took what I had written and just made it better. I also want to thank Dr. Jennifer Priestley and Chris Cusamano at Kennesaw State University for their invaluable help with the statistical analyses conducted for this study.

To Kristy, Evie, Sharon, Johnnie, and Susan: You are the best! You have been great encouragers, professional colleagues, traveling companions, sounding boards, and friends. I have so much respect and admiration for each of you, and I cannot wait until we can all call each other “Doctor.”
GEORGIA HIGH-STAKES TESTING: THE CORRELATION BETWEEN EIGHTH GRADE AND NINTH GRADE ACHIEVEMENT

by Venita L. Bruton

A Dissertation Presented in Partial Fulfillment Of the Requirements for the Degree Doctor of Education

Liberty University, Lynchburg, VA

November, 2011

APPROVED BY:

Dr. Judy Shoemaker, Chair November 11, 2011
Dr. Jose Puga, Committee November 11, 2011
Dr. James L. Curry, Jr., Committee November 11, 2011
Dr. Amanda Szapkiw, Chair of Graduate Studies November 11, 2011
# Table of Contents

ABSTRACT ................................................................................................................................ ii

CHAPTER 1: INTRODUCTION .......................................................................................... 1
  Background ...................................................................................................................... 1
  Problem Statement ........................................................................................................ 1
  Purpose Statement ........................................................................................................ 2
  Significance of the Study ................................................................................................ 3
  Research Questions ....................................................................................................... 4
  Hypotheses ..................................................................................................................... 5
  Identification of Variables ............................................................................................. 5
  Assumptions and Limitations ......................................................................................... 9
  Research Plan ................................................................................................................ 11
  Definition of Core Terms .............................................................................................. 13

CHAPTER 2: LITERATURE REVIEW ............................................................................. 16
  Theoretical Framework ................................................................................................ 16
  History of Standardized Testing ................................................................................... 20
  Accountability .............................................................................................................. 32
  Testing and School Reform .......................................................................................... 49
  Summary ....................................................................................................................... 56

CHAPER 3: METHODOLOGY ......................................................................................... 57
  Research Design .......................................................................................................... 58
  Participants ................................................................................................................... 63
  Setting ........................................................................................................................... 64
  Instrumentation ............................................................................................................ 66
  Procedures ..................................................................................................................... 70
  Data Analysis ............................................................................................................... 71

CHAPTER 4: RESULTS .................................................................................................. 74
  Overview of Problem .................................................................................................... 74
  Restatement of Purpose ............................................................................................... 75
  Instrumentation ............................................................................................................ 76
  Univariate Analyses ..................................................................................................... 76
  Bivariate Analyses ....................................................................................................... 85
  Correlational Analyses ............................................................................................... 93
  Findings ......................................................................................................................... 104
  Disaggregated Data ..................................................................................................... 104
  School-Level Influences on Standardized Testing Achievement ................................ 110
  Summary ....................................................................................................................... 112

CHAPTER 5: DISCUSSION ........................................................................................... 115
  Summary of Findings ................................................................................................... 116
  Discussion ..................................................................................................................... 118
LIST OF TABLES

Table 1.1: Research Student Groups ............................................................................... 12
Table 4.1: CRCT Ethnicity Frequency Table ................................................................. 77
Table 4.2: CRCT Gender Frequency Table .................................................................... 78
Table 4.3: Middle School Frequency Table .................................................................. 78
Table 4.4: CRCT Testing Year Frequency Table ............................................................ 79
Table 4.5: CRCT Descriptive Statistics .......................................................................... 80
Table 4.6: EOCT Ethnicity Frequency Table ................................................................. 82
Table 4.7: EOCT Gender Frequency Table .................................................................... 82
Table 4.8: High School Frequency Distribution ............................................................. 83
Table 4.9: EOCT Testing Year Frequency Table ............................................................ 83
Table 4.10: EOCT Descriptive Statistics ........................................................................ 84
Table 4.11: CRCT Bivariate Analyses by Test Year ........................................................ 85
Table 4.12: CRCT Bivariate Analyses by Ethnicity ....................................................... 87
Table 4.13: CRCT Bivariate Analyses by Gender ........................................................... 88
Table 4.14: CRCT Bivariate Analyses by School ............................................................ 88
Table 4.15: CRCT Bivariate Analyses by IEP ................................................................. 90
Table 4.16: EOCT Bivariate Analyses by Test Year ....................................................... 90
Table 4.17: EOCT Bivariate Analyses by Ethnicity ....................................................... 91
Table 4.18: EOCT Bivariate Analyses by Gender .......................................................... 92
Table 4.19: EOCT Bivariate Analyses by School ........................................................... 92
Table 4.20: EOCT Bivariate Analyses by IEP ............................................................... 93
Table 4.21: Reading CRCT to Literature EOCT Correlations by Year ......................... 94
Table 4.22: ELA CRCT to Literature EOCT Correlations by Year ............................... 96
Table 4.23: REAss/LitSS Parameter Estimates .............................................................. 99
Table 4.24: ELAss/LitSS Parameter Estimates ............................................................... 100
Table 4.25: Correlations by IEP Status ......................................................................... 105
Table 4.26: IEP Prediction Models ................................................................................ 106
Table 4.27: Correlations by Gender ............................................................................. 107
Table 4.28: Gender Prediction Models ......................................................................... 108
Table 4.29: Correlations by Ethnicity .......................................................................... 108
Table 4.30: Ethnicity Prediction Models ...................................................................... 109
LIST OF FIGURES

Figure 3.1: Howell’s Decision Tree for Selecting Statistical Tests ........................................ 62
Figure 4.1: Reading CRCT Distributions ........................................................................ 84
Figure 4.2: ELA CRCT Distributions ............................................................................. 84
Figure 4.3: Ninth Grade Literature EOCT Distributions ............................................... 84
Figure 4.4: ELA CRCT Performance 2006-2010 ............................................................ 86
Figure 4.5: Reading CRCT Performance 2006-2010 ...................................................... 86
Figure 4.6: Ninth Grade Literature EOCT Performance 2007-2011 .............................. 91
Figure 4.7: Scatterplots for REAss/LitSS OLS Regressions ........................................ 103
Figure 4.8: Scatterplots for ELAss/LitSS OLS Regressions .......................................... 104
LIST OF ABBREVIATIONS

Adequate Yearly Progress (AYP)
American Educational Research Association (AERA)
American Psychological Association (APA)
Annual measurable objective (AMO)
Association for Childhood Education International (ACEI)
Bartow County School System (BCSS)
Common Core Curriculum Standards (CCCS)
Council of Chief State School Officers (CCSSO)
Criterion-Referenced Competency Test (CRCT)
Depth of Knowledge (DOK)
Educational Testing Service (ETS)
Elementary and Secondary Education Act (ESEA)
End-of-Course Test (EOCT)
English Language Arts (ELA)
English Language Arts CRCT Scaled Scores (ELAss)
Georgia Department of Education (GaDOE)
Georgia High School Graduation test (GHSGT)
Georgia Performance Standards (GPS)
Graduate Record Examination (GRE)
Individualized Education Plan (IEP)
Institutional Review Board (IRB)
Ninth Grade Literature EOCT scaled scored (LitSS)
National Assessment of Educational Progress (NAEP)
National Council of Measurement in Education (NCME)
New American Standard Bible (NASB)
No Child Left Behind (NCLB)
Northwest Georgia Regional Educational Service Agency (NWRESA)
Performance Level (PL)
Reading CRCT Scaled Scores (ReaSS)
Scholastic Aptitude Test (SAT)
Socioeconomic status (SES)
Southern Association of Colleges and Schools (SACS)
Standard error of measurement (SEM)
Texas Assessment of Knowledge and Skills (TAKS)
CHAPTER 1: INTRODUCTION

High-stakes testing mandates and the *No Child Left Behind Act* of 2001 compel teachers to understand the implications of standardized testing. Moreover, it is important to use the test data analysis that is provided by the testing authority to impact instruction and, ultimately, students’ learning. The state of Georgia, along with the other 49 states, is mandating student assessment be used as the canon for measuring student achievement and for holding school systems, schools, and teachers accountable (Gabriel, 2010; Scot, Callahan, & Urquhart, 2009; Valli & Croninger, 2007; Vanderhaar, Muñoz, & Rodosky, 2006).

**Background**

The Georgia Department of Education (GaDOE) requires the Criterion-Referenced Competency Test (CRCT) in third through eighth grades, with benchmarks established for third, fifth, and eighth grades. It is composed of five subtests: (a) Reading, (b) English Language Arts (ELA), (c) Mathematics, (d) Science, and (e) Social Studies. All eleventh grade students must take the criterion-referenced Georgia High School Graduation Test (GHSGT) in the five subject areas: (a) ELA, (b) Writing, (c) Math, (d) Science, and (e) Social Studies. During their high school years, all Georgia students must also take the End-of-Course Test (EOCT) in eight subjects: (a) Ninth Grade Literature, (b) American Literature, (c) Math I, (d) Math II, (e) Physical Science, (f) Biology, (g) U.S. History, and (h) Economics.

**Problem Statement**

The problem is that educators in Georgia do not know if relationships exist between scaled scores on the eighth grade CRCT in Reading and ELA and the scaled
scores on the Ninth Grade Literature EOCT. The tests are important as they affect promotion/retention decisions and graduation requirements. In the transition from the CRCT in the eighth grade to the EOCT in ninth grade, it is important to understand the implications of this transition in order to support students as they begin high school.

Currently, no data are available that provide information on how well the CRCT relates to or can predict achievement on the EOCT. When students transition to high school, it is important to identify struggling learners and any gaps in content knowledge they have. If the CRCT is significantly correlated to the EOCT, and if any prediction can be made about EOCT performance based on CRCT achievement, then interventions can be designed to address the needs of struggling learners. Further, test data can help inform curricular and instructional decisions and fill any knowledge gaps. Ultimately, the information can be used to support student learning and increase students’ high school graduation rates.

**Purpose Statement**

The purpose of this study was to determine if there was any correlation in or predictive value of students’ achievement on state-mandated standardized testing from the CRCT in their eighth grade year to EOCT administration at the end of their ninth grade year. Middle school students must take the CRCT, which is comprised of subtests in Reading, ELA, Mathematics, Science, and Social Studies, in May of their eighth grade year. High school students are required to take both the GHSGT and the EOCT. The GHSGT is comprised of a writing subtest given to eleventh graders in September and four subtests in ELA, Mathematics, Science, and Social Studies, all of which are administered in March of the students’ eleventh grade year. The EOCT is given as a summative assessment in Ninth Grade Literature, American Literature, Math I, Math II,
Physical Science, Biology, U.S. History, and Economics across the four years of high school. Ninth graders, the participants in this study, are assessed on EOCTs in Ninth Grade Literature, Math I, and Physical Science.

**Significance of the Study**

The GHSGT is not given until the eleventh grade, mitigating a school’s ability to intervene quickly in addressing students’ achievement or students’ decisions about staying in school. The EOCT needs to be studied as an interim measure of student progress. Both the CRCT and the EOCT were developed to assess students’ mastery of the Georgia Performance Standards (GPS), which have been implemented over the past six years in a staged roll-out. The process of rolling out Georgia’s content area standards began during the 2005-2006 school year, with additional content area standards and tests added in each successive school year. Consequently, GPS-based test data are only available on a limited basis starting in 2006 until each content area’s standards were implemented.

Due to the newness of the GPS-based standardized assessments, it is important to study both aggregate and subgroups’ test results to identify if relationships exist between eighth and ninth grade achievement, especially to determine if there is any predictive value from the data analyses. If eighth grade standardized test achievement correlates with or can predict ninth grade standardized test achievement, then interventions can be developed for those at-risk eighth grade students who do not meet state minimum requirements for proficiency so that they can be supported toward academic success in ninth grade. Interventions could be developed for eighth grade students whose academic achievement is below the state’s minimum requirements and then provide students academic support in the ninth grade. The interventions could result in (a) an increase in
students’ test scores and grades and (b) a reduction in the dropout rate between students’
ninth and tenth grade years.

In the targeted school system, it is uncommon for high schools to have systemic
formative assessment or data-rich progress monitoring. The GHSGT’s eleventh grade
administration makes it impossible for educators to determine which students require
intervention. By this time, students may have dropped out. Thus, student intervention is
constrained by the administration of the GHSGT in the eleventh grade, which is
incompatible with reducing schools’ dropout rates.

Another important consideration is the state of Georgia’s proposed transition in
graduation accountability measures from the mandated GHSGT’s graduation
requirements to a series of EOCTs (“Georgia Department,” 2010). Georgia has agreed to
implement national Common Core Curriculum Standards (CCCS) and will use EOCTs to
assess them. The transition begins in the 2012 school year and culminates in school year
2015 (see Appendix A).

Research Questions

A research study on the relationship between achievement on students’ eighth
grade standardized tests and their ninth grade standardized achievement tests needs
careful design. This study addressed three research questions:

Research question 1. What is the relationship between students’ achievement on
the eighth grade Reading CRCT and the same students’ achievement on the Ninth Grade
Literature EOCT?

Research question 2. What is the relationship between students’ achievement on
the eighth grade ELA CRCT and the same students’ achievement on the Ninth Grade
Literature EOCT?
Research Question 3. If a relationship exists, what is the predictive value of students’ eighth grade achievement as measured by the Reading or ELA CRCT on the same students’ achievement on the Ninth Grade Literature EOCT?

In addition to these, implications for instruction will be discussed.

Hypotheses

Each research question was associated with a research hypothesis, so there are three hypotheses that were evaluated:

Research hypothesis 1. A statistically significant correlation will exist between eighth grade students’ achievement on the Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT. $H_{01}$: There will be no significant correlation between eighth grade students’ achievement on the Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

Research hypothesis 2. A statistically significant correlation will exist between eighth grade students’ achievement on the ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT. $H_{02}$: There will be no significant correlation between eighth grade students’ achievement on the ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

Research hypothesis 3. Students’ eighth grade performance on the Reading ELA CRCT will be predictive of the same students’ performance on the Ninth Grade Literature EOCT. $H_{03}$: There will be no predictive value between eighth grade performance on the Reading or ELA CRCT and the same students’ performance on the Ninth Grade Literature EOCT.

Identification of Variables

Two variables were identified for this study. First, there is the dependent,
criterion variable of the students’ scaled scores on the ninth grade EOCT in Ninth Grade Literature. The predictor variable is the students’ scaled scores on the eighth grade CRCT in Reading and ELA. Data were collected for each year of GPS-based test administration, which began in 2006 for the assessments in CRCT Reading, CRCT ELA, and EOCT Ninth Grade Literature.

Each eighth grade group will be paired by subject area with Ninth Grade Literature scores in the ninth grade. Eighth grade CRCT Reading scores were paired with Ninth Grade Literature EOCT scores, and eighth grade CRCT ELA scores were paired separately with the Ninth Grade Literature score. Students scaled scores were analyzed for each year’s subject area group and also collectively for all years of a subject area test. For example, eighth graders who took the Reading CRCT in spring 2007 and took the Ninth Grade Literature EOCT as ninth graders in spring 2008 were one group and were analyzed as one group.

There are five Reading groups for different paired testing years, and they were each analyzed separately. Each subject area’s aggregate data for all combined years of GPS-based testing were studied. For example, all data for students’ CRCT Reading and EOCT Ninth Grade Literature were combined into one dataset for analysis. These data were disaggregated by test year, school, gender, ethnicity, and students with disabilities. The study groups and data analyses for the ELA CRCT/ Ninth Grade Literature EOCT comparisons were developed in a similar manner as the Reading CRCT/Ninth Grade Literature EOCT dataset.

Prior to beginning analysis, the CRCT and EOCT must be discussed, and the tests’ scoring criteria must be stated. The CRCT was developed by the GaDOE as “part meeting federal requirements for state standards and assessments systems” (“CRCT,”
para. 8). The test was peer reviewed by a team of external experts, convened by the U. S. Department of Education, in the fields of standards and assessments. According to GaDOE, “The CRCT was found to meet nationally recognized professional and technical standards for assessment programs” (para. 8). Raw scores, scaled scores, and performance levels are provided electronically and in hard copy to systems and schools. Parents and students receive a printed individual student report. The GPS version of the CRCT sets the following criteria for scores: (a) 800, meets requirements; (b) below 800, does not meet requirements; and (c) 850 and above, exceeds requirements. Performance levels are identified as Level 1 (does not meet standards), Level 2 (meets standards), and Level 3 (exceeds standards).

Similarly, the EOCT uses the same performance levels, but the cut scores are different. Systems receive raw scores, scaled scores, and performance levels just as with the CRCT; however, the scaled scores are also converted to a grade percentage for ease in weighting them as the grade conversion is 15% of a student’s final grade. Students who score below 400 are categorized at Performance Level (PL) 1 and do not meet standards; scores from 400 to 449 meet standard and are categorized at PL 2; and a score of 450 or higher categorizes a student at PL 3, exceeding standards.

A noteworthy consideration is that the eighth grade is a benchmark year in CRCT test administration. Students must meet the score of 800 or higher in order to be promoted to ninth grade. Two retest opportunities are given during the summer for students who score at PL 1. If a student does not receive a passing score on retests, a committee that is comprised of the school’s principal, the students’ parents, and the students’ teachers convenes during the summer to determine placement for the next school year.
Like the CRCT, the EOCT is a state-mandated, standardized test. In contrast, the EOCT is administered upon completion of each of the eight subject areas tested and counts 15% of a student’s final course grade. As Georgia begins phasing out the GHSGT in the 2011-2012 school year, the EOCT grade weight will increase to 20%. The GaDOE provides electronic and print versions of EOCT score reports, and the student receives both a scaled score and a grade conversion score. The latter is on a 100-point scale and is 15% of the course grade, which must be 70 or above for a student to receive credit for the course.

Even though the GHSGT is the test that students must pass to meet graduation requirements through the 2010-2011 school year, the EOCT may be used as a criterion for receiving a variance for the GHSGT if a student fails to pass a subject area subtest of the assessment. For example, if students fail the English Language Arts section of the GHSGT, their graduation status is threatened unless they have passed the EOCT in either Ninth Grade Literature or their eleventh grade American Literature. With a passing score on one of the two EOCTs, students may exempt the GHSGT in ELA.

Both the criterion variable of CRCT data and the predictor variable of EOCT data used the scaled score as the primary measure. Scaled scores, which are common in large assessment programs such Georgia’s, are based on the number of items correct, or the raw score. The GaDOE developed several forms per year for each CRCT and EOCT subject area assessment, and the raw score is interpolated to a scaled score that relates equitable scores on multiple forms of the same test. This is a crucial step in maintaining the validity and reliability of the assessments (“Georgia EOCT,” 2011). The GaDOE’s Testing and Assessment Division provided validity and reliability data for both the CRCT and EOCT, and these reports also verified the similarity in test item construction and
assessments development. This study used archived, post-assessment data to help ensure objectivity.

Assumptions and Limitations

Assumptions

Several assumptions underlie this study. First, the population of students taking the CRCT and EOCT remain the same from their eighth grade year to their ninth grade year. For example, only students who took the Reading CRCT in eighth grade and also took the Ninth Grade Literature EOCT in ninth grade were included. Second, both the CRCT and EOCT are valid and reliable assessment instruments. The validity and reliability data for both tests were provided by the GaDOE Testing and Assessment Department. Third, because both the CRCT and EOCT were developed to assess student achievement of the GPS, it was assumed that test items were developed consistently for both tests. While this third assumption was more difficult to address, the GaDOE corroborated consistent test development in the validity reports.

This study utilized a regression analysis as one of the analyses. Several assumptions specifically underlie regression analysis. First, the sample is representative of the population. Second, variables are normally distributed. Third, there is a linear relationship between the independent and dependent variable, where the regression line’s best fit is a straight line. Fourth, variables are measured reliably and are error-free. Last, the data distributions have the same variance of errors (Osborne & Waters, 2002).

Additionally, there is the assumption that ethical considerations for this study have been addressed. I have a responsibility to respect participants and to acknowledge their contributions. While protecting system and school anonymity in the research process, I feel it appropriate to communicate appreciation to the system superintendent.
and middle and high school principals for their help in completing this study.

Even in a study where only post-assessment data were used, respect for participants’ anonymity had to be ensured. All data were de-identified by an independent statistician so that no individual student could be identified. Another ethical consideration was not to generalize findings to a population where they may not apply. The characteristics of the study population must be described accurately, and the study’s implications must be based on data analyses. Furthermore, the data must not be interpreted as applicable in dissimilar populations.

When data analyses are discussed for possible implications, a researcher must be vigilant in regard to ethical validation—a qualitative term but still applicable in the quantitative research data discussion. Creswell (2007) advised researchers to question their moral, political, and ethical assumptions and provide equitable treatment for all study groups. Further, researchers should provide practical answers to questions which, in quantitative inquiry, can be characterized by implications from the data analyses.

**Limitations**

There are limitations in correlational research studies. Even if data indicate a relationship between the achievement in eighth grade and in ninth grade, I cannot infer that eighth grade achievement causes ninth grade achievement. Correlation does not equal causation, and it is important to proceed cautiously when looking for predictive value. Even though the design cannot prove causation, it can be used “for prediction, to support a theory, to measure test-retest reliability, etc.” (Waters, 2010, para. 1).

According to Gall, Gall, and Borg (2007), “Correlational research can yield useful findings, but ultimately multiple lines of research and theory building are necessary to develop a full understanding” (p. 341).
The third-variable problem—unmeasured variables that are a potential cause of changes in student achievement—is also a limitation of correlational research ("Research," 2010). This problem was addressed two ways. First, there was more than one study group for each subject area because several years of test administration yielded multiple pairs of eighth grade-to-ninth grade study groups. Data analyses provided information for each subject area group as well as the total of all groups in a subject area. Second, the system and school administrators were consulted and asked to identify system-level and/or school-level factors that might explain outliers or other anomalies in the data from school to school or year to year.

**Research Plan**

A non-experimental correlational research design was determined to be most appropriate for this study and followed student groups over two years of test participation. This study attempts to control for internal and external threats to validity and the variables that affect performance. The GaDOE has mandated an EOCT in Ninth Grade Literature, American Literature, Math I, Math II, Physical Science, Biology, U.S. History, and Economics. Ninth graders were routinely enrolled into Ninth Grade Literature and are participants in this research study.

As Table 1.1 shows, a study group was a group of eighth grade CRCT test takers in Reading who also took the Ninth Grade Literature EOCT the following year and a group of eighth grade CRCT test takers in ELA who took the Ninth Grade Literature EOCT the following year. Data for the paired groups were provided for the years that GPS-based test data were available. The data connected eighth grade to ninth grade scores on the same group of students. The data also compared multiple years of testing in each subject area, ensuring more validity and reliability in the research design and results.
Table 1.1

*Research Student Groups*

Identified Student Groups Based on Paired GPS-based Testing

<table>
<thead>
<tr>
<th>CRCT-EOCT Years</th>
<th>8 Reading/9 Literature</th>
<th>8 ELA/9 Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2007-2008</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2008-2009</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2009-2010</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2010-2011</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

*Note.* An “x” indicates paired subject area tests (i.e., 8th ELA CRCT and 9th Literature EOCT, 8th Reading CRCT and 9th Literature EOCT).

System and school administrators were asked to identify and explain potential reasons for differences between whole group and subgroup performance. A form was sent to each to seek their input into system- or school-level factors that may have impacted students’ scores (see Appendix B). For example, one school may have used a computer-based reading intervention, another school may have focused on a whole-school or small-group math intervention, or another school may have implemented an attendance or behavior incentive that accounted for students being in class more often for instruction. Assessment data were scrutinized by school, and administrators provided insight as to why differences in test performance existed.

Test data for all eighth graders and ninth graders taking the GPS-based versions of the CRCT and the EOCT were in the possession of the researcher. However, all data were de-identified by a statistics consultant prior to data analysis. Students who had
continuous enrollment for both their eighth grade year and their ninth grade year and took both the subject-area subtest of the CRCT and the respective EOCT comprised a study group. Several paired groups in each subject area were identified for the multiple years of GPS-based CRCT and EOCT testing. Analyses were conducted for the system’s students by subject area for each year of paired testing. Furthermore, data were disaggregated by school, gender, ethnicity, and students with disabilities. Descriptive, univariate, and bivariate statistics were provided. Pearson’s product-moment coefficient analyses were conducted along with regression analyses. The standardized test scores were analyzed using Pearson’s product-moment correlation to determine correlation coefficients and the degree of correlation. Regression analyses were conducted to determine any predictive values in the criterion variables.

**Definition of Core Terms**

For clarity, it is important to provide definitions of terms that will be used throughout the study:

**CRCT**: Georgia’s Criterion-Referenced Competency Test is mandated for third, fifth, and eighth grade students in the areas of English Language Arts, Mathematics, Science, and Social Studies. It is used as the annual measure of Adequate Yearly Progress (AYP) for third grade through eighth grade in the researched school system.

