

GEORGIA'S IMPLEMENTATION OF END-OF-COURSE TESTS AND STUDENT GRADES

by

NANCY GREENE NASH

(Under the Direction of C. Thomas Holmes)

ABSTRACT

This study examined student grades and Georgia's implementation of end-of-course tests (EOCT) in one secondary school in middle Georgia. While studies exist on high stakes testing and student achievement, few studies inform about the implications of high stakes testing on student grades (Nichols, Glass, & Berliner, 2005). The possibility exists that grades incorrectly inform about student achievement (Hassel & Lourey, 2005). The researcher sought to determine if student grades changed significantly in courses in which the EOCT was implemented. The mean course grades in United States history and American literature for three student groups were compared using a one-way analysis of covariance. Student groups consisted of juniors who took no EOCT, juniors who took a pilot EOCT, and students who took the implemented EOCT. Convenience sampling was used and resulted in a total of 714 students in American literature and 702 students in United States history. The Georgia High School Graduation Test was used as the covariate between groups. The test of between-subjects effects revealed a statistically significant difference ($p \leq .05$) in student grades in American literature, $p = .005$, and no statistically significant difference ($p \leq .05$) in student grades in United States history, $p = .100$. A pairwise comparison of the student mean grade in American literature was conducted to further determine in which year the statistically significant difference occurred. The findings are significant in examining teacher grading practices for grade inflation, assessment for learning, and real versus perceived student achievement. Results also implicate a need for further study

into the attitudes of teachers and students regarding the potential rewards or penalties of including EOCT results in student grades.

INDEX WORDS: End-of-course tests, student grades, teacher grading practices, grade inflation

GEORGIA'S IMPLEMENTATION OF END-OF-COURSE TESTS AND STUDENT GRADES

by

NANCY GREENE NASH

A.S., Middle Georgia College, 1972

B.A., The University of Georgia, 1974

M.S., Fort Valley State University, 1995

Ed.S., Fort Valley State University, 1998

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

DOCTOR OF EDUCATION

ATHENS, GEORGIA

2007

© 2007

Nancy Greene Nash

All Rights Reserved

GEORGIA'S IMPLEMENTATION OF END-OF-COURSE TESTS and STUDENT GRADES

by

NANCY GREENE NASH

Major Professor: C. Thomas Holmes

Committee: John P. Dayton
C. Kenneth Tanner

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2007

DEDICATION

This accomplishment is dedicated to the loving memory of my father, Joseph Daniel Greene, Sr. and to the honor of my mother Ethel Greene Rearden. All that I am today is a result of the unconditional love and respect given to me by my parents. My much beloved father led by example in his sense of fairness, his excellent work ethic, and his humility. I am still a work in progress as I strive to meet these desirable characteristics. To this day, I am in awe of the sacrifices he made for me to obtain my Bachelor's degree from UGA in 1974 while I fluttered about enjoying the life of a free spirited college student. From my mother I gained a sense of altruism as she consistently led by example in assisting others in need. I am proud to have been raised in a family in which integrity, fairness, altruism, and support were the norm. The confidence I have in myself is a direct result of the confidence they had in me. I am very grateful for the love and support of my parents and am proud to dedicate this work to them.

ACKNOWLEDGEMENTS

I am pleased to acknowledge those who helped me complete this dissertation. Without them, I would never have even undertaken the task, much less completed it. First of all, I would like to thank my major professor, Dr. C. Thomas Holmes. His patience and guidance were invaluable. His easy-going manner and sense of humor helped get me through times of high stress and low productivity. Special thanks goes to my committee members, Dr. C. Kenneth Tanner, who asks the hard questions, and Dr. John Dayton, who graciously agreed to come into the process late. Thanks to each of you for your contributions.

I also wish to give a special thanks to my fellow cohort members and co-workers, Dr. Vicki Rogers and Dr. Deondray Farlar. They both encouraged me to join them in the program and continued to encourage me to finish! I wholeheartedly thank my co-workers and English teachers extraordinaire, Karen Mitcham and Judy Morgan for their editing and encouragement. I also thank my long-distance friend Maureen Kurowsky for her invaluable services as an editor. Any errors are mine, as I never gave them the entire document to proof.

To my brothers, Danny and Al, I am grateful to be a part of a family in which laughter, humor, and creativity are a part of everyday life. You and your families are so important to me and I appreciate your support more than you will ever know.

Lastly, I thank my family for their love and support. My husband Ricky and my sons Graham and Jeffrey are entirely the joys of my life. They have sacrificed, supported, encouraged, and validated every degree undertaking I have sought. I hope they know how very proud I am of each of them and that what I am is truly a reflection of them. My heart is filled with love and awe that I am wife and mother to these very special guys.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	viii
CHAPTER	
1 THE PROBLEM	1
Statement of the Problem	2
Purpose	4
Justification	5
Sample and Population	5
Research Questions	6
Method	6
Organization of the Study	7
2 REVIEW OF RELATED LITERATURE	8
Historical Perspective	8
Grade Inflation	10
Teacher Grading Practices	13
Test-Based Accountability	18
Statewide Exit and End-of-Course Exams	26
Assessment	31
Georgia's Education Reforms	35
3 RESEARCH METHODS	41
Purpose	41

	Research Questions	42
	Hypotheses	42
	Design.....	43
	Population and Sample.....	43
	Instrumentation.....	46
	Data Collection.....	48
	Analysis	48
	Limitations.....	49
4	PRESENTATION AND ANALYSIS OF DATA.....	50
	Purpose	50
	Research Questions	50
	Hypotheses	51
	Population and Sample.....	52
	Data Collection.....	52
	Procedures	53
5	SUMMARY, DISCUSSION, AND RECOMMENDATIONS	61
	Summary	61
	Discussion	63
	Recommendations	66
	REFERENCES	69

LIST OF TABLES

	Page
Table 2.1: Comparison of Standards-Based Reform and Test-Based Reform	20
Table 2.2: Types of Exit Exams States Use or Plan to Use	28
Table 2.3: Scale Score Required for E-GHSGT	38
Table 2.4: Comparison of Achievement Levels for Purposes of Assessment	38
Table 3.1: Comparison of Percentage Graduation Rate, Dropout Rate, and Attendance	45
Table 3.2: School, State, and National Comparison of National Tests.....	46
Table 3.3: School, RESA, and State Percent Pass Rate Comparison of HSGT	46
Table 4.1: Descriptive Statistics of Groups	54
Table 4.2: Tests of Between-Subjects Effects for Dependent Variable American Literature	55
Table 4.3: Pairwise Comparison of Mean Difference of Student Groups for Three Years.....	55
Table 4.4: Tests of Between-Subjects Effects for Dependent Variable U.S. History	55
Table 4.5: Descriptive Statistics by Ethnicity.....	56
Table 4.6: Tests of Between-Subjects Effects for Dependent Variable American Literature by Ethnicity.....	56
Table 4.7: Tests of Between-Subjects Effects for Dependent Variable U.S. History by Ethnicity.....	57
Table 4.8: Descriptive Statistics by Gender.....	57
Table 4.9: Tests of Between-Subjects Effects for Dependent Variable American Literature by Gender.....	58
Table 4.10: Tests of Between-Subjects Effects for Dependent Variable U.S. History by Gender.....	58

Table 4.11: Descriptive Statistics by Course Level	59
Table 4.12: Tests of Between-Subjects Effects for Dependent Variable American Literature by Course Level	59
Table 4.13: Tests of Between-Subjects Effects for Dependent Variable U.S. History by Class Level	60

CHAPTER 1

THE PROBLEM

Members of the Georgia State Board of Education and the Georgia General Assembly have implemented testing requirements as prerequisites to earning a high school diploma since 1983. Prior to this time, students earned a general high school diploma by completing a required number of Carnegie units. No exit exam was required to establish minimum competency. The first of these minimum competency-testing requirements was the Basic Skills Test (BST) in 1983. This test remained in place through the entering freshman class of 1990–1991 who subsequently graduated in June 1994. Consisting of tests in language arts, mathematics, and writing, the BST required students to attain a minimum passing score in order to earn a high school diploma. The BST gave way to the Georgia High School Graduation Test (GHS GT) with students who entered grade 9 at the beginning of the 1991 school year who would later graduate in 1995. The GHS GT phased in tests in the five areas of language arts, mathematics, writing, social studies, and science (Georgia Code § 20-2-282; State Board Rule {SBR} 160-4-2-.30, State Education Rules, 2006).

The most recent state-mandated testing required is the End-of-Course Test (EOCT), which became effective in 2004–2005 (Georgia Code § 20-2-281; SBR 160-4-2-.13, State Education Rules, 2006). At this time, the requirement for both the GHS GT and the EOCT are still in place. The GHS GT continues in effect until the EOCT is fully developed and implemented, at which time the Georgia State Board of Education will discontinue the GHS GT on a schedule devised by the state board (Georgia Code § 20-2-281). The EOCT requires students to take cumulative exams in the eight areas of ninth grade literature and composition, American literature and composition, Algebra I, geometry, biology, physical science, economics,

and United States history. Students may not receive credit without taking the exam. These exams, which are designed to test students' proficiencies in these individual subject areas, count for 15% of their final grades in the corresponding courses (Georgia Code § 20-2-281).

While the BST, GHSGT, and the EOCT are similar in purpose, one major difference exists. The BST was strictly a state-administered, stand-alone test. The GHSGT continues to be a state-administered, stand-alone test. A student who passes the tests and meets all other requirements receives a high school diploma. The difference is that while the EOCTs are also state-administered tests, these tests count for 15% of a grade that individual teachers across the state have assessed and computed. Therefore, the EOCT graduation requirement will rely heavily on grades students earn in each of the eight courses to be tested. With the current GHSGT, teachers instruct students on state-determined performance standards. The recently implemented EOCT brings the possibility of altered teacher grading practices, as teachers will be computing 85% of the final grade in courses that determine whether or not a student receives a high school diploma (Georgia Code § 20-2-281; State Education Rules, 2006, SBR 160-4-2-.13).

Statement of the Problem

Some education experts indicate that grade inflation is a sign that high school teachers are assigning high grades for weak work. Some students who earn college scholarships based on high school grade point averages (GPAs) lose their scholarships in college for failure to maintain high grades (Schouten, 2003).

Nationally, The College Board reported evidence of grade inflation in which grade point averages have risen and Scholastic Aptitude Test (SAT) scores have fallen (The College Board, 1997). Fifteen years ago, students with A averages accounted for 28% of SAT test takers; however, today 42% of SAT test takers have A averages. College students of today with higher

grade point averages do not score better on the SAT than A students a decade ago (Schouten, 2003). Not only are grades inconsistent from high school to college, but grades also vary from high-poverty school systems to affluent systems. The U.S. Department of Education (DOE) (1994) reported that students who received mostly A's in high-poverty systems have about the same reading scores as C and D students in the most affluent schools.

Further indication of grade inflation may be a result of recent mandates to improve student achievement from the federal legislation *No Child Left Behind (NCLB)*. Under this legislation, states are required to produce annual state and local report cards that inform about state and school progress under the accountability provision of *NCLB*. This report card is designed to measure state and school annual yearly progress (AYP) based on required state assessments (*NCLB*, 2001). Exit exams are used by some states as a measure of AYP (Gaylor, 2005). Accountability standards are designed to impact all educators from the system level to the school level to the classroom level. If teachers do assign higher grades for weak work, the possibility exists for students to fail to achieve on tests of specific content knowledge such as exit exams or end-of-course tests. Georgia currently requires both an exit exam and an end-of-course test in selected subjects. Georgia is exploring the possibility of using its EOCT as a replacement for the GHSGT (Sullivan et al., 2005). Assuming all other graduation requirements are met, the GHSGT is strictly a pass or fail situation; a diploma is awarded for passing or withheld for failing (State Education Rules, 2005, SBR 160-3-1-.07). Passing or failing the EOCT is not subject to the same condition since teacher-grading practices will account for 85% of the final grade in each subject requiring an end-of-course test.

Georgia Governor Perdue and legislators have debated the issue of grade inflation with regard to the HOPE Scholarship Program (Salzer, 2003). All high school B averages are not the

same. For similar average SAT scores, differences in HOPE eligibility exist across Georgia public school systems (Campbell & Melton, 2001). Critics have argued that grades are inflated in some Georgia schools to ensure that students meet the requirements for the B average needed for HOPE (Salzer, 2003). When computing grades, teachers weigh assessments differently and consider factors other than academic achievement, such as effort, behavior, cooperation, and attendance (Marzano, 2000). At a time when state leaders are debating the inconsistency of grading issues tied to the HOPE scholarship, they are phasing out a stand-alone test as a graduation requirement and turning to the EOCT, which brings the possibility of similar inconsistent grading issues.

Student grades attached to EOCT can either provide rewards if students do well or provide consequences if students do poorly. If teachers feel the need to ensure that EOCT do not cause students to fail their course, teachers could use grading practices that artificially increase student grades. If teachers do assign higher grades for weak work, the possibility exists for students to fail to achieve on tests of specific content knowledge such as exit exams or end-of-course tests yet pass the related course in their high school curriculum. While studies exist on high stakes testing and student achievement, there are few studies informing the implications of high stakes testing on student grades (Nichols, Glass, & Berliner, 2005). The possibility exist that grades incorrectly inform about student achievement (Hassel & Lourey, 2005).

Purpose

The purpose of this study was to examine the impact of Georgia's EOCT on student grades in a middle Georgia secondary school. As colleges report rising grade point averages but lower SAT scores, as evidence mounts that grades assigned in high-poverty school systems and more affluent systems show a discrepancy, and as concerns about grade inflation are debated

with regard to the Georgia HOPE scholarship, Georgia has implemented EOCT as a graduation requirement that relies heavily on grades assessed and computed by individual teachers across the state. This study examined student grades after the implementation of the EOCT.

Justification

This study was significant for establishing the effect of Georgia's implementation of the EOCT, and its results could impact local policy with regard to the best practices of teacher-grading procedures. The purposes of grades as outlined by Airasian (1994) are (1) for administrative purposes, (2) to give students feedback about their progress and achievement, (3) to provide guidance to students about future course work, (4) to provide guidance to teachers for instructional planning, and (5) to motivate students. The state of Georgia has implemented tests that account for 15% of student grades that are a significant part of teacher evaluation and accountability. Also, there is a lack of related literature on the effect of EOCT on student grades when the EOCT is factored in as a percentage of a student's grade.

