# EFFECTS OF STUDENT SELF-CORRECTIVE MEASURES ON LEARNING AND STANDARDIZED TEST SCORES

A Dissertation

Presented to

The Faculty of the School of Education

Liberty University

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

by

Beth D. Poplin

2009

### Effects of Student Self-Corrective Measures on Learning and Standardized Test Scores

by Beth Poplin

APPROVED:

COMMITTEE CHAIR

Beth Ackerman, Ed.D.

COMMITTEE MEMBERS

William Gribbin, Ph.D.

Clarence Holland, Ed.D.

CHAIR, GRADUATE STUDIES

Scott Watson, Ph.D.

#### Abstract

Beth Poplin. EFFECTS OF STUDENT SELF-CORRECTIVE MEASURES ON LEARNING AND STANDARDIZED TEST SCORES. (Under the direction of Dr. Beth Ackerman) School of Education, 2009.

This study examined whether students who graded and corrected their own test papers improved their learning and standardized test scores on the North Carolina end-of-course test in United States History. Four preexisting, intact classrooms of 11<sup>th</sup> grade United States History students in two different high schools formed the basis of this quasiexperimental, Static Group Comparison Design. Two classes formed the control group, and two classes participated in the alternative assessment strategy, with both groups taking the pretest and posttest in United States History. The control group had their weekly tests graded by the classroom teacher and returned to them, while the experimental group self-graded and corrected their test papers by using a predetermined format focusing on the questions' main ideas. As the semester concluded, each class took the state end-of-course test in United States History. After comparing and analyzing scores, using descriptive statistics and the statistical procedure independent samples ttest, this research study determined it was unlikely the treatment had a positive statistical relationship to higher standardized test scores or that students learned more than with teacher-only grading. Finally, the researcher failed to reject the null hypothesis because students in the treatment group did not achieve statistically higher scores on the North Carolina end-of-course test in United States History than students in the control group.

iii

## CONTENTS

Abstract	iii
Contents	iv
Acknowledgements	viii
List of Tables	ix
List of Figures	X

CHAPTER ONE: INTRODUCTION TO THE STUDY	1
Background of the Study	1
Trends	2
Developments	3
Problems	5
Statement of the Problem	11
Research Questions	11
Null Hypothesis	11
Professional Significance	12
Definition of Key Terms	19
Summary	23

CHAPTER TWO: 1	REVIEW OF THE LITERATURE	. 25
Theoretical Review		.25

Constructivist Learning	26
Experiential Learning	28
Mastery Learning	30
Empirical Review	35
Current Learning Trends	35
Performance Strategies	46
Benefits	62

5
3
4
4
8
8
9
1
1
2
2
3
4
5

Procedures Used
Data Analysis91
Data Reduction91
Statistical Reporting and Display92
Analysis: Statistical Tests and Procedures93
Summary of Methodology
CHAPTER FOUR: FINDINGS AND RESULTS
Research Question 1
Research Question 2
Null Hypothesis
Summary

CHAPTER FIVE: SUMMARY AND DISCUSSION10	02
Statement of Problem	102
Review of Methodology	102
Summary of Results	.104
Discussion of Results	.106
Relationship to Current Literature	111
Implications for Practice	.117
Explanation of Unanticipated Findings	119
Limitations	122

aggestions for Further Research
ppendixes126
Appendix A: Liberty IRB Approval Form127
Appendix B: Sample Weekly Test128
Appendix C: Test Correction Criteria135

References13
--------------

#### Acknowledgements

When I began working on this project, I had no idea of the length and depth it would take me into myself and away from everything else. There were times I wondered how I got as far I did or where I would get the energy to finish. Several people I wish to acknowledge because without them my journey would have been much more difficult and possibly the difference that kept me going when things looked the darkest.

I want to first acknowledge and thank my Chair, Dr. Beth Ackerman, whose tireless patience and understanding helped keep me sane with her quick replies and obvious desire for me to succeed. I want to thank the other members of my committee, Drs. Gribbin and Holland, for taking the time to explain and work with me, especially at the beginning of the process when I was still trying to search for a topic and gather my academic wits. My committee always left me feeling challenged, but in such a way that I knew they cared deeply and wanted me to give my best.

I also want to express my appreciation to Liberty University, their graduate degree program, and especially the Graduate Writing Center. I know I would have never found another place that cared so much about the individual. Liberty worked with me because they wanted me to succeed and I knew I could always find help, both from God and from them.

Finally, I want to thank the people close to me who made all the difference. To my loving husband, Tim, who listened and helped me brainstorm long into the nights, who believed in me because he knew how hard I worked and not because he took for granted I would succeed - to him I want to say a truly heartfelt thank you and a deep

appreciation. I also wanted to thank my mother, Jancie, my father, Bob, and my brother, Bert, because without exception they supported my decision to begin the doctoral program and then stood with me until the conclusion.

Finally, I want to express thanks to those friends who helped me when they could with advice, moral support, editing, or just being there to bounce ideas around: Elizabeth Soderberg, Marion Buchheit, Cecil Kemp, Dianne Drum, and Barbara Mounce. A hearty thank you as well, to all the students who have passed through my door over the last several years on their way to academic success because they helped keep me focused on my goals when I was afraid and tired.

## List of Tables

Table 1	
Table 2	
Table 3	
Table 4	

## List of Figures

Figure 1	•••••	 • • • • • • • • • • • • • • • •	•••••	• • • • • • • • • • • • • • • •	•••••	95
Figure 2		 				96

#### CHAPTER ONE: INTRODUCTION TO THE STUDY

This study examined whether students who graded and corrected their own test papers improved their learning and standardized test scores on the North Carolina end-ofcourse test in United States History. This dissertation challenged the belief that there is no merit in student self-grading and correcting (Sadler & Good, 2006). Teacher grading has been the standard that measured a classroom grade and student progress (2006). While educators have always sought reliable means of improving test scores, perhaps adopting different grading practices—practices that include student involvement—on class assessments could be an effective way of improving student learning and thereby raising standardized test scores.

This study used the classroom practice of students self-grading and correcting their test papers. The intended goal was to discover if this treatment could improve student-learning and standardized test scores on the North Carolina United States History end-of-course test.

#### Background of the Study

The following section is a discussion of the background behind this research study. It considers trends, such as a growing importance on standardized test scores that have been the driving force behind education policy in recent years and current developments to elicit change. There is special attention paid to some of the problems educators are facing with the growing importance on standardized test scores. The section concludes with a statement of the problem and the null hypothesis under investigation.

#### Trends

Student achievement and standardized testing have more prominently figured into national and state politics in recent years, especially because of President Bush's No Child Left Behind legislation. The growing trend has been an increasing focus on highstakes testing and tactics to raise standardized test scores (Horn, 2003). Today's schools revolve around high-stakes testing and concurrently showing improvement at district and state levels. Most states, North Carolina included, have devised a standard curriculum, which classroom teachers must follow and implement. Classroom teachers are responsible for covering particular goals and objectives, which correlate with the mandated curriculum. At the end of the course, students take their states' matching endof-course assessments.

The schools are increasingly under pressure to plan test strategies, and this has become a rising trend since the latter part of the 20<sup>th</sup> century. William Hayes (2006), for instance, commented recently on the changes that have affected the nation's schools by comparing the education practices of the past and the emerging strategies of today. Traditionally, he contended, the teacher's job was to impart the information and skill necessary to survive in society. Students were to be passive recipients in their education, and the classroom teachers singularly decided the lesson contents.

As the 20<sup>th</sup> century progressed, so did the idea that students learn best by engaging in activities, which provide hands-on experience, and not by being unmotivated learners. The teacher's role began to evolve and become more like a facilitator of learning to help provide an intense and productive academic experience (Hayes, 2006). The new practice of allowing students a more proactive and involved role in their education had been gaining momentum and support. In recent years, though, ideals that are more traditional have returned because of standardized testing. Hayes argued that the new fashion of standardized testing caused schools to employ four strategies that have had a profound impact: back-to-basics movements, mandated state curriculum standards, high-stakes testing, and increased school accountability.

Standardized testing, changes in curriculum standards, and emphasis on more traditional methods of instruction are winning popularity, not just in the United States, but also in other countries (Phelps, 2000). Phelps gathered research from 31 different countries and determined that in most industrialized nations, large-scale, high-stakes tests are growing trends. He argued that not only is there increasing support for additional testing programs and the importance schools place on them, but there are also developing changes in the styles, types, and reasons for the tests (Gewertz, 2007).

#### **Developments**

Considering the expanded roles of high-stakes testing, there are developments, which seem most promising. In the states of Texas and Maryland, for instance, departments of public instruction have been changing the ways students test for proficiency. Departments of instruction have changed from testing knowledge gained over the four-year high school period with achievement tests, to testing what students should have learned during a particular course, as in end-of-course tests (Gewertz, 2007). Gewertz implied there are ongoing shifts in educational thinking away from the exit examinations required for graduation to end-of-course assessments, which, if rigorously applied and aligned with course content, could help guide and deepen instruction and learning. With the recent interest in testing, foremost in the minds of many educators is how these emerging assessment choices combine with and improve upon students' achievement and learning (Croft, Waltman, Middleton, & Stevenson, 2005).

Shifting focus to a more testing-oriented, educational environment, educators are trying to both improve students' learning and raise test scores to higher levels. State curriculum departments have begun devising thorough, complex objectives that teachers must cover and students must master to achieve proficiency in the course (Croft, Waltman, Middleton, & Stevenson, 2005). Because teachers are under scrutiny and pressure to increase their students' test scores, the use of certain test preparation activities may aid in improved retention of the material. For instance, Croft, et al. (2005) argued that teaching more and working harder can encourage improvement, but other methods, such as correlating classroom content with the curricula and coaching to the tests, may produce actual gain in students' learning.

The researchers (Croft, et. al., 2005) did advise, though, that understanding the types of test preparation practices used is very helpful when accurately trying to interpret score gains. They argued that the goal should not be an artificial gain in students' achievement based merely on higher scores, especially when the real intentions are to broaden the domains of content and skill. The researchers recommended varied-format preparations for testing because instruction needs to relate directly to the tests and should provide other opportunities for students to adapt to new formats. In this research study, for example, students utilized the test-taking strategies of self-grading and correcting their papers, with the intended goals of increased student learning and higher standardized test scores.

At the center of recent developments in testing are President Bush's No Child Left Behind legislation and the debate of whether this will prompt notable gains in students' achievement (Fuller, Gesicki, Kang, & Wright, 2006). While each state has its own scale to determine the level of student learning, there is still a question about how much a student is learning and whether implementing a standardized test can improve knowledge acquisition. Parents, educators, and critics of standardized testing hope to see more promise for students' learning and rising test scores; however, there are emerging concerns with the new reliance on high-stakes testing.

#### Problems

Even with new enthusiasm for student achievement and learning, problems are beginning to surface in most schools concerning high-stakes testing. For instance, too much reliance on testing, teaching to the test, and the possible loss of learned skills at the cost of standardized test preparation are a few of the issues that give educators cause for concern (Au, 2007). In a meta-synthesis involving standardized testing, Au (2007) noted concerns from emerging patterns of over-reliance on testing and greater contradictory results than educators had originally intended. Au determined, "The primary effect of high-stakes testing is that the curriculum content is narrowed to tested subjects, subject area knowledge is fragmented into test-related pieces, and teachers increase the use of teacher-centered pedagogies" (2007, p. 259).

In addition, as schools turn their attentions toward test scores to gauge progress, they may find increasingly difficult problems with showing that students are learning more. For example, as a standardized test score becomes the benchmark from which to measure learning and chart improvement, it may be harder to show Adequate Yearly

5

Progress (AYP) (Olson, 2006). When the number of students tested increases, so does the number of goals schools have to meet. Each year, for instance, an AYP performance target will increase if there was obtainment of the previous year's goal. This means, according to Olson, that schools have challenging tasks of trying to improve test scores with different groups of students every time they participate in testing. He argues there are no easy ways to show continuous improvement from the same groups of students, especially at the high school level, where student groups fragment into different class choices.

Finally, there is the possibility that real skill development will be neglected in favor of time spent on test preparation. While student-learning outcomes are the intended focus of standardized test scores, higher-thinking skills development and analytical writing could be two examples of skills sacrificed at the cost of spending more classroom time on high-stakes testing requirements (Horn, 2003). The current high-stakes environment has produced some cultural effects as well. In North Carolina, for instance, testing data are beginning to suggest that non-white, non-Asian students and students with special needs are the groups most deeply affected by high-stakes testing. High quality instruction could be taking second place to the efforts of improving test scores. It is becoming increasingly difficult for an educator to remember that standardized testing is only a tool for teaching and learning. The tests are not to illustrate competitive improvements among schools and states, but to show growth and academic progress, according to Horn (2003).

Even though there are pressures to show improvement with test scores and students' academic growth, educators and policy makers should be aware of the problems

6

that may be surfacing in most schools because of increased emphasis on testing. For instance, excessive test reliance in evaluating student achievement and teaching skills evaluated only by the test are indicators that test scores are now very important in determining students' success in the classroom. As the standardized test is growing in importance, so is the push for the classroom teacher to encourage student learning and raise the score. Teachers and educators are looking for ways to show improvement, and one way could be how students participate in the day-to-day classroom assessments. Active student involvement in the assessment process forms the foundation of this research study.

Today's education system of a standards-based outcome, where standardized testing measures student learning, has origins in a theoretical and empirical perspective. The basis of this dissertation began with the theoretical perspectives of learning from Vygotsky's constructivism, Kolb's experiential, and Bloom's mastery learning theories. With Vygotsky's constructivism, for instance, students learned information for themselves by first being exposed to information and then applying it to new situations to enrich their learning (Slavin, 2006). Teachers acted more in the capacity of a facilitator or a guide and students took the new information and used it on their path to discovery learning.

Constructivism encouraged students to be active learners and because of that, classroom instruction should be more student-centered (Vygotsky, 1978). Learning, according to Vygotsky (1978), could not effectively occur in a vacuum, but was, by its very nature, social and interdependent. Jarimillo (1996) agreed with Vygotsky's ideas when he stated that the learner preferred being an active participant. Bergstrom and O'Brien (2001) concurred with Jarimillo in stating that when students interacted with each other and were involved in a more discovery-oriented classroom environment, Vygotsky's constructivist theory was realized in the learning process.

Similar to Vygotsky's theory of constructivism was Kolb's experiential learning. For Kolb, learning occurred when students were able to observe and then move to active experimentation where they could process information from multiple points of view (Johns, 2001). In Kolb's cycle, learning was more than an assimilation of unrelated concepts but an active, circular process of personalizing information to arrive at new thoughts and conclusions (1983). With experiential learning, the student was actively engaged in classroom participation (2001), unlike the standards-based instruction that occurs in most schools today (Bergstrom & O'Brien, 2001).

Learning is an active process and requires effort and participation, according to Foley (2002). If students were involved in grading and correcting their own papers, the learners would be following in the paths of both constructivism and experiential learning. For example, Slavin (2006) defined constructivism as a process in which students learn information for themselves by first encountering the new information and then applying it to novel situations to further their learning. In this study, the test correction process will allow students to compare new information against old rules and revise what they have learned as in Vygotsk's top-down processing and active learning approach (1978). Additionally, self-grading and correcting would satisfy Bloom's ideas concerning mastery learning (1968). For instance, Bloom believed that instructional practices should be adapted to the needs of diverse learners. All students should have mastered a certain skill to a predetermined level of competence before they continued to the next topic of study (Block, Efthim, & Burns, 1989). While students learn at different speeds, their level of academic attainment varies. Self-grading and correcting would work towards the premise of Bloom's theory of mastery learning that recognized assessment as a tool and feedback as the foundation of modern learning (1968).

While the present-day emphasis is on standardized testing to assess for learning, the earlier theoretical ideologies of constructivism, experiential, and mastery learning form the theoretical framework of this research study. Several modern-day empirical studies also support the notion of improved student learning with an active classroom environment, specifically the strategies of student self-grading and correcting. Recent empirical studies by Sadler and Good (2006), Stotsky (2005), Au (2007), and Gewertz (2002) reflected on the changes in both instruction and testing which have been used for assessing student learning. While educators have reverted to more direct instruction to cover the state-mandated standards for testing (Phelps, 2000), teachers also looked to the theoretical models of Bloom, for instance, for strategies to improve student learning, such as self-grading and correcting (Hayes, 2006).

Struyven, Docky, Janssens, Schelfhous, and Gielen (2006) conducted a study to determine the effects of end-of-course tests on student learning. The researchers found that students instructed in a standards-driven format and assessed with multiple-choice tests might have learned more than students who were assessed by other means, such as with portfolio assessment. Thompson and Newsome (2002) continued research on testing with their study, which sought to discover if multiple-choice tests could help encourage the use of higher-ordered thinking skills in the classroom. Other researchers, such as Kohn (2000), McNeil (2000), and Yeh (2001), conducted studies that focused on

the effectiveness of assessing higher thinking skills with standardized tests and found positive results with their research.

Dweck (2000), Clymer and William (2006), and Sadler and Good (2006) reached similar conclusions in their studies. The three studies agreed that when students became actively involved in the learning process both weak and strong students benefited, especially when there were performance strategies designed to get all students interacting. Sadler and Good, along with researchers Kirby, Downs, and Collean (2007) conducted research on student self- and peer-grading. In both instances, their studies indicated better student understanding. Falchikov and Boud (1989) and Falchikov and Gold's (2000) meta-analyses found positive correlations between student self-grading and learning.

Using these theoretical and empirical ideologies as a framework, this research study used the concept of student self-grading to determine if the addition of student corrective measures improved learning. In this present research study, 11<sup>th</sup> grade United States History students were grouped into two sections, control and treatment groups. Each class took a pretest and participated in the intended research intervention of either student self-assessment with correction measures or control group selection. The treatment group mainly focused on self-corrective measures, while the control group went in a more traditional fashion of teacher-only grading on their weekly tests. The study concluded with the administration of the North Carolina end-of-course test in United States History. Final end-of-course scores were analyzed using the statistical procedure independent samples *t*-test to determine if the null hypothesis would be rejected or if the researcher would adopt the null.

#### Statement of the Problem

Teachers have traditionally assumed an authoritarian role in the classroom setting, especially in marking test papers and assigning grades (Guskey, 2007). Guskey stated that standardized tests have reinforced the idea of authoritarian roles for educators, as teachers scramble to cover the standard curriculums and prepare students for their end-ofcourse testing experiences. When students assume involved roles and actively participate in grading their tests, final learning outcomes and end-of-course grades improve (2007). *Research Questions* 

- Is there a positive, negative, or equivalent relationship between students grading and correcting their own test papers and a higher score on the North Carolina end-of-course test in United States History?
- 2. Will a comparison between pretest and posttest grades show students learned more with the intervention of student self-correcting and grading than with teacher-only grading?

#### Null Hypothesis

Students who self-grade and correct their test papers will not achieve significantly higher scores than students who do not grade and correct their own test papers on the North Carolina end-of-course test in United States History.

Educators should attempt to maintain a balance between the focus of students' learning and higher test scores. This research study offered a unique opportunity to challenge student learning and further develop a test score strategy from the position of active student participation.

#### **Professional Significance**

This quasi-experimental study of student self-assessment and correction contributes to the knowledge of student learning and testing. Now, more than ever, school administrators and classroom teachers are struggling to find a place between improving students' learning and raising standardized test scores. There is pressure to demonstrate a solid foundation of academic and intellectual achievement for each student, while concurrently providing academic improvement as shown by the standardized test score, especially since the No Child Left Behind (NCLB) legislation.