**Criterion variable**: The criterion variable for this study was the ninth grade EOCT scores in Ninth Grade Literature.

**EOCT**: Georgia’s End-of-Course Tests are state-mandated in eight subject areas during students’ ninth through twelfth grade school years. They are given in Ninth Grade Literature, American Literature, Math I, Math II, Physical Science, Biology, U.S. History, and Economics. EOCTs are weighted as 15% of students’ grades but will rise to
20% beginning with ninth graders entering Georgia high schools in school year 2011-2012 when the EOCTs replace the GHSGT as the measure of Adequate Yearly Progress (AYP).

**GHSGT:** The Georgia High School Graduation Test is administered to all eleventh graders in the spring of their eleventh grade year. Students are tested in ELA, Mathematics, Science, and Social Studies. The GHSGT is the federally-approved, state-mandated measure of AYP but is being phased out in favor of the EOCT beginning in school year 2011-2012.

**GPS:** Georgia Performance Standards were developed to increase the rigor of Georgia’s curricula for K-12 instruction. Implementation began in the 2005 school year in a staged rollout that took several years.

**High-stakes testing:** A test is considered high-stakes when results are used for accountability purposes, positively or negatively, to make decisions that impact students, teachers, school administrators, school systems, and communities (i.e., student promotion/retention, whether or not a student graduates high school, and teacher or administrator tenure).

**Predictor variables:** The eighth grade CRCT score data for all study group participants in Reading and ELA were the predictor variables in this study.

**Study groups:** Study groups were identified as the eighth and ninth graders in the researched school system who took the eighth grade CRCT in Reading and ELA and the Ninth Grade Literature EOCT the following year. Several groups were identified because there are several paired years of GPS-based testing.

**Summative assessments:** Summative assessments are cumulative evaluations that measure student learning after instruction and are usually given at the end of a course.
Some common summative assessments are district benchmarks, final exams, or state-mandated assessments, and scores used as a measure of AYP.
CHAPTER 2: LITERATURE REVIEW

As long as there have been students, there have been assessments to measure their progress. Ravitch (2002) found the origin of student assessment in America can be traced back to the nineteenth century. Yet, the concept of teachers and schools being accountable for students’ achievement is novel. According to Lauer et al. (2005), standardized tests matter, and the research found that mandated, standards-based assessments influence teacher instruction in three ways. First, teachers aligned the scope and sequence of their content curricula to the tested content standards. Second, changes in teachers’ pedagogy were not likely to be influenced by testing programs. Last, standardized testing appeared to encourage teachers’ assessment practices in the classroom.

Since the No Child Left Behind Act (NCLB) was signed into law in 2002, states and their respective public school systems have been accountable to the federal government for demonstrating student progress, which has led to state-mandated, high-stakes testing in most of the United States (Lauer et al., 2005). The pertinent aspects of standardized testing are student testing data, student achievement and learning, and teacher curricula and instructional strategies. The review of literature for this study will focus on the following aspects of standardized testing: (a) its history, (b) teacher accountability, (c) curricular and instructional autonomy, and (d) school reform efforts based on standardized assessment data.

Theoretical Framework

According to Littlejohn (2007), “It is critical for the inquirer to state his or her paradigm(s) when undertaking—and later publishing and sharing—the research findings,
for paradigms inform how the inquirer approaches and frames the research question and proceeds in answering it” (p. 3). The theoretical framework for this study is a Christian theocentric humanist paradigm. The humanist paradigm can be characterized by the belief that learning is a “personal act to fulfill one’s potential” (“Humanism,” 2010, para. 1). Despite the belief that humanism is a godless philosophy, there is a branch, albeit small, called Christian humanism, which is the philosophical foundation for this research study. Clouse (2011) explained Christian humanism as follows:

The Christian humanist values culture but confesses that man is fully developed only as he comes into a right relationship with Christ. When this happens, a person can begin to experience growth in all areas of life as the new creation of revelation. (para. 13)

Robbins (2008), a humanistic psychologist and Editor-in-Chief of Janus Head: Journal of Interdisciplinary Studies in Literature, Continental Philosophy, Phenomenological Psychology, and the Arts, hosts a webpage (http://www.mythosandlogos.com/) with a philosophical focus on humanistic psychology and phenomenological research.

According to Robbins, humanism developed into a positivist framework, and both underpinned a post-positivist framework. The current research study is based on a post-positivist framework and will rely on Rescher’s coherence theory of critical reasoning.

Post-Positivism

Post-positivism “posits a reality that is ‘out there’ to be discovered, but, in contrast, the reality can only be known imperfectly and within probability” (Littlejohn, 2007, p. 5). Littlejohn further stressed that “post-positivist inquirers assume a detached, separated position from the object of their study and such objectivity is important” (p. 11). Post-positivism is necessarily permeated with value considerations. Psychologists
generally agree that humanism informed positivism and that both informed post-positivism (Robbins, 2008). Yet, post-positivism extends humanism and positivism and allows for understanding the significance of values along with the reality of the data. This is the blended, added benefit to post-positivist research. Robbins (2008) stated:

Those who perform research in the realm of positive psychology should, in turn, pay close attention to some of the lessons of history offered by humanistic psychology. First, positive psychology cannot be a value neutral endeavor, and it must take pains to examine its implicit values to make them as explicit as possible. . . . Second, virtues cannot be studied in isolation, but must be approached holistically, or else those virtues risk falling into vice. And, finally, positive psychology will never live up to its promise of articulating the good life until it pays due respect to the central virtue of phronesis, or wisdom. (p. 107)

Coherence Theory of Critical Reasoning

Rescher’s (2001) coherence theory of critical reasoning is appropriate to support quantitative correlational inquiry. Background for Rescher’s theory reaches back to Dewey who, in his 1929 book, The Quest for Certainty, stated: “If we can judge events for indications of other events, we can prepare in all cases for the coming of what is anticipated” (as cited in Demetrion, 2004, “Abstract,” para. 3). Demetrion concluded that, by using Dewey’s model, a researcher can prevent something from happening or intentionally lead to one event over another if reasoned judgment supports the choice to be connected with that which we are after. Dewey believed that one’s observed facts can lead to an idea that can become a possible solution (Demetrion, 2004).

In summary, Rescher’s (2001) post-positivist coherence theory of critical reasoning can be characterized as “a spider’s web in which each item of knowledge is a
node linked to others by thin strands of evidential connection, each weak, but all together, collectively adequate to create a strong structure” (as cited in Demetrion, 2004, “A Network Model,” para. 3). Alcoff (1996), a modern coherence of critical reasoning theorist, also acknowledged the importance of Rescher’s early coherence theory. This study’s correlational research design seeks to establish the coherentist links within the “spider web” of study data.

**Christian Theocentrism**

The overarching worldview for this study is Christianity, which is rarely associated with humanism. Even so, Huitt (2009) referred to a pre-positivist humanism, where theocentric values are accepted in the paradigm. He stated that a small yet important group within the humanist paradigm believes “while humanity is a distinct species, existing separate and apart from all animal species, God or a Supreme Being is the center of humankind’s existence” (para. 4). When God lives within the hearts of His people, He guides them to their best nature, created in His image. God assures His presence to His followers in scripture:

> For this reason I bow my knees before the Father, from whom every family in heaven and on earth derives its name, that He would grant you, according to the riches of His glory, to be strengthened with power through His Spirit in the inner man, so that Christ may dwell in your hearts through faith; and that you, being rooted and grounded in love, may be able to comprehend with all the saints what is the breadth and length and height and depth, and to know the love of Christ which surpasses knowledge, that you may be filled up to all the fullness of God.  
> Eph 3:14-19 (NASB)

For this study, theocentricity, having God as one’s focus, is distinguished as Christianity,
which is the researcher’s faith.

**Synthesis of Worldview**

In any scholarly endeavor, the researcher’s paradigm, or worldview, and theoretical framework must be transparent throughout the work (Creswell, 2007). The theoretical framework for this review of literature is post-positivism, relying on the coherentist critical reasoning theory and embedded within a Christian theocentric humanist paradigm.

**History of Standardized Testing**

Testing in education is not a new idea. As Dawson (2010) said, “As long as there have been teachers, there has been evaluation” (para. 3). Throughout the course of history, societies have instituted testing for military and government positions. For example, tests for military selection date back to 2000 B.C., and the Chinese used test results for civil service qualification in 200 B.C. (Cizek, 2005). The Dead Sea Scrolls recorded tests that were used in Qumran to determine when a man was qualified to become a formal member of the community (Madaus & Russell, 2010/2011). In scripture, the guards from the tribe of Gilead defeated fugitives from the tribe of Ephraim by using a high-stakes test:

> Then Jephthah gathered all the men of Gilead and fought Ephraim; and the men of Gilead defeated Ephraim, because they said, "You are fugitives of Ephraim, O Gileadites, in the midst of Ephraim and in the midst of Manasseh." The Gileadites captured the fords of the Jordan opposite Ephraim. And it happened when any of the fugitives of Ephraim said, "Let me cross over," the men of Gilead would say to him, "Are you an Ephraimite?" If he said, "No," then they would say to him, "Say now, 'Shibboleth.'" But he said, "Sibboleth," for he could not
pronounce it correctly. Then they seized him and slew him at the fords of the Jordan. Thus there fell at that time 42,000 of Ephraim. (Joshua 12:4-7, NASB)

Moon (2009) stated that high-stakes educational testing in America originated in Boston in the late 1840s as policymakers were determining schools’ effectiveness by comparing classrooms and schools. Whether perceived as negative or positive, testing today is an integral part of education in the United States (Madaus & Russell, 2010/2011; Moon, 2009; Phelps, 2005a; Skerrett, 2009; Supovitz, 2009; William, 2010).

**Early Development of Standardized Testing**

Historically, according to Madaus and Russell (2010/2011), the earliest formal assessments were oral, and that oral format continues today in many schools. Testing has evolved into a predominantly paper format and over the past twenty years is increasingly computer-based. In addition to different testing formats, Madaus and Russell found that the requirements of student content mastery shifted from performance measures to standardized exams. In addition, advancements in technology transitioned testing from subjective qualitative assessment to an objective quantitative assessment with a single answer.

In 1900, 60 years after Boston’s policymakers enacted high-stakes testing, 12 university presidents in the northeastern United States established the College Entrance Examination Board to oversee administration of college admissions tests (Lemann, 1995). According to Lemann, the original purpose was to standardize the secondary boarding schools’ curricula so that their students would be well-prepared for college. These admissions tests were essay tests rather than multiple-choice tests and required a long time to score. Lemann reported that the first intelligence tests emerged in the first decade of the twentieth century, and there were several early test creators who changed
the face of standardized testing to much of what is seen today.

In addition to the policymakers’ testing initiative in Boston, the first intelligence tests emerged in the early 1900s. Binet, Mann, Brigham, Terman, and Chauncey developed the earliest intelligence tests. In 1905, Binet collaborated with Simon to publish the *Binet-Simon Test of Intelligence* with an early interest to identify “feeble-minded children who could not profit from the ordinary program of school instruction” (Lennon, 1955, p. 34). Lennon stated that Binet’s work was an early precursor of standardized testing with the intent of classifying students and identifying them for guidance purposes in education. Mann capitalized on the emphasis of quantifying student achievement. In particular, Mann was an advocate of common (public) schools as he sought to replace traditional oral exams with essay-format common exams that could provide objective data on Boston students (Gallagher, 2003). Mann’s approach “effectively introduced the notion of a standardized test” (Madaus & Russell, 2010/2011, p. 24).

Following Mann, other researchers created more standardized assessments and more efficient means of administering them. Brigham developed the first Scholastic Aptitude Test (SAT), modeled on the U.S. Army’s Alpha test, and administered it to over 8,000 high school students in 1926 (Lemann, 1995). The Army used the multiple-choice, paper-and-pencil Army Alpha test developed by the American Psychological Association (APA) initially to identify officer candidates (Solley, 2007). In 1919, Terman transformed the Army Alpha test into the National Intelligence Test for students, selling over 400,000 in its first year. Terman also revised and expanded early intelligence-quotient testing by collaborating with Binet in the development of the *Stanford-Binet Test of Intelligence* in 1916 (Gallagher, 2003).
Throughout the early intelligence test development, the response format varied, but in 1914, multiple-choice formats were invented (Gallager, 2003). As the multiple-choice format became more prevalent, scanners were developed in the 1930s to expedite results from the multiple-choice tests. Optical, high-speed scanners in the 1950s further increased the efficiency of scoring (Lemann, 1995; Clarke, Madaus, Horn, & Ramos, 2000; Madaus & Russell, 2010/2011). Clarke et al. reported that by the 1930s, annual sales of tests from just the Otis/Terman intelligence test and the Stanford Achievement Tests were over $2.25 million, attesting to the growing influence of the testing market. Clark et al. also found that by 2000, most of the U.S. testing market was monopolized by 13 companies, which had revenues of over $15 million each.

Chauncey, who collaborated with or learned from the other pioneers in the testing field, implemented the earliest mass administrations of the SAT in the U.S. According to Lemann (1995), Chauncey is also credited with integrating machine scoring with mass test administration. Lemann further stated that Chauncey orchestrated a multi-site, simultaneous administration of the SAT in 1943 to 325,000 military recruits to determine induction or college deferment. Also, public education was becoming firmly entrenched in the 1940s, and for the first time, a majority of U.S. young people were graduating from high school. Chauncey’s first mass administration of the SAT paved the way for the proliferation of testing throughout the U.S.

While Lemann (1995) found that standardized test numbers were expanding and student enrollments were increasing in public schools, testing technology was impacting all aspects of standardized testing. Computer technology developed quickly and is now used in every aspect of current standardized testing, from test item development to test development to test administration to test scoring and reporting (Clarke et al, 2000).
Standardized testing demonstrated validity and reliability, and educators began mass administering them to American high school students. In 1953, Ludlow stated, “Testing today is truly a ‘big business’” (p. 279). Ludlow’s belief that testing is big business is ironic, especially when it is viewed in the context of the current educational environment. According to Clarke et al., elementary and secondary students in the 1990s took close to 400 million tests a year in the United States, and most states mandated multiple-choice tests, sometimes with a short-answer component.

The standardized testing trend continues with the advent of computer-based tests, which have made access to the tests, test administration, and availability of results much more efficient (Madaus & Russell, 2010/2011). The Educational Testing Service (ETS) administers the **Graduate Record Examination** (GRE) and the **SAT I: Reasoning Test** electronically; the GRE is only administered on computer (Clarke et al., 2000). In Georgia, both the CRCT in elementary and middle grades and the high school EOCT can be taken as a computer-based test.

**Social Policy and Standardized Testing**

In 1965, during the Johnson administration, the **Elementary and Secondary Education Act** (ESEA) was enacted, significantly impacting the standardized testing movement in the U.S. The ESEA brought federal funding into state education for the first time, and states were required to use standardized testing to document the success of programs for which federal funds were used (Solley, 2007). According to Solley, the required documentation process evolved into the notion that standardized testing could evaluate student learning.

In 1983, during the Regan administration, the National Center on Excellence in Education published *A Nation at Risk*, beginning the mandated testing debate and
politicizing the high-stakes testing movement (Nichols, Glass, & Berliner, 2005; Solley, 2007). *A Nation at Risk*, while statistically flawed, criticized public education and nationalized an accountability movement in the United States ("National Commission," 1983). The Commission’s publication was quickly followed by the creation of over 50 state education commissions, increased graduation requirements, and state education reforms that increased states’ standardized testing programs and students’ course requirements (Au, 2009).

In the latter part of the 1980s, the standardized testing debate continued. Acutely aware of the issue’s importance, Republican George H. W. Bush’s 1988 presidential campaign emphasized the importance of education as he supported minimum competency testing to determine graduation or grade retention. Solley (2007) stated that President Bush developed his *America 2000*, with a focus on testing and standards. Following the Bush administration, President Clinton and Vice President Gore continued to focus on *America 2000*’s standards and goals as they called for a national assessment system and a high-stakes test for high school graduation.

As the 21st century commenced, the educational reforms continued. The most notable reform was the *No Child Left Behind Act* (NCLB), which was signed into law by President George W. Bush in 2002. President Bush’s goal was to tie Title I funding from the federal *Elementary and Secondary Education Act* (ESEA) to students’ standardized test achievement (Solley, 2007). Reflecting a focus on high-stakes testing, the ESEA was reauthorized and renamed the *No Child Left Behind Act*. According to Solley, a major premise of NCLB was that “increased pressure to do well on standardized tests, along with a set of rewards and punishments, will increase student learning and achievement” (p. 33). NCLB mandated that by 2006 all students be tested
in Reading and Math in grades 3-8 and once in high school. Then, by 2008, testing in Science would be added and required for all students once in elementary school, once in middle school, and once in high school.

In addition to NCLB’s mandated student assessments, NCLB required that all schools and school systems be evaluated annually for consistent improvement, a measure termed Adequate Yearly Progress (AYP). Results are disaggregated by subgroups of ethnicity, economically-disadvantaged status, and disabilities. Annual measurable objectives (AMOs) are established to set percentage benchmarks for the level of proficiency toward standards. As currently authorized, all schools and systems must meet an AMO of 100% by 2014 (“No Child,” 2001).

NCLB remains the law of the land. It was to be reauthorized by 2008, yet no reauthorization occurred. Until any Congressional reauthorization, NCLB’s current form is law (Klein & McNeil, 2010). President Barack Obama, elected in 2008, has called for educational reform and a rewrite of NCLB in order to give states more autonomy. According to Klein and McNeil, the Obama administration has proposed an outline for NCLB reauthorization that would continue the use of state-mandated student assessment. States would be granted flexibility in calculating a school’s effectiveness in student achievement. In addition to school’s being accountable for student testing, the proposal calls for other accountability indicators, including college and career readiness indicators of course completion, attendance, and school climate (Dee & Jacobs, 2010; Klein & McNeil, 2010). In March, 2011, Obama asked Congress to send him the education bill prior to the start of the 2011-2012 school year (Klatell, 2011).

The Obama administration is advocating for national common standards and national assessments. Despite a climate in 2008 that did not bode well for national
standards, a year later the National Governors Association announced that 42 states agreed on the Common Core Curriculum Standards (CCCS) Initiative (Goldstein, 2009). U.S. Department of Education Secretary, Arne Duncan, has come out in support of the CCCS, which were developed under the auspices of the Council of Chief State School Officers and the National Governors Association. Currently, the CCCS have been adopted for implementation in 49 states and territories (“Common Core,” 2010). Additionally, Obama’s Race to the Top initiative and education reform policy continues a non-partisan focus on high-stakes testing that now spans eight presidential terms.

Testing Issues in the Accountability Era

Standardized testing was touted as a means for predicting a student’s ability to learn and was also used early in the twentieth century to identify college aptitude. American education’s proclivity for ranking and sorting students made it necessary to standardize test administration and interpretation as testing became the measure of ensuring equity (Gallagher, 2003). Ongoing reliability and validity data continued to support the professional and accurate nature of standardized testing. Since the 1970s, unfortunately, test data have been based on comparing cross-sectional data (Kelly & Monczunski, 2007). To illustrate a cross-sectional data comparison, a study would compare the data from one year’s group of fourth graders with the prior year’s fourth graders instead of comparing students’ scores in the fourth grade to their subsequent scores in the fifth grade. Kelly and Monczunski believed that cross-sectional data led to instability in interpreting results.

Test equity was another issue that received scrutiny. According to Gallagher (2003), test critics have made accusations of inequity since the mid-1960s as the advent of the civil rights movement led to a greater awareness of testing inequity. Gallagher
further discussed the Coleman Report of 1966, which validated equity in standardized
testing. While the Report’s claims were found to be in error, it remained an influential
and referenced study for many years. Gallagher found that allegations of test bias and the
failure of tests to account for cultural differences continue into the 21\textsuperscript{st} century.

In the current NCLB era, business and government entities have consistently and
staunchly endorsed test-based accountability as a measure of educator performance.
Standardized test scores have become the bottom line as they are regarded as concrete
and reliable measures of students’ minimum competencies (Gallagher, 2003). Over 80% of
Americans favored test-based accountability and had supported it throughout the years
of NCLB (Hart & Teeter, 2004). NCLB assumed that state-mandated tests provide useful
information to school administrators and teachers.

Yeh (2006) tested this NCLB assumption in a qualitative study investigating
whether or not teachers and administrators utilize test data. Yeh found that high-stakes
tests failed to provide diagnostic information or information about individual student
growth that impacted instructional decisions. Yeh further stated that the administrators
and teachers found little use for the tests in improving instruction or prescribing learning
strategies based on the needs of the individual students. Yeh concluded that the NCLB
assumption was flawed and that NCLB did not manifest the improved student
achievement that it was designed to do.

Studies such as Yeh (2006) revealed there is no ambiguity with regard to high-
stakes testing. It is defended or hated, touted or detested. Love it or hate it, high-stakes
testing has a central role in American schools. According to Cizek (2005), testing critics
assert that tests are responsible for (a) increasing teacher stress, frustration, and burnout;
(b) increasing drop-out rates in high schools; (c) increasing students’ stress and stress-
related illnesses; (d) narrowing the curriculum; (e) not measuring higher-order thinking skills; (f) expanding the achievement gap; (g) demonstrating testing bias; and (h) promoting cheating. Proponents of high-stakes testing assert that any criticisms of testing are not supported in research and are merely the opinions of policymakers, commentators, critics, parents, and even some educators (Cizek, 2005; Geisinger, 2005; Goodman & Hambleton, 2005; Phelps, 2005a; Phelps, 2006b; Sireci, 2005).

Phelps (2005a) summarized 40 years of research that investigated the public’s opinion of high-stakes testing. Phelps incorporated the results of 245 surveys and 67 research studies administered and conducted between 1965 and 2002. Survey respondents included teachers, administrators, board members, professors, politicians, employers, and students. Phelps found that survey items did not address testing until the late 1970s, which coincided with the beginning of minimum competency testing and high school graduation test requirements. The next increase in standardized testing-related survey items occurred in 1984, in concert with the publication of A Nation at Risk. The highest percentage of testing-related survey items occurred in 2000 and 2001, and Phelps stated that the NCLB legislation contributed to the increase.

Throughout Phelps’s (2005a) research, a recurrent theme was the public’s support for high-stakes testing. Despite negative media coverage, Phelps found the public remained undeterred in their support. Between 1965 and 2002, 69% of the public favored administering standardized tests at least once a year, and a group of the respondents wanted additional testing. The author stated that teachers were the only group of respondents who did not support additional testing. Phelps attributed the decline to an accountability model that was student-centered in the 1970s and 1980s but became teacher-centered from 1999 to the present.
Conversely, Perrone (1976, 1991) and Solley (2007) authored position papers in opposition to standardized testing for the Association for Childhood Education International (ACEI). In the 1976 position paper, Perrone called for a moratorium on standardized testing, especially in primary grades. Perrone denounced the practice of utilizing test results to determine school entry, promotion and retention, and program placement. Fifteen years later, Perrone restated ACEI’s position, calling again for a moratorium against standardized testing. Perrone believed that standardized testing caused teachers and students undue stress, especially minority students and students from lower socioeconomic backgrounds.

Similar to Perrone’s 1991 conclusions, Solley (2007) agreed with the need for a testing moratorium in early grades. Solley believed that results from high-stakes testing are used punitively and discriminate against students, teachers, and schools when results should be utilized instead to diagnose students’ learning needs and influence instructional decisions. While ACEI recognizes that assessments are needed to improve instruction and learning, Solley disagreed with those who endorse high-stakes testing. Rather, Solley found that standardized testing (a) does not improve learning; (b) decreases student motivation to learn; (c) narrows the curriculum; (d) limits instruction to rote memorization; (e) bases promotion, retention, and placement decisions; and (f) forces teachers to teach to the test. Luke and Woods (2008) reiterated many of Solley’s concerns, adding the issues of test score manipulation and the utilization of packaged intervention programs that do not have a sound research base.

A 2005 study by Nichols, Glass, and Berliner provided further evidence that high-stakes testing in schools is ineffective. The study found a significant body of evidence that identified the negative, perhaps unintended, effects of high-stakes testing along with
“no convincing evidence that the pressure associated with high-stakes testing leads to increased achievement” (p. 109). Lee (2006) also found no significant gains in student achievement on the National Assessment of Educational Progress (NAEP) since NCLB was signed into law.

Yet, comparisons of year-to-year state-mandated tests revealed that test scores do increase, but the results are not always generalizable to other standardized assessments such as NAEP. Further, scores from high-stakes tests may not be appropriate measures of progress (Fuller, Wright, Gesicki, & Kang, 2007; Mintrop & Sunderman, 2009; Scott, 2008). Nichols, Glass, and Berliner (2005) also cited unprofessional treatment of educators, the distortion of accountability indicators, the inconsistent evidence that intended testing effects happen, and the impossibility of achieving AMOs of 100% in 2014. Like Solley (2007), Nichols, Glass, and Berliner called for a moratorium on high-stakes testing.