Sample and Population

The population for this study was a suburban high school located between two larger metropolitan areas in middle Georgia. The sample consisted of 11th grade students who took the EOCT for American composition and literature or United States history during the school years 2003–2004 and 2004–2005. The sample also included the 11th grade students during the school year 2002–2003. Students at the selected school consistently have performed at or above state level on the GHSGT but below state and national averages on national assessments such as the SAT. The county has one high school, two middle schools, four elementary schools, and one prekindergarten school. The median household income in the county is \$43,712 while the state

average is \$42,421. The racial makeup is approximately 69% White, 30% Black, and 1% other (Georgia Statistics System, n.d.).

Research Questions

1. Did the adjusted mean of United States history final course grades change with the implementation of the EOCT?
2. Did the adjusted mean of American literature final course grades change with the implementation of the EOCT?
3. Did the adjusted mean course grade change by gender after implementation of the EOCT?
4. Did the adjusted mean course grades change by ethnicity after implementation of the EOCT?
5. Did the adjusted mean course grades change by course level after implementation of the EOCT?

Method

The mean course grades in United States history and American literature for three student groups were compared using a one-way analysis of covariance (ANCOVA) to determine whether a significant difference existed among the three student groups. The independent variable was the EOCT, and the mean course grades in United States history and American literature were the dependent variables. The three student groups were 11th grade students in 2002–2003 who had no EOCT, the 11th grade students in 2003–2004 who had a pilot EOCT, and the 11th grade students in 2004–2005 who experienced the actual implementation of the EOCT.

Organization of the Study

Chapter 1 presents an introduction, the statement of the problem, purpose of the study, justification of the study, research questions, sample and population, and method used for the study. Chapter 2 includes a review of the related literature with regard to historical perspective, grade inflation, teacher grading practices, test-based accountability, end-of-course test, exit exams, assessment, and Georgia education reforms. Chapter 3 describes the methodology, and statistical treatment used to analyze the data collected. Chapter 4 reports the findings of the study. Chapter 5 summarizes the findings of the study and includes conclusions and recommendation to be considered.

CHAPTER 2

REVIEW OF RELATED LITERATURE

The following section contains a review of the related literature organized as follows: historical perspective, grade inflation, teacher grading practices, test-based accountability, statewide exit and end-of-course exams, assessment, and Georgia's education reforms.

Historical Perspective

Grading is the “collection and evaluation of evidence on students’ achievement or performance” and is a “professional judgment on the part of teachers” (Guskey, 2002, p. 12). These judgments are reported or communicated to parents, students, and others in terms of grades or marks and may take the form of numbers, letters, words, or symbols. Some have argued that grading is not essential to learning (Frisbie & Walton, 1992). However, teaching and learning are more effective when teachers verify the progress of student learning (Guskey). Teachers’ roles in grading may be twofold: that of advocates as they check the progress of student learning and that of judges as they assign grades (Bishop, 1992).

Kohn (1993) questioned the necessity and value of grading but noted that getting rid of grades would present enormous challenges because grades are such a fundamental part of our education system. Typically, grades are justified as a means to motivate students, to sort students, and to provide feedback to students. In fact, Kohn stated that grades undermine intrinsic motivation and effectively serve as demotivators. Children enter school eager to learn but sometimes lose their motivation and enthusiasm as they focus on external rewards such as grades. Sorting students by grades is usually done badly because the grades are based upon criteria too subjective and upon categories and tests too rigid and too superficial. Finally, Kohn

stated that the better form of feedback is written comments because “grades don’t say enough about people’s performance” (p. 202).

Kohn (2004, p.75) outlined three main effects of grading:

- Grades tend to reduce students’ interest in the learning itself.
- Grades tend to reduce students’ preference for challenging tasks.
- Grades tend to reduce the quality of students’ thinking.

Additionally, Kohn stated that grades are not valid, reliable, or objective; encourage cheating; distort the curriculum; and can spoil students’ relationships with teachers and fellow students.

Guskey (2002, p. 4) reported four developments that promoted change in grading and reporting systems:

- Inconsistencies in the grading policies and practices of elementary, middle, and high school educators should be recognized.
- The growing emphasis on standards and performance assessments makes current reporting practices inadequate.
- Advanced technology allows for efficient reporting of detailed information on student learning.
- Growing awareness of the gap between our knowledge of grading and reporting methods and common practice necessitates change.

Beginning in middle school, classrooms become a “points-driven economy” (Guskey, 2002, p. 21) with teachers setting the currency exchange rate with their grading standards. Students work to accumulate academic wealth and are eventually drawn into a points-driven system in which the value of learning weakens. Within this academic economy grading and reporting are further challenged by teacher subjectivity. Even when school systems have specific

grading policies in place, individual teacher subjectivity can influence the outcome of student grades (Guskey).

Grades resulting in a student's GPA have long been one of the criteria for determining college entrance. Roth (2005) expressed the belief that standardized testing takes precedent over GPAs in determining who is accepted to college because grades are more easily corrupted than test scores. The basis for judging what determines an A varies from school to school and is influenced by many variables. Roth advocates one nationally standardized test more comprehensive than the SAT or ACT for postsecondary school entrance. Such a test would include mathematics, English, history, sciences, foreign language, and geography and would last days not hours. Such tests could be patterned after the French Baccaalaureate or the Swedish Studentent. France, Sweden, and Switzerland rely on standardization to provide the best educational opportunities for students. Relying on grades and class rank in judging students' work to be comparable is problematic when no universal grading standard exists.

Grade Inflation

Hu (2005) proposed a framework under which college grades be examined. Grade inflation, grade increase, grade compression, and grade disparity are four different types of grading problems in higher education. Grade inflation refers to similar quality of academic performance in a given course being awarded a higher grade at the present time than it would have been previously. While grade inflation is the hot topic within all levels of education, this practice should be distinguished from other grade-changing phenomena. Grade increase is the rise of average grades in a given course over time. Grade compression indicates that variations in student course grades are limited so much that grades can no longer distinguish student performance. Grade disparity suggests that similar academic performances may be rewarded

differently in different courses or in different academic disciplines. Grading issues on any level are complex and should be examined in regard to factors such as student characteristics, faculty characteristics, academic disciplines, institutional characteristics, and grade levels.

Hu's (2005) review of empirical evidence from existing literature indicated that, according to individual institutional studies, college grades have risen but that the rise is less obvious among nationwide studies. Grading disparities present a more serious threat to the integrity of college grading than does grade inflation because these disparities affect students' choice of courses and lead to faculty lowering their grading standards. Hu reasoned that grade disparity leads to possible grade inflation and that both grading problems could lead to grade compression.

A 1992 survey conducted by the American Association of Collegiate Registrars and Admissions Officers (AACRAO) found several trends in grades and grading practices in undergraduate higher education. First, the four-point grading system is becoming more detailed and specific. Second, students' failing grades are increasingly being assigned, recorded, and computed in the student's grade point average. Third, the use of transfer grades in calculating grade points and making honors determinations is declining. These practices were viewed as a response to grade inflation. Another conclusion of the AACRAO survey was the movement toward institutions replacing repeated course grades with the grade from the most recent attempt or the grade from the best attempt (Riley, Checca, Singer, & Worthington, 1994). This last conclusion is contradictory to the latest policy of computing eligibility for Georgia HOPE scholarship recipients. Georgia's HOPE Scholarship Program has provided scholarships and grants to residents of Georgia since September 1993. The latest regulations for the high school graduating class of 2007 indicate that each failing and passing grade in attempted coursework

will be counted toward the required 3.0 GPA (Georgia Student Finance Commission, 2005). While colleges appear to be replacing repeated course grades with best attempts that allow students to improve their academic standing, at least one scholarship program is doing the opposite.

Twenty-five interviews with college faculty revealed that perceptions of their own grade inflation differed from the reality. Professors reported lower grades than they actually gave and believed grade inflation was a problem at the university level but less in their department and even less in their grading. The social psychological literature bears out the tendency of individuals to believe that they are better than average and that their situation is unique (McCabe & Powell, 2004). This tendency may help explain professors' attitudes regarding grade inflation.

Hassel and Lourey (2005) reported that grade inflation contributes to a lack of student accountability. Students in their senior year in high school spend only 1 hour studying per day and continue this pattern into college. These students have an adequate GPA in high school and college despite the fact that college professors tell students to expect at least 2 hours of homework for every hour spent in class. Edwards (2000) reported lowering academic standards, distorting students' views of their academic abilities, weakening credibility of universities, and devaluing college degrees as consequences of grade inflation.

Hassel and Lourey (2005) further noted that grades would always be subjective. Institutional and departmental standards are essential for establishing performance parameters. All must buy into these parameters in order to avoid easy versus hard professors. Avoiding grade inflation makes it possible for schools to turn out more rigorously educated students who are better equipped to compete in the job market. Guskey (2002) believed that the problem of grade inflation lies in the meaning of grades assigned rather than students simply receiving higher

marks. He stated that the best way to fight grade inflation “is not to assign fewer grades but to push for clearer standards” (p. 85).

Teacher Grading Practices

Disparities in grading practices by college professors range from the traditional view to the postmodern view. These differing views result in consequences such as inequitable assessment of students (Johnson, 2003). Bilimoria (1995) stated that grade distributions are higher from the postmodern perspective, in part because “the scope of what is legitimate and appropriate knowledge in the academic enterprise is widened” (p. 448). The questioning of tenets traditionally held to be true allows a larger number of students to “legitimate an appropriate grasp of a widened content” under the postmodern view. The postmodern perspective favors higher grade distributions because evaluation criteria are expanded. Students can engage in many acceptable dialogues rather than be limited to a few as in the traditionalist view.

The late Duke University provost Strohbehn convened a faculty committee to investigate the problem of grade inflation. Committee chairman Graham took a broader view and led the committee through a full examination of the problems associated with Duke University’s grading policies. The Committee on Grades concluded that disparities in grading practices, not grade inflation, were responsible for most of the problems associated with lenient grading. Further, inequities in grading practices persist because their consequences are misunderstood and are perpetuated by the following myths (Johnson, 2003, p. 9):

- Student grades do not bias student evaluations of teachers.
- Student evaluations of teaching provide reliable measures of instructional effectiveness.
- High course grades imply high levels of student achievement.
- Student course selection decisions are unaffected by expected grading practices.

- Grades assigned in unregulated academic environments have a consistent and objective meaning across classes, departments, and institutions.

The DUET (Duke Undergraduates Evaluate Teaching) experiment was a Web-based survey instrument used during the 1998–1999 academic year to investigate the relationship between student grades and student evaluations of teaching. Johnson’s DUET experiment provided evidence that higher grades do lead to better course evaluations, that student course evaluations are not good indicators of student learning, that higher mean course grades do not reflect higher levels of student achievement, and that students can manipulate their GPAs by selecting courses with instructors who grade leniently.

Johnson (2003) provided summaries of observational studies that investigated the relationship between student grades and students’ evaluations of teaching. Data in Johnson’s tables were based on comprehensive reviews compiled by Feldman (1976) and Stumpf and Freedman (1979). The mean of the correlation between 36 student-level studies of grades and student evaluations of teaching was approximately 0.21, suggesting a small but not unimportant relationship between these two variables. The summary of these student-level studies suggested a positive relationship between grades awarded to or expected by students within a class and the same students’ evaluations of that class or teacher. Also summarized were 31 class-level studies of grade and student evaluations of teaching. The mean correlation between expected or received classroom mean grades and mean classroom student evaluations of teaching was 0.31, again suggesting, if not a significant relationship between grades and student evaluations of teaching, one worth further investigation.

Behavior will most likely occur when that behavior is expected by others. Researchers suggested grading practices that include high grades for inferior work set forth certain

expectations that may be difficult to overcome as students move through college courses. Instructors should communicate clear expectations with regard to student learning and grading practices (Hassel & Lourey, 2005).

Safer, Farmer, Segalla, and Elhoubi (2005) surveyed 75 sections of undergraduate freshman college algebra courses at California State University, Long Beach, from the fall of 2001 through the fall of 2002 to gather student ratings of the overall effectiveness of their instructor. A multiple regression analysis was used with the dependent variable being the mean of the student evaluations and the following seven independent variables: Instructors, Days (whether the class was 2 or 3 days a week), Time (whether the class was during the morning, afternoon, or evening), Rows (number of rows in the classroom), N (number of students in the class), Mean Grade (average grade of all students in the class), and WeBWork (whether or not students in the class used a Web-based homework system in the class). The results of the multiple regression analyses yielded the dependent variable and three independent variables to be significant. The mean student evaluation of the instructor, the mean grade of the students for the particular class, the instructor of the class, and the number of rows in the classroom were significant at 5% level of significance. Factors that influenced student assessments of the effectiveness of their college algebra teacher included: the greater number of rows in the classroom, the lower the average student evaluation was; the higher the scores on student evaluations were, the higher were the student grades; and individual instructors differed significantly from one another in student evaluations.

Bonesronning (2004) hypothesized that the teacher can manipulate student effort and thus student achievement by choosing the proper grading practices. Empirical analysis demonstrated that students who are exposed to rigid grading perform significantly better than other students

and greatly improve their knowledge and skills. Teachers' grading may establish a difference between real and perceived student achievement. Where grades are used to establish admission to selected programs, perceived achievement might be more important than real achievement. A new climate of teacher accountability as measured by state assessments demonstrates the importance of teacher grading practices as an indicator of real student achievement.

Bonesronning (2004) discussed a Norwegian system in which perceived achievement is important to lower secondary students where grades are used to determine admission to upper secondary school. Easy grading brings about similar effects on student effort as a wage increase has on labor supply in an economic model. The student responds negatively to easier grading by decreasing studying effort. Betts and Grogger (2003) found that test scores rise in schools with high grading standards, particularly for students who are already high-achievers. Higher standards in teacher grading practices could result in students responding negatively. A threshold standard above a students' critical level could elicit no student effort at all. Also, teachers manage students of varying achievement levels and usually turn to a trade-off when determining their threshold levels for grading and student achievement.