For the near future, or at least through the next series of elections, standardized test scores will be important for the President and Congress as they continue to refine the NCLB initiatives. State departments of public instruction, local school districts, school administrators, and classroom teachers are feeling the pressures of the NCLB Act of 2001, in which state academic standards became the center of attention (Stotsky, 2005). According to Stotsky's report, "All states are now required to have demanding academic standards in place and to demonstrate steady student progress toward academic proficiency, as set forth in those standards" (2005, p. 10). The Act now links states? accountability for increasing students' achievement to the quality of their teachers (2005) and requires that all students have access to the general curriculum at their designated grade levels (NCDPI, 2006).

With the publicity surrounding high-stakes testing, North Carolina remains concerned about standardized scores from the perspective of its School-Based Management and Accountability Program, the ABCs (Accountability, Basics, Control) at the local level (NCDPI, 1997). The program directors had a definite goal in mind: The ABCs set standards for student performance and growth, provided for an accountability system, deemed schools and districts accountable for students' education, and insisted local schools and districts involve all parents, teachers, and community representatives to help develop and implement local accountability and program evaluation systems that complement the state ABCs plan. (North Carolina, 2004, p. 10)

Most of the points set by the ABCs plan directly involve measurement by standardized tests, and in North Carolina, measurements are mainly end-of-course tests at the high school level. North Carolina gives end-of-grade tests as well, but typically at the elementary grade levels.

The standardized tests should encourage growth and improvement at local, state, and national levels. For the first time in United States history, mostly because of NCLB, "Key elements of the public education system are joined, such as pass rates on licensure tests by teachers, state accountability, and academic standards that set forth what K-12 students are expected to learn in core subjects" (Stotsky, 2005, p. 10). Students now have a guide to what they will learn, and teachers finally know what they will teach. State standardized test scores matter to all involved in education, both in encouraging student learning and showing continuous growth.

While concerned states now implement their own standard curriculums and tests, there are no national curriculums or federally mandated standardized tests. In 2006, 22 states required students to pass an exit exam to graduate, but only four of those states used end-of-course tests (Gewertz, 2007). The North Carolina end-of-course tests though, are distinguishable from other standardized tests, and this influences the value of this research study for educators, especially in North Carolina.

In distinguishing North Carolina's end-of-course tests, the Department of Public Instruction asserted that "the North Carolina end-of-course (EOC) tests were initiated in response to legislation passed by the North Carolina General Assembly and the North Carolina Elementary and Secondary Reform Act of 1984" (NC Assessment Brief, 2004, p. 1). The Act mandated a standard curriculum for all students in the core content areas, with tests developed for five foundation subjects: English, Math, Science, Social Studies, and Vocational Studies. North Carolina developed end-of-course tests for two reasons: "To provide accurate measurement of individual student knowledge and skills specified in the North Carolina Standard Course of Study, and to provide accurate measurement of the knowledge and skills obtained by groups of students for school, school system, and state accountability" (NC Assessment Brief, 2004, p. 1). The North Carolina end-ofcourse tests, according to Thompson and Newsome (2002), are distinguishable from other states in that "if state tests focused more on higher-order thinking skills, then these tests might actually help teachers improve classroom instruction and assessment by encouraging teachers to include these thinking skills in the classroom" (p. 2). To try to meet this goal, the North Carolina Department of Public Instruction incorporated both Principles and Standards for School Mathematics and Dimensions of Thinking as a theoretical framework for developing the end-of-course exams. Unlike other states with standardized tests, North Carolina's exams try to foster development of higher-order thinking and learning skills in the classroom, while assessing these skills using multiplechoice test questions.

For instance, Thompson and Newsome (2002) reported, "*Dimensions of Thinking* included metacognition, critical and creative thinking, thinking processes, core thinking skills, and the relationship of content-area knowledge to thinking" (pp. 2-3). The researchers argued that the North Carolina Department of Public Instruction (NCDPI) encouraged teachers to use all seven core-thinking skills, which should be the foundation of the questions on the end-of-course exams in daily teaching practices: knowing, organizing, applying, analyzing, generating, integrating, and evaluating. Thompson and Newsome explained that "the North Carolina department of Public Instruction's framework also originated from Bloom's Taxonomy, which includes knowledge, comprehension, application, analysis, synthesis, and evaluation" (2002, p. 3).

From the foundation of Bloom's Cognitive Taxonomy, the Department of Public Instruction further subdivided the question format into three categories representing the varying levels of knowledge acquisition (Thompson & Newsome, 2002). In their report, Thompson and Newsome discussed the different categories and the levels of knowledge:

Category I questions focus on a knowledge and application format, which is recall and simple application. Category II constructed questions around a foundation of organizing and application, which is more complicated, but the premise is the student should already know how to proceed. Finally, there are the Category III questions that focus on analyzing, generating, integrating, and evaluating how to solve the problems that should not be immediately apparent without thought. (Thompson & Newsome, p. 5)

North Carolina has attempted to make its multiple-choice format, end-of-course test experience an exercise in student learning and critical thinking, not merely a set of tests to assess achievement. The *Princeton Review* (2003) ranked North Carolina's endof-course tests at number four in the nation in academic alignment with curricular test quality, ongoing ability to improve, and accountability. Educators in Texas, too, are changing testing formats and moving toward end-of-course tests instead of high school exit exams (Gewertz, 2007). Concurrently, California has more than 10 different examinations in K-12 to determine achievement and proficiency (California State, 2001). While North Carolina's end-of-course tests are unique in their attempts to encourage student learning and higher-ordered thinking skills (NCDPI, 2008b).

Because of the emphasis on higher-ordered thinking with the North Carolina endof-course assessment, this research study utilized the United States History end-of-course test. North Carolina, according to Thompson and Newsome (2002), assesses both standardized pre- and posttests for validity and reliability when administered under the properly prescribed conditions. This research study used the scores students obtained from the initial administration of the pretest at the beginning of the 2009 spring semester. The control and treatment groups then took their final, end-of-course exams at the conclusion of the semester. Grade comparisons from the pretest to the end-of-course exam, in both the control and treatment groups, helped determine rejection or adoption of the null hypothesis by the statistical procedure of independent samples *t*-test.

The goal of the North Carolina end-of-course test in United States History is to measure how well the teacher can cover the North Carolina Standard Course of Study and how well the student can master and understand its content (NCDPI, 2006-07). This research study tried to determine if there were changes in students' test scores because of the treatment, or if the results occurred by chance alone. A final rejection of the null hypothesis would have indicated self-grading and corrective measures were statistically significant and an indication student learning had occurred because of the intervention of the treatment group.

This research study contributed to student learning and the knowledge of testing by investigating whether a state test and student self-grading and correcting could positively affect learning. Several researchers (Kirby & Downs, 2007; Sadler & Good, 2006) have studied possible benefits of student self-grading and correcting, and research findings have typically shown positive results, which might signal viable strategies for the future of testing. Also important, though, are the intrinsic values of increased student learning, a feeling of self-efficacy, student ownership, and empowerment in the classroom, which may come as added benefits. For instance, Kirby and Downs (2007) stated that "worldwide, self-assessment practice has been gaining recognition, and it has been linked to the adoption of deeper approach to learning: self-regulated learning and the development of metacognitive skills" (p. 476). Further, Sadler and Good commented that those students who corrected their own tests improved dramatically in the classroom, and self-grading resulted in increased student learning. Orsmond, Merry, and Callaghan (2004), too, agreed that self-assessment and correction was useful in helping a student reach his or her learning goal.

While some authors did mention that self-assessment encourages critical thinking, analysis, and improvement (Sterling, 2008), others also found more personal rewards might be possible, such as the previously mentioned self-efficacy, ownership, and feeling of empowerment. Guskey (2007) argued that when a student took an active approach in

the classroom, rote exercise and memorization were not the only ways a student participated. A new, nurturing environment began when the student felt more confident in his or her abilities and therefore wanted to participate. Further, Guskey contended that the students felt a personal ownership toward their education. New opportunities for academic social interactions followed and gave even the weaker students feelings of accomplishment and empowerment in the classroom. For example, Tan (2008) argued that student self-assessment had the potential to further lifelong learning and empower, rather than discipline, a student. He ventured that there has always been a degree of tension felt between what the instructor expects from a student and what the student believes the instructor wants (2007). When students become involved in self-grading and correcting, they are more conscious of the set standards for good work and are more keenly aware of what constitutes high-quality work (Andrade & Du, 2007). As a result, student-teacher conflict and anxiety diminish. Students are in a position of awareness and confidence with their work, and become more motivated to take responsibility for their learning (Edwards, 2007).

Strong, Davis, and Hawks (2004) and Edwards (2007) for instance, touted the benefits of self-assessing and correcting because they believed it leads to student selfempowerment. Traditionally, they argued, students have taken a secondary position to the teacher in the classroom hierarchy. The teachers grade and return the tests, and students are passive in the learning environment. The teachers are seemingly like dictators, and students typically accept the grade and progress to the next topic of discussion. In an environment where students take an active role in grading and correcting their own papers, though, they assume responsibility for their learning, education, and motivation (2004). Strong, Davis, and Hawks argued that the students are suddenly active participants and are in a position where ownership of their education becomes apparent; thus, empowerment occurs. Students who are able to correct the questions missed have more responsibility for their learning and a greater sense of control (Strong, Davis, & Hawks, 2004). This feeling of control gives the learners a sense of increased self-worth and self-efficacy. The students become active participants, and suddenly they are even more involved and interested in the classroom environment. Students, according to the researchers, are no longer passive recipients of knowledge, but take an active role in their learning.

Finally, self-grading and correcting gives the students a chance to identify their mistakes, helps reinforce what they have just learned, and allows them to have immediate feedback (Edwards, 2007). The learners quickly profit from their mistakes by not missing an opportunity for reinforcement. In the conventional way of assessment, teachers sometimes taking several days to grade and return papers, the students would most likely have forgotten the questions and lost any desire to pursue the right answers. Students care more about the questions missed if feedback is prompt.

#### Definition of Key Terms

Several key terms are used through the course of this dissertation. Words included in this list are mainly educational terms that need clarification to aid in understanding this experiment. The following terms appear in alphabetical order: *Achievement Levels*: Students' learning and progress appear on North Carolina's end-of-grade and end-of-course tests by achievement level (NCDPI, 2006-07). There are four achievement levels:

*Level I* Students performing at this level do not have sufficient mastery of knowledge and skills in this subject or course area to be successful at the next grade level or at a more advanced course level.

*Level II* Students performing at this level demonstrate inconsistent mastery of knowledge and skills in this subject or course area and are minimally prepared to be successful at the next grade or course level.

*Level III* Students performing at this level consistently demonstrate mastery of grade level subject matter and/or course matter and skills and are poised for the next grade or course level work.

*Level IV* Students performing at this level consistently perform in a superior manner clearly beyond that required to be proficient in this grade level or subject matter and are very well prepared for the next grade level or for a more advanced level in the subject area. (NCDPI, 2006, p. 1)

Adequate Yearly Progress (AYP): According to the NCDPI:

AYP measures the yearly progress of different groups of students at the school, district, and state levels against yearly target goals in reading/language arts and mathematics. All public schools, in North Carolina and throughout the country, must measure and report AYP as outlined in NCLB. AYP is the minimum level of progress in reading/language arts and mathematics proficiency made by students in a year. (NCDPI, 2006-07, p. 1)

*Alternative assessment:* This referred to assessments that measure students' learning in forms other than traditional pencil-and-paper tests.

*Assessment:* This referred to the process of evaluating a student's knowledge or skills in the classroom setting.

*Control group:* This was the group of students receiving no alteration in assessment during the study.

*Curriculum:* This referred to the organization of subject matter taught over a prescribed period of time (NCDPI, 2004).

*End-of-Course (EOC) tests:* All high school students in North Carolina are required to take end-of-course tests for the core subject areas of math, science, social studies, and English. The end-of-course tests are standardized tests and are meant to determine student performance in a particular course, according to the NCDPI:

EOC tests are designed to assess the competencies defined by the North Carolina Standard Course of Study for 10 different subject areas, including United States

History, and must be taken during the last 10 days of school. (NCDPI, 2006, p. 2) *Experimental or Treatment group:* This referred to the group of students that received the experimental, altered-assessment strategy of self-grading and self-correcting. *Grade level, Achievement Level III, and proficiency level:* According to the NCDPI:

Each of these terms refers to student work that meets the achievement standard set by North Carolina. Students scoring at Achievement Level III or Achievement Level IV perform at grade level and are well prepared to meet the demands of the next grade. At the high school level, the term *proficiency level* is more frequently used and refers to students scoring at Level III (83-92 percentile score) or Level IV (93-100 percentile score) on end-of-course tests. (NCDPI, 2006, p. 2) *High-Stakes Testing:* This described the uses of standardized achievement tests that carry serious consequences for students and educators (NCDPI, 2006-07).

*Learning:* This term meant the knowledge or skill acquired by instruction or study (Merriam-Webster, 2008).

*Learning outcomes:* This term described the result of what students may have learned in a unit of study or the whole course. Measured outcomes on North Carolina's standardized test grades fall within the score range of a III or IV achievement level on the end-of-course tests (NCDPI, 2006-07).

North Carolina Standard Course of Study: According to the NCDPI:

The North Carolina Standard Course of Study provides every content area subject a set of competencies for each grade and high school course. Its intent is to ensure rigorous performance standards that are uniform across the state. It sets content standards and describes the curriculum available to every child in North Carolina's public schools. (NCDPI, 2008a, p. 1)

No Child Left Behind (NCLB): According to the NCDPI, NCLB is defined as:

The reauthorization of the Elementary and Secondary Authorization Act and represents a sweeping change in the federal government's role in local public education. NCLB has a variety of goals, but the most dominant ones are for every school to be at 100 percent proficiency by 2013-14 as measured by student achievement on state tests and every child taught by a highly qualified teacher. The law emphasizes new standards for teachers and new consequences for Title I schools that do not meet student achievement standards for two or more consecutive years. (NCDPI, 2006, p. 2) *Pacing guide:* This referred to a written schedule displaying the alignment of concepts, topics, and skills related to a particular curriculum addressed over a defined period of time (NCDPI, 2006).

*Pretest:* This referred to the test given to students at the beginning of the semester before any coverage of the course curriculum has occurred.

*Proficiency:* This term meant the mastery or the ability to do something at grade level (NCDPI, 2006, p. 3).

*Self-assessment and self-grading:* Both of these terms referred to the process of students grading their own test papers using a pre-coded answer key created by the teacher.

Standardized test: According to the NCDPI:

This term meant a test administered and scored in a consistent manner. The tests are designed in such a way that the questions, conditions for administering, scoring procedures, and interpretations are consistent and are administered and scored in a predetermined, standard manner. (NCDPI, 2006-07, p. 2)

*Student Learning:* This term referred to learning that was student-driven or student-led. *Teacher grading:* This term referred to the process of the classroom teacher grading the students' test papers.

*Test corrections:* This referred to the altered assessment strategy in which students analyzed and wrote about the missed questions on their tests.

*Traditional assessment:* This referred to the process of the classroom teacher grading the students' test papers and returning them for the students' inspection.

#### Summary

The first chapter of this dissertation discussed the background, stated the

investigated problem, and probed the professional significance behind the possibility of using student self-grading and corrective measures. The next chapter established a theoretical and empirical foundation beginning with a review of literature. The review of literature first focused on the early theoretical aspects of learning involving the theories of constructivism, experiential theory, and mastery learning. The second part of the literature review focused on current learning trends, performance tactics, and potential strategy benefits. The third chapter of this dissertation detailed the methodology chosen for the research study, while the fourth and fifth chapters discussed the statistical analyses and conclusions.

#### CHAPTER TWO: REVIEW OF THE LITERATURE

A large body of literature on student learning and self-grading provided the foundation for the following research study. This chapter examines both theoretical and empirical studies which demonstrate the evolving nature of today's education system. The first section in this review of literature details the search processes and the different historical theories of learning that influenced this study. These theories have continued to evolve into today's concepts of a standards-based, outcomes-based education in which standardized testing measures students' learning. The theoretical perspectives discussed include constructivism, experiential learning, and mastery learning. Concluding the theoretical discussion is a section on how education theory has changed in recent years, transitioning from the three main theoretical perspectives to the current theory of improving student learning with standardized testing.

The second section of this chapter focuses on a review of empirical studies involving learning and student self-grading and correcting. While standardized testing continues to gain momentum in mainstream education, there is a solid research base indicating that standardized test preparation in itself is not necessarily the most effective way to increase student learning. The empirical studies concentrate on research with alternative means of assessment, such as student self-grading and correcting.

#### Theoretical Review

The review of theoretical literature centers on three different theories: constructivist, experiential, and mastery learning. With constructivist theory, discussion of Vygotsky's perspective occurs through top-down processing and discovery learning.
Second, Kolb's experiential learning theory, in relation to how active student involvement has influenced education, emphasizes the importance of learner involvement and the circle of learning. Finally, the theoretical review concludes with a more in depth discussion of Bloom's mastery learning theory. Bloom's theory emphasizes the importance of varying instruction and evaluation technique, such as using self-assessment and correction as a tool. The empirical review focuses on instructional and assessment variation in the modern classroom of standardized testing and paves the way for this research study.

### Constructivism

According to Slavin (2006), constructivism means that students learn information for themselves by first encountering information and then applying it to new situations to further their learning. For instance, in constructivism, teachers are to assist students in obtaining the information, but the learners have a responsibility to take the new information and discover how to apply it to the things they already know. Students should also learn to realize new ideas and relate them to alternative situations. Slavin argued that in this way students are able to compare new information against old rules and continually revise what they have learned. In this sense, students are much more active with their own learning than in the traditional classroom mentality mentioned in Dewey's setting (1983). In Dewey's traditional classrooms, the desks sat in straight rows and stayed anchored to the floor to keep the students from moving the chairs. The plan was to discourage interaction among the students and prompt stricter attention to the teacher. The teacher delivered instruction, the student passively received instruction, and the idea was that the learner automatically internalized and utilized information to make a learning connection. In the classroom, learning was a passive activity with little interaction between the student and teacher and with minimal social contact, even between learners (Dewey, 1995).

In constructivism (Slavin, 2006) the student is an active learner, and because of that, constructivist method encourages a more student-centered instruction. As modern educational practices have changed since Dewey's time, so, too, have the theories of learning. Piaget (1952) and Vygotsky (1978) pioneered the idea that learning could not effectively occur in a vacuum, but was, by its very nature, social and interdependent within the learning environment. Vygotsky's theories on sociocultural development, according to Jarimillo (1996), actually predated the educational movement of constructivism and found increasing support in the modern views of constructivist learning. Vygotsky (1978), for instance, believed that social experiences shaped students, and being in a group, such as the classroom, encouraged and nurtured individual cognition. According to Jarimillo (1996), internalized concepts, obtained through selfdiscovery, constructed a child's intellectual personality. He argued that the learner was not an empty vessel merely awaiting knowledge from an instructor's lecture, but instead preferred being an active participant involved in hands-on activities that were interesting and challenging. In the classroom, students interacted with and learned from both their peers and teachers.