Volante and Ben Jaafar (2010) summarized a study that found additional consequences of standardized testing. On a positive note, they found that testing motivates students to study and increases student achievement. The greatest gains were in districts with aggressive assessment policies where tests included structured response items along with multiple-choice items. The districts experiencing more success also invested in after-school programs. Volante and Ben Jaafar noted that teachers had higher expectations of students with disabilities post-NCLB, made positive changes to assessment and instruction practices, and demonstrated higher levels of participation more often in content-related professional learning.

As well, Volante and Ben Jaafar (2010) discussed the negative consequences of standardized testing. They concluded that there was little evidence that testing closed or
narrowed the achievement gap between white and minority students. In addition, the authors found that testing led to decreased graduation rates, increased grade retention, and greater stress (particularly for low-performing students). Volante and Ben Jaafar concluded that teachers’ perceptions of standardized testing were (a) high stress and low morale, (b) a tendency to teach to the test, (c) a more rigid instructional focus on test preparation and rote learning, and (d) cheating due to accountability measures. Additionally, Volante and Ben Jaafar found that good, highly-qualified teachers transferred out of low-performing schools that needed them the most.

**Accountability**

Following the publication of *A Nation at Risk*, standardized testing became a greater, more high-stakes force in education. Accountability for results and student learning became a focus for policymakers, politicians, states, and local school systems. Obviously, high-stakes tests matter, especially after the NCLB legislation (Au, 2007; Chiang, 2009; Dee & Jacobs, 2010; Moon, Jarvis, Brighton, & Hall, 2007; Supovitz, 2009). Because NCLB was not reauthorized in 2008, it remains in effect until Congress acts on reauthorization. The NCLB legislation provides a stated purpose to hold schools, local school systems, and states accountable for improving all students’ academic achievement (“No Child,” 2001).

**Federal Oversight**

NCLB holds states, systems, and schools accountable by setting AMOs in both Reading/English Language Arts and Mathematics. The AMO percentage goal for students is pre-determined and will rise incrementally until 2014, when 100% of students are expected to pass all assessments. Daly (2009) stated three major assumptions that underpin NCLB. First, in a short period of time and with no changes to funding, all
subgroups in schools can demonstrate academic success. Second, research-based methods of reform can accomplish this achievement. Last, formal testing programs and sanctions are strong incentives to bring about achievement. Schoen and Fusarelli (2008) found that the impetus for NCLB was to increase a school’s accountability to the public.

In light of the fact that schools are accountable to the public, schools that fail to meet a required AMO for two years will receive sanctions. A school’s AYP score will determine sanctions. A school will receive a warning the first year that it does not meet its AYP benchmark. When a school fails to make AYP for two consecutive years, the state and school system must provide assistance and interventions to the school to improve students’ achievement. This includes school choice and transportation if a student is approved for transfer to a school meeting the AYP benchmark (“No Child,” 2001). Stullich, Eisner, McCrary, and Roney (2006) noted that in 2004-2005 there were 9,000 Title I schools identified as needing improvement, a 50% increase from the prior year. Daly (2009) advised that researchers predict that a majority of U.S. Title I schools could be labeled needing improvement by 2014 due to escalating and increasingly demanding AMOs.

According to the NCLB legislation, when a school fails to make AYP a third consecutive year, students who are economically-disadvantaged get an option for supplemental educational services provided by state-approved instructors outside of the school’s instructional day. Schools that fail to meet AYP for a fourth consecutive year will receive corrective action, which can mean replacing curriculum, replacing or reorganizing staff, increasing the length of the school day, or bringing in consultants. Finally, a fifth consecutive year of not meeting AYP benchmarks could result in the state or a private company taking over the school, or the school could become a charter school.
Mintrop and Sunderman (2009) stated that accountability benchmarks such as AYP are punitive in nature. Because the focus of NCLB is Title I schools, Mintrop and Sunderman asserted that punishing low-performing schools is counterintuitive as the NCLB legislation was designed to assist these schools. Miller, Kerr, and Ritter (2008) conducted a study that evaluated high-stakes testing as a student performance measurement for the impact on equity. The authors concluded that NCLB, which was initially designed to help minority students, disproportionately punished minority schools.

It is important to note that NCLB also provided guidelines for rewarding schools that make AYP, although these are fewer. For example, schools can receive recognition or financial rewards. States must develop Academic Achievement awards, which recognize schools when they make AYP for two consecutive years or when they close achievement gaps, and Distinguished Schools awards that recognize schools that demonstrate the greatest gains in student achievement. States must also provide financial awards for teachers in distinguished schools (‘No Child,’” 2001).

**Governmentality in Testing**

Graham and Neu (2004) relied on Foucault’s work on governmentality and his thoughts on the nature of examinations in order to construct a genealogy and chronology of standardized testing in Alberta, Canada. According to Sauer-Thompson (2004), governmentality is exploring the problem of the state and of politics (i.e., political ideologies, rationales, and the techniques of domination) while also exploring the government of the individual subject from both ethical and sociological perspectives.
Graham and Neu discussed (a) how standardized testing became prevalent, (b) how the tests “function indirectly to achieve the goals of government,” and (c) how the assessments help to “construct governable persons” (p. 297). Graham and Neu suggested that testing appeases politics and policy at the expense of teachers and other related educational personnel.

Madaus and Russell (2010/2011) explained the paradox of testing. High-stakes testing is a means of government control while it also is means of building a quality education system and ensuring accountability. Other studies investigated external factors in standardized testing, and there are many studies that have sought to understand teachers’ perceptions of standardized testing (Baker & Johnston, 2010; Guskey, 2007; Johnson, Yarrow, Rochkind, & Ott, 2009; Mulvenon, Stegman, & Ritter, 2005; Wiliam, 2010). However, few studies look at how test data can be applied to classroom instruction. While school system and school administrators are also accountable for student performances, it is teachers who seem most scrutinized because of their direct interaction with students.

Still, accountability has its benefits. Harris and Herrington (2006) looked at the rise of accountability over a 50-year span that culminated with NCLB. They found that regardless of who is in control politically or their decade of dominance, policies that have increased capacity, resources, exposure to rigorous content, teacher quality, and teacher retention have increased student achievement and reduced achievement gaps among student groups. Sims (2008), like Harris and Herrington, found accountability programs raise test scores. According to Springer (2008), the greatest gains in student performance have been in failing schools, where punitive measures resulted in greater productivity and student achievement among lower-performing students.
From a governmental perspective, Schoen and Fusarelli (2008) stated that NCLB reflects America’s core values of equity and accountability, and it is law until its reauthorization. They further believed that educators need support in developing creative instructional strategies as opposed to teaching under the duress of NCLB sanctions. The authors concluded that NCLB can engender innovation if it is reauthorized in a manner that supports quality instructional practices.

**State Accountability Requirements and Impact**

Regardless of one’s positive or negative perception of testing, high-stakes testing is a fact in the American educational landscape and impacts state, school system, school, and teacher accountability (Harris & Harrington, 2006; Mintrop & Sunderman, 2009; Supovitz, 2009; Wiliam, 2010). All states in the United States accept Title I funding and are required to establish AYP goals for every district and school in their state. Furthermore, states must develop content and achievement standards, assessing at least 95% of their eligible students annually in grades 3-8 and once in grades 10-12. These data reports must be communicated annually to parents and communities in order to provide information on the school’s progress (“No Child,” 2001).

Additionally, provisions of NCLB require states to hold schools and school systems accountable for making AYP. The state must provide the United States Department of Education with schools’ and school systems’ AYP performance, and the Secretary of Education must review the information. Schools and school systems that do not make AYP are subject to legislative sanctions from the state’s department of education. Interestingly, each school system and its associated schools are sanctioned, but the state is not sanctioned. After setting standards, administering assessments, and determining the pass score for each assessment, the state is only required to collect data,
report it to the federal government, and publish assessment results information in the form of a state report card (“NCLB Action,” 2011).

The AYP measure, the federal bottom line yes/no indicator of success, has become the canon for determining a school’s success (Chiang, 2007; Daly, 2009; Jennings, Noblit, Brayboy, & Cozart, 2007; Kelly & Monczunski, 2007). A school’s AYP status is headlines in local newspapers, and a system’s schools are compared favorably or unfavorably based on its results on the high-stakes tests. AYP data are available online through states’ departments of education. Oregon, for example, hosted a FAQ page about AYP results. One question asked how someone moving to Oregon can find information to help select a school. Oregon’s response was a link on their website that used high-stakes test results as the criteria for identifying successful schools. The link on the website allowed a potential resident to review AYP data and associated state report cards, relying on state-mandated standardized test results to demonstrate the success of a school (“ODE,” 2011).

Apparently, AYP results are impacting the real estate market. Dougherty et al. (2007) conducted historical and qualitative research that found “suburban homebuyers’ awareness of public school test scores has become more influential in the private real estate market” (p. 2). While Dougherty et al. acknowledged that over one-third of homebuyers researched a school’s test scores, they found that word-of-mouth, including social networking, communicated a school’s quality to prospective homebuyers. They also noted a real estate agent’s role in providing information about school quality.

To illustrate the importance real estate agents place on quality schools, Internet searches of real estate agents’ home pages provide local school system information (“Manns Choice,” 2011; “Mullinax Team,” 2011). The Mullinax Team home page
actually links to the GreatSchools website, a school comparison site with individual schools’ standardized test results. The site provides a search engine to locate and/or compare schools, and an overall rating of a school between 1 (low) and 10 (high) is assigned. Their school ratings “provide an overview of a school’s test performance by comparing the school’s state standardized test results to those of other schools in the state” (“GreatSchools,” 2011, para. 5).

**School System and School Accountability**

Understanding the role of standardized testing on accountability as it relates to schools and school systems is necessary. Accountability issues affect curriculum decisions, instructional strategies, school resource allocation, and formal evaluations of administrators and teachers. Vanderhaar, Muñoz, and Rodosky (2006) studied school system and school leadership, and their study confirmed prior research that students’ socioeconomic status (SES), students’ prior achievement, and teacher experience were the strongest predictors of student achievement. Vanderhaar, Muñoz, and Rodosky believed that their finding called accountability based on high-stakes testing into question.

In regard to how accountability affects curricula and pedagogy, Au (2007) provided a qualitative metasynthesis of 49 studies with an inquiry focus on how high-stakes testing affects teachers’ content and instruction. Au identified the primary effects of high-stakes testing: (a) Curricular content is narrowed to just what is tested, (b) content knowledge is fragmented into knowledge that is tested, and (c) instruction becomes more teacher-centered. Au further stated that this was not true in a significant minority of the 49 cases and believed that future research should explore the difference in that minority of teachers who expanded and enriched student-centered learning.
Horn (2009) conducted a qualitative study of teachers’ perceptions about instructional changes that occurred as a result of the Texas Assessment of Knowledge and Skills (TAKS). Horn identified and discussed the variables that affect students’ TAKS scores while opining as to whether standardized testing could determine teacher quality. Paris and Urdan (2000) discussed the impact of high-stakes testing on teachers, administrators, and parents, where each group surveyed believed too much emphasis is placed on test scores. The authors developed a list of assessment reform practices that could improve education: (a) Reduce the amount of high-stakes tests; (b) interpret scores clearly; (c) prohibit decisions based on a single score; (d) use high-stakes testing as incentives and not as punishments; and (e) create non-political committees to oversee educational assessments.

In a personal narrative, Wasserman (2001) discussed the use of standardized testing across a career as well as society’s reliance on numbers and data. He felt that educators trust a subjective analysis of data that is then used in subjective evaluations of teachers. According to Wasserman, numbers only have a meaning given to them rather than value in and of themselves. He concluded that test data only determine the mastery of facts and cannot evaluate real student learning and competencies.

As a result of NCLB, it is difficult to evaluate teachers without considering their students’ performance that is based on numerical score data. Gabriel (2010) cited the pressure of the teacher evaluation process due to NCLB’s emphasis on student achievement. As school improvement requirements increase and more schools face sanctions, administrators and teachers could lose their jobs. Gabriel reported that several states are using test scores as the criteria for determining teacher tenure, and “many school districts already link teachers’ bonuses to student improvement on state tests”
Similarly, Valli and Croninger (2007) argued against test accountability as part of teacher evaluation in a study of 18 schools, 69 teachers, and over 1,500 students. They examined one premise of accountability that gains in student achievement in one year can be attributed to individual teachers. The authors found that multiple factors have a role in student learning and that it cannot be ascribed to one teacher. Valli and Croninger cautioned against using standardized testing as a major component of teacher accountability as their study suggested that data-based decision making and teacher collaboration create an environment of collective rather than individual teaching.

Unfortunately, Scot, Callahan, and Urquhart (2009) found that the high-stakes testing environment is impervious to collaborative instruction according to their study involving 500 teachers who participated in an online professional development project. The authors found that standardized testing restricted curricular and instructional practices as it focused on rote learning rather than critical thinking skills. Curriculum standards were mandated and pacing guides were limiting. The researchers reported themes of teacher disempowerment, teaching to the test, and teaching to the lowest denominator. According to Scot et al., many of the teachers reported that their administrators used “bullying, threats, and coercive tactics to influence the achievement of higher test scores” (p. 50). Scot et al. noted further that forced curriculum guides and strict evaluation processes further reinforced teachers’ perceptions of coercion.

Webb (2005) found additional evidence of teacher dissatisfaction as a result of standardized assessment data. The author cited a case study of an elementary school where surveillance and accountability practices at the district level and within the school threatened teachers with punitive consequences for poor student test performance. Webb
stated that schools should develop teachers’ learning skills and that teachers’ trust was broken when their interests are ignored. Webb opined that teachers, who had an acute knowledge of the students and who were held accountable, were subjected to ill-advised policies.

Additionally, Webb (2005) found that the educational bureaucracy was negatively impacting teachers. O’Day (2002) agreed and posited a rationale that “schools are nested within larger systems and environments” and that external forces “seek to influence from the outside what goes on inside schools” (p. 3). Jennings, Noblit, Brayboy, and Cozart (2007) studied schools and their larger bureaucracies of school districts and state departments of education. According to Jennings et al., school systems in the past directed the schools, but now the state has usurped school systems in school accountability issues. The authors found that federal accountability policies and the resulting high-stakes testing implemented by the state had the effect of focusing the state’s attention on individual schools.

A study by Sims (2008) illustrated the findings of Jennings et al. Beginning in the 1990s, when state-mandated high-stakes testing ballooned in importance, school districts were allowed to move their start dates from September to August. In particular, Sims found low-performing schools believed an early start to the school year would give students additional test preparation time. Yet, parents, farmers, and tourism representatives in Wisconsin protested the shift in start dates. Sims found their protests led to a state law that would not allow school to begin before September 1.

While school calendar decisions are usually left to local school systems, Sims indicated that the state law had the effect of leveling the playing field, and the state took on what was once a school systems’ decision, reinforcing the state-to-school
accountability tie. School districts were initially developed to monitor and control schools’ compliance to the state’s policies and procedures. After NCLB, school systems shifted from monitoring and controlling school functions to facilitating them according to state laws and education policies (Jennings et al., 2007).

Perceptions play a large role in policy decisions. Regardless of whether administrative expectations and judgments of school and teacher quality come from internal or external forces, standardized test data can be used incorrectly and inappropriately. Unfortunately, testing has been used as a single, punitive measure rather than one piece of what should be many data sources for judging teacher performance and, most importantly, student learning (Gabriel, 2010).

**Accountability and Ethics**

There are accountability concerns regarding the ethics of standardized testing. Educators must adhere to their state’s codes of ethics. In Georgia, the Code of Ethics for Educators specifically addresses testing violations: “Standard 11: Testing - An educator shall administer state-mandated assessments fairly and ethically. Unethical conduct includes but is not limited to: 1. committing any act that breaches Test Security; and 2. compromising the integrity of the assessment” (“Georgia Code,” 2009, p. 5).

Consequences for violating Georgia’s Code of Ethics range from a warning to short-term or long-term suspension of teaching certification to dismissal from employment. The state is currently investigating cheating and testing violations in several school systems and their schools (Resmovits, 2011; Torres, 2010; Vogell, 2011a). One school system transferred students in their system to an alternative school setting (also in their system) so the transferring schools would have indicators of higher student performance. Former Georgia Governor Sonny Perdue authorized a special investigation
into alleged cheating within a large metropolitan school system in north Georgia, and the report was delivered to his successor, Governor Nathan Deal, on June 30, 2011. The special investigators found that some educators in that system erased and changed students’ answers. Georgia is enforcing various consequences, including certificate revocation and termination (Bowers, Wilson, & Hyde, 2011; Vogell, 2011b). As a result of the special investigation, some administrators and teachers may also face criminal charges (Vogell, Judd, & Rankin, 2011).

Unfortunately, Georgia is not the only state dealing with the ethical dilemma of cheating (Gabriel, 2010). Toppo, Gillum, and Bello (2011) reported possible widespread cheating on standardized tests in the District of Columbia’s public schools. Federal investigators from the U.S. Department of Education have been assigned to look into the allegations. According to Rothschild (2011), New Jersey investigated erasures for possible indicators of cheating. In an effort to deter cheating, New Jersey state officials turned in eight teachers to the state’s teacher licensing office for investigation.

Investigations of possible cheating have also taken place in Colorado, Indiana, Massachusetts, Nevada, Texas, and Virginia. Gabriel (2010) found instances where teachers pointed out wrong answers to students, utilized overhead projectors (at their principal’s direction) to show student answers to a leaked copy of a test. There was also evidence that teachers previewed a test booklet and distributed a study guide for students. According to Gabriel, “Experts say the phenomenon is increasing as the stakes over standardized testing ratchet higher—including, most recently, taking student progress on tests into consideration in teachers’ performance reviews” (para. 3).

Mulvenon, Stegman, and Ritter (2005) noted that the teachers surveyed in their study who were required to administer state-mandated standardized tests demonstrated
higher stress levels. Consequently, the high stress level could cause some teachers or administrators to make an unethical decision. Mulvenon et al. noted, “One interesting aspect of this survey that may have future implications in education was the willingness of teachers to admit a readiness to violate testing protocols” (p. 54). Lai and Waltman (2008) investigated teacher ethics in relation to their test preparation practices and found that some teachers were willing to practice actual test items with students. Lai and Waltman also suggested that “determination of the appropriateness of a given practice may have very little to do with whether that practice is consistent with professional ethics” (p. 41). High stress levels can cause some students and teachers to violate test rules in high-stakes testing environments. Stress can be the catalyst in decisions to violate test rules, regulations, and stated procedures, especially in the poorest-performing schools (Chiang, 2009; Sims, 2008; Smolin & Clayton, 2009).

In summary, opponents of high-stakes testing believe it narrows curricula, restricts teachers’ instructional creativity, decreases overall learning, and increases stress levels among teachers and students (Cizek, 2005; Solley, 2007). Additionally, critics of standardized testing assert that teachers are unfairly evaluated if students’ test scores are the sole criterion for determining effectiveness (Gabriel, 2010; Gallagher, 2003). Proponents of high-stakes testing counter that high-stakes testing that leads to teaching to the test is fine as long as the tests are valid and reliable for measuring achievement (Sims, 2008). Sims also found that proponents believe that a greater emphasis on testing brings about higher scores. Both sides agree, however, that accountability based on high-stakes test performance is not going away.

**Teacher Autonomy**

According to Hyslop and Sears (2010), the teaching profession has three key
pillars: (a) recognizing the professional autonomy of teachers, (b) grounding teacher autonomy in a professional community of practice, and (c) engaging teachers and their professional community in public dialogue about education. Skerrett (2009) asserted that NCLB has forced teachers to align curricula and instructional strategies with the test. Supovitz (2009) cautioned that changing instructional strategies and covering specific content due to high-stakes testing are superficial measures for improving instructional practice. High-stakes testing has affected teacher autonomy since the publication of *A Nation at Risk* in 1983, and it influences teachers’ curriculum decisions (Boote, 2006). Boote felt that standardized testing sought the “*remote control of teachers*” and affected their curricular and instructional decisions, thereby reducing their autonomy (p. 462).

Likewise, in Tennessee, Vogler (2006) conducted a quantitative study that incorporated 141 teachers’ responses to a 48-item survey. Vogler found that teachers used predominantly teacher-centered practices in response to preparing students for testing. The author identified a correlation between teachers’ choices of instructional practices and the time they spent on test preparation. When the use of teacher-centered practices increased, more time was spent on test preparation. Au (2009) evaluated a body of empirical research and concluded that “high-stakes tests exert some level of control over teachers’ instructional practice, and that this control often times contradicts what many teachers feel is good pedagogy” (p. 46). Other researchers found a correlation between teacher autonomy and NCLB (Luke & Woods, 2008; Madaus & Russell, 2010/2011). When consequences are punitive, teachers match their content and pedagogy to the test, resulting in a loss of teachers’ autonomy in instructional decision-making (Gayler, 2005).

Instructional autonomy has been discussed for almost one hundred years. An
article with a unique approach to teachers’ perceptions of testing compared Rugg’s (1920) article that provided teachers’ comments on standardized testing in the early twentieth century with five modern teachers’ comments on testing. The unidentified author demonstrated that “things do not change very much” (“Social studies,” 2003, p. 199). In 1920, Rugg found that most tests with standardized items were fact-based rather than requiring thought, inference, reasoning, and judgment. In addition, Rugg stated that test items should assess students’ critical thinking skills. The anonymous author demonstrated that teachers felt the same way in the 2003 article.

Webb (1997, 2002, 2005, 2007) developed the Depth of Knowledge (DOK) framework to address the evaluation of students’ critical thinking skills. Webb identified four levels of questioning that stratify questions by the degree to which higher-order thinking skills are required to answer them. According to Webb, each DOK level increases in the complexity of thinking skills required. The four DOK levels are (a) level 1, simple recall questions; (b) level 2, skill/concept questions; (c) level 3, strategic thinking questions; and (d) level 4, questions to extend thinking. Levels 3 and 4, then, would require the thought, inference, and reasoning that Rugg believed are needed in test items.

Webb’s theory was employed in the design of test items for Georgia’s standardized test program (Barker, 2008; Forte & Paek, 2008). According to Faircloth (2009), GaDOE’s Northwest Georgia Regional Educational Service Agency (NWRESA) promoted Webb’s DOK as a recommended best practice for classroom teachers in designing questions and test items. NWRESA provided professional learning on the use of the DOK framework in designing classroom assessments, supporting teachers in the understanding of standardized test item design and toward the autonomy they need for
designing assessments for instruction. Rugg (1920) believed that the use of standardized assessments would improve classroom instruction as long as test items went beyond rote questioning, which was what NWRESA advised in their professional learning. Rugg’s belief was incumbent on using assessment data to inform students’ needs and allowing teachers to make instructional decisions at the classroom level.

Similarly, Lennon (1955) felt that standardized testing should respect students and lead to “individualization of instruction” and “differentiation of goals and curricula” (p. 35). Rugg’s and Lennon’s words have relevance today in research-based practices such as assessment for learning, individualized instruction, and differentiated instruction (Chappuis & Stiggins, 2002; Landrum & McDuffie, 2010; Murawski & Hughes, 2009; Stiggins, 2005, 2008; Tomlinson, 1999; Tomlinson, Narvaez, & Brimijoin, 2008).

Yet, a modern teacher said that the summative tests do not and cannot “adjust for the course material not yet covered when the test is administered” (“Social Studies,” 2003, p. 200). This mirrors Rugg’s (1920) belief that a summative assessment often tests material which has yet to be covered in the school year. Both in Rugg’s time and now, teacher autonomy is impacted by how assessments are utilized and the inability to adjust them for what has been taught (“Social Studies”). Even though 83 years separate the articles, Rugg in 1920 and the teachers in the modern article believed that the tests do not measure what effective teachers are actually teaching.

On the one hand, standardized tests are criticized because they are summative as opposed to diagnostic, do not provide timely results, and impact teachers’ instructional creativity. Crocco and Costigan (2007) found that teachers felt their “personal and professional identity thwarted, creativity and autonomy undermined, and ability to forge relationships with students diminished” due to scripted lessons and mandated curricula
associated with high-stakes testing across America (p. 513). Similarly, Scot, Callahan, and Urquhart (2009) asserted that NCLB accountability policies undermine teacher autonomy and create “paint-by-number teachers teaching cookie-cutter students” (p. 51).

Conversely, standardized tests are commended for their adherence to controlled administration and uniform scoring, which does not limit teachers’ autonomy. Buck, Ritter, Jensen, and Rose (2010) found that standardized testing did not stifle teacher creativity. They stated that teachers had positive opinions and beliefs in regard to testing as the teachers asserted that standardized testing positively impacts instructional decision-making. Grant (2007) concluded that test-based instructional practices like lecturing and rote memorization can co-exist with class discussions, projects, and debates, which require greater critical thinking skills that may not be measured on high-stakes tests.