McMillan, Myron, and Workman's (2002) summarization of existing literature on elementary classroom assessment practices indicated that few questions used for assessment are of the type that taps into students' higher level thinking skills. Nonachievement factors such as effort, ability, and conduct are important to teachers in their grading practices. McMillan et al. surveyed hundreds of teachers of grades 3–5 in urban, suburban, and rural schools to investigate assessment and grading practices. Elementary teachers appeared to use six grading components, of which academic performance was the most important, but effort, participation, and extra credit work was also important. Conversely, disruptive student behavior and the grading practices of

other teachers had little effect on grading practices. Teachers demonstrated considerable variations in how they weighed different factors for grading. Classroom assessment and grading practices are highly individualized from teacher to teacher. High variability of grading practices among teachers might be due to different teaching styles or educational philosophies, student characteristics, and curriculum. Teachers who reward effort within their grading practices might give students and parents deceptive information about a student's true competency.

Padgett-Harrison (2000) found through 80 teacher surveys and 10 teacher interviews that the most significant influence on teacher grading practices was other teachers. Teachers relied on colleagues for information on how to establish grading practices. The second most influential factor was school board policy and the third was school administrators. Padgett-Harrison's study further revealed that factors such as effort and participation are an element in teacher grading practices.

In a newsletter from the Southern Regional Education Board reporting on the *High Schools The Work (HSTW)* conference in Nashville, Tennessee, a summary of Canady's "Are the Grading Practices Clear?" is presented (*HSTW*, 2005). Canady, professor emeritus at the University of Virginia, offered nine beliefs related to school policies and grading practices at the 19th Annual *HSTW* Staff Development Conference. High-achieving students are more likely to be motivated by a low grade than low-achieving students. Historically, schools have emphasized what students don't do rather than what they do. Teachers should distinguish between student work to be graded and student work to be used for re-teaching and opportunities for students to raise grades. Before accountability measures became a part of education, schools were taken with the idea that students must adhere to due dates regardless of mastery of skills. Flexibility in due dates allows for more time for skill mastery. The school environment must be safe and

nurturing in order for students to make significant changes. Large percentages of students experience failure at an early age with approximately 27 percent of students giving up academically by the end of the fifth grade. Students at the top and bottom of the achievement ladder are not the ones in the most need. Students in the middle are the one who benefit most from an adult mentor, peer tutoring, and extended learning time. Canady did not suggest that teachers water down the curriculum, accept inferior work, or make excuses for students who do not perform. He suggested that teachers use strategies that allow students to complete high-level assignments and pass rigorous assessments.

Test-Based Accountability

The reauthorization of the *Elementary and Secondary Education Act* under the *NCLB Act* of 2001 imposed test-based accountability for all 50 states. Tests have historically served to provide knowledge on which students and schools were succeeding and to identify students and schools in need of additional help. Testing advocates also argue that large-scale testing can influence instructional practices. High-stakes testing measures all students in core academic subjects, rewards students and schools who do well, and pressures those who do not. A test-based accountability system contains the four interrelated components of goals, measures, targets, and incentives and may use norm-referenced or criterion-referenced tests in order to accomplish these actions. Norm-referenced tests are commonly reported and indicate the relative standing of students with respect to a large group. Scores are reported as percentile ranks, standard scores, and grade equivalents. Criterion-referenced tests in the context of accountability determine a student's status with respect to specific content standards. A student is measured by the degree to which he or she has mastered a specific body of knowledge. Test outcomes are

directly interpretable in performance levels such as advanced, proficient, or basic (Hamilton, Stecher, & Klein, 2002).

Evers (2001) proposed that state tests are the gauge by which the accuracy of student grades be judged. His proposal was prompted by students in Texas being given passing grades in Algebra I despite failing state tests, particularly in schools with large numbers of poor and minority students. Evers proposed that these grades cheat students and are an example of the “soft bigotry of low expectations.” Opponents of standardized tests stated that teachers are correct and that the tests are wrong. Evers likened eliminating standardized test to getting rid of the diagnostic tools in the medical profession. In fact, with the call for increased accountability, standardized testing appears to be on the rise.

Since the 1990s, standards-based reform has expanded so that in 2004, 20 states had exit examinations, accounting for more than half of all public school students in the United States. While some examinations such as the ITBS or the SAT-9 are nationally accepted, many are state specific and have titles such as exit exams, end-of-course tests, benchmark exams, high school assessments, or writing proficiency tests (National Center for Educational Statistics, 2004).

The call for accountability and the resultant high-stakes testing is a powerful movement that has reached nationwide proportions (Thompson, 2001). The single movement of standards-based reform is in actuality two movements, with “test-based reform” or “high-stakes, standardized, test-based reform” one side of the movement and the “authentic, standards-based reform” the other side. The two movements are distinguished by their focus on instruction and equity issues. Academic progress such as grade promotion and graduation rates based on a single indicator tends to narrow a curriculum and diminish instruction to prepping for a test. True standards-based reform stresses equity for all learners through high-level, authentic teaching and

learning. Test-based reform and standards-based reform are seemingly both about moving from an educational approach that values input to one that values outputs or results. When the important characteristics of the two movements are considered, however, the differences become apparent. Table 2.1 provides a comparison of some of the essential features of the two movements.

Table 2.1

Comparison of Standards-Based Reform and Test-Based Reform

	Standards based	Test based
Learning standards	Involves teachers, parents, and others as active participants	Expert psychometricians design tests as single indicators; tests are the real standards
Learning standards	Describes what all students learn at each level	Sets standards at specific levels
Professional development	Promotes high-quality development for teachers and administrators to support teaching standards	Teachers and administrators are pressured to raise test scores; student are drilled
Student assessment	Aligned with standards; many opportunities to demonstrate that standards were met	Single state or national test determines grade promotion or graduation
Support for student learning	High quality; individualized	Tends to focus on test-taking techniques
Educational purpose	Learning community with goal of helping students reach intellectual, social, and personal potential	Raising test scores

Table adapted from Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan*, 82(5), 358–362.

While exit exams appear to be positively influencing curriculum and instruction, there are concerns that such tests are unfair, that they might increase the dropout rate, and that they negatively affect student performance. Several states have expressed specific concerns about state-required exit exams. Career and technical education (CTE) advocates in Massachusetts

argued that CTE students were at a disadvantage on the Massachusetts Comprehensive Assessment System because of their hands-on curriculum as opposed to academic students who spent 100% of their time studying subjects covered on the tests. Delaware parents objected to a proposed three-tier diploma system, arguing that it is unfair to base a “basic,” “standard,” or “distinguished” diploma on one test without considering grade point average or other high school assessments. The states of Alaska, Oregon, Massachusetts, and California face lawsuits filed against them on behalf of students with disabilities with regard to discrimination, testing accommodations, and alternative assessments. In addition, Florida legislators stepped in when 12,500 students faced the possibility of not graduating due to failure on the state exit exam. While many states plan to add an exit exam or end-of-course tests, other states are amending, postponing, or re-examining their state requirements (Emeagwali, 2004).

Maryland field-tested English 1, Algebra 1, geometry, biology, and American government end-of-course exams in the spring of 2001 (Olson, 2001). Some experts believe that the newer end-of-course tests are an improvement over the minimum-competency exit exams. The best known end-of-course tests and those that have been around the longest are the Advanced Placement and International Baccalaureate exams. These exams are used to determine the awarding of college credit for high school students who follow a prescribed syllabus. Advocates of end-of-course tests for courses such as algebra, chemistry, or English believe that such tests improve instruction, better motivate students, bring greater equity to the content across different schools and classrooms, pose more difficult questions and problems, signal multiple levels of performance as opposed to pass/fail on an exit exam, and increase teacher responsibility for how well students do on specific subject exams. Conversely, some parents argue that such

tests can actually lower standards in high performing schools by causing teachers to narrow and simplify their instruction.

The debate over exit exams continues as some question whether the exams are fair, whether the minimum competency is too low, whether the exams cause higher dropout rates, and the overall purpose of the exam. Low-income, Black, and Hispanic students, English-language learners, and students with disabilities lag as much as 40% behind White and Asian students on initial exit exam pass rates. While most states claim that their exit exams are at or near the 10th grade level, examination of six states' exams found math and English tests to be the equivalent of seventh, eighth, and ninth grade levels. It is true that exit exams do stop some students from earning a diploma; however, there is no clear consensus by researchers. Exit exams are intended to motivate improvement among students, schools, and districts. While some exit exams are achieving this goal, high-stakes testing can also discourage low-performing schools, narrow curriculum, and increase professional development costs (Research Corner, 2005).

Nichols, Glass, and Berliner (2005) reported that pressure created by high-stakes testing designed to hold schools accountable under the *NCLB Act* has had almost no important influence on student academic achievement. Researchers measured the impact of high-stakes testing pressure on student achievement by creating a pressure rating index (PRI) using both quantitative and qualitative methods. Correlations between the PRI and National Assessment for Educational Progress (NAEP) results from 1990 to 2003 in 25 states were analyzed. Some of the results associated with increased testing pressure revealed that high-stakes testing disproportionately affected minority students, increased the possibility of eighth and tenth grade students being held back or dropping out of school, and revealed no gains in NAEP reading scores at the fourth or eighth grade levels. After replicating previous analysis and correlating their PRI and NAEP

results, the researchers concluded that there is no convincing evidence that the pressure associated with high-stakes testing leads to any important benefits for student achievement. The researchers listed the following reasons for calling for a halt on policies forcing public education systems to rely on high-stakes testing: unprofessional treatment of educators working in high-stakes testing situations, the possible corruption of indicators used in accountability systems that rely on high-stakes testing, data from studies that question the intended and unintended effects of high-stakes testing, and problems with reaching the goals set by the *NCLB* act in a reasonable time frame.

Even as some studies revealed problems with *NCLB*, the U.S. Department of Education reported that in the last 5 years the nation's students (particularly minority students) have made gains. Much of the progress in reading has been made in the last 5 years. Examples of student progress reported are as follows: America's 9-year-olds had their best scores in reading and math in the history of the 30-year Report Card; White, African-American, and Hispanic 13-year-olds have made significant gains in math in the past 5 years; and 13-year-olds have gained the highest math scores ever recorded. Ullman (2005) suggested that *NCLB* is indeed working.

The nonpartisan Civil Society Institute sponsors the Citizens for Community Action project at *NCLBGrassroots.org*. A report released in August 2005 from this project stated, "47 of 50 states are in some stage of rebellion against *NCLB*" (§ 1). Following are the key findings of the report:

- Legislation. Twenty-one states have considered bills critical of *NCLB*, seven states have actually passed legislation critical of *NCLB*, fifteen states have weighed options to opt out of *NCLB*, and four states have considered bills prohibiting use of state resources for *NCLB* implementation.

- Litigation. Some states have directed their attorneys general to explore litigation. Lawsuits have already been filed on behalf of English language learners. Also, lawsuits declared that *NCLB* is an unfunded mandate and that some provisions of *NCLB* conflict with the Individuals with Disabilities Education Act.
- Opting out, waivers, and exceptions. All but 10 states have sought an exemption or waiver from, or some accommodation for, *NCLB*'s rigid requirements.
- Cost studies. Fourteen states participated in the Council of Chief State School Officers *NCLB* cost consortium in order to assess the cost of implementation. Other states have individually released their own studies that reported significant funding gaps in *NCLB*.
- "Failing school" studies. Nine states released reports specifying the detrimental impact that federal testing requirements have on AYP. Schools have been penalized and sanctions have been imposed on schools failing to meet AYP. Even states that have led the way in standards-based reform and have students performing in the top tier on the SAT will be deemed "failing" by 2014 under *NCLB*.

Rose and Gallup (2005) noted several significant findings in their 37th Annual PDK/Gallup Poll of the Public's Attitudes Toward the Public Schools. Ninety percent of the public believed closing the achievement gap is very or somewhat important. Results of the poll indicate that the public attributed the gap to factors other than the quality of schooling (75%), believed parents and students have more to do with whether students learn (63%) than teachers do, but still believe it is the responsibility of the schools to close the gap (59%). Other findings yielded strong public opinion regarding *NCLB*. In the 2003 poll, 24% of the public said they knew a great deal or a fair amount about *NCLB*. This figure grew to 40% in 2005, while 59%

said they know very little or nothing at all. The public's view of the strategies used in *NCLB* are summarized in the following:

- Sixty-eight percent said a single test could not give a fair picture of a school.
- Eighty percent said testing in English and math only could not fairly picture a school.
- Eighty-two percent were concerned that *NCLB*'s emphasis on English and math means less emphasis on art, music, history, and other subjects.
- Seventy-nine percent preferred offering help to a student in a "needs improvement" school compared to 16 percent who preferred transfer to another school.
- Sixty-eight percent believed special education students should not be required to meet the same standards as other students.
- Sixty-two percent disagreed with *NCLB*'s practice of including scores of special education students in determining AYP.
- Fifty-eight percent believed the emphasis on standardized tests will cause teachers to teach the test.
- Fifty-four percent believed teaching the test is a bad thing.

Overall results imply that the public is firmly committed to its public schools and wants to see them get better. These results could be beneficial to policy makers at all levels in making a collaborative effort to improve public education.

The National Governors Association and Achieve, Inc., a bipartisan, nonprofit organization, hosted the National Education Summit on High Schools at which governors of 13 states formed a coalition, the American Diploma Project. The participating states—Arkansas, Georgia, Indiana, Kentucky, Louisiana, Massachusetts, Michigan, New Jersey, Ohio, Oregon, Pennsylvania, Rhode Island, and Texas—educate more than a third of all U.S. students and are

committed to lowering the dropout rate and improving the knowledge and skills needed for success. The association identified the top 10 steps to be taken to put states on a path to redesign their high schools. One of these steps was the development and funding of supports to help students pass high school exit exams through the use of online tutorials, intensive intervention programs, summer academies, and multiple opportunities for students to take the exams (Dessoiff, 2005b).

Statewide Exit and End-of-Course Exams

The National Center for Education Statistics (NCES) (NCES, 2005) reported on statewide types of promotion and graduation practices based on statewide exit and end-of-course exams for 2004–2005. NCES detailed eight practices as follows: promotion contingent upon performance on statewide exams; graduation contingent on performance on exit exams or EOCT, appeals process for students failing exams, alternative for earning a diploma if exams are failed, nonstandard or tiered diploma for students who fail exams, exit exams or EOCT based on 10th grade or higher standards, state required remediation for students failing exams, and state finances remediation. Twenty-nine states reported “yes” to at least one of the reforms, while 23 reported “no” to all eight reforms. Of the 23 who reported “no,” the District of Columbia and Washington plan to have graduation conditional on a student’s performance on statewide exit exams or EOCT beginning with the class of 2008. Florida and Georgia reported the highest number of testing reforms, each answering “no” to only one reform. Florida does not have an appeals process in place while Georgia does not provide alternative criteria for students to earn standard diplomas if they fail the exit exam.