According to Kozulin (1998), for instance, Vygotsky believed that particular learning activities provided a framework for guided instruction. An example was Vygotsky's top-down processing strategy. In the classroom setting, the teacher began with a problem, sometimes presented by the students themselves, and then students

27

worked to discover how to solve the problem (1998). If actively involved in discovery, students might have developed problem-solving skills and engaged in socio-cultural learning experiences. This was the point Vygotsky (1978) called cognitive scaffolding, which reflected the cultural process of assistance through cooperation and collaboration.

Top-down processing, social interaction, problem-solving, and an active learning approach are indicative of Vygotsky's theories on learning (1978), and all of these ultimately lead to discovery learning by the students (as cited in Slavin, 2006). Discovery learning is part of the evolving process of Vygotsky's theory (1978), which has become more prevalent in recent years (as cited in Bergstrom & O'Brien, 2001). The premise behind discovery learning, according to Slavin, was to encourage students to learn and discover mostly on their own, with the teachers ultimately assuming the role of facilitators instead of leading the classroom process. As students discover principles for themselves, the pupils actively engage in the learning process and assume more responsibility for their own learning. In the case of today's classroom, a student who self-assesses and corrects his or her paper is beginning the self-discovery and learning process. From the standpoint of Vygotsky's original theory, constructivism has evolved and incorporated such strategies as top-down processing and discovery learning.

# Experiential Learning

Another theoretical perspective is that of experiential learning theory and the work undertaken by Kolb. According to Johns (2001), Kolb's experiential learning theory emphasized the importance of learner involvement in education. For Kolb (1983), active learning meant more than just having the student enter the classroom of aligned desks and remain in the position of a passive learner. The student needed to be involved in his or her own learning and discovery process. Kolb meant for students to assume an active role in the classroom. Johns discussed the beginning of experiential learning theory, which had its start with the human and cognitive development research that began in the 1870's with the concept of pragmatism.

While Kolb wanted educators to shift to a more experience-based approach, he highlighted the need for learner involvement and developed a model, which he termed *the* cycle of learning (1983). In Kolb's cycle of learning, there are two dimensions to education, which are the gathering of facts and the processing and personalizing of information (1983). Thus, in his cycle, students move from reflective observation to active experimentation where the students can absorb and process information from multiple points of view. Kolb, according to Johns, believed that for learning to be a complete process, incoming knowledge must travel in a circular pattern, his cycle of learning, which meant moving from an experience, to reflection, and finally to a generalization and application of the learned concepts (2001). Kolb believed that learning did not merely proceed in a linear fashion from one seemingly unrelated concept to another, but was an active process of assimilating information. The progression would finally combine an active process with previously learned and experienced concepts to arrive at new thoughts and conclusions. Kolb believed that learning was a circular process that continued to fashion itself in a reoccurring pattern. Learning is not a fixed process, but is shaped through experience and further exposure to detail (1983).

In experiential learning, then, the pupil has the opportunity to fashion his or her education experience with active construction and participation in the classroom setting, which forms the basis of Kolb's cycle of learning (as cited in Johns, 2001). A problem in today's schools, according to Kolb's model, is the emphasis on a standards-based instruction. For instance, by focusing mostly on state-mandated standards, students are only receiving a surface approach to learning as a means to achieve an end: higher standardized test scores. Student motivation may be merely superficial, such as fulfilling a testing requirement, and might lead to a situation in which a student is doing just enough work to obtain the extrinsic objective for that class period. If Kolb's cycle of learning were applied to the typical classroom setting, then the instructor would need to incorporate activities, which would strengthen student involvement. The students' motivational outlook would then change from an extrinsic acquisition of standardsrelated material to an intrinsically motivated, active approach to excellence.

Learning is an active process and requires effort and participation on the part of the students, as well as interaction with the instructor (Foley, 2002). Teachers should promote interaction between the students and the intended instructional materials, instead of encouraging students to become passive learners. Students should have the opportunity to become involved in activities that simulate Kolb's four learning modes, as explained in his cycle of learning: direct experience, reflection and observation, theory and principle, and application to practice (as cited in Johns, 2001). Involving students in grading, and more importantly, correcting their own papers should fulfill the needs of Kolb's cycle of learning by encouraging a student to be part of his or her own learning and discovery process. Concurrently, the pupils would become actively engaged in the constructivist approach of top-down processing advocated by Vygotsky.

### Mastery Learning

A final historically evolving theory for consideration is that of Bloom's mastery

learning concept. While Kolb's cycle of learning posed an active, cyclical approach to knowledge acquisition, Kolb mentioned additional elements such as the students' knowledge base, procedural skills, self-regulation of learning, and motivation and affect, which were also part of the learning process (as cited in Vermunt & Vermetten, 2004). While Kolb focused on the active participation of the student, Bloom believed that other aspects, such as using assessment like a tool, were necessary for consideration (Guskey, 2007).

According to Slavin (2006), Bloom's mastery learning defined a process of adapting classroom instructional practices to the needs of diverse learners. The premise was to make certain that all, or nearly all, of the students mastered a certain skill to a predetermined level of competence before they continued onto the next skill level (Block, Efthim, & Burns, 1989). In the traditional classroom, most students had the same amount of instructional time to master the objectives before the class moved to the next series of topics. Bloom theorized, and several other contemporaries of his agreed (Feuerstein, 1980; Gardner, 1983; Arrendondo & Block, 1990), that if each student had as much instructional time as he or she needed to master the concepts, every pupil would eventually arrive near the same level of competence. Students who had trouble keeping up needed even more instructional time and assistance. Slavin (2006) stated that the premise of Bloom's theory was that almost every student could learn a subject's essential skills, while the student and teacher acted in the appropriate roles to bring about learner success.

In support of this research study is Bloom's (1968) contention that assessment is a tool and that feedback, self-correction, and enrichment should become the cornerstone of

modern mastery learning (Guskey, 2007). Bloom argued that in the traditional classroom over 90 percent of students were able to master what teachers were trying to teach, but there were also limitations to learning. He theorized that teaching all students the same way and giving the learners the same amount of instructional time would produce much variation in students' learning. Bloom cautioned that after most classroom assessments, teachers found initial instruction had not been appropriate for all students, and the assessments did little more than verify that there were discrepancies in learning. To combat this problem, Bloom suggested that instead of teachers using classroom assessments to determine inequalities in learning, the tests could diagnose individual learning problems and help design remediation schedules (as cited in Guskey, 2007).

The premise was for the student to work through the problem again to make certain he or she learns from the mistake (Guskey, 2007). Bloom also advocated the use of assessments as tools but further mentioned there were more responsibilities for classroom teachers than merely correcting the assessments and handing papers back to the students. Immediate feedback is important, but the student must have the opportunity to engage in an active, corrective activity for each formative assessment. The corrective measure, self-correction on a returned test paper for example, means a student has detailed direction of how to master the skill of each objective. Guskey further commented that, if appropriate, corrective initiatives should occur in the classroom. The corrective activities would catch minor problems and prevent them from later developing into major learning difficulties. The instructor has the ability to change and reorganize his or her instruction, which might prevent the same learner misunderstandings during future instruction. After classroom corrective measures conclude, Bloom (1968) recommended that students take another formative assessment, which might vary slightly. The second classroom assessment would serve to verify whether the corrective measures had been effective in assisting mastery of the concepts, and the follow-up test would give the students another chance at success, perhaps to improve their achievement motivation in class. Theoretically, the strategy of self-corrective measures could help in the modern classroom to assist students working through initial misunderstandings of the class lesson content. Second, self-corrective measures could aid in utilizing the state's end-of-course assessment process to determine if any improvements in learning have occurred from the initial classroom formative assessment process.

Through formative classroom assessments and correction of learner errors, Bloom believed that all students could learn more than with traditional approaches in the classroom (1968). Guskey (2007) reiterated Bloom's message that feedback by itself was not enough to improve student learning and that criticism paired with a corrective measure would offer guidance and suggest how to manage improvement in progress. Guskey further argued that correctives in themselves were not good enough, but needed to be qualitatively different from the instruction which the learners had initially received. Bloom (1968) also stated that teachers should routinely accommodate different characteristics of student learning styles. After the students have received feedback, worked through the corrective process, and engaged in additional assessment to check for understanding and improvement, they should show increased learning through formative assessments; thus, enrichment would have occurred. Bloom (1968) cautioned against the students engaging in the corrective measures outside class. He said that when students used corrective measures, the learners needed to have the teacher's direction and input for guidance during the classroom experience. He believed that when instructors have the students merely do corrective activities outside class, learners rarely experience the same degree of success. If teachers have students complete the corrective activities in class, the end results should be an increase in students' confidence in learning situations and in initiating corrective activities on their own (Guskey, 2007). Block, Efthim, and Burns (1989) also cautioned about the time needed to implement Bloom's advocated corrective measures. They stated that while corrective measures were effective with enhancing students' learning and skills, teachers found it difficult to plan additional tasks during an already tight schedule.

A study by Whiting, Van Burgh, and Render (1995) and a meta-analysis by Kulik, Kulik, and Bangert-Drowns (1990) reached similar conclusions about the apparent effectiveness of mastery learning. The 1995 study found that students who had engaged in mastery learning were more likely to have positive results on test scores, better grade point averages, and even better attitudes toward school. Similarly, in all programs studied by the 1990 meta-analysis where students participated in the mastery learning strategy, the students made impressive gains academically. Results from mastery learning have been consistently positive.

In the present age of President Bush's No Child Left Behind legislation, the importance of curriculum-based outcomes measured through standardized tests continues to grow. Theoretical literature from researchers, such as Vygotsky, Kolb, and Bloom, tends to support learning strategies that are active and student-oriented. Learning,

34

described as a circular format that includes instruction, student involvement, and active corrective initiative carried out in a classroom setting, occurs with teacher assistance. Previous research studies, such as the 1990 meta-analysis (Kulik, et al., 1990), have indicated that students actively engaged in the learning process performed at a higher degree of consistency and mastery than students who were not actively engaged.

### **Empirical Review**

Organized into three categories, this section shows the differences between the evolving, theoretical aspects of learning discussed in the last section and the ways educators evaluate learning and performance in today's schools: current learning trends, possible classroom strategies, and benefits from self-grading. These three factors work together to help determine how a teacher instructs and even how a student learns.

# Current Learning Trends

While the previously discussed theories focused on Vygotksy, Kolb, and Bloom, modern theories emphasize different instruction techniques for improving learning. While theories of learning are still evolving among educators, there are differences in modern schools' beliefs of how students learn and the best ways of assessment. Several studies have examined characteristics of learning in today's schools, such as being standarddriven, test-oriented, and learner-active.

With the passage of the No Child Left Behind Act in January 2001, state academic standards and standardized curriculums became an essential part of the education process. According to Stotsky (2005), once the legislation went into effect, states needed to have demanding academic standards in place and ways of assessing academic progress. Schools wanted to demonstrate continued academic success and progress toward proficiency goals, as defined in the standardized curriculums mandated by the states or districts. The legislation also prompted an accountability aspect for each school to show increasing student achievement. Now, as compared to educational practices in previous decades, states have a responsibility for establishing a set of consistent standards and objectives students should master. The schools have a responsibility to provide all students with the highest quality teachers available, and students are expected to learn the standards and be able to perform at the states' required levels of proficiency to show that learning has occurred.

In addition to the No Child Left Behind Act, some states have also devised specific, standards-driven plans to assist in implementing changes. North Carolina, for instance, has a program called the School-Based Management and Accountability Program (the ABCs), which has set state standards for students' performance, made school districts accountable, encouraged parental involvement, and recognized the public's need to keep abreast of educational happenings (North Carolina, 2004). Stotsky (2005), along with Au (2007), became intrigued by the new standards-driven learning approach, but both had questions concerning the effectiveness of, and perhaps over-emphasis on, summative evaluations at the conclusion of courses. Stotsky wondered about the quality of the standards: if the principles were demanding enough, if the instructions were clear, if the teachers' training was sufficient, and if the instructors' knowledge was current for the demands of a standards-driven course.

Au (2007) hoped to answer some of these questions in his meta-analysis of 49 studies that focused on the effects of the standards-driven curriculum and how the resulting tests affected the classroom. Au's findings seemed to offer contradictory results

36

from what he was expecting. For instance, instead of discovering that the standardsdriven curriculums increased the exposure to varying teaching practices, Au saw even less student exposure to new ideas and concepts. He discovered that the content narrowed to cover only the curriculum objectives, which the students focused on during the formative evaluation process. Further, he argued that in most of the studies, subject knowledge was fragmented into pieces of specific, testable information, rather than delivered in a more holistic approach in which the students could develop new ideas and make assumptions which might carry over to different subjects. Au and Stotsky (2005) both found that most studies demonstrated an increase in teacher-centered pedagogies.

In support of the standards-driven curriculums of today Au did find that in some of the studies a more defined set of standards has led to content expansion, integration of knowledge, and more student-centered cooperative learning opportunities (2007). Clymer and William (2006) also supported the use of the standards-driven curriculum. They argued that assessment systems derived from the standards supported learning because students were obtaining information in incremental steps instead of all at once. Students, according to Clymer and William, learn that *smart* is not necessarily something they are, but something they become.

While several researchers (Clymer & William, 2006; Au, 2007; Stotsky, 2005) have discussed the changing structure of today's educational environment, there is also the trend of the test-oriented classroom. Some of the initiatives the No Child Left Behind Act required were intended for school districts to show increasing student achievement, as states were now accountable for progress (2005). To fulfill this requirement, most states are now focusing more on standardized tests, such as North Carolina's ABCs plan (North Carolina, 2004). Several states, Texas for instance, have traditionally given assessments based on the knowledge students have accumulated over the years, but that is changing, too (Gewertz, 2007). Gewertz argues that more states are transitioning to standards-based tests and end-of-course exams, which assess what students learn in the course. She contends the tests are meant to be rigorous, while aligning with course content to deepen students' understanding and knowledge of the content standards.

A test-oriented school experience affects all students in North Carolina (Horn, 2003). While schools are stressing the standards-driven curriculums and the end-ofcourse tests, researchers and educators alike are attempting to find the most effective form of testing to assess students' learning. Au (2007) conducted a qualitative metasynthesis, which analyzed 49 studies to discern how standardized tests affected curriculums, measured contents and the types of knowledge learned, and determined the pedagogies chosen for the classrooms. He cited contradictory trends in the results and argued that students received more narrowly defined curriculum content than was expected to be on the standardized tests. While he recognized the inherent need for assessing classroom progress, he questioned the effectiveness of such heavy reliance on testing. There appeared to be less active learning and more teacher-centered activities, such as direct instruction.

Au (2007) found that in only a few of the studies the standardized end-of-course tests overtly caused an expansion in classroom direction, instruction, and more studentcentered activity. Au argued that the majority of classroom situations he studied seemed to fall into the category described by Hayes (2006), as a back-to-basics movement. Hayes described American education as being almost in a regressive movement because teachers were reverting to strategies and classroom instructional tactics used in previous decades. Hayes cited examples, such as more teacher-oriented approaches and fewer cooperative learning activities available to students, to demonstrate the decline in classroom instructional variety. Hayes also noted more emphasis on lecturing in order to cover the materials required by the curriculums. Most states now claim, according to Fuller, Gesicki, Kang and Wright (2006), that they have more students than before scoring at or above the average annual levels of improvement.

There are still considerable debates over the improvements in student learning. Some of these debates may have occurred in response to the standardized curriculums or the end-of-course assessments students are taking. Phillips argued (2000), in support of Fuller, et al. (Policy Analysis, 2006), that his study of national and international trends indicated a concerted move to increase the emphasis on a test-oriented school environment. Phelps maintained that in his analysis on the continuing trend of standardized testing, he found that in 31 countries there was large-scale testing to assess student learning or achievement. He stated that in 28 countries, the number of subjects and the frequency of students tested had increased over a ten-year period, in comparison to only three countries that had stopped using standardized tests.

Concurring with the upward testing trend discussed in Phelps' study is a report released by the California State Postsecondary Commission (CSPC) (California State, 2001). CSPC reaffirmed the analysis in Au's (2007) research, which indicated a growing trend in state standardized testing, in addition to the complexity and diversification of the overall formative assessment. For instance, the CSPC report began with the early history of California's standardized testing program and then summarized changes and additions, which have continued to occur in recent years. The CSPC report mentioned that even in recent years, student testing had become so diverse and inclusive that, at the time the article appeared, California used more than 10 different forms of standardized tests to assess student achievement and proficiency. Texas now utilizes 12 end-of-course tests targeting different grades and subjects (Gewertz, 2007), while additional states, such as Iowa, continue to expand their testing programs (Deeter & Prine, 1998). Mississippi, Tennessee, Maryland, and North Carolina lead the nation, according to Gewertz and Horn (2003), on the most research into standardized testing. The researchers wanted to determine how a standardized test could benefit the multiple groups of students assessed yearly. Gewertz & Horn investigated North Carolina, which claimed to have created end-of-course tests for all content areas (North Carolina, 2004). The tests were meant to assess the implementation and learning achievement in every curriculum.

Struyven, Dochy, Janssens, Schelfhous, and Gielen (2006) conducted a study on the overall effects of end-of-course tests on student performance. They argue that even though school educators have many more resources than in previous years, the standard mode of assessment still involves traditional evaluation techniques, such as written and oral exams. The researchers looked at several different means of assessment including portfolios, peer assessments, and multiple-choice evaluation formats. They used a data collection format which employed both pre- and posttest designs. The authors wanted to see if after administering standardized testing on the two differing occasions, they could assume that a multiple-choice test serves the purpose of measuring knowledge acquisition and knowledge construction.

In their study, after administering the two tests to students, the researchers

(Struyven, et al., 2006), discovered through ANOVA and the Bonferroni comparisons that the groups which had taken multiple-choice tests outperformed the groups which participated in other assessment structures, such as portfolio and case-based assessments. The researchers also discovered that students assessed with the multiple-choice formats outperformed all of the other categories tested, except for those engaged in problem-solving activities. In the problem-solving activities, the students assessed in the multiple-choice formats still scored in an average range. In their conclusions, the researchers argued that the students who engaged in the multiple-choice question formats performed to a higher degree than the students evaluated during the class in an alternative format, such as with the portfolios. In fact, the researchers found that students assigned portfolio work typically waited until the last minute to do most of the project, and therefore the authors attributed some of the testing success to the last burst of learning that went into building the portfolio.

In further support of the multiple-choice format test and the summative evaluation technique of end-of-course testing the researchers (Struyven, et al., 2006), concluded that the enormous amount of content knowledge students had to learn prior to taking the assessments put them into the position of being able to focus solely on the final assessment. The students were not spending their time searching for the answers, contrary to the students who had been working in the portfolio format. The researchers concluded that multiple-choice testing was supportive of student performance rather than some of the other alternative assessment methods, such as portfolio and peer assessment. They determined that student-led, activity-based learning effectively competed with the multiple-choice format tests; however, results were inconclusive about the processes involved in students' learning. The researchers (Struyven, et al., 2006) surmised, though, that the type of assessment does make a difference. They recommended further research into the assessment process using triangulation of methods, searches for assessing student learning outcomes, and standards taught from the curriculums.

While Struyven, et al. (2006) studied the effectiveness of multiple-choice, end-ofcourse assessments, Thompson and Newsome (2002) wanted to discover if multiplechoice tests could also help facilitate the use of higher-ordered thinking skills in the classroom environment. Thompson and Newsome worked from the premise of three studies: those of Kohn (2000), McNeil (2000), and Yeh (2001), who collectively wanted to see what would happen if state tests focused more on higher-ordered thinking skills. Perhaps classroom teaching and assessment processes could improve because of including strategies for the sake of end-of-course assessments. Thompson and Newsome conducted workshops throughout North Carolina and involved more than 60 teachers who both analyzed and wrote end-of-course items. The result, according to the study, was a grouping of the testing items into specific categories, which highlighted different levels of thinking skills. The study helped state officials facilitate the inclusion of higherthinking skills and prove the testing framework could be a viable tool for classroom assessment (2002).