Teachers’ perceptions of testing are also affected by the amount of time they have been in the profession. Winkler (2002) examined new and veteran teachers’ perceptions of Virginia’s Standards of Learning assessment. Pursuant to Virginia’s testing requirements, Winkler concluded that experienced teachers cited both a loss of power and professionalism while new teachers cited a loss of professional collaboration and pedagogical freedom.

Clearly, standardized test data and mandates to improve students’ achievement are galvanizing the revisions of curricula and classroom instruction. The environment of testing accountability often drives change, and teachers reported that since the implementation of high-stakes standardized assessment, they have lost the instructional autonomy they had prior to high-stakes standardized testing (Au, 2007; Gallagher, 2003; Graham & Neu, 2004; O’Day, 2002). For this reason, student testing and teacher creativity appeared to be mutually exclusive although researchers found that pockets of
student-centered instruction existed (Au, 2007; Lai & Waltman, 2008; Vogler, 2006). The question becomes whether or not there is a way to prepare students for testing that also allows teachers to ensure a student-centered learning environment.

If the goal is student learning, then educators should make curricular and instructional decisions as well as design school reform efforts that focus on student learning. Moreover, Hyslop and Sears (2010) concluded that professional autonomy for teachers is not just a fundamental requirement for educational improvement but also for ensuring that students become an active democratic citizenry. When comparing the early years of education with the current high-stakes environment, little has changed (“Social Studies,” 2003; Ballard & Bates, 2008). According to Ballard and Bates, the difference now is that teachers are accountable for students’ results and for state-wide and systemically-mandated instructional reform.

**Testing and School Reform**

What then can be determined about how standardized testing impacts school reform? Every state in America uses high-stakes testing to meet NCLB requirements (Baker & Johnston, 2009). The purpose of NCLB is to hold schools, local school systems, and states accountable for improving all students’ academic achievement (“No Child,” 2001). States, school systems, and schools are required to address the needs of low-performing students, up to and including compliance with accountability sanctions of NCLB if a school is found to need improvement based on AYP indicators. NCLB focuses on Title I qualification, which is based on SES. Generally, the percentage of students who qualify for free or reduced-price lunch determines Title I status. Baker and Johnston found that Title I schools significantly underperform non-Title I schools. Furthermore, when comparing SES and ethnicity, the authors found that students’
economic disadvantage had a stronger correlation to low student achievement.

**Perspectives of Administrators and Teachers**

Teachers’ perceptions of reform efforts have garnered researchers’ attention. In a survey of 900 teachers, Johnson, Yarrow, Rochkind, and Ott (2009) found that teachers fell into three categories: disheartened, idealists, or contented. Johnson et al. looked at how teachers’ perspectives differed, the atmosphere and leadership in their schools, and ideas for reform. They concluded that reform efforts in general, and reform efforts based on test results, will not be as successful if teachers are disheartened.

Guskey (2007) provided additional data on administrators’ and teachers’ perceptions of state-mandated testing. Guskey’s quantitative study surveyed 314 educators in three states and attempted to determine if teachers’ and administrators’ perceptions are different. Guskey stated that significant differences exist between the two groups. The author also noted that implementing instructional change and student interventions is difficult when teachers are not given standardized test data within a specified time frame. Guskey found that there was general agreement that testing is needed, but it should include multiple measures trusted by various stakeholder groups. However, there was neither identification of the specific measures needed nor how they would be communicated.

Communication and collaboration are important factors in school reform efforts. Schools are organizational systems with important stakeholders: students, parents, teachers, administrators, staff, and community members. Therefore, each needs a voice in reform efforts. In Georgia, recent reform efforts focused on a systematic accreditation process that requires the involvement of the school system and school stakeholders. The state of Georgia’s accrediting body is AdvancED, the parent organization for the
Southern Association of Colleges and Schools (SACS) and the North Central Association. AdvancED focuses on the continuous school improvement process, requiring adherence to strict standards as it accredits 27,000 schools in America and in schools in 69 other countries (“AdvancED,” 2011).

After successfully completing the accreditation process, schools are accredited for five years and during that five years their improvement efforts are monitored (“AdvancED,” 2011). SACS accreditation is required by the GaDOE. The AdvancED continuous school improvement process can also be aligned to goal-setting for federally-required Title I consolidated applications. Title I schools are required to submit the consolidated application as part of NCLB documentation. All Georgia public schools’ Title I consolidated applications are submitted to GaDOE and provided to the U.S. Department of Education as evidence of school improvement strategies and implementation (“Title I,” 2006). Private accrediting institutions, along with state and federal governments, oversee public school reform, which is expected by a school’s stakeholders (Reese, 2007).

**Other Factors in Test-Based School Reform**

In 1904, Dewey stated that the tendency of educational development and school reform was “to adopt for one year or for the term of seven years, this or that new study or method teaching, and then as abruptly to swing over to some new educational gospel” (as cited in Dondero, 1997, p. 218). Dewey’s words are applicable in the 21st century, too. Resources must be allotted for each new method adoption, and one factor impacting reform from high-stakes testing and accountability has been resource allocation (Chiang, 2009). For example, Chiang found that the threat of NCLB sanctions led to increased spending on instructional technology, curriculum development, and teacher training.
Chiang indicated that the focus on raising test scores and the change in expenditures led to increased math performance in fifth grade. That increase in performance was found in the same group of students into their second year of middle school. Similar to Chiang, Dee and Jacob (2010) also found that achievement gains could be attributed to increased resource allocation in their study, which focused on direct instruction.

Another factor of test-based reform is the alignment of curriculum, standards, instruction, and assessment. In particular, the alignment of the assessment to the mandated curriculum helps ensure test reliability, which in turn can help determine treatment validity of implemented interventions (Decker & Bolt, 2008). Several alignment models have been developed and are recommended by the Council of Chief State School Officers (CCSSO), the same group that co-developed the national CCCS Initiative. The CCSSO-approved models are (a) the NAEP ESSI Webb Procedures; (b) the HumRRO Model; and (c) the Surveys of Enacted Curriculum Model, which was developed by CCSSO (Vockley, 2009).

Some researchers studied the CCSSO models and conducted a study on the alignment the Indiana kindergarten assessment to the state’s kindergarten content standards, using Webb’s DOK framework (Roach, McGrath, Wixson, & Talapatra, 2010; Roach, Niebling, & Kurz, 2008). Documenting the alignment of curriculum, instruction, and assessment is federally-mandated, and Roach et al. (2008) stressed the need for increasing alignment research and practice. Roach et al. (2010) extended prior research in implementing an alignment design that the researchers hoped would be replicated to other assessments and standards beyond standardized testing.

Professional learning is another factor in test-based school reform. Many professional learning courses address understanding test data and using high-stakes test
results to inform instructional decisions. Henning (2004) investigated how test data impacted instruction. In an analysis of the Iowa Test of Basic Skills, the author studied 24 elementary and middle school lead teachers at their home schools. Training on four types of data analysis was provided, and participants were allowed to choose the data analysis they preferred. The types were (a) comparing to the norm, (b) analyzing trends, (c) correlating data, and (d) disaggregating data. The study’s data consisted of a short, written data analysis reports from each of the 24 research participants.

Utilizing qualitative a priori coding of the participants’ written observations, Henning (2004) found that the teachers analyzed their data in a variety of ways and that two teachers utilized data applications that were not provided in the training. Teacher leaders made effective use of only one year of data by disaggregating and correlating the data. Henning found no evidence that the lead teachers enacted instructional change or impacted student achievement based on their analysis of their school’s test data.

Similarly, Moon, Brighton, Jarvis, and Hall (2007) noted that only a minority of teachers developed rich, challenging curricula in an environment of high-stakes testing. Moon et al. found for the majority of teachers, “Data from both the national survey and the subsequent qualitative component of this study indicate that teachers' curricular and instructional practices are substantially shaped by the high-stakes associated with testing” (p. xi). Teachers reported an escalating emphasis on the mandated test results from year to year. They felt the emphasis on results led to narrowed curricula and increased teacher-centered instructional practices.

Unfortunately, even though Moon et al. indicated that testing has led to curricular and instructional reform, those reforms may not be the best practices for student learning. Furthermore, Giles and Hargreaves (2006) studied the pattern of reform sustainability
over time and found that even innovative schools that use a professional learning
community model reverted to traditional schooling patterns due to accountability aspects
of high-stakes testing.

Another factor in school reform is teaching to the middle, which focuses on the
average learners and marginalizes the lowest-performing and the highest-performing
students. Test results in Georgia identify students’ at one of three performance levels and
illustrate the factor of teaching to the middle. For example, students in grades 3-8 must
take the CRCT. Student performance levels are 1, did not meet standards; 2, met
standards; and 3, exceeded standards. The focus is on the group at performance level 1.
Levels 2 and 3 are lumped together on published AYP reports. Reback (2008) found that
this pass/fail type of system encourages schools and school systems to improve the
academic achievement of students who are closest to Level 2. Reback noted that when
students feel that their test scores matter, low-achieving students perform better than
higher-achieving students. In response to their test results and accountability concerns,
schools in the Reback study reallocated resources in order to target specific students and
subjects, which resulted in disproportionate spending.

Student retention is another factor of student reform that impacts all stakeholders.
Roderick and Nagaoka (2005) found no evidence to support schools’ decisions to retain
students even though grade level retention is a component of NCLB. Students are
retained when they do not pass their state’s high-stakes test in specific benchmark years.
Students in the Roderick and Nagaoka study who were retained continued to struggle in
the subsequent year and were more likely to be considered for special education
placement. The authors found it was inconclusive as to whether retention increased
student achievement, but for some students, retention resulted in lower grades.
In Georgia, retention decisions are based on a student’s performance on gateway standardized testing administered in grades three, five, and eight. Prior to retention decisions, test data are reported to schools and parents are contacted. In compliance with NCLB requirements, students who do not pass Reading or Math portions of the test are retained (“Promotion,” 2002). Student retention is part of NCLB and state accountability requirements, even though research indicates that retention has far more negative consequences than positive ones based on students’ subsequent academic achievement in high school (Bonvin, Bless, & Schuepbach, 2008; Jimerson, Pletcher, Graydon, Schnurr, Nickerson, & Kundert, 2006; Patterson & Beltyukova, 2007; Silberglitt, Jimerson, Burns, & Appleton, 2006; Stearns, Moller, Potochncko, & Blau, 2007).

Clearly, school reforms impact a school’s stakeholders—administrators, teachers, parents, and students. Each group has a perception of and experience with high-stakes testing. Yet, are parents able to understand a school’s testing policies and test data, or are they able to interpret their child’s individual report? Is a student retained if he or she does not meet grade level competencies? Parents reported that even though they supported test administration and were interested in their children’s results, there was poor communication about testing and test data from school administrators, teachers, and counselors (Mulvenon, Stegman, & Ritter, 2005).

According to Reese (2007), Americans love to reform public schools and have since Thomas Jefferson advocated state-assisted schooling in 1781. Since public schools emerged in the U.S., Americans have presumed it is their right to have their education concerns addressed because they pay yearly taxes. Reese discussed how policymakers have acquiesced to taxpayer demands. Schools have become multi-purpose institutions that provide academic instruction, athletic activities, career electives, character education
programs, and various clubs. A recent poll revealed that Americans want the school funding crisis resolved, teacher pay correlated to performance, charter school options increased, government involvement in education reduced, and teachers provided with additional professional development opportunities. When asked what schools must do to improve, respondents wanted schools to (a) help students be successful, (b) improve the quality of teaching, and (c) implement a rigorous curriculum (“PDK/Gallup,” 2010).

**Summary**

For American educators, accountability measures such as AYP will continue. Federal and state educational legislation will mandate that students pass standardized assessments. Throughout its history in the U.S., standardized testing has affected teachers’ instructional decisions as they attempt to ensure equity and educational attainment for all students. Unfortunately, standardized testing policies have also created an environment where educators, especially public school teachers, are held accountable for testing results which really should only be one measure used to determine teacher effectiveness. Standardized test results should inform school reform and improvement decisions, but the testing results should be only one component of a comprehensive analysis of students’ learning.

Based on the information that supports the use of high-stakes testing to help inform decisions on curriculum, instruction, and assessment, the current study strives to augment the understanding of standardized testing in Georgia and the best use of testing data. This study examined the correlation between students’ performance on the eighth grade CRCT and the ninth grade EOCT and any predictive value of 8th grade scores on ninth grade achievement.
CHAPTER 3: METHODOLOGY

Similar to the other states in America, Georgia is mandated to administer student assessments due to the accountability measures of the *No Child Left Behind Act of 2001* (NCLB). Hence, developing an understanding of the data provided by the testing authority will benefit school systems and their respective schools. Moreover, it is important for schools to utilize the student data to impact student learning and instruction. Georgia’s Criterion-Referenced Competency Test (CRCT) is mandated for third, fifth, and eighth grades, and many systems administer it to all students in grades three through eight. It is comprised of five subtests: (a) Reading, (b) English/Language Arts (ELA), (c) Mathematics, (d) Science, and (e) Social Studies. At the high school level, all eleventh grade students in Georgia who entered ninth grade prior to the 2011-2012 school year must take the criterion-referenced Georgia High School Graduation Test (GHSGT) in the following subject areas: (a) ELA, (b) Writing, (c) Math, (d) Science, and (e) Social Studies.

In addition to the GHSGT, Georgia high school students in grades 9-12 must take a state-mandated End-of-Course Test (EOCT) in eight pre-identified subjects. Beginning in the 2011-2012 school year, the GaDOE will phase out the GHSGT for high school students and replace it with the EOCT. GaDOE will use the EOCTs as the AYP accountability measure for NCLB for ninth graders entering high school in the 2011-2012 school year (Barge, 2011). In a press release, Georgia State School Superintendent Barge said, “The EOCTs are much more rigorous, and they test a student immediately following a course, rather than waiting until a student’s Junior year to determine whether or not he or she has mastered the content of our curriculum” (as cited in Cordoza, 2011, para. 2).
The purpose of this study was to determine if there was any correlation in or predictive value of students’ achievement on state-mandated standardized testing from their eighth grade year to their ninth grade year. This correlational research study followed student groups over two years of test participation and attempted to control for internal and external threats to validity and for variables that affect performance.

The methodology included information regarding research design, study participants, the study setting, instrumentation, procedures, and data analysis. The results from the administrations of the CRCT and EOCT, both aggregate and by subgroups, can inform Georgia educators of any relationships that may exist between test scores on the CRCT and test scores on the EOCT. Further, if eighth grade standardized test achievement can predict ninth grade standardized test achievement, then interventions can be developed for at-risk eighth grade students whose achievement does not meet state minimum requirements for meeting standards so that they can be supported toward academic success in ninth grade. Additionally, interventions for these at-risk students may decrease the incidence of dropping out of high school, which often happens between ninth grade and tenth grade. Ultimately, successful interventions may increase high school graduation rates. The high school graduation rate becomes more than a number; it represents people—the students who see their education through graduation and into post-secondary options.

**Research Design**

Throughout the research process, I have sought understanding and the wisdom to make appropriate judgments, relying on God’s promise, “But if any of you lacks wisdom, let him ask of God, who gives to all generously and without reproach, and it will be given to him” (James 1:5, New American Standard Bible [NASB]). His wisdom is
characterized in James 3:17, “But the wisdom from above is first pure, then peaceable, gentle, reasonable, full of mercy and good fruits, unwavering, without hypocrisy” (NASB). His wisdom has provided guidance through the research process and beyond.

This study utilized a correlational research design to define the relationship between two variables. Gall, Gall, and Borg (2007) stated, “The basic design in correlational research is very simple, involving nothing more than collecting data on two or more variables for each individual in a sample and computing a correlation coefficient” (p. 335). “The basic research question for correlation research is - What is the relationship between two or more variables for a given set of subjects” (“Correlational Research,” 2010, para. 2). This study sought the answers to three research questions.

The first research question is stated as follows: What is the relationship between students’ achievement on the eighth grade Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT? Study data were a database of standardized test data for all eighth graders and ninth graders taking the CRCT and the EOCT for each of the years of GPS-based testing from 2006 to 2011. The analyses compared eighth grade students’ scores on the CRCT with their scores on the EOCT in their ninth grade year. The data were combined into one spreadsheet and exported into BASE SAS 9.2 statistical software program for study. Data for the eighth grade Reading CRCT and the Ninth Grade Literature EOCT were separated from other content areas and identified by testing year for analysis. Determining the strength of the correlation provided the answer to the research question.

The second research question follows: What is the relationship between students’ achievement on the eighth grade ELA CRCT and the same students’ achievement on the
Ninth Grade Literature EOCT? The analysis procedure for studying the eighth grade ELA CRCT and the Ninth Grade Literature EOCT mirrored the process described for Research Question 1. Using the same combined spreadsheet exported to BASE SAS 9.2, the ELA CRCT data and the Literature EOCT data were separated from the other contents for analysis.

Last, the third research question is stated as follows: If a relationship exists, what is the predictive value of students’ eighth grade achievement as measured by the Reading or ELA CRCT on their achievement on the Ninth Grade Literature EOCT? The first two research questions sought to determine the strength of correlations between the scores from eighth grade CRCT tests to ninth grade EOCTs. The third question, however, addressed the interest in any predictive value in the eighth grade scores. The content-specific datasets were exported to the BASE SAS 9.2 statistical software program, where a regression analysis was conducted to identify if any predictive value was evident.

The nature of a correlational research design can become complicated as other aspects are added, such as disaggregating data or incorporating additional study groups from prior years into the study (Gall, Gall, & Borg, 2007). Perhaps the most important aspect of correlational research is that it examines the strength of relationships or the direction between two or more variables. In addition, if variables are correlated, the one variable may predict the other. One variable does not cause the other and vice versa; a researcher must ensure that correlations between variables are not characterized as such.

In correlational research, the measured relationship between two variables has both a degree and a direction. The degree is identified between -1 and +1 in a decimal measure called the correlation coefficient. This is sometimes referred to as Pearson’s product moment coefficient or Pearson’s $r$. The closer to -1 or +1, the greater the
relationship, and a zero indicates that no relationship exists. The direction of the relationship is indicated by the positive or negative sign (- or +). “A negative correlation means that as scores on one variable rise, scores on the other decrease. A positive correlation indicates that the scores move together, both increasing or both decreasing” (Davis, 1997, para. 6).

Davis continued, “If there is a correlation between two variables, and we know the score on one, the second score can be predicted. Regression refers to how well we can make this prediction” (para. 8). If there is a correlation, for example, between eighth grade ELA CRCT achievement and Ninth Grade Literature EOCT achievement, then, using the regression model, the researcher can draw inferences about any predictive value of the CRCT criterion variable. Caution must be exercised on assigning any predictive value, remembering that correlational research does not assign cause and effect between variables.

In addition to determining correlation, I analyzed the data to see if the observation was real or just chance by testing the null and the alternative hypotheses. For this study, the null hypothesis was that there was no relationship between the eighth grade CRCT and ninth grade EOCT variables (H₀: r = 0). The alternative hypothesis was that a relationship existed between the two variables (H₁: r ≠ 0).

At the α = .05 significance level, a researcher can conclude that the odds are less than or greater than 5 out of 100 and whether this was a chance occurrence. If the Pearson’s product-moment coefficient indicates significance, the conclusion can be made that it was not a chance finding and that the correlation was statistically significant at which time the researcher can reject the null hypothesis and accept the alternative (Davis, 1997, para.16). Conversely, if α < .05 and the correlation coefficient is not significant,
then the null hypothesis can be accepted.

Several questions had to be considered before deciding on this study’s
correlational research design. Howell (2007) provided a graphical representation of a
decision tree and stated that it is “designed to help you consider the relevant issues
involved in selecting a statistical test (the issues of the type of data, the question of
relationships versus differences, the number of groups, and whether variables are
independent or dependent)” (p. 519). Howell’s decision tree is shown in Figure 3.1; the
highlight has been added by this researcher.

Figure 3.1: Howell’s Decision Tree for Selecting Statistical Tests

Using Howell’s decision tree organizer and the research questions developed for
this study, three issues had to be addressed in order to identify the type of research
design. First, the type of data had to be defined. For this study, the data observations
were quantitative standardized test scaled scores. Then, the type of question must be
considered. Because the study is determining the relationship between test scores in eighth and ninth grades, a researcher can follow the decision tree to the “relationships” block on the organizer, which indicates the research design is correlational.

Next, even though several groups were being used, only one group’s scores were being compared at a time; thus, there was one measurement, and data were continuous rather than ranked. For this study, both analyses for correlation coefficients and regression analyses were conducted. Data were analyzed using Pearson’s $r$ to determine the degree of correlation to address the first two research questions. The third research question considers the nature of any possible predictive value of the eighth grade CRCT in Reading and ELA on the respective Ninth Grade Literature EOCT. The regression analyses were conducted to allow the “prediction of one variable from knowledge of one or more other variables” (Howell, 2007, p. 212). The regression analyses provided a foundation from which to contemplate any predictive values in the variables of CRCT performance.

This correlational study followed student groups over a paired eighth and ninth grade test participation. Five groups were identified for each subject area of the standardized tests over years of test administration from 2006 to 2011. For this study, a study group was a group of eighth grade students who took the CRCT in Reading and/or ELA and who also took the Ninth Grade Literature EOCT in their ninth grade year for each the years 2006-2011 that GPS-based test data are available.

**Participants**

In 2006, GPS-based assessments began for the CRCT in Reading and ELA and for the EOCT in Ninth Grade Literature. This study’s population was all eighth grade and ninth grade students in a rural north Georgia school district from school year 2006 to
school year 2011. Approximately 2,000 students were enrolled into eighth and ninth grades annually in the district. Students who took the eighth grade Reading CRCT and/or the eighth grade ELA CRCT and then took the Ninth Grade Literature EOCT in their subsequent ninth grade year were considered a sample group. Several groups were identified by subject for all years of GPS-based testing. For example, students who took the ELA CRCT in 2006 as eighth graders and then took the EOCT in Ninth Grade Literature in 2007 as ninth graders were considered a group, representing approximately 1,000 students.

Table 3.1 shows the groups that were identified for each of the paired two-years of CRCT/EOCT data. There were five reading/literature and five ELA/literature groups, and each group had about 1,000 students. Students enrolled as repeaters in ninth grade courses were excluded. The test data were analyzed for paired subject area groups as well as for each subject area group independent of the others. Subgroup membership was provided in the student achievement data from the school system. These data were disaggregated by school, gender, ethnicity, and students with disabilities to determine if there was a difference in subgroup performance.

**Setting**

According to data provided on the researched school system’s website, Best County is located in north Georgia in the foothills of Georgia’s Appalachian Mountains. Best County is on the Interstate 75 corridor and is in a prime location for growth in industry, agriculture, and manufacturing. Best County is home to 100,157 residents and has always maintained a strong sense of community and tradition (Wagner, 2011).
Table 3.1

Research Student Groups

<table>
<thead>
<tr>
<th>CRCT-EOCT Years</th>
<th>8 Reading/9 Lit</th>
<th>8 ELA/9 Lit</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006-2007</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2007-2008</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2008-2009</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2009-2010</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>2010-2011</td>
<td>x</td>
<td>x</td>
</tr>
</tbody>
</table>

Note. An “x” indicates paired subject area tests (i.e., 8th Reading CRCT and 9th Literature EOCT, 8th ELA CRCT and 9th Literature EOCT).

Two post-secondary campuses are in the county. Also, many students attend one of four colleges and universities close by, all of which are only a 30-minute drive from the center of Best County (BCSS, 2010).

The median age for the Best County population is 35.6 years, and 21.2% of residents are age 15-19. An average family size is 3.14 persons. The county population is 84.5% White, 10.4% Black, 4.9% Hispanic, and 0.2% other. The median household income is $54,346, and per capita income is $22,683. The current unemployment rate is 11.4%. The percentage of Best County residents with a high school diploma or with a bachelor’s degree or higher is 83.2% and 15.0%, respectively (BCSS, 2010).

Best County is the 25th largest school district in Georgia and provides education to approximately 14,500 students. The system and each of its 20 schools are accredited by the Southern Association of Colleges and Schools. There are 12 elementary schools,
four middle schools, and three high schools, as well as a state-funded pre-K center. Student demographics are 80% White, 8% Black, and 7% Hispanic. Students who are economically disadvantaged comprise 52% of the total student population, and all elementary and middle schools are identified Title I schools. There are 12.5% of students receiving special education services, and 4% percent are English Language Learners (BCSS, 2010).