The Center of Education Policy (CEP) grouped state exit exams into three categories based on states’ own descriptions of their tests:

- Minimum competency exams (MCE) generally focus on basic skills below the high school level.
- Standards-based exams (SBE) are aligned with state standards and generally targeted at the high school level.
- End-of-course exams (EOC) are aimed at assessing whether students have mastered the contents of specific courses at the high school level; EOC exams are usually standards-based and taken by students after completing a specific course.

Several states reported phasing in exit exams as they continue to move away from minimum competency exams to more challenging standards-based exams and EOCT. By 2009, if states continue as intended, three states will administer minimum competency exams, eighteen will administer standards-based exams, and five will administer EOCT exams. Some states, such as Georgia, Indiana, Maryland, South Carolina, and Texas, use a standards-based exit exam upon which graduation is contingent as well as from one to eight EOCTs upon which neither graduation nor promotion is contingent. Georgia is exploring using its EOCT as a possible replacement for the GHSGT, but no timetable exists for this possibility (Sullivan et al., 2005). Table 2.2 shows the trend in states' use of exit exams where states appear to be moving away from MCE to SBE.

Table 2.2

Types of Exit Exams States Use or Plan to Use

	MCE focused on basic skills below the high school level	SBE aligned with state standards and targeted at the high school level	EOC tied to a specific course at the high school level
In 2002 (18 states)	FL, MD, MN, MS, NV, NM, OH, SC, TN, VA	AL, GA, IN, LA, NJ, NC, TX*	NY, TX*
In 2005 (19 states)	AK, MN, NM	AL, FL, GA, IN, LA, MA, NC, NJ, NV, OH, SC, TX	MS, NY, TN, VA
In 2009 (25 states)	AK,, UT	AL, AZ, CA, FL, GA, ID, IN, LA, MA, MN, NV, NJ, NM, NC, OH, SC, TX, WA	MD, MS, NY, TN, VA

Source: Sullivan et al., Center on Education Policy, July 2005.

* Texas students had option of SBE or EOC.

Indiana currently administers end-of-course assessments in Algebra I and English II with voluntary pilot administrations in Algebra II, Biology I, and United States history in spring 2006. The requirement for students to pass the EOCT is a local decision. Schools choosing to incorporate the EOCT into local grading do not use the assessment for more than one third of a student's course grade. There are no "stakes" for students at the state level (Indiana Core 40 EOC Assessments, 2006).

According to Mississippi State Board of Education policy, the Subject Area Testing Program (SATP) was phased in with ninth graders in the 2001–2002 school year to replace the Functional Literacy Exam (FLE). The SATP consists of EOCT tests in Algebra I, Biology I, English II, and United States history from 1877. The FLE has gradually been phased out and the SATP simultaneously phased in so that the graduating class of 2006 must pass all four EOCT in order to graduate (Mississippi SATP, 2006). While Mississippi has an appeals process, the state

has neither an alternative criteria for students to earn diplomas nor a nonstandard or tiered diploma system (NCES, 2005).

New York students may earn either a Regents or local high school diploma, both of which are contingent upon scores made on the Regents exam in the following subjects: English, mathematics, science, United States studies, and global studies. For the purposes of a Regent-endorsed diploma, a score of 65 is considered passing. For a local diploma, a score of 55–64, as determined by the local school, also may be considered passing up through the 2007–2008 school year (New York Department of Education, 2006).

Students in North Carolina are required to take the North Carolina Competency Tests of Reading or Verbal and/or Mathematics Tests. All students entering the ninth grade in 1994–1995 and beyond and who follow the career preparation, college technical preparation, or college/university preparation courses of study must meet the competency standards to receive a high school diploma. North Carolina also administers EOCT in the areas of English I, Algebra I, geometry, Algebra II, civics and economics, United States history, physical science, biology, chemistry, and physics. Tests results must be used for at least 25% of the student's final grade for each respective course and the results will also be included in the student's permanent records and high school transcript (North Carolina Testing Program Overview, 2005).

In compliance with Tennessee state code and at the recommendation of the High School Testing Advisory Committee in 1998, the Tennessee State Board of Education designated 10 high school courses for the development of EOCT. Beginning with students entering the ninth grade in 2001–2002, students must pass EOCT, known as Gateway Exams, in three subjects: Algebra I, Biology I, and English II. Seven other EOCTs are administered but not required for graduation: math foundations, geometry, Algebra II, physical science, chemistry, English I, and

United States history. Tennessee reported the following purposes and rationale for end-of-course testing:

- Improvement of student learning in core content areas: Tests will align with curriculum standards and will include content knowledge and critical thinking skills.
- Preparation for further learning: Tests will occur early in a student's high school program of study.
- Diagnostic information: Tests will help improve student performance and preparation for ACT, SAT, and Work Keys.
- School and program improvement: Test data will be used to improve instruction.
- Accountability: Tests will provide for teacher, school, and school system accountability as called for in the Education Improvement Act.

Results of individual student performance on all 10 of the EOCT count not less than 15% of the student's grade in the course and semester in which the test was administered. Additionally, all students must achieve a passing score on the EOCT in Algebra I, biology, and English II in order to graduate. Passing scores are determined by the number of items answered correctly and are reported as "advanced" or "proficient." Previously, students took the Tennessee Competency Test only in mathematics and language arts. The last scheduled competency test was spring 2004, after which time all Gateway EOCTs must be passed for a regular diploma (Tennessee.gov, 2005).

The requirement that Virginia students pass end-of-course standards of learning (SOL) tests in order to graduate became effective with students who entered the ninth grade during the 2000–2001 school year. Students may earn either a standard diploma or an advanced studies diploma. When testing for the standard diploma was implemented, requirements stated that a

student must earn at least 22 standard units of credit and pass 6 verified units of credit, which are earned by passing the course and its related end-of-course SOL test. The six required EOCTs consisted of two tests in English and four tests of the student's choice. This requirement was modified with the entering ninth grade class of 2003–2004 to be two English tests, one mathematics test (Algebra I, Algebra II, or geometry), one science test (earth science, biology, or chemistry), one history/social studies test (United States history, World History I, World History II, or world geography), and one test in the subject of a student's choice (computer science, technology, or other board-prescribed areas). Students may receive an advanced studies diploma by earning 24 units of credit and taking and passing nine end-of-course SOL tests: two in each area and one test of a student's choice. Tests may be retaken if the student does not achieve a passing score (Virginia DOE, 2005).

Assessment

Course grades and standardized tests are forms of evaluating student learning. Course grades may be formative or summative assessments; standardized tests are summative assessments. Assessing student learning can be distinguished by the use of two mere prepositions but with different approaches and outcomes. Assessment *of* learning—summative—determines what students already know, usually in the form of standardized tests. Assessment *for* learning—formative—is an instruction-oriented approach designed to help students learn at the time at which learning is taking place (Black & William, 1998; Popham, 2006; Priestley & Sime, 2005; Stiggins, 2002). The GHSGT and EOCT are assessments *of* learning. Grades assigned by teachers and earned within the classroom by students consist of both assessments *for* learning and assessments *of* learning.

Stiggins (2002) reported on the assessment crisis, stating that, although we currently have ample assessments *of learning* (standardized tests), we lack adequate assessments *for learning* (day-to-day classroom assessments made by teachers). Politicians staunchly believe and resolutely focus on standardized test scores as the answer to student achievement, giving rewards and punishments to increase teacher and student effort, and as a gage for accountability. Assessment *for learning* usually results in student grades earned by students and assigned by teachers.

As early as 1996, the Mid-Atlantic Equity Consortium in their “Educate America” publication called for a closer look at student assessment and testing. Politically powerful advocates of “outcome-based” education called for high standards and a national system of testing in order to achieve educational improvement that led up to Bush’s legislative NCLB initiative in 2001. Such standardized testing for the sake of student assessment shapes curriculum and teaching practices. The role of student assessment in educational improvement can only be constructive if the assessments help teachers to identify student’s strengths and weaknesses and to determine the most appropriate and effective means of helping students to learn and achieve (Denbo, Grant, & Jackson, 1994).

Currently, teachers assess student learning at the end of a marking period. This assessment only allows for a small amount of testing time for each standard or skill and often provides information too late to be useful. Leahy, Lyon, Thompson, and Wiliam (2005) suggested a change from *quality control* in learning to *quality assurance*. The quality control approach is similar to manufacturing, in which defective learning might just be ignored or tossed away. The quality assurance approach is assessment *for learning* and involves adjusting teaching as needed while learning is taking place. The researchers are associated with Educational Testing

Service and have designed two workshop models to assist teachers in introducing techniques of assessment *for* learning. They promote the following assessment-for-learning strategies:

1. Clarify and share learning intentions and criteria.
2. Engineer effective classroom discussion.
3. Provide feedback that moves learners forward.
4. Activate students as owners of their learning.
5. Activate students as instructional resources for each other.

Accountability relies on large-scale assessment, standardized testing, and student achievement. Policy-makers who constantly push for accountability need to take into account how large-scale assessment can adversely affect student motivation and contribute to a higher dropout rate. School improvement and improved student learning require more than the implementation of large-scale assessment programs. Volante (2006) offered five considerations to provide a framework to promote authentic teaching and learning. First, educational reforms cannot be implemented in a top-down only manner. Reform must be a collective process balancing top-down and bottom-up input. Second, comprehensive achievement targets must include more than large-scale student achievement testing. Such tests limit the scope of higher order thinking skills by focusing on what is easily measurable in a multiple-choice test. Such tests narrow the curriculum. A third consideration is to provide a more equitable balance between classroom and large-scale assessment. Just as doctors rely on many diagnostics tools to determine the health of a patient, so should school systems rely on many forms of assessment to determine the healthy status of a school district. Also, educational reformers must strike a balance between short-term and long-term goals. Long-term goals must sometimes be delayed in order to achieve more progressive short-term objectives. Finally, the focus of education should

remain on all parts of the student population. When accountability rests on large-scale tests' scores, the tendency is to concentrate on “bubble students”—those who can raise overall test scores. Students who are unlikely to achieve at a higher level become an acceptable casualty of the battle to raise scores.

Black and Wiliam (1998) reported on the work of their colleagues in the United Kingdom's Assessment Reform Group (ARG). Student learning is driven by what teachers and students do inside the classroom, yet most school reform taking place in the United States and other countries does not address this most basic principle. The classroom is treated as a black box. Certain inputs—students, teachers, parent concerns, high-stakes testing, rules, etc.—are fed into the black box with certain expected outputs—competent and knowledgeable students, better test scores, and satisfied teachers. But once the classroom door closes, what goes on inside? Their meta-analysis of previous research regarding the effects of classroom assessment for learning indicated test-score improvements for students on classroom assessments as well as external assessments when assessment for learning was the focus as opposed to assessment of learning.

The Assessment Reform Group (ARG, 1999) distinguished between assessment for learning and other current practices of classroom assessment. “Teacher assessment” is assessment carried out by teachers that may or may not be *for* learning and further does not give the purpose of the assessment. The assessment is assumed to be formative (*for*), but in fact is summative (*of*). This assumption leads some schools to believe that the assessments taking place in their classrooms are adequate when in fact they are not. The ARG provided a summary of characteristics of assessment that promotes learning. This type of assessment (p. 7):

- is embedded in a view of teaching and learning of which it is an essential part,

- involves sharing learning goals with pupils,
- aims to help pupils to know and to recognize the standards they are aiming for,
- involves pupils in self-assessment,
- provides feedback that leads to pupils recognizing their next steps and how they take them,
- is underpinned by confidence that every student can improve, and
- involves both teacher and pupils reviewing and reflecting on assessment data.

Georgia's Education Reforms

Georgia's history of minimum competency testing began in 1983 with the BSTs of reading, mathematics, and writing. In 1991 the Georgia General Assembly established new requirements that all students seeking a Georgia high school diploma must pass a new set of GHSGTs in three areas of language arts, mathematics, and writing with the entering freshman class of 1990–1991 who would later graduate in 1994. By the graduating class of 1998, the GHSGT had phased in tests in the five areas of language arts, mathematics, writing, social studies, and science. The GHSGT is a curriculum-based assessment derived from the Quality Core Curriculum (QCC) (Quality Core Curriculum, 2005). The original QCC was brought into being by the *Quality Basic Education (QBE) Act* in 1985. The *QBE* required that the state maintain a QCC to be revised every 4 years; however, the first actual revision did not occur until 1996–1997 (Georgia Code § 20-2-282; Quality Basic Education Act, 2005). This revision was minor. In 2002, an audit by Phi Delta Kappa International's (PDKI) Curriculum Management Center concluded that the QCC lacked depth, did not meet national standards, and would take 23—not 12—years to cover. The results of this audit led to the current implementation of the Georgia Performance Standards (GPS) (PDKI, 2004).

The reauthorization of the *Elementary and Secondary Education Act (ESEA)* under the *NCLB Act* of 2001 extensively raised accountability levels for states. On January 8, 2002, President Bush signed into law the expectation that all students will meet or exceed state standards in reading and math by the year 2014 (Herring, 2005). One of the foundations of *NCLB* is adequate yearly progress (AYP), which is measured by yearly student achievement on statewide assessments. Schools are required to meet standards in two areas: student participation in the assessment of English/language arts and math, and level of academic performance on the assessment of English/language arts and math. If one of these standards is not met, a second indicator not related to the assessment is required (Tifig, 2005). With the arrival of *NCLB*, Georgia began a comprehensive examination of ways to meet AYP standards. Major initiatives started were Georgia Performance Standards (GPS), Enhanced Georgia High School Graduation Test (E-GHSGT), and EOCT.

The GHSGT content descriptions were revised to reflect the transition from QCC to GPS in English/language arts and science. In spring 2006, the GHSGT will align to both the GPS and the QCC because this assessment is based on core content and skills students have had an opportunity to learn during their years in high school. The transition of curriculum content to be assessed will be in effect for the school years 2005–2006 and 2006–2007. A solely GPS-aligned GHSGT in English/language arts and science will be given in spring of 2008, at which time the first-time test-takers in the 11th grade will have had only GPS instruction in high school. Provisions will be made for students who must retest under the GPS and QCC standards (GHSGT, 2005).