The growing trend in recent years, especially in North Carolina, has been that of a standards-driven curriculum, implemented in each classroom. The No Child Left Behind Act has pressured states and districts to develop accountability measures, which determine student progress. Classroom activities often focus on lessons with a test-oriented mentality, knowing that students, teachers, and even schools are accountable for

meeting minimum proficiency goals. In addition to the curriculum standards and end-ofcourse tests, there are still pressures to give students a learner-active education. Emphasizing a standard curriculum, assessing student progress with an end-of-course test, and implementing a learner-active approach in the classroom may provide a more complete learning environment for students.

Each student in the classroom learns differently, but the present test-oriented accountability system demands that every student test in the same format, with the multiple-choice end-of-course assessment. Guskey (2007), in agreement with Bloom's models on mastery learning, commented that teaching all students the same way, giving them the same timeframe in which to learn, and then testing for proficiency with the same format created diverse results in student learning. Bloom (1968) argued that each student could learn to the same competency level, but the process, strategy, and time needed for that to happen were different for each person. Bloom's mastery learning emphasized variation in learning tactics and highlighted that it was the teacher's responsibility to structure the delivery and assessment process to accommodate the dissimilarity in learning styles. Other studies (Gardner, 1983) have supported the need for varying instructional techniques, but with the increasing emphasis on standardized testing and accountability measures, teachers have to incorporate alternative learner activities. One of the practices mentioned in the previous theoretical section Bloom advocated, was that of a more learner-active approach.

Active student learning, according to Walberg (1986), involves implementing plans, which help students actively participate in the learning process. Activities may include cooperative learning, group presentation, peer- or self-assessment, and mistake correction after receiving feedback from the teacher. Bloom (1968), and later Guskey (2007), both advocated using a classroom much as a laboratory setting. Instead of students being passive recipients of knowledge, for instance, they would actively participate in activities. After gathering and compiling data from 7,000 high school students, Whiting, Van Burgh, and Render (1995) determined that students had better classroom and standardized test grades after having actively participated in the learning process. Concurring with Whiting, et al. (1995) was a meta-analysis conducted by Kulik, et al. (1990), which found there were fewer educational treatments that consistently demonstrated the level of student achievement than those which actively involved student participation in the learning process, such as receiving feedback and correcting mistakes. The researchers' results were in agreement with Walberg (1986), who argued that students using feedback and corrective measures learned more with less variation in achievement outcomes.

Struyven, et al. (2006), who researched the effectiveness of testing, also supported an active-learning approach to teaching, especially when evaluating students in a multiple-choice, end-of-course test format. In their study, the researchers commented that a student's performance on knowledge construction was higher than when a student merely acquired information for a test. The authors also stated, concurring with Bloom (1968), that weaker students performed significantly better and learned more when engaged in active classroom environments during monitored work time. Michlitsch and Sidle (2002), when advocating an active approach to learning, suggested arranging the classroom learning structure around a case-based assessment approach. These are problem-solving assignments, which might include real-life cases or problems, for which the students would do the appropriate research, discovering the information and contexts behind the legal and social aspects of the assignments. The researchers argued that this was one of the best strategies not only to get students actively involved in their own learning, but also to help further develop higher-order thinking skills in the process.

Clymer and William (2006) studied strategies of how standards-based grading systems improved learning expectations for students. The authors commented that classroom learning should be a dynamic process with the aim of involving each student, not a static environment that encouraged shallow learning. The researchers contended that if students understood the learning and testing process, then the pupils would have a deeper understanding of what they had learned, and by the end of the term, the learners would be more actively engaged in their own education.

To reinforce their theories, Clymer and William (2006) gathered survey responses from students on grading and classroom instruction techniques. The researchers determined, as did Dweck (2000) in a similar study, that when students assumed a more prominent role in their own education, the learners tended to develop a deeper understanding of the targeted curriculum goals. Students took more of an interest in what they were doing and became more involved in the education process. Suddenly education was an ownership issue for the student. The classroom-learning environment could be a place where the teacher was more of a coach than an instructor (2006).

When students become actively involved in the learning process (Clymer & William, 2006; Dweck, 2000; Black & William, 1998), there are also opportunities for interaction between weaker and stronger students. The students who benefit most from the active learning process are the highest- and lowest-achieving students, who have

increased exposure to one another and therefore have the opportunity to interact with and learn from others (1998). While researchers and educators alike advocate a more active approach for students in the classrooms, the question still remains as to which techniques are the most effective for both encouraging student learning and increasing assessment performance, especially since the recent emphasis on improving standardized test scores remains a major issue in education. A more current shift to active student participation involved the performance strategies of self-grading and correcting, which was one of the focus points of this study.

#### Performance Strategies

There are several performance strategies, such as cooperative learning, which have received careful attention from educators, but only recently have researchers begun to study student self-grading and corrective measures as a possible performance strategy. Traditional educational practices, as previously mentioned in this study during the time of Dewey (1968), meant the classroom arrangement had desks in straight rows, and the teacher formally assumed a dictatorial position of authority and information dissemination. Students received information, answered test questions with memorized answers, and received feedback only when the teacher returned the graded papers.

To assess and return papers quickly, some teachers currently allow students to grade quizzes, and peer grading is more common than self-grading (Kirby, Downs, & Colleen, 2007). Teachers traditionally viewed self-grading in a less positive light because of the possibility of cheating (Sadler & Good, 2006; Edwards, 2007; Strong, Davis, & Hawks, 2004). In recent years, some researchers have more carefully examined self-assessment as a strategy for deeper learning and a possible improvement for testing.

Through the years of 1961-1989, Falchikov and Goldfinch (2000) conducted studies on alternative assessments, most notably self-grading, and the possible benefits of tests as learning tools. Falchikov and Goldfinch argued that universities have utilized self-assessment for years, as the collegiate settings have encouraged active learner experiences grounded in the philosophies of Piaget and the constructivist thoughts of Vygotsky. Falchikov and Goldfinch (2000) argued that self-grading would result in detailed self-examination of the learner's progress, which would naturally lend itself to a learning experience. Universities increasingly utilize self-assessment techniques, especially in the business and medical fields, where there is necessity for a student to effectively examine and analyze his or her performance (Falchikov & Boud, 1989).

Falchikov and Boud (1989), for instance, conducted a meta-analysis involving 48 studies focused on student self-grading. At the onset of their study, the researchers' intentions were to examine the differences between students and teachers' grading marks when assessing the same student performance. The authors concluded that the students who had the most educational experience, graduate students, for instance, were able to self-grade with the greatest degree of accuracy. The researchers also discovered that the more experienced students were most likely to underestimate their own performance.

While the underlying assumption was that self-grading students inflated grades, Falchikov and Boud (1989) determined that in most of the studies there was no overall consistent tendency for students to under- or overestimate their performance. The authors found, though, that stronger students were more able to assess their grades than weaker students, who tended to inflate marks (1989). Building from the studies of Falchikov and Boud's meta-analysis, Strong et al. (2004) conducted a case study on selfgrading and determined self-assessment was beneficial to the learning experience. The authors performed their study on a college-level, general education class and hoped to determine whether the self-grading systems were appropriate.

Their case study (Strong, et al., 2004) spanned a period of two semesters and involved 480 students. In both semesters, groups of students took the same number of quizzes, projects, and tests while taught in the same format. The classes taught in the same classroom had every aspect duplicated as nearly as possible, such as the instructors, teaching assistants, and grading policies. The research study began with the intended delivery of instruction followed by the formative assessment. During each semester, the instructor followed this structure then concluded with the final course exam.

At the end of the first semester, the instructor assessed the students' total performance in the course and determined class grades based on the same grading practices used for the second semester; however, the instructor did not report grades to the students. Strong, et al. (2004) then had each student meet with one of the teaching assistants to discuss the course and his or her grade. During the conference, the student learned his or her ranking in class. This was meant to make the student aware of his or her relative position in class grade rank and in the shape of the general grading curve.

As the meeting concluded between the teaching assistant and students, the students were given a self-evaluation form to complete, which was meant to help provide structure and evenness to the self-grading process (Strong, et al., 2004). The students were to use the self-evaluation form to review their performance in class and assign grades. After the meetings, the teaching assistant compared the grades the students had given themselves to the marks the instructor had assigned. The same course delivery, test, and grade assignment methods became standard the second semester, except only the instructor awarded grades.

The researchers gave participants in the second semester course a questionnaire, which was meant to survey students' opinions of the grades they received and their view on the value of self-assessment (Strong, et al., 2004). The researchers initially worried about grade inflation, and their study did support the concern. The authors found that 57 percent of students rated their total course performance in the A range, whereas the instructor-assigned grades only reflected 31 percent had achieved that score. The researchers surmised that student self-grading might improve the effectiveness of certain assessment practices, such as multiple-choice tests.

The final determination of their research led Strong, et al to conclude that selfassessment, if properly implemented, could effectively assess student learning (2004). While grade inflation was a problem, the researchers recommended using self-assessment in a smaller classroom setting with students properly trained to perform to the standards of the instructor. Concluding, the researchers found student self-grading was a positive learning experience for the students, and as a result, recommended in the future that the instructor shift more to teaching and self-grading activities than to just assigning grades.

While researchers (Falchikov & Boud, 1989) have commented positively on the benefits of student self-assessment, such as enhanced critical-thinking analysis and improved student learning (Freeman, 1995; Sterling, 2008; Sadler & Good, 2006), Stefani (1994) conducted a research study to determine the reliability of student-assigned marks with potential learning benefits in self-grading exercises. In her study, the subjects involved were two first-year undergraduate biology classes engaged in lab activities. At the beginning of the study, the students determined the assessment scale, thereby assuring a high degree of student ownership throughout the process.

When the laboratory activities for the two classes were completed, the students had seven days to finish and submit their reports. The teaching assistants graded the students' work, but did not release the grades to the participants. After the seven-day period had passed, the assistants returned the work to the students. The students self-graded their papers and returned the work to the teaching assistants. After examining the results, Stefani (1994) determined that when the students self-graded their papers, the grades were more stringent than when the teaching assistant assessed the work. There was also an indication that students who received higher marks from the assistant were likely to have underrated themselves more frequently than students receiving lower marks.

Stefani's final determination (1994) was that the use of student self-grading, in place of the assistant's scoring, resulted in a similar scoring pattern for most of the grades, with only a small tendency towards underestimating. She did note that students seemed more motivated and interested in the lab assignments than usually observed, and she wondered if this had to do with a greater sense of student involvement. Stefani commented at the conclusion of her study that one characteristic of an effective learner was that he or she had a realistic view of personal strengths and weaknesses, and she argued that learning to self-assess was a valuable part of the education process. She also reported that "the correlation between the students who self-assessed during the course of the semester and the outcomes of their final exams had an r value of 0.71, while the correlation between the grades assigned by the teaching assistant and the final outcomes

on those exams were given an r value of 0.58" (p. 73). Stefani (1994) explained that this statistic was an intriguing result because it suggested that when self-grading alone determined exam results, the outcome was moderately similar to the grades obtained when the teaching assistant was responsible for assessing the final exams.

Stefani reported that after the study had concluded, students completed a questionnaire concerning their perceived experience with self-grading. She commented that almost 100% of the students who responded said that the self-grading procedures made them think about the responsibilities and requirements of the course more, and 85% of the students claimed that they learned more than when engaged in the traditionally structured classroom environment. While Stefani still questioned the validity of relying on student self-grading to determine marks in every instance, she argued that if students were to become accustomed to self-assessment early in the class, the integrity of the grading system might be manageable as students became familiar with the procedures.

While there have been numerous studies on student self-assessment (Falchikov & Goldfinch, 2000) at the college level, only recently have the emphases on standardized tests sent educators scrambling to find alternative classroom methods to increase students' learning. Using the premise of Stefani's study on the benefits of student self-assessment, Sadler and Good (2006) built on the idea that self-assessment not only increased student performance and learning, but also helped teachers preparing for standardized tests. For their research study, Sadler and Good involved four middle school science classes. They intended to compare grades awarded by the teacher to grades the students both assessed themselves and peer-assessed, to determine if the results were comparable. Like Stefani, Sadler and Good stated that there were

considerable benefits for students who took an active role in the classroom learning process, and one of the best ways to do this was for the learners to become involved in grading. Sadler and Good argued that using self-grading as part of the students' learning experience may have benefits that transcend the subject-matter acquisition. Students are able to look deeper in their own learning and begin to become aware of their own strengths and weaknesses, as Stefani (1994) mentioned.

Sadler and Good (2006) also argued that bringing the students into a learneractive environment could make classroom activities more productive, friendlier, and more encouraging for students to work in a cooperative role together. Additionally, the researchers contended that the reasons for, and value of, testing became more apparent to the students, as did their desire to work with a deeper sense of motivation and purpose. The authors reasoned that when students worked within the grading structure of the classroom, then ideas and any negative emotional responses they felt toward testing began to disappear as the testing experience became less threatening and mystifying. Students were suddenly partners in the learning process instead of testing subjects.

While Stefani's research study focused on the self-grading capabilities of a college level science class, Sadler and Good (2006) stated that very few studies have actually focused on elementary and high school age children. Instead, all of the studies they found, even the ones included in Falchikov and Boud's 1989 and Falchikov and Goldfinch's 2000 meta-analyses, occurred at the college level. With their 2006 study, Sadler and Good evaluated the possibility of classroom teachers using self-assessment in the K-12 range, too, in hopes of finding self-grading to be beneficial for both teachers and students.

When their study began, Sadler and Good (2006) worked to determine how close grades assigned by the teacher were to the grades determined by the students. The researchers wanted to discover if student self-grading could be a substitute for teacher grades and if student grading could be a tool for increased learning. The study was conducted in four heterogeneous science courses where the class means for prior tests and quizzes were all within 0.5% of 85%. The participating classes were assigned to one of the four groups: the control group, in which there was teacher-only grading; a group that self-graded; and two groups which engaged in peer grading.

The teacher had constructed the test, which contained both multiple-choice and essay questions. The students used their notebooks during the test. When classes finished with the assessment, the teacher conducted a discussion with the students to allow input in devising the grading rubric. One week after administering and grading the first test, the teacher gave an identical test to the class with the same conditions as the first assessment. The researchers analyzed 386 test grades. They mainly used descriptive statistics, such as establishing means, standard deviations, *t* tests, and ANOVA (Sadler and Good, 2006).

In their analysis, Sadler and Good (2006) determined that self-graded papers correlated with the teacher-awarded grades (r = 0.976), which demonstrated a high interrater reliability. When compared to the students who peer-graded another's paper, the researchers discovered that the self-assessing students tended to average about five points above the peer-graded students in grades, not because of grade inflation but because of higher learning gains. According to Sadler and Good, the self-grading students tended to make larger achievement gains at all ability levels than the other group whose teacher graded the papers. The researchers determined that self-grading was the more effective technique at all performance levels.

In attempting to answer their early hypotheses, Sadler and Good (2006) responded that the results supported the fact that self-grading students' scores correlated close enough with the teachers' marks to be a reliable substitution, even by seventh graders. The researchers argued that "students at all levels appear to benefit from self-grading, with significant gains at the lower and middle levels" (p. 25). Sadler and Good (2006) concluded their study by commenting:

Student grading is not an isolated educational practice, but is a part of the system of learning and assessment carried out in a teacher's classroom. It involves sharing with students some of the power traditionally held by the teacher, the power to grade. When used responsibly student grading can be highly accurate and reliable, saving teachers' time. In this study, self-grading appears to further student understanding of the subject matter taught. (p. 28)

Research studies conducted by Sadler and Good on student self-grading demonstrate that particular technique can improve learning and understanding. Student self-assessment in the K-12 setting could help students become more actively involved in their education and help in preparation for another situation, such as encountering the states' end-of-course tests. While Sadler and Good (2006) argue that student self-grading is helpful in encouraging learning achievement, it is only one part of the grading and performance process. Self-grading is very important in encouraging active student involvement in classroom activities and in learning outcomes (Stefani, 1994). This research study took the idea of self-grading from Sadler and Good and tried to determine if the addition of students correcting their own papers would improve learning, which may positively affect students' scores on the North Carolina end-of-course test in United States History.

Even though there have been numerous studies conducted on the effects of student self-grading (Stefani, 1994; Sadler & Good, 2006; Falchikov & Boud, 1989), there have been few studies on the effects of student corrective measures in the learning process. According to Forbes, Popard, and McBride (2004), teachers who both encourage students to make corrections on class work and stress the value of it are teaching the students to be independent problem-solvers. The researchers' area of study was teaching reading, so they knew the value of allowing students to make mistakes, providing the opportunity to work through problems, using the mistakes as a way for students to realize errors, and then working to correct the misconceptions. The researchers argued that learning to read is learning from correcting mistakes, and therefore correcting mistakes is a very necessary part of the classroom experience (2004).

Forbes, et al. (2004) commented, "Correction is an observable behavior from which we can infer the reader has engaged in monitoring and searching strategies" (p. 2). The authors argued that when students began learning from their own mistakes, the learners also benefited from self-instruction and felt intrinsically rewarded. The authors found that students who are high-achievers corrected themselves much more frequently than lower achieving students and that self-correcting behavior probably has a tutorial value for struggling students as well. The researchers believed that students who routinely self-corrected were more likely to have developed metacognitive skills indicative of the progress older students made after having learned to read. The findings indicated that when students realized they made mistakes, the learners would oftentimes be eager to correct the errors and then use what they knew in application to other learning situations. Finally, the researchers supported the belief that teachers who valued corrective behaviors in children indeed helped the students in problem-solving skills, which the pupils took into other subjects and opportunities for learning.

While the idea of using corrective measures encouraged students when they were learning to read, Clymer and William (2006) supported Forbes, et al. (2004) in the notion that when students corrected themselves, learning improved, especially in standardsbased grading systems. In their study of grading practices in the science classroom, Clymer and William found that students performed better when given feedback that did not just say they had done a good job, but offered corrective advice on how to learn from the incorrect responses. Bloom (1968) agreed, in his discussion of mastery learning, that one the biggest problems for a classroom teacher was that the instruction technique was simply not appropriate.

Bloom (1968), for instance, believed teachers should use their classroom assessments as learning tools. Teachers should provide a learning environment where students can receive immediate feedback and use that feedback to guide themselves in correcting errors. Bloom recommended that tests become part of the classroom-learning environment, where identification and remediation of student problems follow. Bloom recommended the use of feedback and corrective measures in his mastery learning, and as Guskey also mentioned (2007), students would have an opportunity to overcome their difficulties and then have a second chance at success.

Guskey (2007) argued in support of Bloom's (1968) theories when he commented that in the normal course of classroom testing, coupled with the students incorporating corrective measures on their errors, all students could learn more and receive a better education than was typical under traditional methods of teaching. Guskey (2007) further stated, "By itself, however, feedback does little to help students improve their learning. Significant improvement requires feedback to be paired with correctives: activities that offer guidance and direction to students on how to remedy their learning problems" (p. 16). In the case of students utilizing teacher feedback for corrective purposes, Guskey believed in handling corrections differently than from the original delivery of instruction and integrating different learning strategies. For instance, Guskey commented that merely giving papers back and letting students rework the missed problems did little to help them learn from their mistakes. For a student to really improve, learning feedback activities needed to be structured in such a way that the student would receive guidance and direction from the teacher as well as from another student who would act as a tutor (2007). Students can learn from the testing process, especially when incorporating selfassessment and using their completed work to guide revision efforts (Andrade & Du, 2007).