**Instrumentation**

The CRCT and EOCT assessments were the instruments utilized in the research study. Before analyzing the data, it was important to understand scoring for both the CRCT and EOCT. The CRCT was developed by the Georgia Department of Education (GaDOE) as “part meeting federal requirements for state standards and assessments systems” (“Criterion-Referenced,” para. 8). The test and content area subtests were peer reviewed by a team of external experts, convened by the U. S. Department of Education, in the fields of standards and assessments. According to GaDOE, “The CRCT was found to meet nationally recognized professional and technical standards for assessment programs” (para. 8).

Scaled scores are provided electronically and in hard copy to systems and schools and are based on the number of correct items. Parents and students receive a printed individual student report. The GPS version of the CRCT sets 800 as the minimum scaled score needed to meet requirements. The cut scores are (a) below 800, do not meet the requirements; (b) 800-849, meet requirements, and (c) 850 or above, exceed requirements. Performance levels are identified as Level 1, does not meet standards; Level 2, meets standards; and Level 3, exceeds standards.

Germaine to Georgia high-stakes testing is the fact that the eighth grade is a
benchmark year in CRCT test administration. Students must meet the cut score of 800 or higher in order to be promoted to ninth grade. Two retest opportunities are given during the summer of the eighth grade year for students who score at Performance Level 1. If a student does not receive a passing score on retests, a committee comprised of the middle school’s principal, the parents, and teachers is convened over the summer to determine where the student will attend the next school year.

Like the CRCT, the EOCT is a state-mandated, standardized test administered at the completion of each of the eight required courses. The GaDOE provides electronic and print versions of EOCT score reports to systems and schools. Students receive an assessment report that provides the scaled score. Students who score below 400 are identified at Performance Level 1 and do not meet standards. Scores from 400 to 449 meet standards and are identified at Performance Level 2. Last, a score of 450 or higher identifies a student at Performance Level 3, exceeding standards. These scaled scores are converted to a percentage for ease in computing students’ grades and on individual student reports provided to parents. Students also see a grade conversion score on their individual reports. The grade conversion score is on a 100-point scale and is entered as 15% of the final course grade, which must be 70 or above for a student to receive credit for the course. As Georgia phases out the GHSGT, the EOCT grade weight will change to 20% for those students entering ninth grade in the 2011-2012 school year.

Even though the GHSGT is the test that students must take to meet graduation requirements, the EOCT may be used as a criterion for receiving a variance for the GHSGT if a student fails to pass the GHSGT in the comparable subject area. For example, if students fail the English Language Arts section of the GHSGT, graduation status is threatened. Yet, if the students have passed the EOCT in either Ninth Grade
Literature or eleventh grade American Literature, then they meet graduation requirements and do not have to retake the GHSGT in ELA.

For this study, both the predictor variable of CRCT data and the dependent, criterion variable of EOCT data used the scaled score as the primary measure. Both of these tests have validity and reliability data provided by the Testing and Assessment Division of the GaDOE (“CRCT,” 2010, “EOCT,” 2010). All study data were post-assessment and de-identified for student and teacher anonymity, which helped ensure objectivity.

The GaDOE oversees all aspects of state-mandated testing, including the CRCT and the EOCT. There is rigid adherence to the Standards for Educational and Psychological Testing that have been established by the American Psychological Association (APA), the National Council on Measurement in Education (NCME), and the American Educational Research Association (AERA) (AERA, 2008; APA, 2011; NCME, 2011; “Standards,” 2009). The CRCT and EOCT assess the overall quality of education in Georgia (“CRCT,” 2010, “EOCT,” 2010).

Validity is the foremost consideration, but it cannot be assured without high measures of reliability, an understanding of testing context and degree of validity, and collections that are measured over time. The GaDOE has taken several steps to ensure that the CRCT and the EOCT are valid instruments. The first evidence of validity is clearly providing the test’s purpose. According to the GaDOE, the purposes of the statewide standardized testing program are (a) to measure student progress toward mastery of the Georgia Performance Standards (GPS), the state’s mandated curriculum; (b) to identify struggling learners; (c) to provide data and data analyses to inform instructional decisions; and (d) to identify strengths and weaknesses that school systems
use in educational planning (“CRCT,” 2010, “EOCT,” 2010). Additional goals for the CRCT are to demonstrate accountability and meet the requirements of NCLB.

Second, validity is ensured in the each aspect of CRCT and EOCT development, from initial curriculum alignment to eventual test data. Test development is a multi-step process beginning with aligning the curriculum and identifying content descriptors that will be tested. Committees made up of content specialists, contracted test designers, and Georgia educators are involved throughout the development process to create test and item requirements. Field testing occurs by embedding sample test items in the operational versions of the CRCT or EOCT for committee review. In multiple reviews, the test development committees decide if field test items are approved or rejected for future test forms of the CRCT or EOCT. Multiple forms are developed by content specialists and psychometricians. These forms used in the same year or in subsequent test administrations are statistically equated to make sure each form of the test is of equal difficulty (“CRCT,” 2010; “EOCT,” 2010). This was an important consideration for the current research proposal because multiple years of testing are being studied.

For a test to be valid, it must also be reliable. The 2010 CRCT reliability was measured by Cronbach’s alpha reliability coefficient and by the standard error of measurement (SEM). GaDOE reported Cronbach’s alpha in Reading at .86 and in ELA at .89. The SEM for each was 2.38 and 2.70, respectively (“CRCT,” 2010). Conditional SEMs were also provided. Cronbach’s alpha coefficients ranged from 0 to 1, and the reported coefficients indicated reliability for students’ test performance. The SEMs reinforce the CRCT’s reliability and consistency, and the reliability indices support the test’s overall validity.

For the Spring 2010 EOCT administration, GaDOE reported one Cronbach’s
alpha reliability coefficient of .89 for all eight test administrations rather than one coefficient for each of the eight testing areas. The range of coefficients for all eight tests fell between .87 and .93, all of which indicate a high degree of reliability. Furthermore, a SEM was used to quantify test precision on the two forms of the Spring 2010 EOCT. Confidence intervals for Ninth Grade Literature for Form 1/Form were 3.28/3.35 (“EOCT,” 2010). GaDOE addressed validity in test and item development and EOCT administration and provided adequate statistical data to establish reliability.

**Procedures**

I gained approval from the school system and school to acquire and use the test data (see Appendix C). I also secured approval for Liberty University’s Institutional Review Board (see Appendix D). Participants were all students enrolled in the researched school system in eighth grade since 2006 who took a GPS-based CRCT in Reading or ELA and also took the Ninth Grade Literature EOCT the following year. A statistician de-identified the standardized test data to ensure participants’ anonymity and to control for researcher bias. All students’ GPS-based standardized test data for the CRCT in Reading and ELA and the EOCT in Ninth Grade Literature were exported into the BASE SAS 9.2 statistical software program for analysis.

Data analyses provided descriptive statistics, univariate and bivariate statistics, Pearson r correlation coefficients, and regression models. In addition, statistical analyses provided assumption testing that addressed distribution normality, sufficient sample size, and outliers. Data were disaggregated by test year, school, gender, ethnicity, and students with disabilities. Using the correlation coefficients, the data analyses were interpreted to identify the strength of correlations between the study groups’ CRCT and EOCT scaled scores. The regression analyses were conducted to determine if there was any predictive
value of the CRCT to the EOCT. The results of the data analyses were interpreted, and implications for educators and students will be discussed.

**Data Analysis**

The test data were analyzed for all of the researched district’s study groups in each subject area and test year to determine if eighth graders’ achievement correlates with ninth grade achievement and if eighth graders’ achievement predicts their ninth grade achievement on the identified standardized test measures. The standardized test database for the school system’s eighth and ninth graders was combined into one spreadsheet that was disaggregated by subject area. The data were also disaggregated by testing year, school, gender, ethnicity, and students with disabilities. Further, data analysis addressed each research question.

Research question 1 was stated as follows: What is the relationship between students’ achievement on the eighth grade Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT? The spreadsheet containing the Reading CRCT and Literature EOCT data were analyzed using the BASE SAS 9.2 statistical software program to identify Pearson’s product-moment coefficients, or Pearson’s $r$. Data were disaggregated by testing year, school, gender, ethnicity, and students with disabilities for the implications of subgroup performance.

Research question 2 was stated similarly and follows: What is the relationship between students’ achievement on the eighth grade ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT? Using BASE SAS 9.2, Pearson’s $r$ coefficients were identified for the data from the ELA CRCT and Literature EOCT spreadsheet. As with the Reading CRCT and Literature EOCT data, the coefficient analysis helped determine if correlations existed and the strength of any correlations.
Research question 3 was stated as follows: If a relationship exists, what is the predictive value of students’ eighth grade achievement as measured by the Reading or ELA CRCT on their achievement on the Ninth Grade Literature EOCT? Unlike the data analyses for correlations, this question sought to determine if there was any predictive value in eighth grade scores on ninth grade performance. A regression analysis was conducted on each content-area dataset to determine if any predictive value exists as well as the strength of predictive value if present.

Data were studied using BASE SAS 9.2 software. Both univariate and bivariate descriptive statistics were provided. Using statistical software, data analyses were run to identify any extreme values and to determine skewness. Further, data were analyzed for Pearson’s $r$ to the degree to which the variables are correlated, and regression analyses were also conducted to determine any predictive values in the criterion variables. Both the Pearson’s correlation coefficients and the regression analyses were used to determine the degree to which two variables are correlated and to identify any possible predictive value of one criterion variable on its associated predictor variable (Howell, 2008). Not only were the degree of correlation between eighth grade CRCT and ninth grade EOCT scores identified using Pearson’s $r$, but also the predictive value of the eighth grade CRCT scores could be inferred using regression analyses. Data results were visually represented using tables, histograms, and scatterplots.

There were three null hypotheses for the study:

- $H_{01}$: There is no statistically significant correlation between eighth grade students’ achievement on the Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.
- $H_{02}$: There is no statistically significant correlation between eighth grade
students’ achievement on the ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

- $H_{03}$: Students’ eighth grade performance on the Reading or ELA CRCT will not be predictive of the same students’ performance on the Ninth Grade Literature EOCT.

The null hypotheses that there was no correlation between eighth grade CRCT scores and ninth grade EOCT scores can be shown as $H_{01}: r = 0$, $H_{02}: r = 0$, and $H_{03}: r = 0$. Gall, Gall, and Borg (2007) discussed tests of statistical significance and provided information to avoid a Type I error where one would reject the null when it is true or a Type II error where one would accept the null when it is false by selecting the most appropriate significance level. The null hypotheses in the study will be rejected at $\alpha < .05$. If identified when the statistical analyses are conducted, the appropriate $p$ value of less than .0001, .01, .05, or .10 will be stated as the measure of significance.

Howell (2008) also provided factors that affect correlation. One is range restriction, which will not be an issue in this study because the range will not be artificially limited. Nonlinearity is another factor that Howell identified; however, a linear relationship is an underlying assumption of the current research design and will be tested in the data analyses. The third factor is heterogeneous subsamples. This study’s research design will control for heterogeneous subsamples by studying disaggregated results by testing year, school, gender, ethnicity, and students with disabilities.
CHAPTER 4: RESULTS

The Best County School System (BCSS) in Georgia requires the Criterion-Referenced Competency Test (CRCT) in third through eighth grades, with third, fifth, and eighth grades identified as benchmark years for determining grade placement. The CRCT is the instrument used to measure Adequate Yearly Progress (AYP) for elementary and middle schools in Georgia (“Criterion-Referenced,” 2011). The End-of-Course Test (EOCT) is given as a summative assessment in Ninth Grade Literature, American Literature, Math I, Math II, Physical Science, Biology, U.S. History, and Economics across the four years of high school. Ninth graders, participants in this study, are assessed on EOCTs in Ninth Grade Literature, Math I, and Physical Science. Georgia students take the CRCT in May of their eighth grade year and take the EOCT in May of their ninth grade year.

Overview of Problem

The problem investigated in this study is that educators in Georgia have not examined if relationships exist between scaled scores on the eighth grade CRCT in Reading and ELA and the scaled scores on the Ninth Grade Literature EOCT. The GaDOE has not conducted studies on correlations between the high-stakes CRCT and EOCT or on any predictive value in the CRCT on EOCT performance. Scores from both the CRCT and the EOCT are used in making decisions about promotion and retention, course and program placement, and progress toward high school graduation. Additionally, schools’ performance on each assessment is published in school system communications as well as in local media. The transition to high school is an important educational milestone for students, and it is important to identify struggling learners and
any gaps in their content knowledge. Understanding the implications for struggling students will help educators support them in the important transition to high school and the shift from the CRCT in the eighth grade to the EOCT in ninth grade.

If the CRCT is significantly correlated to the EOCT, and if any prediction can be made about EOCT performance based on CRCT achievement, then interventions can be designed to address the needs of struggling learners. Moreover, test data analysis could impact students’ instruction and, ultimately, students’ learning. Test data can inform curricular and instructional decisions so that educators can fill students’ knowledge gaps in mastering the tested curriculum standards. Thus, the information can be used to support student learning and high school graduation rates.

**Restatement of Purpose**

The purpose of this study was to determine if there was any correlation or predictive value in students’ achievement on state-mandated standardized testing from the CRCT in their eighth grade year to EOCT administration at the end of their ninth grade year. Both the CRCT and the EOCT were developed to assess students’ mastery of the Georgia Performance Standards (GPS), which were implemented in middle school Reading and ELA and in high school Literature during the 2005-2006 school year. Additional content area standards and tests were added in subsequent school years in a staged roll-out.

Due to the limited number of years the GPS-based standardized assessments have been administered, it was important to study both aggregate and disaggregated data to identify if eighth and ninth grade achievement were correlated and, particularly if the eighth grade data have any predictive value for the ninth grade achievement. If correlations were evident, then students at risk for graduating can be identified early, and
interventions can be developed to support these students toward academic success in ninth grade. The interventions could result in (a) an increase in students’ test scores and grades, (b) a reduction in the dropout rate between students’ ninth and tenth grade years, and (c) an increase in the number of students graduating high school.

**Instrumentation**

Due to the fact that the CRCT and EOCT were the instruments utilized in the current research study, it was important to establish their validity and reliability. The Testing and Assessment Division of the GaDOE provides assessment and accountability briefs. For the 2010 test administration, GaDOE reported Cronbach’s alpha in Reading at .86 and in ELA at .89. In addition, the $SEM$ for each is 2.38 (Reading) and 2.70 (ELA) (“CRCT,” 2010). Conditional $SEMs$ were also provided. The $SEMs$ reinforce the CRCT’s reliability and consistency, and the reliability indices support the test’s overall validity.

In regard to the Spring 2010 EOCT, GaDOE reported one Cronbach’s alpha reliability coefficient of .89 for all eight test administrations as opposed to having a coefficient for each of the eight testing areas. The range of coefficients, using Cronbach’s alpha for all eight tests, fell between .87 and .93, indicating a high degree of reliability. Confidence intervals for Ninth Grade Literature for Form 1/Form 2 were 3.28/3.35 when using a $SEM$ to quantify test precision on the two forms of the Spring 2010 EOCT (“EOCT,” 2010).

**Univariate Analyses**

A set of CRCT scores and a set of EOCT scores were provided for all years of GPS-based standardized assessments (2006-2011). These data were assimilated into one data file, which were de-identified by a statistics consultant prior to analysis. This
ensured confidentiality and anonymity for participants. Data included scaled scores for 5,495 CRCT test-takers and 5,085 EOCT test-takers from 2006 through 2011. A univariate analysis was conducted in order to study characteristics of the research population of all students from 2006 through 2011 who took the CRCT in Reading or English Language Arts (ELA) in their eighth grade year and also took the Ninth Grade Literature EOCT in their ninth grade year.

**CRCT Data**

Ethnicity data included six values: 1, Asian/Pacific Islander; 2, African American/Non-Hispanic; 3, Hispanic; 4, Native American/Alaskan Native; 5, White/Non-Hispanic; and 6, Multiracial. The frequency was dominated by the value 5, which represented White students. Table 4.1 illustrates ethnicity frequency, where 81.03% of participants are White. When data results were provided based on ethnicity, the student samples were categorized as White or non-White for discussion purposes.

**Table 4.1**

**CRCT Ethnicity Frequency Table**

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>31</td>
<td>0.56</td>
<td>31</td>
<td>0.56</td>
</tr>
<tr>
<td>2</td>
<td>522</td>
<td>9.51</td>
<td>553</td>
<td>10.07</td>
</tr>
<tr>
<td>3</td>
<td>381</td>
<td>6.94</td>
<td>934</td>
<td>17.02</td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td>0.53</td>
<td>963</td>
<td>17.54</td>
</tr>
<tr>
<td>5</td>
<td>4448</td>
<td>81.03</td>
<td>5411</td>
<td>98.58</td>
</tr>
<tr>
<td>6</td>
<td>78</td>
<td>1.42</td>
<td>5489</td>
<td>100.00</td>
</tr>
</tbody>
</table>
Data studied according to gender can be seen in Table 4.2, which shows 49.03% of the participants were female and 50.97% male.

Table 4.2

*CRCT Gender Frequency Table*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>2,694</td>
<td>49.03</td>
<td>2,694</td>
<td>49.03</td>
</tr>
<tr>
<td>Male</td>
<td>2,801</td>
<td>50.97</td>
<td>5,495</td>
<td>100.00</td>
</tr>
</tbody>
</table>

This study’s participants were enrolled in one of four middle schools in the BCSS, and Table 4.3 illustrates the distribution of data in the four middle schools. School B represented the largest percentage of participants at 31.45%, and School C had the smallest percentage at 18.80%.

Table 4.3

*Middle School Frequency Table*

<table>
<thead>
<tr>
<th>School</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>School A</td>
<td>1,200</td>
<td>21.84</td>
<td>1,200</td>
<td>21.84</td>
</tr>
<tr>
<td>School B</td>
<td>1,728</td>
<td>31.45</td>
<td>2,928</td>
<td>53.29</td>
</tr>
<tr>
<td>School C</td>
<td>1,033</td>
<td>18.80</td>
<td>3,961</td>
<td>72.09</td>
</tr>
<tr>
<td>School D</td>
<td>1,534</td>
<td>27.92</td>
<td>5,495</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Of the eighth grade students who participated in this study, 11.45% are students with disabilities and are served by an Individual Education Plan (IEP).
CRCT data were provided for the GPS-based standardized testing years 2006 through 2010, and the data were evenly distributed across the testing years as illustrated in Table 4.4.

Table 4.4

<table>
<thead>
<tr>
<th>CRCT Testing Year Frequency Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>2006</td>
</tr>
<tr>
<td>2007</td>
</tr>
<tr>
<td>2008</td>
</tr>
<tr>
<td>2009</td>
</tr>
<tr>
<td>2010</td>
</tr>
</tbody>
</table>

Test data were reported as scaled scores. For the CRCT, the GaDOE set 800 as the minimum scaled score needed to meet requirements and, thus, pass the test. Scores below 800 do not meet the requirements; scores of 850 or above exceed requirements. Table 4.5 provides the descriptive statistics for the scaled scores on the Reading CRCT (REAss) and the ELA CRCT (ELAss).

Figure 4.1 and Figure 4.2 are histograms showing the distributions of the standardized scores for the Reading CRCT and the ELA CRCT. Both illustrate a normal distribution, which is one of this study’s assumptions.
Table 4.5

*CRCT Descriptive Statistics*

<table>
<thead>
<tr>
<th>CRCT Subject</th>
<th>N</th>
<th>M</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Lower 95% CL for M</th>
<th>Upper 95% CL for M</th>
</tr>
</thead>
<tbody>
<tr>
<td>REAss</td>
<td>5,482</td>
<td>827.57</td>
<td>828.00</td>
<td>22.47</td>
<td>755.00</td>
<td>920.00</td>
<td>826.97</td>
<td>828.16</td>
</tr>
<tr>
<td>ELAss</td>
<td>5,464</td>
<td>829.53</td>
<td>829.00</td>
<td>27.59</td>
<td>739.00</td>
<td>950.00</td>
<td>828.50</td>
<td>829.97</td>
</tr>
</tbody>
</table>

Note: CL = confidence limit

*Figure 4.1:* Reading CRCT Distributions
Both the Reading CRCT and the ELA CRCT distributions can be characterized as normal; however, each has potential outliers at the high end on the right of each histogram.

**EOCT Data**

Like the CRCT data, the ethnicity data for the EOCT contained six values. Similar to the CRCT data, the ethnicity values were dominated by value 5, which represents White students who took the Ninth Grade Literature EOCT. As Table 4.6 illustrates, 78.35% of students who took the EOCT were White ninth grade students in the BCSS. This aligned with the demographics of the school system student population, where White students represent 80% of the total student population.

In addition, gender data for the EOCT were comparatively even. Males had only a slightly higher percentage represented as shown in Table 4.7.
Table 4.6

*EOCT Ethnicity Frequency Table*

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>34</td>
<td>0.67</td>
<td>34</td>
<td>0.67</td>
</tr>
<tr>
<td>2</td>
<td>546</td>
<td>10.78</td>
<td>580</td>
<td>11.45</td>
</tr>
<tr>
<td>3</td>
<td>386</td>
<td>7.62</td>
<td>966</td>
<td>19.07</td>
</tr>
<tr>
<td>4</td>
<td>23</td>
<td>0.45</td>
<td>989</td>
<td>19.52</td>
</tr>
<tr>
<td>5</td>
<td>3,969</td>
<td>78.35</td>
<td>4,958</td>
<td>97.87</td>
</tr>
<tr>
<td>6</td>
<td>108</td>
<td>2.13</td>
<td>5,066</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Table 4.7

*EOCT Gender Frequency Table*

<table>
<thead>
<tr>
<th>Gender</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>2,447</td>
<td>48.17</td>
<td>2,447</td>
<td>48.17</td>
</tr>
<tr>
<td>Male</td>
<td>2,633</td>
<td>51.83</td>
<td>5,080</td>
<td>100.00</td>
</tr>
</tbody>
</table>

In the BCSS, the four middle schools feed into three high schools. Table 4.8 shows the three high schools’ frequency distributions. Students with disabilities represented 9.68% of all students who took the EOCT, which is below the BCSS’s overall students with disabilities population (12.5%).
Table 4.8

*High School Frequency Distribution*

<table>
<thead>
<tr>
<th>School</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>School AA</td>
<td>1,122</td>
<td>22.50</td>
<td>1,122</td>
<td>22.50</td>
</tr>
<tr>
<td>School BB</td>
<td>1,784</td>
<td>35.80</td>
<td>2,906</td>
<td>58.30</td>
</tr>
<tr>
<td>School CC</td>
<td>2,079</td>
<td>41.70</td>
<td>4,985</td>
<td>100.00</td>
</tr>
</tbody>
</table>

The EOCT data included students’ scaled scores for the years 2007 through 2011 of GPS-based testing. Table 4.9 provides the EOCT data for each testing year and demonstrates that the frequency of observations increased after 2008. This is due to the staged GPS roll-out in Georgia, which added Physical Science scores in 2009 and Math scores in 2010.

Table 4.9

*EOCT Testing Year Frequency Table*

<table>
<thead>
<tr>
<th>Year</th>
<th>Frequency</th>
<th>Percent</th>
<th>Cumulative Frequency</th>
<th>Cumulative Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>627</td>
<td>12.33</td>
<td>627</td>
<td>12.33</td>
</tr>
<tr>
<td>2008</td>
<td>634</td>
<td>12.47</td>
<td>1,261</td>
<td>24.80</td>
</tr>
<tr>
<td>2009</td>
<td>989</td>
<td>19.45</td>
<td>2,250</td>
<td>44.25</td>
</tr>
<tr>
<td>2010</td>
<td>1,379</td>
<td>27.12</td>
<td>3,629</td>
<td>71.37</td>
</tr>
<tr>
<td>2011</td>
<td>1,456</td>
<td>28.63</td>
<td>5,085</td>
<td>100.00</td>
</tr>
</tbody>
</table>
This study isolated the Ninth Grade Literature EOCT data from the data file for analysis.

GaDOE sets specific cut scores for meeting standards on the test, which are as follows:
(a) below 400 does not meet standards, (b) 400 to 449 meet standards, and (c) 450 or higher exceed standards. Descriptive statistics for the Ninth Grade Literature EOCT scaled scores (LitSS) are provided in Table 4.10.

Table 4.10

EOCT Descriptive Statistics

<table>
<thead>
<tr>
<th>Subject</th>
<th>N</th>
<th>M</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Lower 95% CL</th>
<th>Upper 95% CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>LitSS</td>
<td>4,234</td>
<td>418.87</td>
<td>419.00</td>
<td>36.98</td>
<td>200.00</td>
<td>600.00</td>
<td>417.86</td>
<td>420.09</td>
</tr>
</tbody>
</table>

Note: CL = confidence limit

The histogram for the LitSS distributions is found in Figure 4.3 and illustrates a normal distribution.