“The *A+ Education Reform Act* of 2000, O.C.G.A §20-2-281, mandates that the State Board of Education adopt end-of-course assessments in grade nine through twelve for core

subjects to be determined by the State Board of Education” (EOCT, 2005, ¶ 1). Eight content area assessments comprise the EOCT program and include assessments in Algebra I, geometry, United States history, economics, biology, physical science, ninth grade literature and composition, and American literature and composition. EOCTs are criterion-referenced tests that assess the statewide curriculum of specified high school courses to ensure that high academic standards as being met. The EOCTs provide diagnostic information about the achievement of Georgia students in these eight core areas. A student’s final grade in a course requiring an EOCT is determined by using the student’s grade as stated by the course teacher as 85% and the EOCT score as 15% of the final grade. Students must earn a 70 or higher as the final course grade, but are not required to pass the EOCT in order to earn credit toward graduation. Three stated purposes of the EOCT are to align with the Georgia curriculum standards and assess specific knowledge and content, to provide diagnostic information to identify strengths and weaknesses in student learning, and to provide data to evaluate the effectiveness of classroom instruction (EOCT, 2005).

Under the QCC standards, the GHSGT and the EOCT test overlap in United States history, biology, physical science, English/language arts, and Algebra I. The GHSGT in science under the transitional GPS/QCC standards lists knowledge of and skills in such areas as cells and heredity; ecology; structure and properties of matter; energy transformation; and forces, waves, and electricity rather than the broad QCC areas of biology and physical science. The knowledge and skills on the GHSGT in English/language arts in the GPS/QCC transitional version include reading comprehension, literary analysis, and conventions and writing as opposed to the QCC version that includes reading/literature, critical thinking, and writing/usage/grammar (GHSGT, 2005).

Changes continue to be made to the GHSGT as a result of *NCLB*. The English/language arts and mathematics tests have been enhanced each spring to meet *NCLB* standards and federal peer review of Georgia's accountability system. While the standards used to determine a student's eligibility for a diploma have remained the same, the enhanced tests are used to establish AYP at the school level. These scale scores are set to become increasingly higher through the year 2014. The E-GHSGT has three levels of achievement (see Tables 2.3 and 2.4). The new proficiency standard of 511 for English/language arts and the proficiency standard of 516 for mathematics will be used to measure AYP for Georgia's high schools; individual student performance, upon which a diploma is issued, will remain at 500 (GHSGT, 2005).

Table 2.3

Scale Score Required for E-GHSGT

	Basic	Proficient	Advanced
English/language arts	400–510	511–537	538–600
Mathematics	400–515	516–524	525–600

Source: GA DOE, Testing Division.

Table 2.4

Comparison of Achievement Levels for Purposes of Assessment

Purpose of assessment	Levels of achievement		
Student diploma	Fail	Pass	Pass plus
School/system/state AYP	Basic	Proficient	Advanced

Source: GA DOE, Testing Division.

Students who have not passed the GHSGT can seek waivers from the state Board of Education. In the past, requested waivers were rarely granted; the board approved only 41

requests over the past 4 years. This number represented a rejection rate of three fourths of student requests. The state legislature threatened to intervene and the board has conceded to add some flexibility to a process viewed by many as flawed. Attendance, academic performance, and other factors have been introduced into the appeals process. While the state legislature requires students to pass the GHS GT in order to receive a high school diploma, they are also instructing the state Board of Education to implement conditions to receive a high school diploma other than by passing the graduation test (MacDonald, 2005).

The GPS curriculum is being phased in with a K–12 plan. Grades 9–12 phase-in began with English/language arts and science year I in 2005–2006 and year II in 2006–2007. Social studies year I phase-in is 2006–2007 with year II to follow in 2007–2008. Year I of mathematics phase-in begins in 2007–2008. Year II of math phase-in begins in 2008–2009 with ninth graders. As these students progress through high school, the phase-in continues through the year 2011–2012. The GPS implementation will also affect state-mandated testing GHS GT and EOCT, resulting in the E-GHS GT and EOCT based on new curriculum standards. The December 2005 administration of the EOCT in English/language arts and science was based on GPS, while all other content areas were aligned to the QCC. The GHS GT in English/language arts and science will be combination of QCC and GPS since students received instruction based solely on QCC in some content areas (GPS, 2005).

The GPS version of the English/language arts tests and the science tests set new benchmarks, and for this reason the GPS test will not be equated to the former QCC versions. Therefore, comparisons of student performance in previous years will not be valid or appropriate. In the spring of 2006, the GHS GT reflected a transitional blueprint in

English/language arts and science that aligns to both the GPS and the QCC (Testing Update Memo, 2005).

One of the most recent attempted reforms enacted by the Georgia Senate called for the “Grade Integrity Act of 2007.” Senate Bill 9 is an attempt to protect teachers from pressures by administrators and parents to change student grades. Pressure from administrators could result in an ethics report to the Professional Standards Commission. Grades may be discussed and even changed, but any changes must be “clearly indicated in the student’s school records.” The bill unanimously passed the Senate and has moved on to the House of Representatives (Fain, 2007).

CHAPTER 3

RESEARCH METHODS

The review of literature revealed a variety of practices regarding student grades, state testing, and the incorporation of student grades within state testing. While some research exists that inflation of student grades occurs, there is also evidence that grade inflation is not a problem. While testing measurements are standardized, grading practices from teacher to teacher, school to school, and system to system are not. This study was carried out to determine the effects on teacher-assigned grades as a result of the implementation of Georgia's EOCT.

Purpose

Georgia students are required to take EOCT in eight subjects. The purpose of this study was to examine the impact of two of these EOCTs on student grades in a middle Georgia secondary school. According to the information brochure, *Georgia's End-of-Course Test (EOCT)*, the EOCT affects a student's grade in the following manner:

The EOCT is the final exam for an EOCT course. The student's final grade in the EOCT course will be calculated using the course grade as 85% and the EOCT score as 15%. The student must have a final course grade of 70 or above to pass the course and to earn credit toward graduation. (pp. 2-3)

Georgia has implemented EOCT as a graduation requirement that relies heavily on grades assessed and computed by individual teachers across the state. This study examined student grades before and after the implementation of the EOCT in United States history and American literature (Cox, 2006).

This study was conducted to determine the effect of the EOCT on students' final averages in two courses requiring the EOCT. Two of the eight courses requiring EOCT were examined:

United States history and American literature. Students take these two courses in the 11th grade in the same year in which they take the GHS GT. These two EOCTs were selected so that the GHS GT could be used as a covariate.

Research Questions

1. Did the adjusted mean of United States history final course grades change with the implementation of the EOCT?
2. Did the adjusted mean of American literature final course grades change with the implementation of the EOCT?
3. Did the adjusted mean course grade change by gender after the implementation of the EOCT?
4. Did the adjusted mean course grade change by ethnicity after the implementation of the EOCT?
5. Did the adjusted mean course grade change by course level after the implementation of the EOCT?

Hypotheses

The five hypotheses tested were as follows:

H_{01} There will be no statistically significant difference among adjusted mean student grades in United States history for the three student groups.

H_{02} There will be no statistically significant difference among adjusted mean student grades in American literature and composition for the three student groups.

H_{03} There will be no statistically significant difference among adjusted mean student grades by gender for the three student groups.

Ho₄ There will be no statistically significant difference among adjusted mean student grades by ethnicity for the three student groups.

Ho₅ There will be no statistically significant difference among adjusted mean student grades by course level for the three student groups.

Design

The quasi-experimental design consisted of three student groups, two dependent variables, one independent variable, and a covariate. Student group one consisted of first-time 11th grade students who took United States history and/or American literature and received a final grade. Student group two consisted of 11th grade students who took United States history and/or American literature and composition, received a final grade, and also took the pilot administration of the EOCT. The third student group was composed of 11th grade students who took United States history and/or American literature, received a final grade, and also took the EOCT in the year of implementation. All three student groups took the GHSGT, which was used as a covariate.

Population and Sample

The school in which the study took place is located in a suburban community between two larger metropolitan areas in middle Georgia. Total student enrollment in grades 9–12 was 1,394 in 2002–2003, 1,458 in 2003–2004, and 1,476 in 2004–2005. Nonprobability sampling technique was used to select all students in the 11th grade class from the school years 2002–2003, 2003–2004, and 2004–2005. Students in the sample were in the 11th grade for the first time and were attempting the HSGT for the first time. Students in the Program for Exceptional Children (PEC) were excluded due to nonstandard test administration, instructional techniques,

and curriculum content. In the final data analysis, some students were excluded due to missing test scores or grade information.

The county in which the study took place has one high school, two middle schools, four elementary schools, and one prekindergarten school. The population for the county was 23,639 in 2000, 26,235 in 2004, and is projected to be 29,634 in 2010. The per-capita income was \$19,126 in 1999, \$23,660 in 2002, and \$23,761 in 2003. In 2002, 12.4% of children ages 5–17 were below poverty level, ranking this county number 18 of 159 where the lowest level of children in poverty was 4% and the highest level was 37.7%. The racial composition of the population is approximately 69% White, 30% Black, and 1% other (Georgia Statistics System, n.d.; U.S. Census Bureau, 2006)

The racial framework of the high school in 2002–2003 and 2003–2004 was 27% Black, 72% White, and 1% other compared to the student population in the state of Georgia of 38% Black, 52% White, and 10% other in 2003 and 38% Black, 51% White, and 11% other in 2004. In 2004–2005, the school population was 25% Black, 74% White, and 1% other; the state student population was 38% Black, 49% White, and 13% other. Students in the “other” category include Asian, Hispanic, Native American/Alaskan Native, and multiracial (Governor’s Office of Student Achievement, 2004).

In 2004, there were 77 teachers in the high school—26 male, 51 female, 7 Black, and 70 White. Thirty held Bachelor’s degrees, 37 held Master’s degrees, and 10 held Specialist’s degrees. The average years of experience were 16.31, compared to the state average of 12.49. In 2005, there were 79 teachers—29 male, 50 female, 9 Black, and 70 White. Thirty held Bachelor’s degrees, 41 held Master’s degrees, and 7 held Specialist’s degrees. The average years

of experience were 14.76; the state average was 12.55 (Governor's Office of Student Achievement, 2005; data for 2003 were not available using this same Web site).

The school graduation rate (see Table 3.1) is below that for the state of Georgia and the dropout rate is above that for the state. The percentage of state attendance is generally better in both the high school and the system within which the school is located when comparing ≤ 5 days and >15 days in attendance; however, the high school and its system average better attendance than the state in the 6–15 day range. Overall, the state attendance is better than the attendance of the high school in this study. When comparing this school's performance on nationally administered tests such as the SAT, ACT, and AP exams, state and national averages are higher in every category than student averages at this high school (see Table 3.2). This high school outperforms its local RESA area and the state average when the GHSGT is used for comparison (see Table 3.3). This school falls behind state and national assessments in graduation rate, dropout rate, attendance, and performance on nationally standardized tests; the only category in which this high school outperforms others is the GHSGT.

Table 3.1

Comparison of Percentage Graduation Rate, Dropout Rate, and Attendance

	2004–2005			2003–2004			2002–2003		
	School	System	State	School	System	State	School	System	State
<u>Grad. rate</u>	64.6		69.4	61.5		65.4	61.		63.3
<u>Dropout rate</u>	5.4		5.0	7.6		5.1			5.5
<u>Attendance</u>									
≤ 5 days	43.3	45.1	54.1	39.7	42.8	52.6	41.8	50.2	49.9
6–15 days	39.6	41.7	35.3	40.4	41.8	35.4	41.1	36.8	36.2
>15 days	15.1	13.3	10.3	19.9	15.4	12.1	17.2	13.0	13.8

Source: Governor's Office of Student Achievement; Annual Report Card.

Table 3.2

School, State, and National Comparison of National Tests

	2004–2005			2003–2004			2002–2003		
	School	State	Nation	School	State	Nation	School	State	Nation
<u>SAT</u>									
Highest	978	1007		968	999		957	997	
Most recent	964	989	1020	956	981	1017	947	980	1016
<u>ACT</u>									
Composite	21	20	20.9	16.9	20	20.9	19.5	19.8	20.8
<u>AP exams</u>									
% students scoring 3 or better	50	55	—	41.2	56.5	—	—	—	—

Source: Governor’s Office of Student Achievement; Annual Report Card.

Note. — = data not reported

Table 3.3

School, RESA, and State Percent Pass Rate Comparison of HSGT

	2004–2005			2003–2004			2002–2003		
	School	RESA	State	School	RESA	State	School	RESA	State
ELA	97	93	95	97	94	93	99	96	95
Math	93	90	92	95	92	92	96	91	91
Science	85	80	83	84	81	82	85	80	81
Soc. studies	72	63	67	66	64	67	70	66	60
Writing									

Source: GA DOE; Office of Testing.

Instrumentation

The two testing instruments used in this study were the Georgia EOCT and the Georgia HSGT. The state legislature of Georgia mandated the development and adoption of the EOCT with the *A+ Education Reform Act* of 2000 (Georgia Code §20-2-281). The State Board of Education adopted EOCT in grades 9–12 for eight core high school subjects in 2000. The purposes of the EOCT are to ensure all students have access to a rigorous curriculum and to improve student achievement. The tests provide feedback to students and parents about student

achievement throughout grades 9–12 and also inform educators about instructional practices. Prior to the EOCT, the GHSGT was the only standardized benchmark indicating student success on the state curriculum. The EOCT allows areas of student and instructional strengths and weaknesses to be addressed early (Cox, 2005).

The development of the EOCT includes a contracted nationally recognized test development company and Georgia educators. All test items on the EOCT are approved by Georgia teachers as suitable and relevant to each course. Each EOCT includes a variety of items ranging from basic understanding to high achieving. This broad range of test items aids in the assessment of student knowledge and instructional strategies. The validity of the EOCT begins with test development and involves continuous review by content experts and Georgia educators for alignment and quality. Georgia DOE states that the reliability coefficient for all EOCTs is well above .70. The DOE cites the following examples on the spring 2004 EOCTs: .86 on the physical science test and .94 on the ninth grade literature and composition test (Cox, 2005).