Arguably, current searches into the literature indicated a growing interest toward utilizing student self-corrective measures in the classroom, but most of the recent studies focused on self-corrective measures for spelling words, foreign languages, and students with disabilities. The research studies focused on corrective measures in spelling and students with disabilities (Alber & Walshe, 2004; Viel-Ruma, Houchines, & Fredrick, 2006) have centered attentions mainly on elementary school students. Conversely, studies involving the use of corrective measures for foreign language acquisition have concentrated mostly on college-age or adult learners (Yoshida, 2008; and Hall, 2007). Without exception, all of the preceding studies recommended the possibility of conducting further research with high school students, which was the intended focus of the present research study.

One of the problems facing instructors utilizing the corrective efforts, though, is the question of how soon after receiving feedback students should begin to analyze and correct their errors. Mathan and Koedinger (2005) contend, in their study on delayed versus immediate feedback, it is important in the learning process that corrective measures begin as soon as possible after the actual test. The researchers argue that students rely on feedback more and more as the learners begin the self-corrective measures, and timely feedback is important to guide the error modification process.

Mathan and Koedinger (2005) conducted a research study to determine if immediate feedback followed by corrective measures seemed to work better for learning and achievement. The researchers selected participants from a local temporary employment agency. The experimental session lasted for three days and involved several sessions. The experiment consisted of a pre- and posttest design, with sections containing problem-solving, conceptual understanding, and multiple-choice questions. At the conclusion of the pre-test, one group of students received immediate feedback and engaged in corrective measures at that time, while the other group had to wait three days for the same feedback and resulting corrections. After both groups had completed identifying and correcting their errors, the participants tested again. The researchers determined the group which had received immediate feedback and promptly undertaken the corrective measures performed at a significantly higher rate of success on the final test than the group that had waited three days before beginning corrections.

Conversely, Mahan and Koedinger (2005) argued that immediate feedback and corrective measures, such as student self-grading and correcting, could hinder the learning process because learners might not exercise the new skills outside the testing environment. Their research, however, supported the notion that immediate feedback and correction methods are more effective than grade feedback alone.

Concurring with Mathan and Koedinger's research was the study conducted by Grobe and Rendle (2007), on finding and fixing errors in worked examples. The researchers contended that when students worked through examples, such as in mathematics, it was important for the learners to recognize mistakes, correct errors, and apply the concepts. Grobe and Rendle (2007) argued the necessity to study the benefits of an incorrect solution for three reasons: an error is an inherent part of human life, previously understood knowledge is persistent, and the probability of having a right answer can be increased by reducing the chance of getting a wrong answer (p. 21). The researchers conducted a study in which students either were given incorrect answers to worked problems or had problems with incorrectly worked steps (2007). The participants then worked at solving the problems. Grobe and Rendle concluded from the outcomes that while students did not know if feedback from the instructor was incorrect, in either the solution or the steps, students seemed to benefit most when they were attempting to write self-explanations. The authors determined that a mixture of correct and incorrect solutions enhanced the final student outcomes when comparing test results. The

researchers also concluded that when students had to explain which part of the worked problem was erroneous, the quality of the students' self-explanations improved.

Further research by Yoshida (2008) and Hall (2007) determined that learning improves with self-corrective measures in the classroom, and if there were more time available for students to work with corrections as a guided activity amid teacher support, students could indeed learn more. The researcher continued by saying that in the classroom students do not always pay attention to what the teacher communicates. When the teacher gives students an opportunity to work through and correct items for themselves, the students may have more interaction time with the instructor. Yoshida concluded with the comment that even though finding time for students to work through corrective measures in class was difficult, teachers should try to arrange opportunities for more self-correction.

While research studies have supported the use of corrections in foreign language classes, there has been growing support for the use of self-corrective strategies with students in elementary school who are struggling with disabilities or having difficulties with spelling. Shelia and Walshe (2004), for instance, conducted a research study that focused on six students and their weekly spelling words. The research study involved giving the six students a weekly list of 20 spelling words, which were divided into two lists of 10 words. The students practiced writing one group of words, then went back and corrected the few misspellings after completing the list. With the other group of words, the students wrote the list and immediately self-corrected if there were mistakes.

Shelia and Walshe (2004) determined from their study that all six students spelled a higher percentage of words correctly when the learners had practiced self-correction. The researchers argued there was a difference when corrective measures occurred, as the findings indicated that the longer a teacher waited to begin corrections the less effective was the learning process. The authors argued the importance for students to receive immediate feedback after obtaining new skills so they did not practice items with errors.

Viel-Ruma, et al. (2007) conducted another research study working with spelling and students with disabilities. Like Shelia and Walshe (2004), they found that immediate feedback from the teacher, followed by self-corrective measures from the students, increased student learning and understanding. Viel-Ruma, et al. asked three students with learning and spelling disabilities to participate in their research study. The students received 16 vocabulary words every week. During the first week, students learned the words through the traditional method of writing the word three times while looking at its correct form. The second week students used an error self-correction strategy of writing the word and then checking its spelling accuracy (2007).

The researchers determined, as did Shelia and Walshe (2004), that the use of student self-corrective measures was more effective at improving spelling performance with the students; however, Viel-Ruma, et al. (2007) also noted several difficulties, which they had not anticipated encountering. For instance, the researchers found that while the students reported learning more using corrective measures, the students did not claim to prefer any particular strategy. Additionally, the authors commented that the students had little desire to engage in self-corrective measures and recommended further research into ways of making self-corrective strategies more desirable (2007).

Most of the research studies examined here involving self-grading and selfcorrecting have cited increased student learning, both in problem-solving skills and from
the perspective of today's standards-driven classroom environment. Educators are trying to increase standardized test scores, especially since the No Child Left Behind legislation, because of the desire to increase student learning. While there are current trends, such as classrooms structured on a standards-driven, test-oriented path, performance strategies, such as student self-grading and correcting, may improve test scores and increase student learning. Research in the areas of self-grading and correcting have shown those strategies to be promising as classroom reteaching-retesting strategies (Sadler & Good, 2006), but there are other potential benefits as well, such as increased learning, enhanced self-efficacy, and higher standardized test scores.

### Benefits from Self-grading and Correcting

The previous sections in the empirical review have emphasized current learning trends and performance strategies, which could improve classroom testing and learning outcomes. Increased student learning, enhanced self-efficacy, and the challenge of a higher standardized test score comprise this concluding portion of the empirical review of literature.

In their 2006 study on self- and peer-grading, Sadler and Good cited several benefits to students grading their own papers. While the authors argued that self-grading were beneficial for the teacher on several levels, they also found that when students tested a second time there was an increase in understanding. Sadler and Good stated that even in previous studies, such as Falchikov and Boud's 1989 research on self-grading, when students received quicker feedback their understanding about a topic was deeper, and the learners became more aware of their own academic strengths and weaknesses. Continuing, Sadler and Good (2006) found that students were also more interested in the learning process when grading their own papers, and therefore were perhaps more willing to spend extra time preparing for class and studying. From the results of the study, the researchers claimed that students who self-graded routinely performed better at higher-level skills than students who received graded tests from the teacher only. Students who graded their peers' papers, according to the researchers, did not seem to have gained any further understanding than from the usual experiences of teacher-graded tests. The study found that students who self-graded their test papers did appear to have a better understanding of the material.

Additionally, other research studies agreed with Sadler and Good's assessment. Stefani (1994), for instance, concurred with Sadler and Good's (2006) argument that students experienced an increase in learning when engaged in the process of selfassessment. She stated that when self-grading papers, students tended to have more of a realistic perception of their own abilities and, as a result, became more self-aware and critical of shortcomings. The author stated that knowledge of a weakness could strengthen academic standing once a student became aware there was a problem. She found when students self-assessed their own tests that almost 100% of the time the learners said it made them think more, and 85% of the students claimed that they had learned more through self-grading than when traditionally assessed by the teacher (1994).

Additional research studies by Freeman (1995) and Struyven, Dochy, Janssen, Schelfhout, and Gielen (2006) supported the notion that students who utilized selfgrading and even self-corrective measures in the classroom learned more than students who only received graded papers from the teacher. Freeman's research found that when teachers incorporated self-assessments into the classroom environment, learning increased and standardized test scores for the group were better. He believed that selfassessment motivated students to learn at a deeper level, and their thinking and learning skills became more enhanced by the experience. He further stated that the assessment system a teacher used highly affected the performance and outlook for a class of students.

Struyven, et al. (2006) argued in favor of using self-assessment in the classroom, as they saw a significant positive relationship between student performance and the resulting effect it had on the end-of-course assessments. The researchers found that when students took tests based on the multiple-choice formats, such as an end-of-course test, performance seemed to improve because the learners were more actively engaged in their own instruction and assessment, as with self-grading activities. In addition, the research findings also concurred with Freeman's (1995) conclusions that the students' perception of the assessments, coupled with their involvement in the learning process, influenced learning. Struyven, et al. stated that if educators wanted students to learn in deeper, more meaningful ways, then students needed to participate in assessment activities that would challenge and help them want to learn.

Struyven, et al. (2006), found that while the active participation of students in the assessment process affected learning, the type of assessment might not actually produce an effect on student learning. In the report, not all of the types of assessments studied showed comparable results. The study pointed favorably to the multiple-choice tests, though, because the authors argued the possibility of students learning for understanding, even when involved in an objective-testing format. The result, the researchers reported, was that the students received high scores on both knowledge acquisition and knowledge

test items. This helped the authors determine that the choice of a particular testing method and the actual assessment chosen made a difference in learning. The researchers stated that it was possible to conclude from their results that multiple-choice testing was more beneficial to student performance and learning, in comparison to some of the other testing methods, such as portfolio assessment.

Agreeing with Struyven, et al. (2006) were Falchikov and Boud (1989) and Kitsantas, Reiser, and Doster (2004) in their studies on self-grading and the potential affects it had on students. In their meta-analysis, Falchikov and Boud stated that "lifelong learning requires that individuals be able not only to work independently, but also to assess their own performance and progress" (p. 395). Kitsantas, et al. determined from their study that when students became more self-regulated learners, such as when they participated in activities like self-assessment, there were a variety of positive outcomes including a higher degree of skill acquisition and satisfaction. They believed that students who routinely engaged in self-evaluation during activities usually outperformed students not encouraged doing so. Kitsantas, et al. (2006) stated, "Research has shown that students who evaluate their own work are more likely to attribute poor performance to strategy deficiency rather than to effort or ability and, thus, search for new ways to enhance their learning" (p. 271). Student performance and attitude, according to Kitsantas, et al. (2006), affected learning, while Strong, Davis, and Hawks (2004) additionally stressed the need of having a learning environment where students felt free to experiment creatively with learning.

Strong, et al. found, too, that some of the students in the study believed selfgrading affected their desire to learn more. Students also felt encouraged to try different ways to learn. Guskey (2007) expressed belief in self-grading by asserting that all students could have better grades on the end-of-course formative assessments, gradepoint averages, and attitudes toward learning all school activities when students used selfassessing techniques. Walberg's findings (1986) agreed that self-grading and corrective measures helped students learn and had the potential of closing achievement gaps in testing.

While researchers and educators alike have been trying to discover the best series of strategies to increase learning and raise standardized test scores, their studies into student self-correcting and grading have yielded additional benefits, such as increased motivation and a positive sense of self-efficacy (Andrade & Du, 2007). For instance, Andrade and Du found during their study of undergraduate students' experiences with criteria-referenced self-assessment, that students reported having more of a positive attitude toward class work, teacher expectation, and the course of study. Additionally, the research subjects reported that the experience with self-assessment made them feel there had been improvements in their quality of work and motivation to learn.

Concurring with the findings of Andrade and Du (2007) was Locker and Cropley's (2004) study on the effects of testing anxiety in male and female adolescents. They found that with the increased emphasis on standardized testing, students were reporting more anxiety, more stress to succeed, and a greater need to perform to expected standards. Locker and Cropley first wanted to measure how much anxiety students felt during the classroom instructional day and testing time. The researchers determined that when students had engaged in a more active role during the typical classroom day, such as utilizing the strategy of self-grading, they reported feeling less anxious and stressed when taking standardized tests (2004). Students reported more feelings of confidence and self-esteem, agreeing with the findings of Stefani's reported benefits of self-grading (1994).

Stefani's research (1994) indicated that when students utilized self-assessment during instruction, the learners had a more realistic perception of their abilities and knew which tendencies were strengths and weaknesses. The researcher noted that students tended to be more highly motivated and more interested in the learning tasks when engaged in self-grading strategies. She also argued that a problem teachers have with allowing students to participate in self-grading activities is the traditional notion that instructors should have the power in the classroom. The author used grading as an example of the teacher exercising their power and control over the learning environment. When an instructor began class with teaching students to use self-assessment strategies, according to Stefani, the students associated the teacher with the role of a facilitator, and tended to feel more comfortable, both in the classroom environment and in their relationship with the instructor.

Current research studies, too, concurred with Stefani's findings, such as the study by Strong, et al. (2004) which compared student grades over two semesters between classes that had incorporated self-assessment and classes in which the teacher assigned all of the grades. They found in the classes where the students had engaged in selfassessment practices that students reported feeling more positive about their total classroom experiences. Students indicated on a questionnaire, and when interviewed by the teaching assistants, they felt more motivated to learn because of self-grading and experienced a greater sense of responsibility for their own learning. Some students mentioned wanting to learn more about the subject once the classes had concluded, and 53% of the students claimed they had a better understanding of the material. Finally, the majority of the class agreed that involvement in self-assessment helped them work harder on assignments. The participants also claimed self-assessment increased the quality of their thoughtfulness and made class a more enjoyable experience.

Strong, et al. (2004) concluded in their study that students who engaged in selfgrading found it to be effective, fair, and appropriate. The researchers (2004) determined, "Self-assessment opens doors for increased student interest, motivation, creativity, learning, and retention, thus improving the possibility of having successful academic experiences" (p. 55). To them, when students participated in the selfassessment process, grading itself dropped in importance, and teaching became the focus.

In another study with positive outcomes concerning student self-assessment, Pajares (1996) stated that self-grading was a key component in student learning and motivation and that increased self-efficacy was a major benefit and often overlooked. Pajares spoke of self-efficacy as an individual's perceived abilities to accomplish and achieve specific results. He stated that self-efficacy influenced self-regulated learning goals, and the student who was confident of his or her abilities in the classroom would feel more motivated and inclined to put forth additional effort in his or her academic pursuits.

Edwards (2007) agreed with Pajares on the positive effects of self-efficacy when students graded their own papers. Edwards experimented with self-grading at the undergraduate level and found that students had more interest in a class when they actively participated in the grading process. The majority of Edwards' students reported feeling better about the quality of work they were producing, their experiences in class, and their understanding of expectations as learners. The author found students expressed more self-confidence, self-acceptance, and self-esteem when they had graded their own papers versus when the author had marked the assessments. She concluded by saying that one of the traditional problems underlying most classrooms were the student-teacher conflicts which occurred because of grade expectation and anxiety. Her students responded favorably at the end of the course in which they actively participated in selfassessment by commenting that classroom anxieties diminished and the relationship with the instructor improved. A more relaxed and enjoyable atmosphere greeted them upon entering the class each time. Students reported feeling more positive about the class and their role as learners when allowed to self-assess the work.

Kitsantas, et al. (2004) stated, too, in their study on self-regulated learners and goal setting, that students experienced a greater sense of self-efficacy and satisfaction with their progress when in a class that encouraged self-grading and correcting techniques. The researchers found students had a higher level of skill acquisition, based on higher grades obtained in these classes, and, in addition, the learners rated instruction more positively. The authors discussed how students involved in the self-assessment process tended to outperform students who do not take an active part in grading. The research subjects, like those of Pajares (1996), also reported heightened feelings of selfefficacy, competence, and satisfaction.

Finally, Sadler and Good (2006) in their study on self- and peer-assessment, found benefits in student motivation and self-efficacy when they studied how students took an active role in the grading process. The researchers determined that students at all levels of performance benefited from self-grading and felt it was a valuable activity. Sadler and Good, like Kitsantas, et al. (2004) also claimed that students who engaged in self-grading reported being more confident in their classroom abilities and in willingness to attempt newer, more difficult tasks when provided with the chance to do so. Other researchers, too, like Andrade and Du (2007) and Tan (2008), reported finding that when students actively participated in their own learning, as with self-assessment, they did feel a greater sense of motivation and self-efficacy. Andrade and Du stated that students' grades improved, as did the sense of what the learners considered quality work. Students were more keenly aware of the true meaning of the classroom standards and teacher expectations. Tan concurred by saying that self-assessment enhanced a student's desire to further his or her lifelong learning and empowered rather than disciplined the student.

While student self-grading and correcting have been shown to benefit learning, enhance motivation, and improve self-efficacy, there are research studies which support self-assessment's potential in raising standardized test scores. Davis and Rand (1980) studied the effects of self-grading versus instructor grading on the performance of two classes of graduate psychology students. At the end of the course, the authors compared the results of the final semester and exam grades and determined there was no significant difference in student performance between the self-graded and teacher-graded classes on the quizzes or paper, but there was a significant difference on the exam.

Concurring with Davis and Rand (1980), Guskey (2007) found that when teachers used student self-grading and correcting measures in their classrooms, the students' grades on formative assessment measures were higher than with teacher-only grading. Additionally, Whiting et al. (1995) conducted research among 7,000 students and collected data, which suggested that when students engaged in actively grading and correcting their own work, there was a positive influence on the test scores and grade point averages.

While there have been studies involving the effects of student self-grading on testing, most of the studies, as mentioned previously, have focused on the college student or in the business world and not on grade school or high school students (Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000). Sadler and Good's (2006) study was one of the few which undertook research at the adolescent level because traditional thought held that only older students were reliable enough to self-assess accurately. Falchikov and Boud's meta-analysis concluded that most studies on self-grading typically included students in professional programs, such as medical students, who learned to analyze their progress and self-assessment methods and strategies. Most studies, Falchikov and Boud reported, allowed self-evaluation for projects, posters, and group work. When core academic courses instituted the strategy of self-grading, the most common subjects, according to Falchikov and Boud, were college science classes because of a more definitive assessment series typically involving objective-formatted tests.

This research study resumed where some of the other studies have ended. It attempted to discover if students in a United States History class, who engaged in selfgrading and correcting over the course of a semester, would show growth in learning and demonstrate improvement on the North Carolina end-of-course test in United States History. This review of literature has attempted to show a rich history in the theory of active student involvement and the resulting increase in learning. The empirical literature has demonstrated a foundation of practice and research in modern learning trends, performance strategies, and the potential benefits of self-grading and correcting. This study will add to the body of knowledge by increasing the scope to encompass high school juniors and the subject of United States History. Student self-assessment and correction could be effective strategies to help classroom teachers improve students' learning and the scores on the end-of-course tests, both of which are of paramount concern for today's educators. The next section discusses in detail the methodology utilized in this research study, followed by a presentation of the findings and the final summary of the dissertation's outcome.