![Histogram](image)

*Figure 4.3: Ninth Grade Literature EOCT Distributions*
Bivariate Analyses

Prior to conducting statistical tests of significance, a series of bivariate analysis were conducted on the variables of (a) testing year, (b) ethnicity, (c) gender, (d) school, and (e) students with disabilities. The analysis was conducted on each set of test data. For example, the Reading CRCT data were analyzed separately from the ELA CRCT data for the purpose of bivariate analysis.

CRCT Bivariate Analyses

Table 4.11 illustrates comparisons by testing year for the Reading and ELA CRCT scores.

Table 4.11

CRCT Bivariate Analyses by Test Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>Median</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>REAss</td>
<td>1,159</td>
<td>822.73</td>
<td>823.00</td>
<td>20.13</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>1,154</td>
<td>825.40</td>
<td>825.00</td>
<td>29.02</td>
</tr>
<tr>
<td>2007</td>
<td>REAss</td>
<td>1,091</td>
<td>823.89</td>
<td>825.00</td>
<td>22.67</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>1,089</td>
<td>825.31</td>
<td>826.00</td>
<td>25.97</td>
</tr>
<tr>
<td>2008</td>
<td>REAss</td>
<td>1,092</td>
<td>826.09</td>
<td>825.00</td>
<td>21.31</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>1,086</td>
<td>828.48</td>
<td>829.00</td>
<td>27.99</td>
</tr>
<tr>
<td>2009</td>
<td>REAss</td>
<td>1,053</td>
<td>830.32</td>
<td>831.00</td>
<td>22.23</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>1,051</td>
<td>831.50</td>
<td>831.00</td>
<td>25.30</td>
</tr>
<tr>
<td>2010</td>
<td>REAss</td>
<td>1,087</td>
<td>835.24</td>
<td>835.00</td>
<td>23.62</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>1,084</td>
<td>835.81</td>
<td>835.00</td>
<td>27.94</td>
</tr>
</tbody>
</table>
As indicated in Figure 4.4 and Figure 4.5, mean scaled scores for both Reading and ELA increased over the five-year testing period from 2006 through 2010.

Figure 4.4: ELA CRCT Performance 2006-2010

Figure 4.5: Reading CRCT Performance 2006-2010

An analysis of variance (ANOVA) conducted on the REAss data indicated that the change from 2008 to 2009 and from 2009 to 2010 was significant ($p < .01$). For the
ELAss, the 2009 to 2010 difference was the only increase that was statistically significant
\((p < .01)\).

In regard to ethnicity, Table 4.12 reveals slight differences between the REAss and ELAss scores.

Table 4.12

*CRCT Bivariate Analyses by Ethnicity*

<table>
<thead>
<tr>
<th>Ethnic</th>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>REAss</td>
<td>31</td>
<td>824.03</td>
<td>19.92</td>
<td>777.00</td>
<td>866.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>30</td>
<td>831.20</td>
<td>26.04</td>
<td>755.00</td>
<td>864.00</td>
</tr>
<tr>
<td>2</td>
<td>REAss</td>
<td>521</td>
<td>819.10</td>
<td>20.93</td>
<td>760.00</td>
<td>920.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>519</td>
<td>822.13</td>
<td>26.56</td>
<td>758.00</td>
<td>916.00</td>
</tr>
<tr>
<td>3</td>
<td>REAss</td>
<td>373</td>
<td>822.69</td>
<td>21.96</td>
<td>767.00</td>
<td>920.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>373</td>
<td>823.02</td>
<td>25.28</td>
<td>755.00</td>
<td>913.00</td>
</tr>
<tr>
<td>4</td>
<td>REAss</td>
<td>29</td>
<td>829.10</td>
<td>27.31</td>
<td>770.00</td>
<td>920.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>29</td>
<td>829.52</td>
<td>30.78</td>
<td>765.00</td>
<td>893.00</td>
</tr>
<tr>
<td>5</td>
<td>REAss</td>
<td>4,444</td>
<td>829.07</td>
<td>22.42</td>
<td>755.00</td>
<td>920.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>4,429</td>
<td>830.63</td>
<td>27.70</td>
<td>739.00</td>
<td>950.00</td>
</tr>
<tr>
<td>6</td>
<td>REAss</td>
<td>78</td>
<td>823.72</td>
<td>20.33</td>
<td>777.00</td>
<td>870.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>78</td>
<td>826.49</td>
<td>27.58</td>
<td>757.00</td>
<td>892.00</td>
</tr>
</tbody>
</table>

According to initial ANOVA results, several REAss had statistical differences. In the
following pairs, the higher value has an *\((p < .01)\).

6 versus 5*, 6* versus 2
5* versus 2, 5* versus 3

4* versus 2

3* versus 2

According to initial $t$ tests, statistically significant differences existed between gender scores for ELAss ($p < .01$). Yet, differences may be attributable to inflated statistical power due to the large number of observations. These data are illustrated in Table 4.13.

Table 4.13

**CRCT Bivariate Analyses by Gender**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Variable</th>
<th>$N$</th>
<th>$M$</th>
<th>$SD$</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>REAss</td>
<td>2,689</td>
<td>829.51</td>
<td>22.02</td>
<td>767.00</td>
<td>920.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>2,683</td>
<td>833.26</td>
<td>26.93</td>
<td>739.00</td>
<td>950.00</td>
</tr>
<tr>
<td>M</td>
<td>REAss</td>
<td>2,790</td>
<td>825.71</td>
<td>22.73</td>
<td>755.00</td>
<td>920.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>2,778</td>
<td>825.35</td>
<td>27.67</td>
<td>746.00</td>
<td>950.00</td>
</tr>
</tbody>
</table>

Similarly, when comparing the schools’ REAss and ELAss scores, there was little difference, as seen in Table 4.14.

Table 4.14

**CRCT Bivariate Analyses by School**

<table>
<thead>
<tr>
<th>School</th>
<th>Variable</th>
<th>$N$</th>
<th>$M$</th>
<th>$SD$</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>REAss</td>
<td>1,198</td>
<td>829.07</td>
<td>22.66</td>
<td>760.00</td>
<td>920.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>1,191</td>
<td>830.14</td>
<td>26.46</td>
<td>749.00</td>
<td>916.00</td>
</tr>
<tr>
<td></td>
<td>REAss</td>
<td>ELAss</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---------</td>
<td>---------</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>1,719</td>
<td>1,715</td>
<td>825.63</td>
<td>825.77</td>
<td>22.55</td>
<td>27.02</td>
</tr>
<tr>
<td>C</td>
<td>1,032</td>
<td>1,028</td>
<td>825.45</td>
<td>828.95</td>
<td>21.76</td>
<td>28.56</td>
</tr>
<tr>
<td>D</td>
<td>1,533</td>
<td>1,530</td>
<td>829.99</td>
<td>832.60</td>
<td>22.38</td>
<td>27.98</td>
</tr>
</tbody>
</table>

Initial ANOVA tests of REAss showed that the following pairs are significant at the $p < .05$ level. The higher value has an *.

- D* versus B
- D* versus C
- A* versus B
- A* versus C

Similar to the REAss data, the ANOVA for ELAss scores indicated several pairs were statistically significant at $p < .05$. Again, the higher value has an *.

- D* versus B
- D* versus C
- A* versus B
- C* versus B

In contrast, students with IEP status had more significant differences. IEP status represents the students with disabilities subgroup. As shown in Table 4.15, students without an IEP scored significantly higher than the students with an IEP ($p < .01$).
Table 4.15

CRCT Bivariate Analyses by IEP

<table>
<thead>
<tr>
<th>IEP</th>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>REAss</td>
<td>4,855</td>
<td>830.27</td>
<td>21.27</td>
<td>763.00</td>
<td>920.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>4,840</td>
<td>832.85</td>
<td>26.02</td>
<td>739.00</td>
<td>950.00</td>
</tr>
<tr>
<td>Yes</td>
<td>REAss</td>
<td>627</td>
<td>806.62</td>
<td>20.36</td>
<td>755.00</td>
<td>882.00</td>
</tr>
<tr>
<td></td>
<td>ELAss</td>
<td>624</td>
<td>801.20</td>
<td>22.94</td>
<td>746.00</td>
<td>868.00</td>
</tr>
</tbody>
</table>

EOCT Bivariate Analyses

Table 4.16 illustrates the comparison of the mean LitSS scores by year.

Table 4.16

EOCT Bivariate Analyses by Test Year

<table>
<thead>
<tr>
<th>Year</th>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007</td>
<td>LitSS</td>
<td>627</td>
<td>404.80</td>
<td>33.61</td>
<td>200.00</td>
<td>509.00</td>
</tr>
<tr>
<td>2008</td>
<td>LitSS</td>
<td>634</td>
<td>407.80</td>
<td>33.99</td>
<td>317.00</td>
<td>510.00</td>
</tr>
<tr>
<td>2009</td>
<td>LitSS</td>
<td>657</td>
<td>415.17</td>
<td>37.58</td>
<td>313.00</td>
<td>538.00</td>
</tr>
<tr>
<td>2010</td>
<td>LitSS</td>
<td>1,173</td>
<td>424.79</td>
<td>33.91</td>
<td>323.00</td>
<td>518.00</td>
</tr>
<tr>
<td>2011</td>
<td>LitSS</td>
<td>1,143</td>
<td>429.58</td>
<td>38.39</td>
<td>282.00</td>
<td>600.00</td>
</tr>
</tbody>
</table>

Similar to the CRCT REAss and ELAss for years 2006-2010, Figure 4.6 shows that LitSS scores increased for the testing years 2007-2011, especially from 2008-2011.
Figure 4.6: Ninth Grade Literature EOCT Performance 2007-2011

In Table 4.17, the LitSS scores are compared by ethnicity, and the statistical analyses illustrated in Table 4.18 are comparisons by gender. As with the gender CRCT scores, initial t tests comparing gender on the LitSS scores revealed differences that were statistically significant ($p < .01$). Yet, sample sizes are large, and the difference may be a result of inflated statistical power.

Table 4.17

EOCT Bivariate Analyses by Ethnicity

<table>
<thead>
<tr>
<th>Ethnic</th>
<th>Variable</th>
<th>N</th>
<th>$M$</th>
<th>$SD$</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LitSS</td>
<td>24</td>
<td>419.13</td>
<td>31.40</td>
<td>359.00</td>
<td>472.00</td>
</tr>
<tr>
<td>2</td>
<td>LitSS</td>
<td>434</td>
<td>407.36</td>
<td>31.96</td>
<td>317.00</td>
<td>543.00</td>
</tr>
<tr>
<td>3</td>
<td>LitSS</td>
<td>329</td>
<td>411.08</td>
<td>35.82</td>
<td>317.00</td>
<td>543.00</td>
</tr>
<tr>
<td>4</td>
<td>LitSS</td>
<td>20</td>
<td>420.45</td>
<td>38.82</td>
<td>324.00</td>
<td>493.00</td>
</tr>
</tbody>
</table>
Several pairs of scores are statistically significant at the $p < .05$ level, with the higher value noted with an *.

5* versus 6
5* versus 3
5* versus 2

Table 4.18

**EOCT Bivariate Analyses by Gender**

<table>
<thead>
<tr>
<th>Gender</th>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>LitSS</td>
<td>2,026</td>
<td>423.58</td>
<td>36.22</td>
<td>317.00</td>
<td>600.00</td>
</tr>
<tr>
<td>M</td>
<td>LitSS</td>
<td>2,203</td>
<td>414.73</td>
<td>37.20</td>
<td>200.00</td>
<td>435.00</td>
</tr>
</tbody>
</table>

When comparing LitSS by school, as illustrated in Table 4.19, initial ANOVA testing found that School CC is significantly different from the other two high schools.

Table 4.19

**EOCT Bivariate Analyses by School**

<table>
<thead>
<tr>
<th>School</th>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>LitSS</td>
<td>980</td>
<td>416.10</td>
<td>37.49</td>
<td>200.00</td>
<td>600.00</td>
</tr>
<tr>
<td>BB</td>
<td>LitSS</td>
<td>1,462</td>
<td>415.33</td>
<td>37.09</td>
<td>302.00</td>
<td>543.00</td>
</tr>
<tr>
<td>CC</td>
<td>LitSS</td>
<td>1,792</td>
<td>423.53</td>
<td>36.15</td>
<td>282.00</td>
<td>543.00</td>
</tr>
</tbody>
</table>
The EOCT bivariate analyses also showed differences based on students’ IEP status. Similar to the CRCT scores, non-IEP students scored significantly higher in the EOCT LitSS scores \((p < .01)\). The \(t\) tests confirmed the difference, and the results can be found in Table 4.20.

Table 4.20.

**EOCT Bivariate Analyses by IEP**

<table>
<thead>
<tr>
<th>IEP</th>
<th>Variable</th>
<th>(N)</th>
<th>(M)</th>
<th>(SD)</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>LitSS</td>
<td>3,798</td>
<td>422.91</td>
<td>35.73</td>
<td>200.00</td>
<td>600.00</td>
</tr>
<tr>
<td>Yes</td>
<td>LitSS</td>
<td>436</td>
<td>384.74</td>
<td>29.25</td>
<td>282.00</td>
<td>484.00</td>
</tr>
</tbody>
</table>

**Correlational Analyses**

Correlational analyses determined relationships between the eighth grade CRCT scores and the Ninth Grade Literature EOCT scores. All standardized test data were exported into BASE SAS 9.2 software for analysis. Correlational analyses were conducted to determine Pearson’s product-moment coefficients (Pearson’s \(r\)). These analyses were only conducted on students who took a CRCT in their eighth grade year and also took the EOCT the following year as ninth graders. These data were discussed by paired testing years. For example, eighth graders who took the Reading CRCT in 2008 and then took the Ninth Grade Literature EOCT the following year will comprise one paired group. There will be five paired groups of Reading CRCT to Ninth Grade Literature EOCT (REAss/LitSS) and five paired groups of ELA CRCT to Ninth Grade Literature EOCT (ELAss/LitSS). Utilizing these paired scores reduced sample sizes. The sample sizes were still sufficient, which is one of this study’s assumptions. The
discussion will be framed using the three research questions that support this study.

**Research Question 1**

*What is the relationship between students’ achievement on the eighth grade Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT?* The hypothesis for Research Question 1 stated that a statistically significant correlation will exist between eighth grade students’ achievement on the Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT. As well, the null hypothesis ($H_{01}$) stated that there will be no significant correlation between eighth grade students’ achievement on the Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

**Evidence for relationships.** Reading CRCT and Ninth Grade Literature EOCT scores were analyzed by the paired years of testing. The scores of students who took the CRCT as eighth graders, and who also took the EOCT as ninth graders, were paired to determine any correlation. The Pearson’s correlation coefficient analysis was conducted for each paired testing year to determine if a correlation existed. The results of the analyses are provided in Table 4.21.

Table 4.21

*Reading CRCT to Literature EOCT Correlations by Year*

<table>
<thead>
<tr>
<th>Paired Year</th>
<th>N</th>
<th>Pearson’s r</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006 REAss/2007 LitSS</td>
<td>1,129/629/539</td>
<td>.664</td>
<td>$p &lt; .0001$</td>
</tr>
<tr>
<td>2007 REAss/2008 LitSS</td>
<td>1,159/629/539</td>
<td>.664</td>
<td>$p &lt; .0001$</td>
</tr>
<tr>
<td>2008 REAss/2009 LitSS</td>
<td>1,094/657/570</td>
<td>.729</td>
<td>$p &lt; .0001$</td>
</tr>
<tr>
<td>2009 REAss/2010 LitSS</td>
<td>1,078/1,173/943</td>
<td>.751</td>
<td>$p &lt; .0001$</td>
</tr>
</tbody>
</table>
Notes: (1) Prob > |r| under H0: Rho = 0. (2) N = REAss scores/LitSS scores/number of paired scores of students who took both the CRCT in 8th grade and the EOCT in 9th grade.

Findings. As shown in Table 4.21, the correlation between the REAss and the LitSS are fairly strong, with the correlation coefficients ranging from .664 to .765. Each correlation coefficient is also highly significant at $p < .0001$. Because each correlation is strong and highly significant, $H_{01}$ can be rejected and the research hypothesis accepted as follows: A statistically significant correlation exists between eighth grade students’ achievement on the Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

Research Question 2

What is the relationship between students’ achievement on the eighth grade ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT? The hypothesis for Research Question 2 stated that a statistically significant correlation will exist between eighth grade students’ achievement on the ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT. The null hypothesis ($H_{02}$) stated that there will be no significant correlation between eighth grade students’ achievement on the ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

Evidence for relationships. The analyzed data were only from students who took the eighth grade ELA CRCT and then took the Ninth Grade Literature EOCT the following year. This limited the number of participants, but the number of observations was still sufficient as seen in Table 4.22, which provides the correlation coefficients for the ELAss/LitSS analyses.
Table 4.22

<table>
<thead>
<tr>
<th>Paired Year</th>
<th>N</th>
<th>Pearson’s r</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006 ELAss/2007 LitSS</td>
<td>1,154/629/539</td>
<td>.729</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>2007 ELAss/2008 LitSS</td>
<td>1,154/629/539</td>
<td>.729</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>2008 ELAss/2009 LitSS</td>
<td>1,088/657/570</td>
<td>.745</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>2009 ELAss/2010 LitSS</td>
<td>1,076/1,173/942</td>
<td>.737</td>
<td>p &lt; .0001</td>
</tr>
<tr>
<td>2010 ELAss/2011 LitSS</td>
<td>1,087/1,143/938</td>
<td>.778</td>
<td>p &lt; .0001</td>
</tr>
</tbody>
</table>

Notes: (1) Prob > |r| under H0: Rho = 0. (2) N = ELAss/LitSS/number of paired scores of students who took both the CRCT in 8th grade and the EOCT in 9th grade.

**Findings.** As Table 4.22 shows, Pearson’s correlation coefficients range from .729 to .778 across the paired testing years, providing evidence of strong correlations between the ELAss and the LitSS. Each correlation coefficient is also highly significant at p < .0001. Because each paired testing correlation coefficient shows strong and highly significant relationships, then H_02 can be rejected, and the research hypothesis that a statistically significant correlation exists between eighth grade students’ achievement on the ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT can be accepted.

**Regression Analysis**

**Research Question 3**

If a relationship exists, what is the predictive value of students’ eighth grade achievement as measured by the Reading or ELA CRCT on their achievement on the Ninth Grade Literature EOCT? The hypothesis for Research Question 3 stated that
eighth grade performance for the same subject area CRCT will be predictive of ninth grade performance on the respective EOCT. The null hypothesis (H₀) stated that there will be no predictive value between eighth grade performance for the same subject area CRCT and the ninth grade performance on the respective EOCT. An Ordinary Least Square (OLS) Regression was conducted on REAss/LitSS observations and on ELAss/Litss observations for paired years of testing. An OLS is appropriate because the criterion variable and both predictor variables are continuous, and the relationship between them is linear (Gall, Gall, & Borg, 2007). Linearity is evidenced in Figures 4.7 and 4.8.

**Assumption Testing**

This study utilized regression analysis to determine if predictive value existed from the CRCT scores to the EOCT scores. Several assumptions underlie regression analysis. First, sample sizes are sufficient. According to Gall, Gall, and Borg (2007), correlational research should have a minimum of 30 participants. The descriptive statistics provided evidence that sample sizes were sufficient for the research study, with 5,495 CRCT test-takers and 5,085 EOCT test-takers between 2006 and 2011. When conducting correlational analyses for paired testing years, sample sizes were reduced but still sufficient. As shown in Table 4.4, no paired testing year data for all subject area EOCTs had fewer than 1,053 observations. Similarly, Table 4.21 and Table 4.22 showed that there were no fewer than 539 paired REAss/LitSS or ELAss/LitSS scaled scores.

Another assumption of regression analysis is that variables are normally distributed. This is evidenced in Figures 4.1, 4.2, and 4.3, which show a characteristic bell-curved shape. Outliers, which can represent errors in the data, are present at the positive ends of both the Reading and ELA distributions. The values in the dataset were
determined to be the scores of the few students who scored at the highest levels on the assessments, but the few outliers were not excluded from the data analysis due to the significant number of total observations.

The next assumption underlying regression analysis for this study is that there is a linear relationship between the independent and dependent variables, where the regression line’s best fit is a straight line. By using scatterplots of data observations, the straight regression line is illustrated in Figures 4.7 and 4.8, which show the regression lines for the observations of REAss/LitSS and ELAss/LitSS for each paired year of testing are straight lines.

Also, this study assumes that variables are measured reliably and are error-free. The GaDOE provides reliability data for the CRCT and the EOCT, which shows them to be valid and reliable instruments (“CRCT,” 2011; “EOCT,” 2011). GaDOE scaled score data were imported into the BASE SAS 9.2 statistical software program rather than entered by hand to control for data entry error, and the program was utilized to conduct all data analyses. In this study, several years of paired tests and sufficient numbers in each paired year help support reliable and error-free measurement.

A final assumption is that the data distributions have the same variance of errors. Gall, Gall, and Borg (2007) stated, “Reliability coefficients should be sufficient for making a decision to select a particular test if you keep in mind that no single type of reliability coefficient can isolate all the possible sources of systematic measurement error” (p. 203). Further, Gall, Gall, and Borg concluded that the greater the value of the coefficient, the lower the standard error of measurement. Pearson’s product-moment coefficients established the degree of reliability to account for error variance.
Evidence for relationships in REAss/LitSS. An OLS Regression was conducted for each paired REAss/LitSS as illustrated in Table 4.23. Additionally, the REAss/LitSS OLS Regression results indicate the following prediction models, listed by year:

- **2006-2007**: 9th grade LitSS = -543 + 1.157 * 8th grade REAss
- **2007-2008**: 9th grade LitSS = -479 + 1.081 * 8th grade REAss
- **2008-2009**: 9th grade LitSS = -593 + 1.228 * 8th grade REAss
- **2009-2010**: 9th grade LitSS = -474 + 1.085 * 8th grade REAss
- **2010-2011**: 9th grade LitSS = -584 + 1.217 * 8th grade REAss

To illustrate the prediction models’ importance to this study, consider the formula for 2006-2007. For every 1 point increase in a student’s score on the 8th grade Reading CRCT, the student would expect to increase his or her Ninth Grade Literature EOCT score by 1.157 points. This increase was consistent in all five paired years of testing.

Table 4.23

**REAss/LitSS Parameter Estimates**

<table>
<thead>
<tr>
<th>Paired Years</th>
<th>df</th>
<th>Parameter</th>
<th>Estimate</th>
<th>SE</th>
<th>t value</th>
<th>Pr &gt;</th>
<th>95% CL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2006-2007</td>
<td></td>
<td>Intercept</td>
<td>-543.103</td>
<td>43.022</td>
<td>-12.62</td>
<td>&lt;.0001</td>
<td>-627.616 -458.590</td>
</tr>
<tr>
<td></td>
<td></td>
<td>REAss</td>
<td>1.157</td>
<td>0.052</td>
<td>22.06</td>
<td>&lt;.0001</td>
<td>1.054 1.260</td>
</tr>
<tr>
<td>2007-2008</td>
<td></td>
<td>Intercept</td>
<td>-479.652</td>
<td>36.432</td>
<td>-13.17</td>
<td>&lt;.0001</td>
<td>-551.222 -408.082</td>
</tr>
<tr>
<td></td>
<td></td>
<td>REAss</td>
<td>1.081</td>
<td>0.044</td>
<td>24.40</td>
<td>&lt;.0001</td>
<td>0.994 1.168</td>
</tr>
<tr>
<td>Paired Years</td>
<td>df</td>
<td>Parameter Estimate</td>
<td>SE</td>
<td>t value</td>
<td>Pr &gt;</td>
<td>t</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>----</td>
<td>--------------------</td>
<td>-----</td>
<td>---------</td>
<td>------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.882</td>
<td>0.034</td>
<td>26.33</td>
<td>&lt;.0001</td>
<td>0.816</td>
<td>0.948</td>
</tr>
<tr>
<td>REAss</td>
<td>1</td>
<td>0.938</td>
<td>0.039</td>
<td>24.28</td>
<td>&lt;.0001</td>
<td>0.862</td>
<td>1.014</td>
</tr>
<tr>
<td>2007-2008</td>
<td></td>
<td>-364.284</td>
<td>31.862</td>
<td>-11.43</td>
<td>&lt;.0001</td>
<td>-426.875</td>
<td>-301.693</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.938</td>
<td>0.039</td>
<td>24.28</td>
<td>&lt;.0001</td>
<td>0.862</td>
<td>1.014</td>
</tr>
<tr>
<td>REAss</td>
<td>1</td>
<td>0.938</td>
<td>0.039</td>
<td>24.28</td>
<td>&lt;.0001</td>
<td>0.862</td>
<td>1.014</td>
</tr>
</tbody>
</table>

Note: CL = confidence limit

**Evidence for Relationships in ELAss/LitSS.**  An OLS Regression analysis was conducted on the paired ELAss and LitSS. Table 4.24 provides the statistics for the regression analysis.