In 1991, the Georgia General Assembly passed a law requiring curriculum-based assessments to be administered in grade 11 for graduation purposes (Georgia Code §20-2-281). Students who entered ninth grade in 1991 were required to pass tests in writing, English/language arts (ELA), and math; social studies was added in 1996 and science in 1997. The GHSGT was designed to certify high school students for graduation, with the State Board of Education retaining the authority to award or withhold a high school diploma based upon a student's score on the GHSGT. The tests are based upon the state-adopted curriculum and help to ensure that students have mastered the content necessary to justify the awarding of a high school diploma. Students have five opportunities to pass the HSGT before they graduate. Students may also continue to take the test after they have met all other graduation requirements and have left

school. Test results are reported on a scale of 400–600, with 500 indicating a passing score (Cox, 2005).

The GHSGT serves as the state's accountability measure for AYP under the *NCLB* act of 2001. In 2000 the U.S. Department of Education reviewed the GHSGT in ELA and math and found the tests did not adequately reflect the curriculum. Georgia DOE responded with an enhanced version of the test to be used for accountability purposes under *NCLB*. The Enhanced GHSGT was given for the first time in the spring of 2004. The spring administration is used for AYP reporting; thus, all retest forms are not enhanced. The test scores used in this study are from spring administrations when all juniors are tested for the first time. The Enhanced GHSGT was reviewed again by the U.S. DOE in summer 2004 and was found to successfully meet their requirements. The tests will receive a final review in spring 2006 (Cox, 2005)

Data Collection

Data for this study were collected from the SASI computerized student information system, a system used to collect and archive individual student data such as demographic, grade, testing, and attendance information. The data collected for this study were the student grades in United States history, the student grades in American literature and composition, the scores on the ELA HSGT, and the scores on the social studies HSGT. Data were exported from the SASI system to a Microsoft Excel spreadsheet. Attached to each score were student gender, ethnicity, and graduation program of study to enable the researcher to further disaggregate the data.

Analysis

A one-way analysis of covariance (ANCOVA) was used to determine whether the adjusted group means for student grades in United States history and American literature and composition differed significantly from group to group. The independent variable for this study

was the EOCT. The dependent variables were the mean student grade in United States history and the mean student grade in American literature and composition. The covariate was the GHSGT. Using SPSS (Statistical Software Package for the Social Sciences), the ANCOVA *F* test was performed to evaluate the effect of the EOCT on student grades in United States history and American literature and composition. A significance level of .05 was used. The data were further disaggregated by gender, ethnicity, and graduation program of study.

Limitations

Limitations for this study include the design of the study, the sampling technique, and different teaching styles and grading procedures among teachers. The quasi-experimental design of this study introduces threats to internal validity such as maturation, selection, and mortality (Creswell, 2002). The between-group design is problematic since the three student groups occurred in three separate school years. While the convenience sampling can provide useful information with regard to 11th grade students at this school, I cannot state with confidence that this sample is representative of all 11th grade students. Finally, the student grades in the two courses, United States history and American literature and composition, were a result of approximately 10 different teachers who taught these subjects during the 3 years of the observed data. Teacher instructional and grading techniques could have been a factor in accounting for variation in student grades among the three groups.

CHAPTER 4

PRESENTATION AND ANALYSIS OF DATA

The purpose of this chapter is to present the analysis of data, including the restatement of the purpose of the study, the research questions, an evaluation of the null hypotheses, and the population and sample. An overview of the data collection process and procedures is presented. Tables and summaries are included to aid in the interpretation of data.

Purpose

The purpose of this study was to examine the impact of two of Georgia's EOCTs on student grades in a middle Georgia secondary school. As colleges report rising grade point averages but lower SAT scores, as evidence mounts that grades assigned in high-poverty school systems and more affluent systems show a discrepancy, and as concerns about grade inflation are debated with regard to the Georgia HOPE scholarship, Georgia has implemented EOCT as a graduation requirement that relies heavily on grades assessed and computed by individual teachers across the state. This study examined student grades in American literature and United States history after the implementation of the Georgia EOCT.

Research Questions

The following research questions were asked for this study:

1. Did the adjusted mean of United States history final course grades change with the implementation of the EOCT?
2. Did the adjusted mean of American literature final course grades change with the implementation of the EOCT?
3. Did the adjusted mean course grade change by gender after implementation of the EOCT?

4. Did the adjusted mean course grades change by ethnicity after implementation of the EOCT?
5. Did the adjusted mean course grades change by course level after implementation of the EOCT?

Hypotheses

Five hypotheses tested were as follows:

Ho₁ There will be no statistically significant difference among adjusted mean student grades in United States history for the three student groups.

Ho₂ There will be no statistically significant difference among adjusted mean student grades in American literature and composition for the three student groups.

Ho₃ There will be no statistically significant difference among adjusted mean student grades by gender for the three student groups.

Ho₄ There will be no statistically significant difference among adjusted mean student grades by ethnicity for the three student groups.

Ho₅ There will be no statistically significant difference among adjusted mean student grades by course level for the three student groups.

Using an alpha of .05, Ho₁, Ho₄, and Ho₅ are accepted and Ho₂, Ho₃ are rejected. There was no statistically significant difference in the adjusted mean student grades in United States history, with $p = .100$. Also, there was no statistically significant difference in ethnicity and course level in either subject. There was statistical significance in the adjusted mean student grades in American literature, with $p = .005$. By gender, there was a statistically significant difference in American literature, with $p = .025$. No significant difference in gender was found in United States history.

Population and Sample

The population for this study was a high school in a suburban community between two larger metropolitan areas in middle Georgia. The sample consisted of 11th grade students who took the American composition and literature or United States history EOCT during the school years 2002–2003, 2003–2004, and 2004–2005. Students in 2003–2004 and 2004–2005 participated in the appropriate EOCT, while students in 2003–2003 did not take an EOCT. The sample was further designed to include only 11th grade students who were first-time test-takers of the Georgia HSGT. A student might have a grade in United States history or American literature, but if he did not have an HSGT score as a first-time test-taker within the year under study, he was not included in the sample. Also, students with disabilities were excluded due to a change in the curriculum for these students over the 3 years under study. The final sample of students comprised 11th grade students who were first-time test-takers of the Georgia HSGT and who had a final grade in United States history or American literature courses during the 3 years under study.

Data Collection

Data for this study were collected from the SASI computerized student information system, from the Georgia state report of HSGT results for the school in the study, and from individual student records. SASI is a computerized student information system used to collect and archive individual student data such as demographic, grade, testing, and attendance information. The data collected for this study were the student grades in United States history, student grades in American literature and composition, the individual student scores on the ELA HSGT and the scores on the social studies HSGT, gender, ethnicity, and level of instruction of the history and literature classes. Student gender, ethnicity, and the final teacher-assigned grade

were exported from the SASI system to a spreadsheet. For students in the study years 2002–2003 and 2003–2004, all final grades were teacher computed and assigned. For students in 2004–2005, the final teacher-computed and -assigned grade represented 85% of the student’s final grade prior to computing the final average, with 15% included for the EOCT. The teacher-computed and -assigned grade was the only grade used for this study. Individual student scores for the ELA and social studies HSGTs were input by hand using the student achievement roster provided by the testing division of the Georgia DOE. The resulting spreadsheet was organized with the following column headings: student number, year, gender, ethnicity, United States history grade, level, American literature grade, level, ELA HSGT score, and social studies HSGT score. The researcher visually inspected the data to make sure they were clean and there were no input errors or out-of-range scores. Students were eliminated if any part of their data was missing or could not be confirmed. The number of students eliminated was negligible and did not affect the statistical analysis of the remaining data.

Procedures

The SPSS statistical software was used to compute a univariate analysis of variance, adjusted for differences on a covariate (ANCOVA) in order to determine if there were any statistically significant differences in the mean course grades in American literature or United States history among the three student groups who had no EOCT, a pilot EOCT, and the implemented EOCT. The independent variable was the EOCT. The dependent variables were the changes in grades in literature or history. The Georgia HSGT was used as a covariate as a means for equating the student groups used over 3 different years. The HSGT is directly related to the two courses in the study and thus to student performance in those courses. The test of between-subjects effects revealed a statistically significant difference ($p \leq .05$) in student grades in

American literature and no statistically significant difference ($p \leq .05$) in student grades in United States history.

Table 4.1 displays the total number of participants in each student group. Also noted is the adjusted mean grade for literature and history. The literature adjusted mean grade remained constant for 2 years and increased the year of EOCT implementation. The history grade remained more constant, while taking a dip in the pilot year. This is consistent with the statistically significant difference found in the literature grade but not the history grade.

Table 4.1
Descriptive Statistics of Groups

		2002–2003	2003–2004	2004–2005
American literature	n	250	258	206
	X	80.56	79.75	81.64
	S	9.45	8.97	8.10
	adj X	79.92	79.91	82.20
U.S. history	n	245	256	201
	X	78.96	77.64	79.39
	S	8.34	8.66	8.99
	adj X	79.06	77.78	79.09

The student mean grade in American literature is presented in Table 4.2. The analysis of variance revealed a significant difference at an alpha of .05, $p = .005$. Table 4.3 presents a pairwise comparison of the student mean grade in American literature to further determine in which year the statistically significant difference occurred. The mean difference from year one to year two was negative .01. The mean difference from year one to year three was 2.27 and from year two to year three 2.28.

Table 4.2

Tests of Between-Subjects Effects for Dependent Variable American Literature

Source	Sum of Squares	df	F	Sig.
Am. literature	6721.75	1	96.13	.000
Year	750.98	2	5.37	.005
Error	49648.29	710		
Total	4692375.00	714		
Corrected total	56778.41	713		

R Squared = .126 (Adjusted R Squared = .122).

Table 4.3

Pairwise Comparison of Mean Difference of Student Groups for Three Years

Year	1	2
2	-.01	
3	2.27*	2.28*

* $p < .05$

Table 4.4 displays the student mean grade in United States history that was not statistically significant at $p = .05$.

Table 4.4

Tests of Between-Subjects Effects for Dependent Variable U.S. History

Source	Sum of squares	df	F	Sig.
U.S. history	11175.25	1	189.81	.000
Year	272.54	2	2.31	.100
Error	41095.93	698		
Total	4389559.00	702		
Corrected total	52662.52	701		

R Squared = .220 (Adjusted R Squared = .216).

Table 4.5 displays the mean grades by course and ethnicity. The mean grades for White students were highest during the implementation year of the EOCT in both subjects. The mean grades for Black students were highest in both subjects in 2002-2003, the year of no EOCT. The mean grades of Black and White students were about the same for literature for the first 2 years,

showing a wider margin during year 3; the difference in grades for history remained more consistent with Whites outperforming Blacks all 3 years.

Table 4.5

Descriptive Statistics by Ethnicity

		2002–2003		2003–2003		2004–2005	
		Black	White	Black	White	Black	White
American literature	n	68	178	59	199	54	150
	x	80.46	80.59	80.05	79.66	79.70	82.27
	sd	6.98	10.30	7.33	9.41	7.35	8.31
U.S. history	n	65	176	61	195	53	146
	x	77.88	79.40	75.56	78.29	76.91	80.26
	sd	8.43	8.36	8.72	8.56	6.77	9.59

There was no statistically significant difference ($p \leq .05$) in course grades for either American literature ($p = .459$) or United States history ($p = .950$) by ethnicity, as displayed in Tables 4.6 and 4.7.

Table 4.6

Tests of Between-Subjects Effects for Dependent Variable American Literature by Ethnicity

Source	Type III sum of squares	df	F	Sig.
Corrected model	7646.62 ^a	10	10.94	.000
Intercept	9.61	1	.14	.711
Am. literature	6718.84	1	96.14	.000
Year	49.72	2	.36	.701
Ethnicity	300.38	4	1.08	.368
Year*ethnicity	181.40	3	.86	.459
Error	49131.80	703		
Total	4692375.00	714		
Corrected total	56778.42	713		

^aR Squared = .135 (Adjusted R Squared = .122).

Table 4.7

Tests of Between-Subjects Effects for Dependent Variable U.S. History by Ethnicity

Source	Type III sum of squares	df	F	Sig.
Corrected model	11698.48 ^a	10	19.73	.000
Intercept	991.81	1	16.73	.000
U.S. history	10294.16	1	173.65	.000
Year	259.52	2	2.19	.113
Ethnicity	87.02	4	.37	.832
Year*ethnicity	20.88	3	.12	.950
Error	40964.04	691		
Total	4389559.00	702		
Corrected total	52662.52	701		

^aR Squared = .222 (Adjusted R Squared = .211).

Table 4.8 displays the mean grades by course and gender. The mean grades for female students were higher in both courses than those for male students with the exception of history grades for the implementation year, where males' grades were slightly higher than females'. While the grades for females remained fairly consistent, the grades for males showed a greater improvement from the pilot year to the implementation year in both courses.

Table 4.8

Descriptive Statistics by Gender

		2002–2003		2003–2003		2004–2005	
		Female	Male	Female	Male	Female	Male
American literature	n	121	129	146	112	116	90
	x	82.95	78.31	82.29	76.44	82.18	80.93
	sd	8.26	9.97	8.53	8.45	8.11	8.09
U.S. history	n	118	127	143	113	112	89
	x	80.27	77.73	78.97	75.96	79.00	79.88
	sd	7.72	8.74	8.04	9.15	8.40	9.71

Using an alpha of .05, there was a statistically significant difference by gender in American literature at $p = .025$, as noted in Table 4.9, and there was no statistically significant difference by gender in United States history with $p = .392$ as shown in Table 4.10.

Table 4.9

Tests of Between-Subjects Effects for Dependent Variable American Literature by Gender

Source	Type III sum of squares	df	F	Sig.
Corrected model	10533.42 ^a	6	26.84	.000
Intercept	23.35	1	.36	.550
Am. literature	6532.58	1	99.87	.000
Year	798.31	2	6.10	.002
Gender	2653.40	1	40.57	.000
Year*gender	487.77	2	3.73	.025
Error	46245.00	707		
Total	4692375.00	714		
Corrected total	56778.42	713		

^aR Squared = .186 (Adjusted R Squared = .179).

Table 4.10

Tests of Between-Subjects Effects for Dependent Variable U.S. History by Gender

Source	Type III sum of squares	df	F	Sig.
Corrected model	13115.92 ^a	6	38.42	.000
Intercept	1391.30	1	24.45	.000
U.S. history	11717.93	1	205.94	.000
Year	353.80	2	3.11	.045
Gender	1329.23	1	23.36	.000
Year*gender	106.72	2	.94	.392
Error	39546.60	695		
Total	4389559.00	702		
Corrected Total	52662.52	701		

^aR Squared = .249 (Adjusted R Squared = .243).