#### CHAPTER THREE: METHODOLOGY

The following section explains the methodology used to determine the impact of students self-grading and correcting their test papers on learning and standardized test scores. This research study occurred at the high school level focusing on United States History, where assessments are typically in the multiple-choice format. The study utilized a population which involved four 11<sup>th</sup> grade United States History classes. Each class took a pretest and participated in the intended research intervention of either student self-assessment with correction measures or control group selection. The study concluded with the administration of the North Carolina end-of-course test in United States History. Two United States History teachers participated with their classes, while both teachers utilized grading from treatment and non-treatment groups.

#### Design of the Study

In this study, the research perspective was quantitative, and the design type was quasi-experimental research. This study proceeded in preexisting, intact classrooms and followed the subtype of Static Group Comparison Design (Ary, Jacobs, Razavieh, and Sorensen, 2006). This study used inferential statistics in the data analysis, with consideration for validity factors.

### Statement of the Problem

Teachers have traditionally assumed an authoritarian role in the classroom setting, especially in marking test papers and assigning grades (Guskey, 2007). Guskey stated that standardized tests have reinforced the idea of authoritarian roles for educators, as teachers scramble to cover the standard curriculums and prepare students for their end-ofcourse testing experiences. When students assume involved roles and actively participate in grading their tests, final learning outcomes and end-of-course grades improve (2007). *Research Questions* 

1. Is there a positive, negative, or equivalent relationship between students grading and correcting their own test papers and a higher score on the North Carolina end-of-course test in United States History?

2. Will a comparison between pretest and posttest grades show students learned more with the intervention of student self-correcting and grading than with teacher-only grading?

### Null Hypothesis

Students who self-grade and correct their test papers will not achieve significantly higher scores than students who do not grade and correct their own test papers on the North Carolina end-of-course test in United States History.

Educators should attempt to maintain a balance between the focus of students' learning and higher test scores. This research study offered a unique opportunity to challenge student learning and further develop a test score strategy from the position of active student participation.

### The Research Context

The site of this research study was two high schools, East Side and West Side, in rural, northwest North Carolina. (East Side and West Side are fictitious names, which will preserve confidentiality). The selection of these two particular high schools was convenient because they were close in proximity. Both schools had similar student numbers and demographics and aligned exactly in their pacing guides with the North Carolina Standard Course of Study. The two schools administered the same pre- and posttests devised by the North Carolina Department of Public Instruction. East Side and West Side High Schools are located in the Northwest Piedmont region of North Carolina. The population in this county is 79.1% white, 19.6% African-American, and 4.5% Latino (Census Bureau, 2007). The East Side and West Side High School areas have a lower socioeconomic population base with an average per capita income of \$17,120, as compared to the state of North Carolina, which has an average per capita income of \$20, 307 (2007). Traditionally the Northwestern Piedmont region of North Carolina has been primarily oriented toward manufacturing, but in recent years, a sizable number of employees have become victims of outsourcing. As a result, this county's unemployment rate is somewhat higher (7.4%) than in the neighboring counties (4.8%) (2007).

East Side High School is a public school with a student body numbering 1,111, encompassing an ethnic makeup of 77.1% white, 15.5% African-American, and 7.4% other (ESHS, 2007). The school is composed of four grade levels: ninth, tenth, eleventh, and twelfth. East Side High School has a support staff of 40 and a teaching staff of 72, with 36% having obtained a master's degree or higher. Fifteen percent of the teachers have National Board Certification (2007). Of the teaching staff, 68% have 20 or less years of teaching experience (2007). Since Adequate Yearly Progress (AYP) began in 2000-01, East Side has consistently scored in the Met Expectations range until the 2006-07 school year, when it failed to make AYP for the first time (2007). During the 2006-07 school year, the average daily attendance rate for East Side High School was 94 percent. There were also 75 student retentions combined for all grades in the 2005-06 school year (2007). West Side High School is also a public school with a student body numbering 1,096, encompassing an ethnic makeup of 75% white, 19% African-American, and 6% other (WSHS, 2007). The school is composed of four grade levels: ninth, tenth, eleventh, and twelfth. West Side High School has a support staff of 53 and a teaching staff of 77, with 32% having obtained a master's degree or higher. Eight percent of these teachers have National Board Certification (2007). Of the teaching staff, 47% have 20 or less years of teaching experience (2007). Since AYP began in 2000-01, West Side has consistently maintained a score of Expected Growth until the 2006-07 school year, when it, too, like East Side High School, failed to make AYP (2007). During the 2006-07 school year, the average daily attendance rate for West Side High School was 90 percent. There were also 82 student retentions combined for all grades in the 2005-06 school year (2007).

This research study commenced in four 11<sup>th</sup> grade United States History classes covering a period of one academic semester of 90 days, from late January 2009 until the end of May 2009. Both East Side and West Side High Schools operate on the block schedule, with classes beginning and concluding on the semester system. The data was processed and analyzed at the conclusion of the spring semester. These were averagesized area classrooms for both schools, and each had 35 student desks arranged in seven rows of five facing the front chalkboard. This particular subject is appropriate because 11<sup>th</sup> grade students in North Carolina are required to take United States History, and the end-of-course tests are in the multiple-choice format. The North Carolina Department of Public Instruction requires students to score at the proficiency level to pass the course. This means that a student must score at grade level, which is equivalent to a Level III or Level IV on the end-of-course tests (Glossary of Terms, p. 2).

The United States History classes, chosen for this research study assessment, were heterogeneous in nature and were examples of the make-up of this particular county's cross-section of population. Represented in this study were students with all ranges of ability, social class, ethnicity, and gender. The instructors for the course followed the North Carolina Standard Course of Study for the 11<sup>th</sup> grade United States History curriculum, and the four classes in the study covered the same material and assessed in identical fashion, with the pretest, weekly tests, and the end-of-course exam. One of the instructors was a white female, who has a bachelor's degree in history, and this year is her 30<sup>th</sup> year on the teaching staff at West Side High School. The other instructor was a white male, who holds a bachelor's degree in history, with this being his 27<sup>th</sup> year at East Side High School. These two classroom teachers participated because their scores were neither significantly higher nor lower than any other teacher in the respective academic departments, and because they were the only two teaching United States History classes the spring semester that had similar educational backgrounds and years of teaching experience. Additionally, the two teachers (neither of which was the researcher) chosen for participation in this experiment have been consistently involved in the countywide, cooperating efforts to align the pacing guide between schools. Because the four classes of students chosen were from two different, yet demographically similar, high schools, this helped control for the possibility of threats to internal validity, such as history, maturation, testing, instrumentation, and equivalency.

#### The Research Participants

After receiving permission to undertake this study from the Liberty University IRB Review Board (see Appendix A) and the involved principals, the researcher selected potential research participants. Participants were selected from two high schools. Participation in the research study was strictly voluntary, and inclusion began when a signed permission form had been returned from all students' parents.

### Selection Process

The population of interest in this study consisted of 11<sup>th</sup> grade United States History students in a rural, North Carolina community who attended East Side and West Side High Schools. Four intact United States History classes were the basis for the research. The four classes chosen represented a cluster sample of the typical history classroom in these particular high schools and were a heterogeneous section of the 11<sup>th</sup> grade. Teacher subjects and class population assignments were made official during the summer months while East Side and West Side High Schools were out of session. In both East Side and West Side High Schools, the guidance departments assigned students to their courses before the first day of classes. Typically, few schedule changes occur. Each high school has four guidance counselors who are responsible for 25% of the student body based on last name alphabetical listing. The teachers participating in this study had a completed classroom roster when school began. Each counselor was responsible for placing an alphabetical portion of the student body in class based on the student's registration, when the classroom was obtainable, and when an instructor was available. Counselors also had to be mindful of the need to keep student enrollment under 32, which is the maximum limit. Counselors are typically unaware of which

students belong to which teacher or of the other members of the classroom population. The counselors try to match schedule cards the students have filled out with the available courses, using student numbers. Since there could not be a true random sample represented, a coin toss determined which of the four intact groups was assigned to the experimental or to the control groups (Ary, et al., 2006).

All of these participating students had taken objective, multiple-choice tests on numerous occasions, both in the typical classroom assessment setting and in a standardized testing format, before entering class with these particular instructors and the subject matter. The history students had also been accustomed to the multiple-choice test format from other courses, especially where the state end-of-course tests had been the final assessments.

The participating students had had numerous opportunities to grade their own papers in high school. Starting from the first day of class in the spring semester, the two teachers chosen for this study also gave students the chance to grade their quizzes and multiple-choice homework assignments. The students became comfortable with assessing papers and had become accustomed to doing so in the format which this study required.

### **Research Subjects**

Four sections of heterogeneously grouped United States History students participated under two different teachers, involving two classes that constituted the control group and two which utilized the treatment. East Side High School's Class A was composed of 32 students with 13 males and 19 females between the ages of 16 to 18 years old. Class B had 32 students with 18 males and 14 females varying in age from 16 to 19 years.

West Side High School's Class C was composed of 32 students with 14 males and 18 females between the ages of 16 to 18 years old. Class D had 32 students with 16 males and 16 females varying in age from 16 to 20 years. The students in these four classes represented a variety of social, economic, and cultural backgrounds. For instance, 29 of the students were African-American, eight were Hispanic, two were Native American, and 89 were Caucasian. These classes contained honors, regular, and exceptional level students (see Table 1).

### Table 1

Stud	ent Num	bers fo	or East	Side	and	West	Sid	e Higl	ı Scl	hool
------	---------	---------	---------	------	-----	------	-----	--------	-------	------

High Schools	Males	Females			
1. East Side					
Class A	13	19			
Class B	18	14			
2. West Side					
Class C	14	18			
Class D	16	16			
Total:	61	67			

#### Instruments Used in Data Collection

In order to determine if a student grading and correcting his or her test paper would show improvement in overall classroom performance throughout the semester, several tests and measurements were necessary in the data collection for this quasiexperimental study. The school district requires teachers to give their students the countywide pretest at the beginning of every semester. The North Carolina Department of Public Instruction developed the pretest that United States History teachers used in this research study. The four United States History classes took it during the first week of the spring semester, so all of the classes involved in this study participated in the county pretesting exercise. Teachers received a grade report after the pretest and another concluding grade report with the end-of-course test grade results at the end of the semester.

# Tests and Measurements

Besides the standardized pre- and posttests, weekly teacher-generated tests, aligned with the North Carolina Standard Course of Study's specific learning goals, were administered to all participating students in both control and treatment groups. The tests were objective and consisted of released end-of-course test questions (see Appendix C). The cooperating teachers gave the same weekly tests. The reason for this particular format was to try to ensure as much uniformity as possible between pre- and posttest designs. The tests' 50 multiple-choice questions were checked for both validity and reliability at the state level before being included in the weekly goal assessments. All United States History courses in Northwestern Piedmont County were paced and aligned with the curriculum guide, so all students were tested on the same day, and every class took the exact same assessment. These weekly, teacher-generated tests contained only multiple-choice questions.

Typically, student grading has several variations, but only two major forms concerned the present study. *Teacher grading* describes what has become the typical classroom practice of the instructor assessing student performance on a test, while *selfgrading* refers to a student grading his or her paper (Sadler and Good, 2006). In the case of this research study, the students self-graded using a pre-coded answer key so there would be no question concerning whether an answer was or was not a correct response. The grading keys used were specific, and this should have enhanced test reliability. *Performance Assessment and Tasks* 

The students involved in the treatment groups graded their own papers, and then followed the test correction format (see Appendix D) to conclude the testing and treatment process every week. This study's results were controlled by measuring the scores of two spring semester classes that were the control group (Classes A and C), and any changes in scores for the treatment group (Classes B and D). Any corresponding change in end-of-course scores was detectable.

#### Measurement Guidelines

The end-of-course assessment was the final evaluation for all four United States History classes. During the last week of school, all of the county's United States History classes took the 100-question, multiple-choice, end-of-course test. The end-of-course assessment process was tightly controlled and proctored, and all state guidelines for administration followed to ensure testing validity. After the assessment process, the county's central office scored the tests and reported the scores to the schools and to the specific classroom teachers. An analysis of the data between the control and treatment groups occurred at that point. In this research study, none of the classes had permission to use notebooks or textbooks during the weekly tests, but students in the treatment group were encouraged to do so during the work on the test corrections.

#### *Internal validity.*

While the control and treatment groups had been carefully chosen to ensure as much equivalency as possible, the researcher also had to be mindful of possible threats to internal and external validity. There were threats to internal validity to control for during the course of the research, such as history, maturation, testing, and instrumentation. As for history and maturation, there may have been events that happened during the spring semester's experimentation that could have affected the posttest results. In order to help control for these, the researcher attempted to increase equivalence among the four groups that were part of the experimental situation. Counselors placed students in their respective classrooms, and the participants should have represented a cluster sample. The researcher determined class assignment to either control or treatment group by flipping a coin. This helped to ensure statistical equivalence and lessen the possibility of experimenter bias (Ary, et al., 2006).

Additionally, this researcher attempted to control for internal validity by using homogeneous selection. Because of the tendency for history as a subject to lend itself to numerous interpretations and alternative viewpoints, every group selected for this study was made up of 11<sup>th</sup> grade United States History students. While generalizing any effect the treatment may or may not have on history students, this strategy does decrease the extent to which the findings generalize to other populations (Ary, et al., 2006). Further

studies might determine if other students in different subject or grade areas would show positive results after taking the posttest.

In this research experiment, the threat of the testing effect was not a significant problem concerning validity. While there was an administration of a pretest and posttest for each group, the pretest was in a similar or equivalent form to the end-of-course test students take. The administration of the posttest occurred approximately four months after the pretest and was not threatening to the test's validity. Additionally, instrumentation was not a threat to internal validity because the format and structure of the pre-and posttest remained the same. The two test administrations involved tests that were equivalent, and the tests were both multiple-choice format, approximately the same difficulty level, and involved the students marking answers on an identical answer bubble form.

#### *External validity.*

While controlling for threats to internal validity, there were also threats to external validity that needed attention, such as the setting-treatment interaction and experimenter effects. The setting-treatment interaction received attention in this proposed study because the groups involved in the experiment were located at different schools. In this case, as Ary, et al. (2006) would contend, "If results are found to be similar in both settings with their different populations there is reasonable confidence that generalizations are valid" (p. 319). While the researcher made every attempt to choose schools that were as nearly alike as possible in terms of student body numbers, demographics, and socioeconomic level of the schools' populations, they were still different environments, and any interaction of the treatment with the experimental settings may have limited generalization of the results (Ary, 2006). The populations of the two schools were still somewhat different, as were the onsite facilities. To control for this particular threat to external validity, the research study occurred in two settings, East Side and West Side High Schools. Replicating the research study at different schools helped to control for external validity.

Additionally, there was the potential problem of experimenter effects. One of the control problems developing from an interaction of treatment with experimenter effects, according to Ary, et al. (2006), was the possibility of the experimenter intentionally or unintentionally giving cues which could have influenced the participants. Ary, et al. (2006) asserted that "sometimes the presence of observers during an experiment may so alter the normal responses of the participating subjects that the findings from one group may not be valid for another group or for the broader population, and it would be hazardous to generalize the findings" (p. 318). To control for the experimenter effect, the researcher had a meeting with the cooperating teachers before the beginning of the treatment. The goal was to provide the participating teachers with instructions and clearly stated operational parameters for all variables related to the experiment. While the researcher had to contend with the experiment's validity, there was also the question concerning the actual end-of-course test's validity factors.

#### Instrument Reliability and Validity

The North Carolina Department of Public Instruction (NCDPI) has strict guidelines for the multiple-choice development process to provide for reliability and validity. According to the NCDPI Accountability Services Division (2008b), for instance, the test development process consists of six phases and takes four years to complete. Phase 1 consists of the test specifications blueprint (2008b) that includes outlining the purpose of the test and the test specifications for the grade levels and content areas to be assessed. Phase 2 (2008b) is the item development and review section that seeks to insure that the questions focus on the curriculum objectives. Classroom teachers have reviewed the items for clarity, correctness, potential bias, and curricular appropriateness (2008b). Phase 3 (2008b) is the field test development and administration section. In this phase, "the use of classroom teachers from across the state as item writers and developers ensures that instructional validity is maintained through the input of professional educators with current classroom experiences" (NCDPI, 2008b, p. 2). The intent is to verify that there is a valid representation by objectives and construction validity. The field test is assembled, reviewed, and administered to a stratified random sample of students (2008b). State testing officials want to make certain that the administration of the field test forms follow the routine that will mimic the statewide administration of an end-of-course test.

Continuing, Phase 4 is the phase in which the pilot test is assembled in equivalent and parallel forms to help ensure reliability equivalency (NCDPI, 2008b). The pilot test "is formed from disassembled field test forms and is meant to mimic an administration of the operational test in every way" (NCDPI, 2008b, p. 4). In Phase 5, operational test development and administration occurs, where the "test is given statewide, following all policies of the State Board of Education, including the North Carolina Testing Code of Ethics, while standardized test administration procedures must be followed to ensure the validity and reliability of test results" (NCDPI, 2008b, p. 5). Finally, Phase 6 concludes the multiple-choice test development process with reporting the test results. Finally, internal consistency for the North Carolina end-of-course test in United States History should exhibit a reliability coefficient of at least 0.85 if any decisions are made based on test data (NCDPI, 1996, p. 44). The item-level values of coefficient  $\alpha$  for the pretest were 0.85, utilizing the Spearman-Brown Prophecy Formula, while the NC end-of-course test were 0.92 (NCDPI, 1996, p. 44). The standard error of measurement for the range of scores on the North Carolina end-of-course test in United States History is for students to score within two standard deviations of the mean (95%), with the standard errors typically 2-3 points (NCDPI, 1996, p. 45).

### Procedures Used

While emphasis continues to grow on the importance of standardized testing, this research study attempted to determine if students who grade and then correct their test papers learned more and scored higher than students who experienced teacher-only grading with no corrective measures, as evidenced by scores on the end-of-course test in United States History. The researcher initially secured all the necessary approvals to complete the research study from the cooperating teachers and the respective school administrators. After securing preliminary access from school personnel, permission was granted from the Liberty University IRB Review Board to undertake the research study. At that point, concerns turned to the research participants.

The four classes chosen represented a cluster sample of the typical history classroom in these particular high schools and were a heterogeneous section of the 11<sup>th</sup> grade. In order to ensure the treatment occurred as intended, the researcher worked with the two teachers in the planning process, so the participating teachers felt a sense of ownership in the research procedure. These two teachers have been working together as

a team with the other social studies teachers in this county. They were very positive at the prospect of participating. The cooperating teachers both expressed an interest in this study and of the potential for using the outcomes to improve their teaching program. The assisting instructors were both motivated to follow the methods prescribed in this proposed study.

During the semester, the two instructors employed similar teaching techniques, such as the direct instruction method of curriculum delivery and concurrent testing dates. At the start of the semester, all teachers of United States History were instructed to administer the countywide pretest, and then the research study treatment was put in place. More specifically, for Classes A and C, the teacher administered the weekly, multiplechoice tests to his or her 11<sup>th</sup> grade United States History classes. The tests were 50 questions, which matched the corresponding goals in the North Carolina Standard Course of Study and originated from released end-of-course tests. The questions and format of every test were indicative of the North Carolina end-of-course test all students must take at the end of 11th grade United States History. After every test administration for the control group, Classes A and C, the teacher collected the tests, graded them using a previously coded key, recorded the grades in the grade book, and then returned the test papers to the students. When each student had his or her own paper, the teacher read every question and immediately said aloud the correct answer choice. After giving the correct response, the teacher commented on it briefly, such as any surrounding detail that might have made the question challenging, and then answered any student question that arose. The instructor continued in this format, eventually covering the entire test. The teacher answered any questions the students might have had; then class resumed on the

next topic in the standard course of study.