Table 4.24

**ELAss/LitSS Parameter Estimates**
The regression analysis results indicated the following prediction models, listed by paired testing years:

2006-2007: 9th grade LitSS = -320 + 0.882*8th grade ELAss
2007-2008: 9th grade LitSS = -364 + 0.938*8th grade ELAss
2008-2009: 9th grade LitSS = -372 + 0.957*8th grade ELAss
2009-2010: 9th grade LitSS = -364 + 0.952*8th grade ELAss
2010-2011: 9th grade LitSS = -447 + 1.052*8th grade ELAss

As with the REAss/LitSS predictions models, the ELAss/LitSS models were important to this study. To understand the prediction models, consider the model for the 2006-2007 testing year: 9th grade LitSS = -320 + 0.882*8th grade ELAss. During the 2006-2007 paired testing years, for every 1 point increase in a student’s eighth grade ELA CRCT score, the student could expect an increase of 0.882 points on his or her Ninth Grade Literature EOCT score. As the models demonstrate, the increase can be expected in each of the five paired testing years.
Goodness of Fit

Utilizing the OLS Regression, the scatterplots for each testing year were created. Figure 4.7 provides the scatterplot for each of the five paired testing years for REAss/LitSS. In each REAss/LitSS scatterplot, the positive correlation is evident, and the regression line of best fit is a straight line. One important statistic provided in each scatterplot is $r^2$, which by paired testing year are (a) 2006-2007, .4759; (b) 2007-2008, .5291; (c) 2008-2009, .5362; (d) 2009-2010, .5465; and (e) 2010-2011, .5855. Therefore, in 2010-2011, 58.55% of the variability in the Ninth Grade Literature scores can be captured by the prediction model, and the remaining 41.45% of the variability is attributed to other factors.
Figure 4.7: Scatterplots for REAss/LitSS OLS Regressions

The OLS Regression analysis yielded scatterplots for each of the five paired years of ELAss/LitSS data. Figure 4.8 provides the five scatterplots for ELAss/LitSS observations.
Figure 4.8: Scatterplots for ELAss/LitSS OLS Regressions

As with REAss/LitSS, each ELAss/LitSS scatterplot shows a positive correlation, with the regression line of best fit being a straight line. The $r^2$ of each prediction model, by paired testing year, are (a) 2006-2007, .5639; (b) 2007-2008, .5266; (c) 2008-2009, .5555; (d) 2009-2010, .5435; and (e) 2010-2011, .6046. For example, the prediction model in 2010-2011 captures 60.46% of the variability of Ninth Grade Literature scores, and the other 39.54% of the variability is attributed to other factors that are not included in the model.

Findings

In each prediction model for the five paired years of REAss/LitSS and ELAss/Litss, for every one point increase in the CRCT score in Reading or ELA, an increase ranging from 0.882 to 1.228 can be expected on the EOCT score in Ninth Grade Literature. The stated increase was consistent in every paired testing year. Thus, $H_{03}$ can be rejected, and the research hypothesis can be accepted that predictive value exists between eighth grade performance on the Reading and ELA CRCT and the ninth grade performance on the Ninth Grade Literature EOCT.

Disaggregated Data

Data were disaggregated by IEP status for students with disabilities (IEP and non-IEP), by ethnicity (White and non-White), and by gender. In order to assess for
correlations between the REAss/LitSS and ELAss/LitSS, analyses were conducted to
determine Pearson’s $r$ and assess for correlations between REAss/LitSS and for
ELAss/LitSS. As with the aggregate data, students were only included if they took the
Reading CRCT in eighth grade and took the Ninth Grade Literature EOCT in their next
school year or took the ELA CRCT in eighth grade and took the Ninth Grade Literature
EOCT the following school year.

Students with disabilities were identified by their IEP status, which meant that
they had an IEP and received special education services during the school year. Table
4.25 provides the correlation coefficients for REAss/LitSS and ELAss/LitSS for each
paired testing year. Relationships were significant in each paired testing year for both
Reading and ELA IEP scores, although slightly less so for the IEP data observations
compared to Non-IEP observations.

Table 4.25

Correlations by IEP Status

<table>
<thead>
<tr>
<th>Paired Year</th>
<th>REAss/LitSS</th>
<th>ELAss/LitSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson’s $r$</td>
<td>Pearson’s $r$</td>
</tr>
<tr>
<td></td>
<td>IEP</td>
<td>Non-IEP</td>
</tr>
<tr>
<td>2006/2007</td>
<td>.569</td>
<td>.640</td>
</tr>
<tr>
<td>2009/2010</td>
<td>.590</td>
<td>.735</td>
</tr>
<tr>
<td>2010/2011</td>
<td>.711</td>
<td>.735</td>
</tr>
</tbody>
</table>

Notes: (1) Prob $> \mid r \mid$ under H0: Rho = 0; (2) $p < .0001$ significance level for each
correlation.
Subsequent analysis revealed that IEP status affected predictive regression models in every paired testing year for both REAss/LitSS and ELAss/LitSS. The prediction models are provided in Table 4.26. To illustrate how meaningful the prediction model is for this study, refer to the 2006-2007 prediction model for REAss/LitSS: 9th grade LitSS = -494 + (1.099*8th grade REAss) – (11.580*IEP). For every 1 point increase in a student’s Reading CRCT score, a student could expect a 1.099 increase in the Ninth Grade Literature EOCT score. However, for a student served by an IEP, the Ninth Grade Literature EOCT score can be expected to decrease by 11.580 points. In each prediction model in Table 4.26, an increase can be expected for aggregate scores. When disaggregated by IEP status, students served by an IEP evidenced a decrease in every paired testing year for both REAss/LitSS and ELAss/LitSS ranging from -6.15 to -14.46 points. This outcome was logical and expected based on bivariate analysis results.

Table 4.26

IEP Prediction Models

| Paired Years | 9th grade LitSS = | 8th grade REAss \* | (1.099*8th grade REAss) – (11.58*IEP) |
|--------------|-------------------|-------------------|
| 2006-2007    | -494              | 1.099             |
|               | 9th grade LitSS = | (.850*8th grade ELAss) – (8.710*IEP) |
| 2007-2008    | -431              | 1.024             |
|               | 9th grade LitSS = | (.899*8th grade ELAss) – (9.603*IEP) |
| 2008-2009    | -561              | 1.190             |
|               | 9th grade LitSS = | (.933*8th grade ELAss) – (6.150*IEP) |
| 2009-2010    | -441              | 1.046             |
|               | 9th grade LitSS = | (10.720*IEP)     |

106
9th grade LitSS = -334 + (.918*8th grade ELAss) – (10.320*IEP)

2010-2011  9th grade LitSS = -540 + (1.165 *8th grade REAss) – (12.370*IEP)

9th grade LitSS = -418 + (1.018*8th grade ELAss) – (8.810*IEP)

Yet, data analysis disaggregated by gender is not as clear cut. Correlations vary slightly as shown in Table 4.27. Additionally, when regression analyses were conducted, gender had no effect on prediction models in all paired test years for ELAss/LitSS and in three paired test years for REAss/LitSS. The $p$-value associated with the gender variable was $>.1$, with $p$-values ranging from .11 to .99 on the eight paired years where no effect was noted. Gender had a significant effect in all five of the ELAss/LitSS paired testing years but in only two of the five paired testing years in REAss/LitSS. The prediction models where an effect was identified for the paired REAss/LitSS years can be found in Table 4.28.

Table 4.27

_Correlations by Gender_

<table>
<thead>
<tr>
<th>Paired Year</th>
<th>REAss/LitSS</th>
<th>ELAss/LitSS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pearson’s $r$</td>
<td>Pearson’s $r$</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>Female</td>
</tr>
<tr>
<td>2006/2007</td>
<td>.611</td>
<td>.723</td>
</tr>
<tr>
<td>2007/2008</td>
<td>.683</td>
<td>.776</td>
</tr>
<tr>
<td>2008/2009</td>
<td>.718</td>
<td>.745</td>
</tr>
</tbody>
</table>

Notes: (1) $\text{Prob} > |r|$ under H0: Rho = 0; (2) $p < .0001$ significance level for each
correlation.

Table 4.28

*Gender Prediction Models*

<table>
<thead>
<tr>
<th>Paired Years</th>
<th>Prediction Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>2007-2008</td>
<td>9th grade LitSS = -476 + (1.074<em>8th grade REAss) + (4.880</em>gender)</td>
</tr>
<tr>
<td>2010-2011</td>
<td>9th grade LitSS = -579 + (1.210<em>8th grade REAss) + (4.690</em>gender)</td>
</tr>
</tbody>
</table>

For example, in the 2007-2008 paired testing years, for every 1-point increase in a student’s Reading CRCT score, an increase of 1.1 points could be expected on the Ninth Grade Literature EOCT. For females, though, an increase of 4.880 points could be expected. Likewise, in 2010-2011, female students could expect an increase of 4.690 points.

Similarly, when disaggregating data based on ethnicity, the data for only two of the 10 years demonstrated a significant effect. Data were analyzed by a numerical ethnicity value of 1 through 7. Students in the researched school district are 80% White, represented by the value 5 in the data. Data were analyzed for White (value 5) and Non-White (all others), and results can be found in Table 4.29. As results indicated, correlations varied only slightly based on ethnicity and are strongly significant in both groups of students.

Table 4.29

*Correlations by Ethnicity*

<table>
<thead>
<tr>
<th>REAss/LitSS</th>
<th>ELAss/LitSS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paired Year</td>
<td>Pearson’s r</td>
</tr>
</tbody>
</table>

108
Ethnicity had no effect on prediction models in all five paired testing years for REAss/LitSS and in three of the five ELAss/LitSS paired testing years. The p-value associated with the ethnicity variable was > .1, with p-values ranging from .17 to .99 on the eight paired years where no effect was noted. A significant effect based on ethnicity was found in only two paired testing years for ELAss/LitSS, 2009-2010 and 2010-2011. The prediction models for the paired ELAss/LitSS years where an effect was identified can be found in Table 4.30.

Table 4.30

<table>
<thead>
<tr>
<th>Paired Years</th>
<th>Prediction Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009-2010</td>
<td>9th grade LitSS = -359 + (.947<em>8th grade ELAss) - (3.910</em>ethnicity)</td>
</tr>
<tr>
<td>2010-2011</td>
<td>9th grade LitSS = -440 + (1.040 <em>8th grade ELAss) - (5.070</em>ethnicity)</td>
</tr>
</tbody>
</table>

These data demonstrate that in 2009-2010, for every 1-point increase in a student’s ELA CRCT score, an increase of .947 points could be expected on the Ninth Grade Literature EOCT. For Non-White students, though, a decrease of 3.910 points could be expected.
Likewise, in 2010-2011, Non-White students could expect a decrease of 5.070 points.

**School-Level Influences on Standardized Testing Achievement**

As part of this study’s data collection, a form requesting input on standardized testing factors was sent to secondary administrators in the researched school system (see Appendix B). Forms were sent to the school system’s administrator for secondary schools and to the seven principals at the middle schools and high schools. Six of the eight forms were returned. One middle school (School A) and the system administrator did not return the form.

The administrators were asked to provide input on Reading/ELA interventions implemented in their schools over the last three years that they believed impacted Reading/ELA student achievement. They were asked to list each intervention and then place a number rating from 1 (low) to 5 (high) beside it to indicate the degree to which the intervention impacted student achievement on the Reading or ELA CRCT or the Ninth Grade Literature EOCT.

The middle school administrators’ responses revealed differences among purchased programs used as interventions. Two schools (C and D) used *Writing Destinations*, and both rated it a 3. School B identified advanced classes as an intervention and rated it a 2. School C was the only school to list *Writing to Win*, which it rated a 4. School D rated *Corrective Reading* as a 3. In the researched district, all middle schools are Title I schools, and certain reading programs that focused on at-risk, economically disadvantaged students were implemented at the three responding middle schools. Each middle school utilized *Read to Achieve* and *Read 180* for Title I students. *Read to Achieve* was rated 2, 4, and 3 by School B, School C, and School D, respectively. *Read 180* was rated 4, 4, and 5 by the respective schools, demonstrating the
administrators’ perceptions of its impact on students’ scores.

With the exception of their perceptions of Read 180, middle school administrators showed variations in their ratings of interventions and differences in the interventions used. Standardized test results from each school did not evidence significant gains. Table 4.14 provided results that show little difference in student performance on the Reading CRCT and the ELA CRCT across the five years of test data. As Table 4.14 shows, the REAss mean varied from 825.45 to 829.99, with the standard deviation computed from 21.76 to 22.66. The means of the ELAss ranged from 825.77 to 832.60, and the standard deviation from 26.46 to 28.56.

The three high schools demonstrated less consistency among their interventions than the middle schools. Each school enrolled students in Ninth Grade Literature. School BB and School CC also enrolled low-performing eighth graders into another course called Literary Types as ninth graders to support them in their high school literature class. School BB rated the support course as a 5. School CC rated it at 3 but noted the importance of two reading comprehension strategies explicitly taught in the Literary Types course. School CC believed that the RAP Paraphrasing Strategy and the Independent Reading/Reader’s Response Strategy had the most impact on the achievement of their at-risk students. Another intervention only at School BB was a daily, separate block of school-wide literacy time, which was rated as a 2.

Additionally, School AA and School CC listed practices and strategies within the Ninth Grade Literature course that addressed Literature achievement. Both schools used regular EOCT practice, which School AA rated at 3 and School CC at 5. They both listed a focus on standards, rating it 3 and 5, respectively. Additionally, they both identified genre-focused units, rating the intervention at 3 and 5, respectively, and
teaching vocabulary in context, which they rated at 4 and 5.

While the three high schools had diverse intervention strategies, there was still only slight variation in the Ninth Grade Literature EOCT achievement scores by school. Table 4.19 shows the mean LitSS of School AA as 416.10, School BB as 415.33, and School CC as 423.53, with respective standard deviations of 37.49, 37.09, and 36.15. An ANOVA showed that School CC differed significantly \( (p < .01) \) from the other two high schools in LitSS. School CC believed two specific strategies had impact on their students’ achievement: the RAP Paraphrasing Strategy and the Independent Reading/Reader’s Response Strategy. It is unclear if these interventions may have increased scores, but it could receive further consideration.

**Summary**

Analyses were conducted on eighth grade Reading and ELA CRCT data and on Ninth Grade Literature EOCT data. Descriptive statistics, both univariate and bivariate, indicated that the sample sizes were sufficient, the distributions were normal, and there were few outliers. Next, correlation analyses were performed to determine the strength of relationships between (a) eighth grade Reading CRCT scores and Ninth Grade Literature EOCT scores and (b) eighth grade ELA CRCT scores and Ninth Grade Literature EOCT scores.

All of the Pearson’s correlation coefficients were significant, with strong positive correlations ranging from .66 to .78. The results of the correlation analyses were used to develop OLS Regression models and used the eighth grade CRCT score as the single predictor for the EOCT score. In all paired testing years, the models were strong, with \( r^2 \) values that ranged from .4759 to .6046.

Data were then disaggregated by IEP status, gender, and ethnicity. IEP status was
found to have an effect on scores. Prediction models indicated that students with IEPs could expect their scores on the EOCT to decrease from their scores on either the Reading CRCT or the ELA CRCT. Gender had no effect in three of the five ELAss/LitSS paired years. Yet, in the 2007-2008 and 2010-2011 paired REAss/LitSS, females increased their achievement on the Ninth Grade Literature EOCT over males by almost five points (4.88 in 2007-2008 and 4.69 in 2010-2011). Ethnicity showed an effect on scores in only two of the 10 paired testing years as well. Ethnicity effects were noted in the 2009-2010 and 2010-2011 ELAss/LitSS paired testing years. Non-White students were predicted to score lower than White students by 3.91 points and 5.07 points, respectively.

With gender and ethnicity, two effects each may not represent a trend, even with effects in two consecutive years for ethnicity, but they warrant further consideration. According to Gall, Gall, and Borg (2007), identification of trends in data has to occur over a period of years, and this study only shows effects in two paired testing years for gender and two paired testing years for ethnicity. Educators in BCSS need to watch the achievement of these subgroups in case the effects continue and represent a trend.

Based on the correlational analyses that indicated significant correlations for all paired testing years for REAss/LitSS and ELAss/LitSS, the null hypotheses for Research Questions 1 and 2 were rejected and research hypotheses 1 and 2 accepted. The data indicated that there was a strong correlation between students’ scores on the eighth grade CRCT in Reading and ELA and their scores on the Ninth Grade Literature EOCT. Furthermore, the OLS Regression prediction models indicated significant and strong predictive value of the eighth grade Reading and ELA CRCT scores on Ninth Grade Literature EOCT scores. Thus, the null hypothesis for Research Question 3 was rejected
and the research hypothesis accepted that eighth grade performance for the Reading and ELA CRCT were predictive of ninth grade performance on the Ninth Grade Literature EOCT.

Analyses of the CRCT and EOCT standardized test data are the foundation for answering the three research questions and a catalyst for further deliberation. School administrators’ input on school level factors that they believed impacted test performance are also important information to inform implications for classroom practices and future research. Additionally, research trends must be considered as implications of this study and future research are discussed.
CHAPTER 5: DISCUSSION

High-stakes, standardized testing is a fact in American public education and is an important accountability factor for federal education legislation. NCLB is the law until any reauthorization, and President Obama has asked that it be reauthorized with a continuing focus on standardized-testing accountability (Klein & McNeil, 2010). In Georgia, students’ scores on the ELA and Math portions of the GHSGT have been the AYP measurement for high schools. Beginning in the 2011-2012 school year, though, the GHSGT is being phased out in favor of EOCTs in specific content areas. Middle school students must demonstrate proficiency on the CRCT.

The problem for Georgia educators is that they do not know if there is any relationship between CRCT performance and EOCT performance. The purpose of this research study was to determine if there was any correlation in or predictive value of students’ achievement on state-mandated standardized testing from the CRCT in their eighth grade year to the EOCT at the end of their ninth grade year. This study sought to answer three questions about the relationship between CRCT and EOCT performance.

**Research question 1.** What is the relationship between students’ achievement on the eighth grade Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT?

**Research question 2.** What is the relationship between students’ achievement on the eighth grade ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT?
Research question 3. : If a relationship exists, what is the predictive value of students’ eighth grade achievement as measured by the Reading or ELA CRCT on their achievement on the Ninth Grade Literature EOCT?

In addition to research questions, research and null hypotheses were developed for each research question.

Research hypothesis 1. A statistically significant correlation will exist between eighth grade students’ achievement on the Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT. $H_{01}$: There will be no significant correlation between eighth grade students’ achievement on the Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

Research hypothesis 2. A statistically significant correlation will exist between eighth grade students’ achievement on the ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT. $H_{02}$: There will be no significant correlation between eighth grade students’ achievement on the ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

Research hypothesis 3. Students’ eighth grade performance on the Reading CRCT or the ELA CRCT will be predictive of the same students’ performance on the Ninth Grade Literature EOCT. $H_{03}$: There will be no predictive value between students’ eighth grade performance on the Reading or ELA CRCT and the same students’ performance on the Ninth Grade Literature EOCT.

Summary of Findings

Reading CRCT and ELA CRCT data, as well as Ninth Grade Literature EOCT data, were entered into the BASE SAS 9.2 statistical software program for analysis, which included initial descriptive statistics and ANOVA. Both aggregate and
disaggregate data were studied, with a focus on the following subgroups: (a) gender, (b) ethnicity, (c) school, and (d) students with disabilities. The data analysis utilized Pearson’s $r$ to determine whether significant relationships existed. The analysis also included OLS Regression to identify any predictive value of CRCT performance on EOCT performance.

**Findings for Research Question 1**

The correlation between the REAss and the LitSS were strong, with the correlation coefficients ranging from .66 to .77. Each correlation coefficient was also highly significant at $p < .0001$. Because each correlation was strong and highly significant, $H_{01}$ was rejected and the research hypothesis accepted that a statistically significant correlation exists between eighth grade students’ achievement on the Reading CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

**Findings for Research Question 2**

Pearson’s correlation coefficients ranged from .73 to .78 across the paired testing years, providing evidence of strong correlations between the ELAss and the LitSS. Each correlation coefficient was also highly significant at $p < .0001$. Because each paired testing correlation coefficient showed strong and highly significant relationships, then $H_{02}$ was rejected and the research hypothesis accepted that a statistically significant correlation exists between eighth grade students’ achievement on the ELA CRCT and the same students’ achievement on the Ninth Grade Literature EOCT.

**Findings for Research Question 3**

Each OLS Regression generated a prediction model for the five paired years of REAss/LitSS and ELAss/LitSS. For every one point increase in the CRCT score in Reading or ELA, the model showed that an increase ranging from 0.882 to 1.228 could
be expected on the EOCT score in Ninth Grade Literature. The expected increase held true in every paired testing year. Thus, $H_{03}$ was rejected and the research hypothesis accepted that predictive value exists between eighth grade performance on the Reading and ELA CRCT and the ninth grade performance on the Ninth Grade Literature EOCT.

**Discussion**

As Lauer et al. (2005) stated, standardized tests matter to educators. Many researchers have concluded that high-stakes, standardized testing has mattered throughout history and continues to matter in the modern era of accountability and school reform (Au, 2007; Chiang, 2009; Cizek, 2005; Dee & Jacobs, 2010; Harris & Harrington, 2006; Madaus & Russell, 2010/2011; Mintrop & Sunderman, 2009; Moon, Jarvis, Brighton, & Hall, 2007; Supovitz, 2009; Wiliam, 2010). NCLB’s purpose, which is still federal law, is to hold schools, local school systems, and states accountable for improving all students’ academic achievement (“No Child,” 2001). NCLB, which utilizes standardized testing data as its measure for success, ensures a focus on accountability for results and improvements in public school education (Klein & McNeil, 2010; Solley, 2007).

Clearly, standardized testing has its critics and backers. Cizek (2005) found that testing opponents believe that tests (a) increase teacher stress, frustration, and burnout; (b) increase drop-out rates in high schools; (c) increase students’ stress and stress-related illnesses; (d) narrow the curriculum; (e) cannot measure higher-order thinking skills; (f) widen the achievement gap; (g) are biased; and (h) promote cheating. Proponents of high-stakes testing claim that criticisms of testing are not research-based but are only the opinions of policymakers, commentators, critics, parents, and even some educators (Cizek, 2005; Geisinger, 2005; Goodman & Hambleton, 2005; Phelps, 2005a; Phelps, 2006b; Sireci, 2005).

118
Additionally, several researchers found that high-stakes testing was endorsed by business, government, and the general public (Gallagher, 2003; Phelps, 2005a). The federal focus on educational reform, along with the support for high stakes testing among educational stakeholders, created an environment that led to the development of the CCCS, which 49 U.S. states and territories have implemented (Goldstein, 2009). Georgia has agreed to implement the national CCCS and will use EOCTs to assess them. The state of Georgia is currently changing graduation accountability measures from the GHSGT to a series of EOCTs (“Georgia Department,” 2010). Implementing the new CCCS and using EOCTs as accountability for those standards are important reform efforts being addressed by Georgia educators.

Also, researchers have identified benefits of standardized testing programs. Volante and Ben Jaafar (2010) found that high-stakes testing motivated students to study and raised educators’ expectations for students with disabilities. Reback (2008) concluded that when students attached importance to their test scores, they performed better, especially low-performing students. Several researchers also found that adherence to NCLB accountability policies has led to increases in student achievement on states’ high-stakes tests (Harris & Herrington, 2006; Nichols, Glass, & Berliner, 2005; Sims, 2008; Volante & Ben Jaafar, 2010).

Furthermore, Vanderhaar, Muñoz, and Rodosky (2006) noted that prior achievement was one of the strongest predictors of student achievement. The analysis of CRCT and EOCT data for this study revealed that student achievement increased for each GPS-based testing year from 2006-2010 on both the Reading and ELA CRCT and from 2007-2011 in Ninth Grade Literature EOCT (see Table 4.11, Table 4.15, Figure 4.4, Figure 4.5, Figure 4.6). Furthermore, CRCT and EOCT scores were significantly
correlated, and CRCT scores were predictive of EOCT performance.

Implications of the Findings

This study’s findings could have implications for educators and need to be examined in concert with the reviewed literature. Implications relate to accountability, teacher autonomy, and school reform decisions.

Accountability

Policymakers, politicians, states, and local school systems focus on high-stakes test results as evidence of student achievement, especially due to the accountability focus of NCLB. The results of these state-mandated assessments have important implications for educators. (Au, 2007; Chiang, 2009; Dee & Jacobs, 2010; Moon, Jarvis, Brighton, & Hall, 2007; Supovitz, 2009). Schoen and Fusarelli (2008) found that the impetus for NCLB was to increase a school’s accountability to the public. Opponents believe testing narrows what is taught, decreases overall learning, and increases stress in teachers and students (Cizek, 2005; Solley, 2007).