Table 4.11 displays the mean scores of students by course level. The greatest gain in mean grades among technology/career preparatory students was from the pilot year to the implementation year in United States history with a gain of 3.87 points. The mean grades of

honors students showed the biggest increase from the year of no EOCT to the pilot EOCT in American literature with 4.81 points. College preparatory students actually went down by .45 points in history and had their highest mean grade in literature in the 2002-2003 year.

Table 4.11

Descriptive Statistics by Course Level

		2002–2003			2003–2003			2004–2005		
		CP	H	TC	CP	H	TC	CP	H	TC
American literature	n	126	30	94	123	28	106	103	20	83
	x	82.80	83.30	76.67	80.58	88.11	79.58	82.59	87.25	79.10
	sd	7.01	12.33	10.07	8.38	9.88	7.77	8.23	8.05	7.05
U.S. history	n	105	24	116	116	26	114	97	18	83
	x	79.13	86.04	77.33	78.91	85.00	74.68	78.69	87.17	78.55
	sd	8.69	7.50	7.42	7.32	9.69	8.42	8.64	8.98	8.70

Note. CP = college preparatory; H = honors/gifted; TC = technology/career preparatory.

Test of significance by course level in American literature was $p = .053$ which is not statistically significant at the $p \leq .05$ as displayed in Table 4.12. Table 4.13 displays the United States history grades at $p = .202$ which are not statistically significant and is consistent with other findings in the history grades.

Table 4.12

Tests of Between-Subjects Effects for Dependent Variable American Literature by Course Level

Source	Type III sum of squares	df	F	Sig.
Corrected model	9470.99 ^a	9	15.66	.000
Intercept	632.25	1	9.41	.002
Am. literature	2422.44	1	36.05	.000
Year	507.11	2	3.77	.023
Class level	1685.43	2	12.54	.000
Year*class level	629.94	4	2.34	.053
Error	47307.43	704		
Total	4692375.00	714		
Corrected total	56778.42	713		

^aR Squared = .167 (Adjusted R Squared = .156).

Table 4.13

Tests of Between-Subjects Effects for Dependent Variable U.S. History by Class Level

Source	Type III sum of squares	df	F	Sig.
Corrected model	12356.64 ^a	9	23.57	.000
Intercept	541.24	1	9.29	.002
U.S. history	6656.25	1	114.28	.000
Year	142.16	2	1.22	.296
Class level	455.44	2	3.91	.020
Year*class level	348.51	4	1.50	.202
Error	40305.88	692		
Total	4389559.00	702		
Corrected total	52662.52	701		

^aR Squared = .235 (Adjusted R Squared = .225).

CHAPTER 5

SUMMARY, DISCUSSION, AND RECOMMENDATIONS

This chapter contains a summary of the study and its findings, discussion of the findings, and recommendations for further investigation.

Summary

This study was designed to examine student grades and Georgia's implementation of EOCT. I looked at student grades in United States history and American literature to determine if there was any statistically significant difference before and after the implementation of the EOCT. This study was significant because review of the literature revealed some concerns about student grades and how student grades are assigned. Review of the literature also revealed a concern with assessments of learning, such as the EOCT and the GHSGT, for the purposes of accountability required by *NCLB*. This chapter presents a restatement of the problem, summary of the key findings of the study, discussion of the findings and literature, and recommendations as a result of the study.

The purpose of this study was to examine the impact of two of Georgia's EOCTs on student grades in a middle Georgia secondary school. Five research questions were examined: (1) Did the adjusted mean of United States history final course grades change with the implementation of the EOCT? (2) Did the adjusted mean of American literature final course grades change with the implementation of the EOCT? (3) Did the adjusted means course grade change by gender after the implementation of the EOCT? (4) Did the adjusted means course grade change by ethnicity after the implementation of the EOCT? (5) Did the adjusted means course grade change by course level after the implementation of the EOCT?

The first null hypothesis with regard to United States history was accepted. The second null hypothesis regarding American literature was rejected. Using an alpha of .05, there was no statistically significant difference in the adjusted mean student grades in United States history, with $p = .100$. There was statistical significance in the adjusted mean student grade in American literature, with $p = .005$. The adjusted mean grade for literature remained constant for 2 years and then increased by 2.29 points. The adjusted mean grade for history dipped 1.32 points from year 1 to year 2 and then rose 1.75 points in year 3. The difference in the adjusted mean score in history for year 1 and year 3 was .43 points. Disaggregating the data by ethnicity showed the greatest difference in adjusted mean scores to be in the implementation year when Whites outperformed Blacks by 3.35 points in history and 2.57 points in literature.

Females consistently outperformed males every year, with the exception of history in the implementation year, where males outperformed females by .88 points. By gender, there was a statistically significant difference in American literature, with $p = .025$, but not United States history. Mean scores among males in American literature increased from year one to year three by 2.62 points and from year two to year three by 4.49 points. Mean scores for females in American literature from year one to year three changed negative 0.77 and negative 0.11 points from year two to year three. Males also made bigger gains in mean scores in United States history. From year one to year three gains were 2.15 points and year two to year three gains were 3.92 points. Again, female scores did not produce the same increase in scores as males with a negative 1.27 mean difference from year one to year three and .03 difference from year two to year three.

Discussion

Student grades are a fundamental part of our education system, an education system that is being held accountable for student learning. The possibility exists for students to earn a passing grade in a course yet fail a state test in the same content area (Edwards, 2000; Evers, 2001). The Georgia EOCT is currently 15% of final student grades in eight high school courses. At one time O.C.G.A §20-2-281 stated that “local boards of education shall have the option of allowing scores on end-of-course assessments to be counted as part of a student’s grade in the course.” O.C.G.A. §20-2-281(j)(2) currently reads, “The State Board of Education shall adopt rules and regulations requiring the results of core subject end-of-course assessments to be included as a factor in a student’s final grade in the core subject course....” The Georgia DOE Web site states that “a student’s EOCT score is averaged in as 15% of the final course grade.” There are two established purposes for the EOCT- “to ensure all Georgia students have access to a rigorous curriculum that meets high expectations and to provide information to improve student achievement through effective instruction of the standards in the state-adopted curriculum ”(p.9) (Cox, 2007). An end-of-course assessment is at best a diagnostic tool for teachers to assess a student’s current level of understanding of a subject. By the time the test is administered, credit for the course is all but established. No other Georgia mandated tests, Criterion-Referenced Competency Tests (CRCT) nor the GHSGT, are linked to student grades.

The mean student grade in American literature from this study was 82. Using the state formula of 85%, the student grade is 69.7. Theoretically, a student could score as low as 2 on the EOCT and still get credit for the course with 15% of 2 equal to .3 and 69.7 plus .3 equaling 70. The United States history mean grade of 79 could result in an EOCT score as low as 19. Eighty-five percent of 79 is 67.15 and 15% of 19 is 2.85, which equals 70.

Grade inflation is a concern from middle school through postsecondary education (Felton & Koper, 2005; Hassel & Lourey, 2005; Hu, 2005; McCabe & Powell, 2005; Riley, Checca, Singer, & Worthington, 1994). Teacher grading practices vary from classroom to classroom, school to school, and system to system and also differ from elementary, middle, high, and secondary levels (Bonesronning, 2004; Johnson, 2003; McMillan, Myron, & Workman, 2002; Padgett-Harrison, 2000). In this study, adjusted mean grades for American literature did increase at a statistically significant level. Overall, adjusted mean scores were higher in both literature and history during the EOCT implementation year. Grade inflation and individual teacher grading practices are two possible reasons for the rise in mean scores.

The results of this study further indicated differences in mean scores for males and females. Literature indicated that girls in grades 4, 8, and 12 consistently outperform boys in reading and writing (Dessoff, 2005a). Girls value reading and English somewhat more than boys while boys are reported to have higher interest in mathematics and science (Durik, Vida, & Eccles, 2006; Meece, Gilenke, & Burg, 2006). The 2000 study of the U. S. DOE's National Center for Education Statistics reported that girls are closing the gender gap in the area of mathematics while boys are still more likely to have problems that affect their academic performance (Dessoff, 2005a). This study used two courses heavily depend upon reading and writing where males are notably weaker yet males mean grades improved more than females. This could be further indication of grade inflation by a teachers' desire to ensure males pass the courses despite their performance on the EOCT.

For teachers, assessment means grades and grades equate to performance. This can undermine students' interest in learning, their desire to be challenged, and, ultimately, the extent of their achievement. When students are given the opportunity to rewrite rather than be penalized

with a bad grade, the emphasis becomes one of improvement and learning rather than grading (Kohn, 1993). At a time when teachers might be stressing the importance of grades due to such factors as the Georgia EOCT, Kohn offers seven suggestions for teachers to minimize the importance of grades. The suggestions are:

- Limit the number of assignments for which a letter or number grade is given.
- Limit the number of gradations.
- Reduce the possible number of grades to two: A or Incomplete.
- Never grade students while they are still learning something.
- Never grade for effort.
- Never grade on a curve.
- Bring students in on the evaluation process.

The role of student assessment in educational improvement can only be constructive if the assessments help teachers to identify students' strengths and weaknesses and to determine the most appropriate and effective means of helping students to learn and achieve (Denbo et al., 1994). Stiggins (2002) reported that few states require competence in assessment as a condition to gaining a teaching license. Also, there is no examination that verifies competence in assessment. Professional associations have come together to adopt professional competence standards that include an assessment component.

State-mandated testing such as the GHSGT and EOCT are assessments of student learning. This type of high-stakes testing makes formative assessment or assessment for student learning difficult for teachers to implement. Pressure for schools to meet AYP by *NCLB* legislation requires that teachers make certain students score well on standardized test. Teachers struggle to do what is best for student learning to take place. Teaching for understanding gains

better results in long-term retention than rote recall of facts (Nuthall & Alton-Lee, 1995). In the case of the Georgia EOCT, student grades are tied to performance on a standardized test. This further complicates teacher efforts to maintain the integrity of assessing student performance. Teacher knowledge of assessment strategies and techniques is one thing, but application in the classroom is another. Teachers must be provided with the tools and workshops to support teacher development and understanding of assessments for learning. Professional learning is the key to implementation of assessment strategies into practical classroom practice (Leahy, Lyon, Thompson, & Wiliam, 2005).

Recommendations

Student grades are most often used to determine if student achievement or student learning has taken place. This student achievement in turn is measured on large-scale assessments or standardized tests that then determine school systems' AYP. Student grades, teacher grading practices, classroom assessments, and standardized testing are interconnected elements used to determine student achievement. If student achievement is to be the standard by which a school is measured, then educational leaders should develop a plan to raise standards through student achievement. The following recommendations are made with that in mind.

Pedagogic change in assessing student learning must take place through proactive leadership, professional trust, dialog, and a "start small" strategy (Priestly & Sime, 2005). As Georgia moves to the standards-based curriculum, more assessment for learning will take place. In the school in this study, the English and science GPSs are in the second year of implementation. Proactive leadership, professional trust, and dialog must continue to ensure that all teachers fully participate in the GPS curriculum and follow the principles of assessment *for*

learning. These supporting principles must continue to develop as the social studies GPS rollout begins in 2007–2008 and math follows in 2008–2009.

All stakeholders—students, parents, teachers, administrators, and community—should be informed about grades and grading practices that destroy “learning as discovery” (Kohn, 1993). The misconception is that students exert themselves for a grade when in fact they should be seeking knowledge. A grade is an indication of performance and not necessarily knowledge or learning. A paradigm shift should occur from the current attitude about grades as the indicators of learning, to teacher assessment for learning, to authentic student achievement. Educational leaders must direct efforts to decrease the significance of grades alone and enhance the worth of individual student assessment *for learning*.

Local boards of education and policy-makers should investigate the purpose and wisdom of the Georgia EOCT and its inclusion in student grades. The EOCT can be used as a diagnostic tool to help determine student weaknesses and strengths in certain core areas. Further, the EOCT might be an assessment used to predict success on the GHSGT, a subject for a future study. As illustrated in the example on page 58, the EOCT counts a negligible 15% of a student’s grade in one of eight courses; however, the GHSGT determines whether a student receives a high school diploma. As early as 1999, HB 308 suggested that “a student may not receive course credit for Algebra I, American and Georgia Government, American History, American Literature, Biology I, Chemistry I, Geometry, or Writing and Composition until he or she has earned a passing score on the applicable end-of-course assessment instruments” (Rivers, Jr. 1999). Then Georgia Code § 20-2-281-G (i) (3) stated that “local boards of education shall have the option of allowing scores on end-of-course assessments to be counted as part of a students’ grade in the course”. Discussion continued on what function the EOCT would serve as the state BOE discussed

whether the test would count 20% or 15% of a students' final grade. In the past eight years the EOCT appears to have generated time, energy, and expenditures and progressed from a 'students must pass' position of legislators to 15% of a students' final grade. Further, the EOCT has become a part of the NCLB demands upon the testing industry. Georgia is but one state contributing to the oligopoly of four major testing companies (Reichgott, 2007). One must ask if Georgia's expenditures on both the EOCT and the GHSGT are warranted as education cutbacks continue to occur along with increased class size.

A final recommendation is to support teacher development and understanding of learning assessment. Professional learning is the key to the implementation of accepted practices and research-based strategies into practical classroom practice (Leahy et al., 2005). Assessment for learning depends heavily on what individual teachers do in their classrooms. Once the door closes on the "black box," teachers are autonomous and must maintain the integrity of student grades. Teachers must be encouraged to make changes through a supportive environment that involves school leadership, boards of education, policy-makers, and governments (ARG, 1999). Ongoing professional development in the areas of student grades, teacher grading practices, assessment for learning, and assessment of learning is recommended.