These two classes (A and C) constituted the control group, as the teacher was the only one assessing the test papers, and she or he passed the papers back to the students for review and discussion of the correct answers. Students were passive observers of the grading process, and no further treatment was involved in the testing, which has been the traditional way most classroom teachers used the assessment process. Both East Side and West Side High Schools, though, offered students after-hours assistance for any student desiring more instructional time with the teacher. The high schools offered a reteach-retest program, and students often took advantage of this service; thus, every student had the opportunity to learn more and improve through one-on-one time with the teacher. This insured that students in the control group had an equivalent opportunity to excel with the United States History curriculum.

The treatment group, Classes B and D, had the very same curriculum instruction and multiple-choice tests used in the control group. After each weekly assessment, students turned in their test papers, and when every participant had completed the tests, the teacher handed the papers back to the students for grading. In these classes, the students were responsible for grading their own papers, using a pre-coded answer key, and making note of the correct answer. As in the control group, there were time allowances to answer questions and explain the nuances of answer choices. When the self-grading concluded, the teacher collected all test papers and quickly looked over each test while he or she recorded the test grades to ensure students had not cheated by failing to mark an incorrect answer.

The teachers returned the recorded papers, where the students had noted the

correct answer choices beside any questions missed. Then, for each question answered incorrectly, the student wrote a correction based on a pre-established format that required the learner to state the supporting details of the proper answer. For the remainder of the class period, the teacher walked around the room and monitored student progress while the class worked on the treatment, the test corrections. While the corrective measures could have been completed at home, for the first several attempts corrections were done during class time. This gave the students an opportunity to become comfortable with the correction format and allowed the teacher a chance to assess student progress. While initially this added classroom assignment took time from another activity, such as beginning the next goal of study, it may have reduced the amount of time needed for review before starting a new section. The time spent in class for corrective measures more than made up for the time spent in review and is time saved instead of wasted. In this way, classroom review was tailored for individual student needs.

Each correction assignment had its corresponding due date, such as two days from when the test was originally taken, before the modification was returned to the teacher. When he or she received the students' corrected papers, the teacher assessed the work to ensure the proper correction format was used. Then the students' final grades on the test were changed to reflect the completed corrections, meaning one-half of the initial point deduction was returned for every corrected question. The reasons for this grade adjustment were two-fold: the promise of a better test grade is a short-term reward for students to put the needed effort into doing the corrections the right way, and secondly, it provides an additional incentive to lessen cheating. For instance, if the students know they will have the opportunity to better their test grades, they could be more likely to grade the papers more accurately and may be less likely to attempt cheating during the administration of the test. In their study on student self-grading, Sadler and Good (2006) found cheating to be one of the biggest challenges to data collection. This researcher hoped that by offering students an incentive to improve their test grades cheating would be minimized, as Sadler and Good had cautioned against (2006).

At the conclusion of the spring semester, the four classes in United States History took the North Carolina end-of-course test. Testing data was analyzed to determine if there were significant differences in grades between treatment and control groups. The researcher also watched for any specific significant learning outcomes with the two research groups, such as higher scores on the North Carolina end-of-course test.

### Data Analysis

After the pretest was administered, both the control and treatment groups covered the North Carolina Standard Course of Study. The control group's papers were teacheronly graded and the treatment group engaged in both self-graded and self-corrected measures. At the end of the 2009 spring semester both groups of students took the endof-course test in United States History and their posttest scores were recorded and readied for analysis.

### Data Reduction

The data for this study was analyzed using several strategies. First, the data was sorted into either the control or the treatment groups. Pretest scores from the control group (classes A and C) and the treatment group (classes B and D) were recorded after receiving the grade reports from the central office at the beginning of the '09 spring semester. The research study continued throughout the spring semester with the two groups covering the United States History Standard Course of Study goals. Individual student test grades from weekly goal tests were recorded in the teachers' grade books and kept separate from the pre- and posttest grades. At the end of the semester students from both treatment and control groups took the end-of-course posttest in United States History. The exams were graded at the county's central office, and the posttest scores were recorded in either the control or treatment grouping. There were finally two sets of collected grade data—pre- and posttest scores—for both the control group (classes A and C) and the treatment group (classes B and D).

### Statistical Reporting and Display

After reduction of the raw data, the researcher employed descriptive and inferential statistics, utilizing Statistical Product and Service Solutions (SPSS) statistical software. The pretest and posttest grades were gathered and analyzed after the spring semester's administration to determine if any gains occurred between groups, either in student learning or with end-of-course scores. The researcher determined if there were significant differences within any of the comparisons in the sample (George & Mallery, 2006). The posttest scores were the dependent variable, and the pretest scores helped control for differences. Data concerning the value of the independent samples *t*-test was analyzed and reported using the means, standard error of means, standard deviation,  $\rho$ -value (with a predetermined alpha level of 0.05), degrees of freedom, effect size, confidence intervals, and a two-tailed test for significance in computing variables from standardized test scores. Overall differences in means of student performance between the control group and the treatment group were compared (Sadler & Good, 2006).

## Analysis: Statistical Tests and Procedures

As previously stated, in this research study, the data from the control and treatment groups' pretest and posttest scores were analyzed using inferential statistics from the SPSS statistical software program. More specifically, the researcher utilized the independent samples *t*-test for the main test of significance. Pretests were administered to all students involved in the study, followed by either assignment to the control group where teacher-only grading occurred, or to the treatment group in which students graded and corrected their own papers. At the conclusion of the spring semester the summative posttest was given, and scores were reported back to the teachers. Both pre-and posttest scores from the control and treatment groups were compared using the Independent Samples *t*-test. The researcher was then able to determine from the *p*-value and the two-tailed test of significance whether there was enough statistical difference between the treatment and control groups to reject or fail to reject the null hypothesis.

#### Summary of the Methodology

This chapter has explained the methods used in this research study on learning and student self-grading. Students at two different high schools took a pretest at the beginning of the spring semester, followed by students in the treatment group selfassessing and correcting their test papers, through the period of one academic term. At the end of the semester, every student took the North Carolina end-of-course test in United States History. Standardized test scores were analyzed using the SPSS software program and the independent samples *t*-test. The following sections detail the completed research study and present the results obtained from this experiment on student learning and corrective measures.

#### CHAPTER FOUR: FINDINGS

This study examined in detail whether students who self-graded and then corrected their own test papers learned more and scored higher than students who experienced teacher-only grading. The results of the study are reported in this chapter. The Findings chapter is organized by first addressing the two specific research questions posed in Chapter 1 and then focusing on the null hypothesis. The final section of this chapter ends with a summation of the study's results and leads the reader into the remaining chapter with a conclusive summary and discussion.

### **Research Question 1**

The initial research question from Chapter 1 focused on whether there was a positive, negative, or equivalent relationship between students who graded and corrected their own test papers and a higher score on the North Carolina end-of-course test in United States History. At the conclusion of the research study, descriptive and inferential statistics with SPSS statistical software were used to analyze data concerning the initial research question. Eleventh grade United States History students (N = 128) participated in two high schools. At the beginning of the 2008-09 spring semester, students were assigned to either the control or treatment groups. All students participating then took the countywide pretest, followed with regular classroom instruction per their assigned grouping for the semester, and then concluded the study period with taking the end-of-course assessment.

Concerning the first question, the researcher compared the posttest end-of-course scores between the treatment and control groups. Students who participated in the

treatment group of B and D (n = 64) had an overall posttest performance mean of 81.53, (SD = 8.48) whereas the control group of A and C (n = 64) had an overall posttest score mean of 79.23 (SD = 9.67). The mean difference between the treatment group and control group was –2.29. The frequency distribution of the control and treatment group posttest scores were between 0.50 and 2.50, with a mean of 1.50 (SD = 0.50), (see Figures 1, and 2).

Figure 1

Frequency Distribution of Control Group Posttest Scores



# Figure 2

# Frequency Distribution of Treatment Group Posttest Scores



The combined results of the treatment and control groups' posttest scores, where N = 128, seemed to indicate a normal frequency distribution curve (M = 80.38, SD = 9.13) in a comparison between the two groups' final scores. In response to whether the treatment would yield a positive, negative, or equivalent influence on test scores, the outcome of the compared results suggested a statistically equivalent relationship between students who graded and corrected their own test papers and students who participated in the control group on the end-of-course test in United States History.

#### **Research Question 2**

The second research question from Chapter 1 sought to determine if a comparison between pretest and posttest grades showed students learned more with the intervention of student self-correcting and grading than with teacher-only grading. The North Carolina Department of Public Instruction devised the standards used in this study to determine the level of learning per student, based on end-of-course scores. According to the North Carolina Department of Public Instruction, desired learning outcomes are obtained when students exhibit grade-level proficiency by scoring a level III (grade of 83-92) or Level IV (grade of 93-100) on the US History end-of-course test (NCDPI, 2006-07).

When pretest scores were sorted for analysis, the researcher performed an independent samples *t*-test. The sample size for the pretest control group was n = 64, (M = 33.40, SD = 10.85), while the sample size for the pretest treatment group was n = 64, (M = 33.34, SD = 10.08). The mean difference between the control group scores and treatment group pretest was .062, with a 95 % confidence interval from -3.60 to 3.72. When configuring the independent samples *t*-test for pretest scores, the researcher assumed equal variances and a pre-established alpha value of  $p \le 0.05$ . An independent samples *t*-test comparison between the pretest scores of the control and treatment groups, N = 128, indicated (M = 33.38, SD = 10.43), t (126) = .034, p = .27. An independent samples *t*-test on posttest scores showed the sample size for the posttest treatment group was n = 64, (M = 79.23, SD = 9.67), while the sample size for the posttest treatment group was n = 64, (M = 81.53, SD = 8.43). The mean difference between the control group was n = 64, (M = 81.53, SD = 8.43).
interval from -5.48 to .88. When configuring the independent samples *t*-test for posttest scores, the researcher again assumed equal variances and had a pre-established alpha value of  $p \le 0.05$ . An independent samples *t*-test comparison between the posttest scores of the control and treatment groups, N = 128, indicated (M = 80.38, SD = 9.13), t (126) = -1.42, p = .677 (see Tables 2 and 3).

# Table 2

### Pretest Group Statistics for each Participating Group

Group Name	n	М	SD	t	<i>p</i> <
Control	64	33.40	10.85		
Treatment	64	33.34	10.08	0.032	1.97

*Note.* \*p < .05

# Table 3

# Posttest Means, Standard Deviations, and t-tests

Group	n	М	SD	t	p<

Control	64	79.23	9.67		
				1.37	0.174
Treatment	64	81.53	8.48		

Concerning the second research question, which asked if a comparison between pretest and posttest grades showed students learned more with the intervention of selfgrading and corrective measures than with teacher-only grading, the researcher used the previously mentioned North Carolina Department of Public Instruction's proficiency rating score. Of the 64 students who were in the treatment group, zero students scored in the Level III or Level IV range on the pretest, but 25 scored at Level III proficiency and 4 scored in the Level IV range on the posttest. Of the 64 students who composed the control group, zero students scored in the proficient range of Level III or Level IV on the pretest, while 25 scored Level III and two obtained Level IV on the final posttest assessment (see Table 5). The independent *t*-test figures between pretest and posttest scores of the two groups seemed to indicate no statistical difference in how much students learned whether they engaged in teacher-only grading or self-grading and corrective measures.

Table 4

Proficiency Levels of Pre-and Posttest Scores



Control	64	0	0	0
Treatment	64	0	0	0
Posttest Scores				
Control	4	33	25	2
Treatment	3	32	25	4

# Null Hypothesis

The final query from Chapter 1, the null hypothesis,  $H_0$ , stated that students who self-graded and corrected their test papers would not achieve significantly higher scores than students who did not grade and correct their own test papers on the North Carolina end-of-course test in United States History. After undergoing the pretest experience at the beginning of the spring semester, students in the control group participated in teacheronly grading, while students in the treatment group graded their own papers followed by the use of corrective measures. At the end of the semester, all participating students involved in the study took the North Carolina United States History end-of course test.

When the end-of-course tests were graded, scores were sent back to the schools. After performing an independent samples *t*-test with SPSS statistical software, and the researcher having assumed equal variances for posttest scores, results seemed to indicate no significant statistical difference between students in the control group (M = 79.23, SD = 9.67) versus the treatment group, (M = 81.53, SD = 8.48), t (126) = -1.428, p = .677. Assuming a pre-established alpha value of  $p \le 0.05$ , the two-tailed test of significance for the posttest group was .156, with a *p*-value of .677. The mean difference between groups was -2.29, with a 95 % confidence interval from -5.48 to .88 (see Tables 2, 3, and 4). These results seem to imply that the researcher failed to observe a statistically significant difference in posttest scores between the control and treatment groups.

### Summary of the Results

The results presented above included consideration from the two research questions and the null hypothesis. After performing an independent samples *t*-test, the researcher was able to determine the levels of significance for the pretest and posttest results from both the control and treatment groups. Statistical results for both research questions and the null hypothesis tended to support self-correcting and grading as appropriate strategies to encourage student learning and achievement; however, both research questions and the null hypothesis failed to support that strategy over teacheronly grading. Statistical results seemed to indicate that with all three-research queries there was not enough significance between results to determine self-grading and corrective measures were better at helping students learn more or score higher on the North Carolina United States History end-of-course test than teacher-only grading. A more detailed summary and a discussion of the findings are presented in the next chapter.

### CHAPTER FIVE: SUMMARY AND DISCUSSION

In the concluding section of this dissertation, this final chapter restates the research problem and reviews the key methods used in the study. The main sections of this chapter summarize the results and discuss their implications. The final section offers suggestions for further research and concludes with a look to the future about the possible benefits of this study.

### Statement of the Problem

During the course of this study, there were two research questions and the null hypothesis under investigation. First, the researcher wanted to know if there was a positive, negative, or equivalent relationship between students who graded and corrected their own test papers and a higher score on the North Carolina end-of-course test in United States History. Secondly, if a comparison between pretest and posttest grades showed students learned more with the intervention of student self-correcting and grading than with teacher-only grading. Finally, the researcher asked if there was enough statistical significance to reject or adopt the null hypothesis,  $H_0$ , which stated that students who self-graded and corrected their test papers would not achieve significantly higher scores than students who did not grade and correct their own test papers on the end-of-course test in United States History.

### Review of the Methodology

The research study, a quasi-experimental static group comparison design, covered a period of one academic semester of 90 days, from late January 2009 until the end of May 2009. In this study, two teachers from two different high schools agreed to

102

participate with their 11<sup>th</sup> grade United States History classes. The participating students (N = 128) were assigned to either the control group (Class A and C, n = 64) or the treatment group (Class B and D, n = 64) by the flip of a coin. At the beginning of the semester, all students participating in 11<sup>th</sup> grade United States History took the countywide United States History pretest. After taking the pretest, students proceeded with their normal classroom routine of covering the 12 goals required of the North Carolina United States History course. The cooperating teachers gave their students a 50-question multiple-choice test after the completion of each achievement goal.

After every test administration for the control group (Classes A and C), the teacher collected the tests, graded them using a previously coded key, recorded the grades in the grade book, and then returned the test papers to the students. After returning the tests to the students and discussing the correct responses, the teacher then resumed the next topic of study from the standard course of study. Conversely, after each weekly test administration for the treatment group (Classes B and D), the students were responsible for grading their own papers, using the pre-coded answer key, and making note of the correct answer. As with the control group, there were time allowances to answer questions and explain the answer choices. After grading their papers, the students wrote a correction for the questions they missed using the pre-established corrections format. Students in both the control and treatment groups continued with either teacher-only or self-grading and correcting for the duration of one academic semester.

At the end of the 2009 spring semester, students in both the control and treatment groups took the posttest, the state end-of-course assessment for North Carolina United States History. When both sets of tests were scored by the county's central office, grades were then grouped by either control or treatment group and analyzed using descriptive and inferential statistics, more specifically the independent samples *t*-test. After the *t*-test results were gathered, the researcher summarized the results of the findings and found statistical support to imply a need to adopt the null hypothesis.

#### Summary of the Results

Throughout the period of this study, the focus had been on whether student learning and achievement could be affected if a student graded and then corrected his or her own test papers. Research question one asked if there was a positive, negative, or equivalent relationship between students who graded and corrected their own papers and a higher score on the North Carolina United States History end-of-course test. After taking the initial pretest, members of the control and treatment groups scored statistically equivalent on the test, with both groups having all members score in the lowest level of proficiency, Level I. No student in either the control or the treatment group scored in the Level II, III, or IV range on the pretest.

At the conclusion of the spring semester, students took the end-of-course test and in the control group, there were 4 Level Is, 33 Level IIs, 25 Level IIIs, and 2 Level IVs. When the treatment group took the end-of-course test, their scores included 3 Level Is, 32 Level IIs, 25 Level IIIs, and 4 Level IVs. The class mean of the posttest control group was 79.23 (SD = 9.67) with a posttest treatment group mean of 81.53 (SD = 8.48). These figures seemed to indicate that while there was a mean difference in posttest scores of the control and treatment groups of -2.29 in favor of the treatment strategy, the posttest *p*values (p = .677),—where a pre-established alpha value was  $p \le 0.05$ —seemed to imply no statistical difference between the strategy of student grading and correcting and teacher-only grading.

Secondly, research question two sought to determine if a comparison between pretest and posttest grades showed students learned more with the intervention of student self-correcting and grading than with teacher-only grading. To determine the degree of learning that occurred, the researcher used the same proficiency level scores employed by the North Carolina Department of Public Instruction to define student achievement (NCDPI, 2006-07). For instance, as previously mentioned, a student has scored at grade level, which is considered proficient, on the end-of-course assessment if he or she obtained a Level III (percentile grade of 83-93) or Level IV (percentile grade of 94-100) (NCDPI, 2006-07).

At the end of the experimental period of 90 days, 25 students in the control group had scored at the Level III range, while 2 students scored a Level IV. In the treatment group, 25 students scored Level III and 4 students obtained a Level IV proficiency rating. The class mean for the posttest control group was 79.23 (SD = 9.67) and the class mean for the posttest treatment group was 81.53 (SD = 8.48) with a difference of 2.3. While two more students in the treatment group scored Level IVs than in the control group, the number of Level IIIs was equivalent for both groups. Students did learn the curriculum in both the control and treatment groups, but with a pre-established alpha value of  $p \le$ 0.05, the *p*-value of .677 seemed to indicate no statistical difference in strategies between the two groups when the independent samples *t*-test was performed. The researcher found it likely that students learned a statistically similar amount whether they engaged in teacher-only grading or self-grading with corrective measures.

Finally the null hypothesis,  $H_{0}$ , stated that students who graded and corrected their

test papers would not achieve significantly higher scores than students who did not grade and correct their own test papers on the end-of-course test in United States History. After taking the pretest at the beginning of the spring semester, both the control and treatment groups proceeded in the same fashion except with the classroom grading procedures. The variation occurred between the control group using teacher-only grading, and the treatment group who employed self-grading and test corrections. At the end of the semester, both the control and treatment groups took the end-of-course assessment. When the scores were returned the posttest mean for the control group was 79.23 (SD = 9.67), while the class mean for the treatment group was 81.53 (SD = 8.43). When the posttest grade results were calculated for the independent samples *t*-test, with a preestablished *p*-value of p = <.05, the *p*-value was p = .677, while the researcher assumed equal variances. These results seem to imply that the research experiment failed to show a statistical difference in posttest scores between the control and treatment groups, and that the researcher should adopt the null hypothesis.