Conversely, Harris and Herrington (2006) and Sims (2008) found that accountability policies increase student achievement and reduced achievement gaps among student groups. High-stakes testing opponents and proponents agree, though, that accountability based on high-stakes test performance is not going away (Cizek, 2005; Gabriel, 2010; Gallagher, 2003; Sims, 2008; Solley, 2007).

Educators’ perceptions of testing have resulted in many research studies (Baker & Johnston, 2010; Guskey, 2007; Johnson, Yarrow, Rochkind, & Ott, 2009; Mulvenon, Stegman, & Ritter, 2005; Wiliam, 2010). Unfortunately, the pressure to increase students’ test scores has resulted in increasing incidences of cheating (Gabriel, 2010). The researched school system for this study is in Georgia, which has had its share of
national headlines of widespread cheating in a prominent, metropolitan public school system (Resmovits, 2011; Torres, 2010; Vogell, 2011a). Georgia is not alone in dealing with the ethical dilemma of cheating (Gabriel, 2010). Cheating has been investigated in the District of Columbia’s public schools and in Colorado, Indiana, Massachusetts, New Jersey, Nevada, Texas, and Virginia (Gabriel, 2010; Rothschild, 2011; Toppo, Gillum, & Bello, 2011). Thus, another implication for the researched school system, as well as all educators and education policymakers, is not to use standardized test data as the single factor in defining school improvement and teacher performance.

For all the studies about standardized testing and educators’ perceptions of it, few research studies examined how test data can inform decisions at the teacher-student level. This research study provides strong correlations between the CRCT and EOCT and strong predictive value of CRCT performance on EOCT performance. The significance of these relationships has an implication related to feedback provided to educators, especially those in the researched school system. A weakness in the accountability data is that there is no timely feedback for teachers or students (Crocco & Costigan, 2007; Scot, Callahan, & Urquhart, 2009). With strong predictive value in the eighth grade CRCT scores on ninth grade EOCT performance that was found in this study, it is imperative that CRCT data are available to ninth grade ELA teachers prior to the end of students’ eighth grade year. GaDOE and its school systems need to get standardized test data to schools and in the hands of teachers as quickly as possible so teachers can identify students who may struggle in the ninth grade year and begin interventions as early as possible.

Therefore, utilizing eighth grade CRCT data as one measure of student achievement, high school administrators and ninth grade teachers can identify struggling
learners and the gaps in their content knowledge and support these students as they make the important transition into high school. Summative CRCT test data provide a scaled score by content area as well as a breakdown of performance learning strands within each content area. Interventions can then be designed to address struggling learners’ needs.

Ultimately, the information can be used to support student learning and increase the number of students who graduate from high school. Research-based best practices of assessment for learning, individualized instruction, and differentiated instruction can be implemented based on analyses of students’ performance on high-stakes testing (Chappuis & Stiggins, 2002; Landrum & McDuffie, 2010; Murawski & Hughes, 2009; Stiggins, 2005, 2008; Tomlinson, 1999, 2009).

**Teacher Autonomy**

As a result of the accountability of high-stakes testing programs, teachers asserted that they have relinquished their curricular and instructional autonomy (Au, 2007; Gallagher, 2003; Graham & Neu, 2004; O’Day, 2002). Yet, high-stakes testing programs and teacher creativity in student-centered instruction can take place concurrently (Au, 2007; Lai & Waltman, 2008; Vogler, 2006). Grant (2007) concluded that test-based instructional practices like lecturing and rote memorization co-existed with class discussions, projects, and debates that require greater critical thinking skills than measured on high-stakes tests. Educators need to determine if there is a way to prepare students for testing that also empowers teachers to determine the most successful way to ensure student learning in their classrooms.

Because of the significant correlations identified between CRCT and EOCT scores in the researched school district, teachers need to feel that they are the education professionals that can make the most difference in using test results to address the needs
of their students. With support from school and school system administrations, teachers can identify interventions and determine if the interventions positively impact student learning. According to Hyslop and Sears (2010), professional autonomy for teachers is a fundamental requirement for educational improvement. Teachers are the educators closest to students, so they should be in the best position to determine the needs of their students, implement interventions, assess the benefit to students, and revise strategies to make the greatest gains in student achievement.

**School Reform Decisions**

Student learning is the goal of education. As such, high-stakes testing data should be one of several forms of data used to design school reform efforts, and all school reform efforts should focus on ensuring student learning. Every state in America uses high-stakes testing to meet the requirements of NCLB, which was enacted to hold schools, local school systems, and states accountable for improving all students’ academic achievement (Baker & Johnston, 2009). NCLB’s high-stakes testing requirement is impacting all 50 states (“No Child,” 2001).

School reform issues identified in previous research have been resource allocation, curriculum and assessment alignment, instructional rigor as opposed to teaching to the middle, and decisions on student retention. When school systems allocated funding for instructional technology, curriculum development, and professional learning initiatives that targeted increasing test performance, research found that students’ achievement increased (Chiang, 2009; Dee & Jacobs, 2010). Decker and Bolt (2008) found that aligning curriculum and testing standards improve students’ academic achievement and that alignment tools are being utilized more and more frequently by state departments of education. NCLB’s accountability mandates have caused teachers to
“teach to the middle,” which marginalizes high-performing and low-performing students (Reback, 2008). Additionally, NCLB has established retention policies, which are not supported by research, for under-performing students who do not meet standards (Roderick & Nagaoka, 2005).

With the establishment of the national CCCS, the alignment of curricula, instruction, and assessment have become more important as CCCS have been adopted and are being implemented in 49 states and territories (“Common Core,” 2010). In Georgia, where the researched school system is located, CCCS professional learning is taking place in the 2011-2012 school year, with full implementation of the standards in 2012-2013. The implications for curricula, instruction, and assessments will impact both this study’s participants and educators across the state of Georgia. Federal standards will influence state’s curricula decisions, and it is important to empower teachers to develop instructional strategies and assessment practices that support content standards.

Furthermore, school and school system administrators need to ensure an environment that supports professional learning communities where teachers can determine instructional and assessment best practices. Educators also need to engage in public dialogue about education with their schools’ stakeholders (Hyslop & Sears, 2010; Dufour, Dufour, Eaker, & Many, 2010). In the BCSS, School CC demonstrated significant differences in scores compared to the other two high schools. School system and school administrators need to examine the interventions in place in School CC for possible implementation at the other two high schools in the district.

Additionally, in Georgia’s transition from the GHSGT to the EOCT as its AYP accountability measure, the results of the data analyses have strong implications for educators in the researched school system. In two of the five paired testing years, female
students showed an almost 5-point gain over their male peers. Moreover, non-White students were found to score about 4 points lower than White students in two of the five paired years. These results occurred in the most recent testing years of 2009-2010 and 2010-2011, and BCSS educators need to examine future test results carefully. They should also study existing test results in other grade levels and subject areas for differences in achievement according to subgroup performance. If further evidence is found, they will need to design interventions for achievement gaps.

Students with disabilities performed lower every year on the Reading CRCT, the ELA CRCT, and the Ninth Grade Literature EOCT when compared to their grade level peers who did not receive special education services. Since eighth grade performance is predictive of ninth grade performance, it is crucial to address the needs of students with disabilities as they enter and progress through high school. This is true, too, for male students and for non-White students who may be identified as at-risk for grade promotion or graduation. Research-based interventions need to address these groups’ unique needs.

**Limitations**

There are several limitations in this study. First, a correlational study does not equal causation. Correlational statistics can be used for prediction, or to support a theory, but they cannot be used to prove causation (Waters, 2010; Gall, Gall, & Borg, 2007). Even though this study found a significant correlation between the students’ standardized test achievement in eighth grade and in ninth grade, it cannot be inferred that eighth grade achievement causes ninth grade achievement. The regression analysis found that eighth grade CRCT Reading and ELA performance are each predictive of student performance on the Ninth Grade Literature EOCT.

Another limitation is the third-variable problem, which are unmeasured variables
that are the actual cause of changes in student achievement (“Research,” 2010). For example, standardized testing may be affected by (a) a student’s motivation or health on the day of testing, (b) having a different teacher than another student, or (c) receiving extra instruction outside of the regular classroom or school day. The third variable problem was addressed two ways. First, there were five paired testing years for both REAss/LitSS and ELAss/LitSS rather than only using one paired testing year. Second, input was sought from system and school administrators, who provided school-level factors that may have impacted test performance from school to school or over time.

Furthermore, results cannot be generalized. The study population was limited to eighth and ninth graders in one rural school system, and the results cannot be generalized to other school systems or to other grade levels. Similarly, this study only used the eighth grade Reading CRCT, the eighth grade ELA CRCT, and the Ninth Grade Literature EOCT, which are all valid and reliable instruments (“CRCT,” 2010; “EOCT,” 2010). Even though they are valid and reliable tests, results cannot be generalized to the CRCT or EOCT in other grades or subject areas.

Likewise, the results cannot be generalized to other norm-referenced or criterion-referenced assessments. Not all states have interim tests like the EOCT, relying instead on graduations tests like Georgia’ GHSGT that students must pass in order to graduate (“Standardized Testing,” 2011). Last, the researched school system is small, rural, and has limited diversity. Its results should not be generalized to school systems that are large, urban or suburban, or with greater diversity.

Implications for Future Research

There is a need for further research regarding relationships among state-mandated standardized tests in Georgia. First, the study could be replicated in other Georgia school
systems to determine if significant correlations exist within their REAss/LitSS and ELAss/LitSS data. Additionally, the study needs to be replicated in other content areas. Future research needs to determine correlations and predictive value between the eighth grade Math CRCT and the ninth grade Math I EOCT and between the eighth grade Science CRCT and the ninth grade Physical Science EOCT. Multiple regression analyses could be conducted to determine all correlations in all subject areas. For example, Pearson’s product moment coefficients could determine the strength of the relationship between the eighth grade Reading CRCT and the ninth grade Physical Science EOCT. A multiple regression analysis could identify whether performance on the eighth grade Reading CRCT could predict performance on the ninth grade Physical Science EOCT.

Once educators understand the relationships and predictive value of high-stakes testing data, the focus of future research should shift to using assessment to increase student learning. Georgia had no data that identified correlations between the CRCT and EOCT, a void this research study sought to address. Thus, this study was limited to determining correlations between tests and any predictive value of students’ eighth grade performance in Reading and ELA on their ninth grade performance in Literature. Few research studies were located that addressed how to use data analyses to improve student learning. Since correlations and predictive value are evident, future research efforts in the BCSS should focus on identifying and implementing research-based interventions and then analyzing how an intervention impacted students’ achievement.

**Summary**

Significant correlations exist in the researched school system between standardized test scores for eighth grade Reading and ELA CRCT achievement and Literature EOCT achievement in the ninth grade. Furthermore, eighth grade Reading
CRCT scores and ELA CRCT scores hold predictive value for students’ performance on the Ninth Grade Literature EOCT. This is important news for Georgia educators, and future research needs to determine if the findings are applicable in other content areas and in other school districts.

Also, test scores are an important measure of the academic performance of schools, teachers, and students because of the accountability requirements of NCLB. Educators need to ensure that high-stakes test scores are not the only consideration for school improvement decisions. Rather, high-stakes test data should be one part of a comprehensive, multiple-measure plan for school improvement. Accountability based on multiple measures should become the focus of future educational reform efforts and accountability programs.

As Christian educators in public schools, we need to base our decisions on the needs of the students we serve and, moreover, center our decisions in our Christian faith for the good of our students. Christ told us, "See that you do not despise one of these little ones, for I say to you that their angels in heaven continually see the face of My Father who is in heaven" (Matt 18:10 [NASB]). The education of our students matters. What is more, our students matter. As we work with our students, we need to follow Paul’s recommendation, “Let all that you do be done in love” (1 Corinthians 16:14 [NASB]). As God’s love guides us and works through us, we can make a difference in our schools and in the lives and learning of our students.
REFERENCES


Studies, 32(2), 159-181.


for professional learning communities at work (2nd ed.). Bloomington, IN: Solution Tree Press.


Faircloth, K. (2009, April 23). Depth of knowledge. Professional learning presentation to the Bartow County School System, Cartersville, GA.


Gayler, K. (2005, June). How have high school exit exams changed our schools?


134


Horn, B. K. (2009). The perception of English language arts teachers about instructional
changes following the implementation of the Texas Assessment of Knowledge and Skills Test: A doctoral dissertation. (Unpublished doctoral dissertation). University of North Texas, Denton, TX.


services/ 160-4-2-.11.pdf

Proposed secondary assessment transition. (2011, February). [Bartow County School
System assessment training chart]. Cartersville, GA. Copy in possession of
author.


Reback, R. (2008). Teaching to the rating: School accountability and the distribution of
j.jpubeco.2007.05.003


Research in Social Psychology I. Retrieved from the Indiana University-Purdue
University Fort Wayne website at http://users.ipfw.edu/bordens/social/research.htm

Rescher, N. (2001). *Philosophical reasoning: A study in the methodology of

atlanta-public-schools-cheating_n_890526.html

childhood assessment to state kindergarten content standards: Application of a


Supervision and Curriculum Development.


APPENDIX A

Proposed Secondary Assessment Transition Matrix

from the Georgia Department of Education
# Appendix A: Proposed Secondary Assessment Transition Matrix

from the Georgia Department of Education

<table>
<thead>
<tr>
<th></th>
<th>Ninth Graders</th>
<th>Tenth Graders</th>
<th>Eleventh Graders</th>
<th>Twelfth Graders</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>2011/2012</strong></td>
<td>EOCT = 20% of course grade</td>
<td>EOCT = 15% of course grade</td>
<td>EOCT = 15% of course grade</td>
<td>EOCT = 15% of course grade</td>
</tr>
<tr>
<td></td>
<td>Pass the GHSWT</td>
<td>Pass GHSWT</td>
<td>Pass GHSWT</td>
<td>Pass GHSWT</td>
</tr>
<tr>
<td><strong>2012/2013</strong></td>
<td>EOCT = 20% of course grade</td>
<td>EOCT = 15% of course grade</td>
<td>EOCT = 15% of course grade</td>
<td>EOCT = 15% of course grade</td>
</tr>
<tr>
<td></td>
<td>Pass the GHSWT</td>
<td>Pass GHSWT</td>
<td>Pass GHSWT</td>
<td>Pass GHSWT</td>
</tr>
<tr>
<td><strong>2013/2014</strong></td>
<td>EOCT = 20% of course grade</td>
<td>EOCT = 15% of course grade</td>
<td>EOCT = 15% of course grade</td>
<td>EOCT = 15% of course grade</td>
</tr>
<tr>
<td></td>
<td>Pass the GHSWT</td>
<td>Pass GHSWT</td>
<td>Pass GHSWT</td>
<td>Pass GHSWT</td>
</tr>
<tr>
<td></td>
<td>EOCT = 20% of course grade (Science &amp; Social Studies)</td>
<td>EOCT = 20% of course grade (Science &amp; Social Studies)</td>
<td>EOCT = 20% of course grade (Science &amp; Social Studies)</td>
<td>EOCT = 20% of course grade (Science &amp; Social Studies)</td>
</tr>
</tbody>
</table>

Georgia Department of Education  
Dr. John D. Barge, State School Superintendent  
February 2011  
All Rights Reserved

APPENDIX B

Form Requesting BCSS Administrators’ Input
Appendix B: Form Requesting BCSS Administrators’ Input

**Standardized Testing Factors Identified by School/System Administrators**

A research study is being conducted to seek any correlation that may exist between eighth grade CRCT scores in reading and ELA with the ninth grade EOCT scores in 9th Grade Literature. Your input is valuable. Please provide the information requested below and return this form by courier to Venita Bruton at Adairsville High School.

In the chart below, please list innovations/programs in the first column that have been used in your school over the past three years.

<table>
<thead>
<tr>
<th>Reading/ELA Interventions</th>
<th>For each item listed: On a scale of 1 (low) to 5 (high), indicate the degree to which you believe the intervention/program impacted student achievement on the Reading/ELA CRCT or 9th Grade Literature EOCT.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX C

Research Approval Letter
Appendix C: Research Approval Letter

August 15, 2011

Liberty University
Institutional Review Board
Campus North Suite 1582
1971 University Boulevard
Lynchburg, Virginia 24502

Dear IRB:

Venita Bruton, a doctoral candidate at Liberty University, serves as my associate principal. She has my permission to access student achievement data that will be utilized in her proposed correlational study of student performance on the eighth grade CRCT and ninth grade EOCT.

I look forward to her results for the implications they can have for students at Adairsville High School and in our school system. Please feel free to contact me if needed.

Sincerely,

Bruce Mulkey
Principal

Bruce Mulkey
Principal

Adairsville High School
519 Old Highway 41 NW • Adairsville, GA 30103
Phone 770-606-5841 • Fax 770-773-2722
"Character, Community, Excellence"
APPENDIX D

IRB Application
Appendix D: IRB Application

9/07  RESEARCH EXEMPTION REQUEST  Ref. # ___________
Liberty University
Committee on The Use of Human Research Subjects

1. **Project Title:** Georgia High-stakes Testing: The Correlation between Eighth Grade and Ninth Grade Achievement

2. Please list all sources of funding. If no outside funding is used, state “unfunded”: unfunded

3a. **Principal Investigator(s)** [Must be a Liberty faculty member or investigator authorized by the Chair of the Institutional Review Board. If a student is the principal investigator, the student must have a faculty sponsor. Include contact information for both the student and the faculty sponsor as appropriate]:

   Venita L. Bruton, doctoral candidate  vlbruton@liberty.edu

   Name and Title  523 Spring Place Road, White, GA 30184

   Home (770) 386-7702; Cell (770) 655-7274

3b. **Faculty Sponsor**

   Dr. Judy Shoemaker, LU Professor  jshoemaker@liberty.edu

   Name and Title  (863)604-0111 or (863)326-6208

   Dept., Phone, E-mail address

   Anticipated Duration of Study: August 2011__________

   From                To  October 2011__________

4. **Are you affiliated with Liberty University?** YES X NO □

   **If so, in what capacity?** As a doctoral candidate

5. **Do you intend to use LU students, staff or faculty as participants in your study?** If you do not intend to use LU participants in your study, please check “no” and proceed directly to item 6.

   YES □ NO X

   If so, please list the department and/classes you hope to enlist and the number of participants you would like to enroll.

   [Not applicable]

   In order to process your request to use LU subjects, we must ensure that you have contacted the appropriate department and gained permission to collect data from them.

   Signature of Department Chair: [Not applicable]

   Department Chair Signature(s) Date

6. **Briefly describe the purpose of the study.**
The purpose of this study is to determine if there is any correlation in or predictive value of students’ achievement on Georgia’s state-mandated standardized testing in their eighth grade year on the Criterion-Referenced Competency Test (CRCT) in Reading and English/Language Arts relating to their performance on the End-of-Course Test (EOCT) administration in 9th Grade Literature at the end of their ninth grade year. Currently, no data are available that provide information on how well the CRCT relates to or can predict achievement on the EOCT. When students transition to high school, it is important to identify struggling learners and any gaps in content knowledge. If the CRCT is significantly correlated to the EOCT and if any prediction can be made about EOCT performance based on CRCT achievement, then interventions can be designed to address the needs of struggling learners. Further, test data can help inform curricular and instructional decisions to fill any knowledge gaps. Ultimately, the information can be used to support student learning and increase the numbers of students who graduate high school.

7. Provide a lay language description of the procedures of the study. Address ethical issues involved in the study (See the Avoiding Pitfalls in section of the IRB website for helpful suggestions) and how you will handle them. For example, consider issues such as how subject consent will be obtained (or explain why the study meets waiver guidelines for informed consent), how the data will be acquired, and how the data will be stored confidentially once it is collected. Please attach pertinent supporting documents: all questionnaires, survey instruments, interview questions and/or data collection instruments, consent forms, and any research proposal submitted for funding.

The proposed study will use only post-assessment data to determine any correlation or predictive value between the eighth grade CRCT and its associated ninth grade EOCT. The Georgia Department of Education (GaDOE) provides these data annually to systems, and the data have been gathered for all the years that state-mandated testing has been conducted on subjects aligned to the Georgia Performance Standards (GPS), which began implementation in a staged roll-out that began in selected subjects in 2005. The first GPS-based assessments began in selected subjects the following year. The standardized assessment data are in the possession of the researcher as part of employment responsibilities with the school system, and the researcher’s principal has given the researcher permission to access the student data for the proposed study.

The researcher’s principal has consented to the study (see approval letter at the end of this application). After approval from the Liberty University Institutional Review Board, the statistical analyses will be conducted, including Pearson’s product-moment coefficient analyses to determine any correlation and regression analyses to determine any predictive value. The test data will be analyzed for all of the district’s students for all subject area groups for all available years of test data in an attempt to determine if eighth graders’ achievement correlates with ninth grade achievement and if eighth graders’ achievement predicts their ninth grade achievement on the identified standardized test measures. The standardized test database for the school system’s eighth and ninth graders will be combined into one spreadsheet, and data will be disaggregated by subject area, testing year and school. Additionally, data will be studied according to gender, ethnicity, and students with disabilities for the implications of subgroup performance. Additionally, the input of school and system administrators will be sought regarding initiatives that they feel
may or may not have affected student performance (see “Standardized Testing Factors” form at the end of the application). Upon receiving IRB approval, the researcher will contact each building principal to introduce the form, which will then be sent to each principal by the school district’s courier system and then returned by the same. While the forms will include the principals’ names so that the researcher knows which school scores could have been impacted, at no time will the principals or schools be identified in discussions of their responses.

Ethical considerations include confidentiality of the data. First, the raw data spreadsheets as well as all data analyses will be maintained securely in possession of the researcher in electronic form. All data will be housed on a password-protected data drive, which will be secured by the researcher in a stored lock box. All data will be maintained for at least three years per federal regulations. Additionally, the researcher has an obligation to respect participants and to acknowledge their contributions. While protecting system and school anonymity in the research and dissertation process, the researcher will thank the system superintendent and middle and high school principals in writing for their help toward completing the study.

Even in a study where only post-assessment data are used, respect for participants’ anonymity must be ensured. While the data set will provide scores, ethnicity, gender, and student with disability information, the data used for analysis will not include students’ names. A statistician, who is the researcher’s professional colleague, will remove data that can identify students prior to conducting any data analyses. The researcher will not be able to relate any student achievement data to any particular student. As well, no students or teachers will be identified in any results or discussions of data analyses.

Another ethical consideration is not to generalize findings to a population where they may not apply. The researcher must be clear about the characteristics of the study population and about implications; the data will not be construed as applicable in dissimilar populations. When data are discussed for possible implications, the researcher will take care with regard to ethical validation and question personal, moral, political, and ethical assumptions to provide equitable treatment for all participants. Further, the researcher will provide practical answers to questions, which, in the proposed quantitative inquiry, can be characterized as curricular, instructional, and assessment implications from the data analyses.

8. **Will subject's data be gathered anonymously?** YES NO X

The data needed for this study are already in the possession of the researcher, who had access to the data as a school and former system administrator. Data include students’ names and other identifiers, such as teacher name and school name. However, before using any data for the proposed study, the researcher will work with a colleague, a statistician, who will remove all names so that no students or teachers can be identified by the researcher during data analysis. The results and discussion will protect the confidentiality of the data. The anonymity of all students and teachers for whom data were reported to the school system will be ensured since data utilized by the researcher will not include any names. All data will be maintained securely by the researcher on an electronic storage device and housed in a lock box.

9. **Please describe the subjects you intend to recruit.** For example, minors under age 18, adults 18 and over, students, etc. Also, please describe your recruitment procedures. How will you find participants for your study? How will you contact
them? Please be explicit.

The study subjects will be Bartow County School System students who attended eighth grade and ninth grade between school year 2005-2006 and school year 2010-2011 and took the 8th grade CRCT in Reading and ELA and the 9th grade EOCT in 9th Grade Literature the following year. All data are post-assessment data released to school system by the GaDOE after each year’s state-mandated testing, so the study populate will be all students who took the tests. All data are in the possession of the researcher. All data will be deidentified by a statistician colleague of the researcher prior to use for the proposed study. Neither students nor teachers will be identified in results or discussions of the data. Students’ and teachers’ anonymity and confidentiality will be assured in all data reports, and data will be securely maintained. The researcher’s principal has provided approval for the proposed study, and his approval letter is attached.

FOR ALL APPLICANTS:

I have read the Human Subjects “Research Exemption Request Guidelines.”

Venita L. Bruton
Principal Investigator Signature(s)  July 27, 2011

Dr. Judy P. Shoemaker
Faculty Sponsor (If applicable)  August 2, 2011

See application instructions for each above item. Email form and supporting materials to irb@liberty.edu. Also, submit a hard copy of the form and supporting materials to: The Institutional Review Board (IRB), Campus North Suite 1582, 1971 University Blvd, Lynchburg, VA 24502.