REFERENCES

- Airasian, P. W. (1994). *Classroom assessment* (2nd ed.). New York: McGraw–Hill.
- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*, University of Cambridge School of Education, U.K. Retrieved March 1, 2007, from: http://qca.org.uk/downloads/beyond_black_box2.pdf.
- Bett, J. R., & Grogger, J. (2003). The impact of grading standards on student achievement, educational attainment, and entry-level earnings. *Economics of Education Review*, 22, 343–352.
- Bilimoria, D. (1995). Modernism, postmodernism, and contemporary grading practices. *Journal of Management Education*, 19, 440–458.
- Bishop, J. H. (1992). Why U.S. students need incentives to learn. *Education Leadership*, 49(6), 15–18.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 138–148.
- Bonesronning, H. (2004). Do teachers' grading practices affect student achievement? *Education Economics*, 12(2), 151–167.
- Campbell, N. D., & Melton, K. I. (2001). The fable of the “B”s. *Business Quest: A Journal of Applied Topics in Business and Economics*. Ed: Carole E. Scott, Richard College of Business, University of West Georgia ISSN 1084-3981.
- College Board, The (1997, August 26). *News from The College Board*. New York: Author. <http://www.collegeboard.com/press/senior97/970826.html>.
- Cox, K. (2005). *What Georgia educators need to know about Georgia's testing program*. [Brochure]. Georgia Department of Education, Testing Division.
- Cox, K (2006). *Georgia's End-of-Course Tests (EOCT). Questions and answers for parents of Georgia students*. [Brochure]. Georgia Department of Education.
- Cox, K. (2007). *What Georgia educators need to know about Georgia's testing program*. [Brochure]. Georgia Department of Education, Testing Division.
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Merrill Prentice Hall.

- Denbo, S., Grant, C., & Jackson, S. (with Gillette, M., McCune, S., & Slavin, R.). (1994). *Educate America: A call for equity in school reforms*. Chevy Chase, MD: The Mid-Atlantic Consortium, Inc.
- Dessoff, A. (2005a). Girls lead boys in reading, writing. *District Administrator*, 41(3), 27-27. Retrieved April 1, 2007, from <http://www.districtadministration.com/viewarticle.aspx?articleid=668&p=5>
- Dessoff, A. (2005b). High school reform. *District Administrator*, April. Retrieved June 4, 2005, from: <http://www.districtadministration.com/page.cfm?p=1043>.
- Durik, A., Vida, M., Eccles, J. (2006). Task values and ability beliefs as predictors of high school literacy choices: A developmental analysis. *Journal of Educational Psychology*, 98(2), 382-393.
- Edwards, C. H. (2000). Grade inflation: The effects on educational quality and personal well being. *Education*, 43, 483-501.
- Emeagwali, N. S. (2004). As an increasing number of states adopt exit exams, concerns continue to grow. *Techniques: Connecting Education & Careers*, 79(7), 16.
- EOCT. (2005). Retrieved November 14, 2005 from Georgia Department of Education Web site: http://www.doe.k12.ga.us/ci_testing.aspx?PageReq=CI_TESTING_EOCT.
- Evers, B. (2001, August 20-27). What do tests tell us? Hoover Institution. Retrieved February 21, 2005, from: http://www-hoover.stanford.edu/pubaffairs/we/current/evers_0801.html.
- Fain, T. (2007, February 14). Grade integrity act passes Senate. *The Telegraph*, p. 3B.
- Feldman, K. A. (1976). Grades and college students' evaluation of their courses and teachers. *Research in Higher Education*, 4, 69-111.
- Gaylor, K. (2005). How have exit exams changed our schools: Some perspectives from Virginia and Maryland. Retrieved September 4, 2005, from: <http://www.cepdc.org/index.cfm?fuseaction=Page.viewPage&pageId=493&parentID=481>
- Georgia Code. (2006). Retrieved November 4, 2005, from Georgia General Assembly Web site: http://www.legis.ga.gov/cgi-bin/gl_codes_detail.pl?code=20-2-1.
- Georgia Statistics System. (n.d.) Retrieved January 8, 2006, from: <http://www.georgiastats.uga.edu/>.
- Georgia Student Finance Commission. (2005, August, 9). High School Academic Requirements. Retrieved February 19, 2006 from: http://www.gsfc.org/hope/dsp_hop2000.cfm.

- GHS GT. (2005). Retrieved November 14, 2005 from Georgia Department of Education Web site: http://www.doe.k12.ga.us/ci_testing.aspx?PageReq=CI_TESTING_GHS GT
- Governor's Office of Student Achievement. State of Georgia. Online Reports, Report Card. Retrieved January 15, 2006, from: <http://www.gaosa.org/>.
- GPS. (2005) Retrieved November 14, 2005, from Georgia Department of Education Web site: <http://www.georgiastandards.org/>.
- Guskey, T. R. (2002). *How's my kid doing? A parents' guide to grades, marks, and report cards*. San Francisco, CA: Jossey-Bass.
- Hamilton, L.S., Stecher, B. M., & Klein, S. P. (Eds.). (2002). *Making sense of test-based accountability in education*. Santa Monica, CA: Rand.
- Hassel, H., & Lourey, J. (2005). The dea(r)th of student responsibility. *College Teaching*, 53(1), 2-13.
- Herring, C. P. (2005). About the *No Child Left Behind Act* of 2001. Retrieved January 15, 2006, from Georgia Department of Education Web site: <http://public.doe.k12.ga.us/aypNCLB.aspx?PageReq=AboutNCLB>.
- High Schools That Work* (2005). Best practices for implementing HSTW and MMGW. Southern Regional Education Board. July 13-16, 2005.
- Hu, S. (2005). Beyond grade inflation: Grading problems in higher education. *ASHE Higher Education Report* 30(6).
- Indiana Core 40 End-of-Course Assessments. (2006). Retrieved January 15, 2006, from the Indiana Department of Education Web site: <http://www.doe.state.in.us/core40eca/welcome.html>.
- Johnson, V. E. (2003). *Grade inflation: Crisis in college education*. New York: Springer.
- Kohn, A. (1993). *Punished by rewards. The trouble with gold stars, incentive plans, A's, praise, and other bribes*. New York: Houghton Mifflin Company.
- Kohn, A. (2004) *What does it mean to be well educated? And more essays on standards, grading, and other follies*. Boston: Beacon Press.
- Leahy, S., Lyon, C., Thompson, M., & Wiliam, D. (2005). Classroom assessment: Minute by minute, day by day, *Educational Leadership*, 63(3), 18-24.
- MacDonald, M. (2005, October 12). Waiver system put to test. *The Atlanta Journal-Constitution*, pp. D1, D4.

- Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McCabe, J., & Powell, B. (2004). In my class? No. In W. E. Becker and M. L. Andrews, Eds., *The scholarship of higher education: Contributions of research universities*. Bloomington: Indiana University Press. (Faculty Perceptions, Reality of grade inflation tell two different stories)
- McMillan, J. H., Myron, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *Journal of Education Research*, 95(4), 203–213.
- Meece, J., Gilenke, B., Burg, S. (2006) Gender and motivation. *Journal of School Psychology*. 44(5), 351-373.
- Mississippi SATP. (2006). Office of Student Assessment. Retrieved January 15, 2006, from the Mississippi Department of Education Web site:
<http://www.mde.k12.ms.us/acad/osa/satp.html>.
- National Center for Education Statistics. (2004). State education reforms. Names and types of statewide assessments administered, by state: Table 1.5. Retrieved June 4, 2005, from U.S. Department of Education Web site:
http://nces.ed.gov/programs/statereform/saa_tab5.asp?referrer=tables
- National Center for Education Statistics. (2005). *The condition of education 2005*, NCES 2005-094, Washington, DC: U.S. Government Printing Office. Retrieved June 4, 2005, from U.S. Department of Education Web site:
<http://nces.ed.gov/programs/coe/2005/section4/indicator24.asp>.
- NCLBGrassroots.org, Citizens for Community Action. *NCLB left behind: Understanding the growing grassroots rebellion against a controversial law*. Retrieved November 6, 2005, from: <http://www.NCLBgrassroots.org/landscape.php>.
- New York State Department of Education. (2006). Elementary, middle, secondary, and continuing education. Education requirements/exams/tests/assessments. Retrieved January 15, 2006, from: <http://www.emsc.nysed.gov/part100/pages/1005a.html>.
- Nichols, S. L., Glass, G. V., & Berliner, D. C. (2005). *High-stakes testing and student achievement: Problems for the No Child Left Behind Act*. Educational Policy Studies Laboratory, division of Education, Arizona State University. EPSL-0509-105-EPRU. Retrieved November 6, 2005, from:
http://greatlakescenter.org/Center_Funded_Research_2005.php.
- No Child Left Behind. (2001). Retrieved June 4, 2005, from the U.S. Department of Education Web site: <http://www.ed.gov/NCLB/landing.jhtml>.

- North Carolina Testing Program Overview. (2005). North Carolina Public Schools. Accountability Services Division. Retrieved August 9, 2005, from Public Schools of North Carolina Web site: <http://www.ncpublicschools.org/accountability/testing/>.
- Nuthall, G., & Alton-Lee, A (1995). Assessing classroom learning: How students use their knowledge and experience to answer classroom achievement test questions in science and social studies. *American Educational Research Journal*, 32(1), 185–223.
- Olsen, D. R. (1997). *Grade inflation: Reality or myth? Student preparation level vs. grades at Brigham Young University, 1975–1994*. Paper presented at the annual forum of the Association for Institutional Research, May 18-21, Orlando, FL.
- Padgett-Harrison, S. K. (2000). *Influences on secondary teacher grading practices*. Unpublished doctoral dissertation, University of Georgia, Athens.
- Phi Delta Kappa International. (2004, June 1). Phi Delta Kappa (PDK) International's curriculum management center conducts audit of the state of Georgia's proposed performance standards. Retrieved July 7, 2005, from: <http://www.pdkintl.org/icmc/pr0406.htm>.
- Popham, J. (2006). Assessment for learning: An endangered species? *Educational Leadership* 63(5), 82–83.
- Priestly, M., & Sime, D. (2005). Formative assessment for all: A whole-school approach to pedagogic change. *The Curriculum Journal*, 16(4). 475-492.
- Quality Basic Education Act. (2005). Retrieved January 15, 2006, from Georgia Department of Education Web site: <http://www.glc.k12.ga.us/qcc/qbe.htm>.
- Quality Core Curriculum. (2005). Retrieved January 15, 2006, from Georgia Department of Education Web site: <http://www.glc.k12.ga.us/qcc/overview-qcc.htm>.
- Reichgott, M. (2007). 'No child' test demands taxing system. Associated Press. Retrieved March 25, 2007 from: http://www.boston.com/business/articles/2007/03/25/no_child_test_demands_taxing_system/?page=1
- Research Corner (2005). District Administration, April. Retrieved June 4, 2005, from: <http://www.districtadministration.com/page.cfm?p=1043>.
- Riley, H. J., Checca, R. C., Singer, T. S., & Worthington, D. F. (1994). *Current trends in grades and grading practices in undergraduate higher education: Results of the 1992 AACRAO Survey*. Washington, DC: American Association of Collegiate Registrars and Admissions Officers.

- Rivers, Jr., R. E. (1991) Clerk of the Georgia House of Representatives. House Bill 308 Retrieved March 30, 2007 from:
http://www.legis.state.ga.us/legis/1999_00/fulltext/hb308.htm
- Rose, L. C., & Gallup, A. M. (2005). The 37th annual phi delta kappa/gallup poll of the public's attitudes toward the public schools. *Phi Delta Kappa*, 87(1), 41–57.
- Roth, W. F. (2005). Yes, tests still rule. *American School Board Journal*, 192(5), 57–58.
- Safer, A. M., Farmer, L. S. J., Segalla, A., & Elhoubi, A. F. (2005). Does the distance from the teacher influence student evaluation? *Educational Research Quarterly*, 28(3), 28–35.
- Salzer, J. (2003, October 22). Commission recommends changes in HOPE scholarship. *The Atlanta JournalConstitution*. Retrieved February 21, 2005, from: <http://www.ajc.com>
- Schouten, F. (2003, October 21). Grade inflation takes a toll on students: Many need remedial courses in college. *USA Today*. Retrieved January 8, 2005, from:
http://www.usatoday.com/news/education/2003-10-20-grades-usat_x.htm.
- State Education Rules. (2006). Retrieved January 15, 2006, from Georgia Department of Education Web site:
http://public.doe.k12.ga.us/pea_board.aspx?PageReq=PEABoardRules.
- Stiggins, R. J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, 83(10), 758–765.
- Stumpf, S. A., & Freedman, R. D. (1979). Expected grade covariation with student ratings of instruction: Individual versus class effects. *Journal of Educational Psychology*, 71, 293–302.
- Sullivan, P., Yeager, M., Chudowsky, N., Kober, N., O'Brien, E., & Gayler, K. (2005). State high school exit exams: States try harder, but gaps persist. Center on Education Policy, Washington, DC. Retrieved November 1, 2005, from: <http://www.cep-dc.org/index.cfm?fuseaction=Page.viewPage&pageId=493&parentID=481>.
- Tennessee.gov (2005). Assessment, evaluation, and research. Secondary TCAP assessments. Retrieved August 9, 2005, from the official Web site of the state of Tennessee:
<http://www.state.tn.us/education/assessment/tshseochome.shtml>.
- Testing update memo. (2005). Retrieved January 15, 2006, from Georgia Department of Education Web site:
http://public.doe.k12.ga.us/pea_communications.aspx?ViewMode=1&obj=956.
- Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan*, 82(5), 358–362.

- Tofig, D. (2005, July 8). Overview of accountability AYP & SSAS: Merging SCLB & state requirements. Retrieved November 4, 2005, from Georgia Department of Education Web site: http://public.doe.k12.ga.us/pea_communications.aspx?ViewMode=1&obj=918.
- Ullman, E. (2005). Inside the law. *District Administrator*, November. Retrieved November 6, 2005, from: <http://www.districtadministration.com/page.cfm?p=1318>.
- U.S. Census Bureau. (2006, January 12). State & county quickfacts. Jones County, Georgia. Retrieved January 25, 2006, from: <http://quickfacts.census.gov/qfd/states/13/13169.html>.
- U.S. Department of Education, Office of Educational Research and Improvement (1994). *What do student grades mean? Differences across schools*. (Office of Research Report 94-3401). Washington, DC: Office of Educational Research and Improvement. <http://www.ed.gov/pubs/OR/ResearchRpts/grades.html>.
- Virginia DOE. (2005). Project graduation. Retrieved August 9, 2005, from the Virginia Department of Education Web site: <http://www.pen.k12.va.us/2plus4in2004/index.shtml>.
- Volante, L. (2006/2007). Standards-based reform: Can we do better? *Education Canada*. Winter 47(1), 54–56.