### Discussion of the Results

Research question one asked if there was a positive, negative, or equivalent relationship between students who graded and corrected their own papers and a higher score on the end-of-course test. While two more students in the treatment scored at the proficiency Level IV than with students in the control group, there was no statistical difference between the two strategies to indicate anything other than an equivalent relationship. The same number of students (25) in both the treatment and control groups scored at Level III proficiency, and there was only one student difference in both groups for Level I and II achievement.

After analyzing descriptive statistics for research question one, the researcher was able to determine that there was an equivalent relationship between students who graded and corrected their own test papers and their scores on the end-of-course in United States History. The frequency distribution for the treatment group fell within the normal curve range when compared to the control group (SD = 0.50). The researcher concluded that when students graded and corrected their own test papers over the course of an academic semester, their scores were not statistically different from students who had teacher-only grading and no corrective measures. In this instant, the strategy of self-grading and correcting proved no more effective in raising final end-of-course grades than the use of no strategy. While Sadler and Good (2006) had expected the positive results from their science classroom performance to show likewise results in other subjects, with this dissertation study scores were not affected enough to warrant further use of this particular tactic to improve standardized test scores over any other strategy.

Because the veteran teacher who had given inspiration for this study had truly believed in the effectiveness of test corrections, and the grades were slightly higher in the treatment group, the researcher was surprised to find that there was not enough statistical difference between the two methods to determine self-grading and correcting was a positive strategy for improving end-of-course scores. There is much pressure on school districts and individual teachers to improve test scores, and almost all in-service activities are dedicated to end-of-course test strategies in the county where this research study was done. The researcher thought that perhaps this would be a more effective strategy for this purpose, however the results indicated otherwise.

Even likening back to the theories of Bloom (1968), and more recent works such

as Guskey's (2007), it would seem that when students are more involved in classroom activities and assume a participating role, they learn more. Students working with the teacher in the role of a facilitator instead of only an instructor would seem to encourage more of a working relationship, and perhaps reduce the sense of a learned helplessness that some students feel if their progress feels doomed to failure.

Consistent with the results from question one, was when research question two asked whether a comparison between pretest and posttest grades showed students learned more with the intervention of student self-correcting and grading than with teacher-only grading. Initially it was determined that student learning would be assessed by the same standard in which the North Carolina Department of Public Instruction assesses it; by the number of students achieving the proficiency levels of Level III (83-93) and Level IV (94-100) on the end-of-course assessment. After the administration of the end-of-course test in United States History, students in the control group of A and C included 25 Level IIIs and 2 Level IVs. Students in the treatment group of B and D included 25 Level IIIs and had 4 Level IVs.

The two-tailed test of significance (.156) and the *p*-value (.677) both indicated there was not enough statistical significance between the control and treatment groups to determine students had learned more by grading and correcting their papers. While there were admittedly two more Level IVs in the treatment group than in the control group, there was not enough difference in the mean score of 2.3 points to make a difference in the overall *t*-test statistic. The researcher concluded from the data that if students in the classroom were to use the strategy of self-grading and correcting they would learn as much as in the traditional method of teacher-only grading, just perhaps not more. The

outcome results from this research study indicated that students did not learn statistically more grading and correcting their own test papers than with teacher-only grading.

As with the results from research question one, the researcher was surprised at the outcome. While the forefront of education news today is the importance of consistently bettering standardized test scores, the most important aspect of education is helping children become better citizens and learning all they can to be happy, productive members of society. In other words, the very reason for all the work done in schools is to help children learn. The researcher, especially after having read the theoretical and empirical studies on active involvement and increased participation in the classroom, believed that if she could get students to pay close attention to their own papers during the grading process and then do immediate follow-up to reinforce the correct responses, it would dramatically improve the time between testing and feedback. The researcher also believed it would increase cooperation and self-efficacy between all involved, and that students would not just sit in the class waiting for the bell to ring, but feel a sense of excitement at playing a part in the inner workings of the grading and assessment process for this particular class. While grades were only slightly higher from the treatment group, they were not high enough to suggest that students had actually learned more with the self-grading and corrective measures.

Finally, there is the question of whether to reject or adopt the null hypothesis. The null stated that students who self-graded and corrected their test papers would not achieve significantly higher scores than students who did not grade and correct their own test papers. After students were assigned to either the control or the treatment group, they proceeded through the semester with the same instructional technique, only varying when the weekly assessment time approached. Students in the control group experienced teacher-only grading after each test and students in the treatment group first graded and then corrected their own test papers each week. At the end of the spring semester, the end-of-course exam was administered to both groups of students, and when the scores were returned the researcher was able to gather and analyze the testing data.

After performing the independent samples *t*-test, the test failed to reveal a statistically reliable difference between the posttest control group (M = 79.23, SD = 9.67) and the treatment posttest group (M = 81.44, SD = 8.48), t (126) = -1.42, p = .677. After consideration for a Type II error and assuming equal variances, the researcher was able to fail to reject the null hypothesis for this research study on student learning and self-correcting. The researcher concluded that students who participated in the treatment group did not learn statistically more or perform better on the end-of-course assessment than students who were assigned to the control group. While the researcher thought students might learn more if they were involved with a correction process designed to help them learn from their mistakes, adopting the  $H_0$  in this instance was not concurrent with previous research (Sadler and Good, 2006).

In all three instances as the researcher figured and reconfigured the results, she was surprised that for both research questions and the null that there was not enough statistical difference to support student self-grading and correcting over teacher-only grading. While the main premise in Sadler and Good's (2006) study was that when students used self-grading they learned more, the main objective in the present research study was to take that idea a step further and make the self-corrective measures the focal point. The researcher was hoping to show that when students took an active role in

correcting their papers, they would want to become more involved with their own learning. The learners would perhaps begin to research topics and go a step beyond what might ordinarily be expected of them. The researcher thought that by using a format that would encourage students to think about why they missed a particular question, and then how that question fits in with the greater topic and concepts of the standard course of study, the student would begin to take a much more active part in the classroom environment. Students might have more of a reason to talk about history, and for more than a few seconds it could truly come alive for them. The results were interesting and somewhat disappointing because of the amount of research, both classical and modern in nature that seemed to support self-grading and, more specifically, corrective measures as a way of improving student learning and standardized test scores.

#### Relationship to Current Literature

In the first part of the review of literature on student learning and self-correcting, the researcher discussed some of the earliest theoretical perspectives of learning and theories it was built upon. For instance, early learning theorists such as Piaget (1952) and Vygotsky (1978) both concurred that learning should not occur in a vacuum, but was by its very nature social. Vygotsky's (1978) constructivism emphasized a top-down processing classroom setting in which the teacher began with presenting a problem, and then students worked to discover how to solve the problem. In this way, students would have taken an active approach with their own learning, might have developed problemsolving skills, and engaged in socio-cultural learning experiences (1978).

This idea of discovery learning prevalent in Vygotsky's work also mirrored the work of Kolb with his idea of experiential learning. Kolb (as cited in Johns, 2001) meant

111

for a student to be actively involved in his or her own learning and discovery process. For Kolb (1983), there was a two-dimension cycle of learning, which involved the gathering of facts and the processing and personalizing of information. Ideally, students would move from reflective observation to active experimentation, and ultimately arrive at new thoughts and conclusions. Bloom, like Kolb and Vygotsky, also theorized about the components of learning which included even more emphasis on students' active involvement in their own scholarship (Bloom, 1968).

While Bloom concurred with Kolb and Vygotsky about the need for active student participation, he went even further by saying that assessments could even be used as a tool (Bloom, 1968). Guskey (2007) contended that Bloom's mastery learning promoted assessment as a tool and that feedback, self-correction, and enrichment should become the cornerstone of modern mastery learning. Bloom argued that it was a mistake to assume all students should be taught the same way and given the same amount of instructional time to master the information. Naturally, there would be variations in students' learning, and classroom assessments should be used to diagnose individual learning inequalities and help design remediation schedules (2007). When a student's weak areas are identified they would be paired with a tutor, and the two of them would then work together to investigate the mistakes and discuss them for understanding. Bloom suggested that after taking a test, the teacher was responsible for giving students immediate feedback then students must have the opportunity to engage in an active, corrective activity for each formative assessment (1968).

Theoretically, using Bloom's suggestion for corrective measures would help the modern classroom to assist students in concept mastery and end-of-course test

112

improvement. Guskey (2007) agreed with Bloom's message on the value of classroom self-corrective measures to improve learning. He believed that feedback by itself was not enough to improve student learning, but should be qualitatively different from the instruction which the learners had initially received (2007). While Bloom, and later Guskey, supported the advantages of using self-corrective measures to improve student learning, Block, Efthim, and Burns (1989) took a more realistic approach when they agreed with the benefits of student test corrections, but also realized the time constraints and limitations it would impose on a teacher's daily planning routine. Idealistically, individualized instruction and self-corrective measures would be the most effective technique to enhance learning, concurred Whiting, Van Burgh, and Render (1995), but finding classroom time to always offer students individualized instruction during the corrective process would be difficult.

In bridging the theories of learning from earlier times, when Dewey was beginning to define the rigid organization of early classrooms, to the actual theories of learning from Vygotsky, Kolb, and Bloom, the intended outcome of classroom instruction has remained virtually unchanged. Schools and teachers today are still trying to find newer and better ways of teaching, more effective strategies to improve student learning, and valid ways of assessing student learning and progress. For instance, current trends in education brought about by the No Child Left Behind legislation require school districts to be able to demonstrate proof positive through test results that students are learning more and that the school is improving its ministrations toward all cultural, ethnic, achievement, and socio-economic groups through the use of reporting AYP results (Stotsky, 2005). Schools need to demonstrate continued academic success and progress toward standards-driven curriculums, and improved learning and test scores are what matters in today's schools (2005).

While considerable research has been done on different classroom learning strategies, such as cooperative learning and alternative assessment, there may be considerable value in looking at other tactics to affect learning. While Thompson and Newsome (2002) focused their studies on whether multiple-choice tests could help facilitate the use of higher-ordered thinking skills in the classroom, Walberg (1986) had been researching the value of peer- or self-assessment and mistake correction after receiving feedback from the teacher. Walberg argued that students using feedback and corrective measures learned more, while Dweck (2000) continued this line of study and found that students who were the highest- and lowest-achieving students benefited the most from self-grading.

Teachers have traditionally viewed self-grading in a less positive light because of the possibility of cheating (Sadler & Good, 2006; Edwards, 2007; Strong, Davis, & Hawks, 2004). In recent years, though, especially since No Child Left Behind and the pressures on educators to improve learning and test scores, some researchers and teachers alike are beginning to examine its possible benefits. Two meta-analysis by Falchikov and Goldfinch (2000), and Falchikov and Boud (1989) concurred in their findings that there was no overall consistent tendency for students to under- or overestimate their performance. Strong, et al. (2004) also conducted a study involving student self-grading and found that while students reported they had learned more and had enjoyed the classroom experience, the researchers' conclusions disagreed with Falchikov, Goldfinch and Falchikov, Boud's stance that students did not have a tendency to inflate their grades. Strong, et al concluded that even though grade inflation was a problem, if self-assessment were used in a smaller classroom setting with students properly trained with the instructor's desired grading standards, self-grading would be a positive learning experience for students.

While Stefani (1994) also conducted a research study which concluded with positive results supporting self-grading and correcting, perhaps the most influential research study supporting this dissertation was that of Sadler and Good (2006). Using Stefani's study as a starting point, Sadler and Good took the idea of self-grading further by theorizing self-grading increased student performance and learning, and also helped teachers prepare for standardized tests. While most studies to this point had focused on the college student (Falchikov & Goldfinch, 2000; Falchikov & Boud, 1989), Sadler and Good realized that there might be value in working with middle school students.

Sadler and Good's study took place in four middle school science classes (2006). During the course of the experiment students self-graded their papers with the resulting scores correlating close enough with the teachers' marks to be a reliable substitution. Their conclusion was that self-grading appeared to further student understanding of the subject matter taught (2006). The research study conducted by Sadler and Good on selfgrading demonstrated that this particular technique could improve learning and understanding (2006).

While there have been studies conducted on the effects of student self-grading (Stefani, 1994; Sadler & Good, 2006; Falchikov & Boud, 1989), there have been few studies on the effects of student corrective work in the learning process, and fewer still at the high school level. Mathan and Koedinger (2005) conducted a study on delayed

versus immediate feedback and reported that it was important in the learning process that corrective measures begin as soon as possible after the actual test. They also surmised that corrective measures coupled with immediate feedback were more effective than grade feedback alone (2005).

There are more recent studies where self-grading is a positive way to enhance learning (Andrade & Valtcheva, 2009; Ross & Starling, 2008). Andrade and Valtcheva, for instance, studied self-grading but determined that it was better used on homework and drafts than assignments or tests to be graded. Their main contention with not using it for graded assessments was that its purpose would be better served in the area of guided or independent practice, where it would identify areas of strengths or weaknesses (2009). Ross and Starling (2008) found student self-assessment to be beneficial throughout the homework and grading process but also found a connection with self-esteem and selfefficacy. In their study, Ross and Starling determined that when self-assessment was used in the classroom setting there was a 25% increase in overall student achievement. Additionally, they reported that students also commented on an increase in self-efficacy, especially with girls. The researchers hypothesized that perhaps it was a sense of empowerment or a feeling of having more control over their environment and learning that led girls to an overall better feeling toward their educational experience (2008).

Supporting Ross and Starling (2008) and their positive comments concerning selfgrading and achievement were McMillan and Hearn (2008). McMillan and Hearn's recent study took a much more future-oriented approach in their assessment on the value of student self-grading. The researchers approached their study from the standpoint that while there is value in self-assessment from a cooperative and interactive stance, they contend that in this new age of a standards-driven classroom, self-assessment empowers students to guide their own learning and internalize the criteria for judging success (2008). McMillan and Hearn also believed that self-grading encourages self-efficacy and motivation to improve academically.

While previous research studies, (Stefani, 1994; Falchikov & Boud, 1989) as well as more current work (Andrade & Valtcheva, 2009; Ross & Starling, 2008), have positive outcomes and praise for the value of student self-assessment, there have not been many studies employing both self-grading and corrective measures. Falchikov and Boud's (1989) meta-analysis contained the few instances where both were utilized, but they were conducted at the college level. This dissertation's author has not discovered a research study conducted at the high school level that contained both self-grading and corrective measures designed to assess the effect on learning and standardized test scores.

### **Implications for Practice**

The outcome of this study suggests to the author that while student self-grading and correcting has merit as a strategy in the classroom, it may not be the most effective tactic by itself for increasing learning or standardized test scores. Previous research from Sadler and Good's (2006) study on four middle school science classrooms suggested selfgrading had value over peer-grading and teacher-grading when it came to reinforcing concepts and encouraging learning. Taking a cue from Sadler and Good's research study, and combining it with the veteran teacher's idea on test corrections, gave this author cautious optimism that she had found a new strategy to use in her own classroom and even throughout her county.

Even though the researcher found she must adopt the null hypothesis, she also

recognized there were valuable implications that emerged from the research study. For instance, in their study, Sadler and Good (2006) stated that self-grading appeared to further student understanding of the subject matter being taught. They also determined that when students used self-grading that it involved students sharing some of the power traditionally held by the teacher. In their studies (Falchikov & Boud, 1989; Stefani, 1994; Freeman, 1995; Guskey, 2007; Locker & Cropley, 2004), the researchers indicated that self-grading had contributed to feelings of enhanced self-efficacy, a realistic perception of their own abilities, deeper feelings of motivation, increased desire to learn, more positive attitude toward class work, teacher expectations, the course of study, and even less testing anxiety. Stefani (1994) specifically supported self-grading and said that when students realize teachers are there to facilitate instruction rather than to dictate everything they realize that a child knows the teacher has some degree of control, and the learner feels more comfortable in both the classroom environment and in their relationship with the teacher.

The researcher wonders if the implications from this study would follow more along the course of Bloom's (1968) ideas of mastery learning. For instance, Bloom believed that all students could learn, but the approach and time needed might be different for each individual learner. Perhaps the value of student self-grading and correcting, then, would lie with weaker students or students who require more time and reinforcement with the material. Self-grading and correcting might even be a good strategy for high school age, at-risk students who have issues with authority or motivation. More high schools are now offering remediation in an online format and self-corrective measures might be a way to improve understanding and even self-esteem with students who might not be as inclined to tolerate a traditional classroom setting. It seems likely that while student self-grading and correcting did not prove to be any more statistically effective than teacher-only grading for the treatment group, it might be a viable alternative for the at-risk student or a learner on an Independent Education Plan.

### **Explanation of Unanticipated Findings**

There were unanticipated findings the researcher encountered during the course of this study. First, she was surprised that she had to fail to adopt the null. Overall, she had believed the strategy of student self-grading and correcting would prove to be of significant benefit. For example, several years prior to this research study one of East Side High School's veteran teachers spoke openly on the value of using what she termed test corrections. Her standardized test scores were always higher than the other teachers in the social studies department were and even when she retired, the veteran teacher commented that corrective measures were the key to her standardized testing success. In recent years as standardized testing and curriculums become even more important for determining learning and academic success, this researcher has attended many in-service training modules, but none of them ever spoke of using self-grading and corrective measures as a suitable strategy for improving learning and scores. The researcher has always wondered if corrective measures were, indeed, a better way and was excited over the opportunity to finally study the strategy in a controlled experiment other than occasional debates in departmental meetings.

During the semester, though, the researcher did encounter several other issues she had not anticipated that probably had a negative impact on the results. For instance, student scores on the end-of-course assessment might have actually been higher for those learners who graded and corrected their own test papers if, in fact, the students had actually participated in the treatment group the way the researcher had originally envisioned it. While the researcher had anticipated problems with the cooperating teachers not following the experiment's guidelines properly, it seems that the surprise came from the students.

While students readily looked forward to grading their own papers, they would not complete the correction process. On the first test, most students excitedly participated in both the grading and correcting process, but with proceeding assessments the excitement seemingly wore thin, and less than <sup>1</sup>/<sub>4</sub> of the students in both classes B and D of the treatment group would complete the test correction process each time. When the participating teachers gave the students some time in class to complete the work, the students cooperated at first, but after about three weeks, the teachers had a hard time getting any of the students to work on the corrective process. At the end of the spring semester very few of the students had completed corrections for all tests, and fewer still took the time to do them right.

When the cooperating teachers checked them for correctness, they found that after the first two or three sets of corrections students were just rewording the question and trying to submit it for credit that way, or were copying each others' work. It was frustrating and both teachers at the different high schools had the same problem with the participating students. When the teachers asked the students about it, the learners replied that they found that there were too many corrections to do each time and they were tired of them and got very bored. The students also remarked that they would rather do some sort of extra credit assignment outside of school than do anymore test corrections as they took too long to do and were boring. The researcher was disappointed that the students were no more interested in their general term period grades and the overall learning that was supposed to be tested.

While the researcher had tried to formulate the corrections criteria in a way that would reinforce the standard course of study, perhaps the required format itself was too involved. The researcher did not take into account how students might feel if they routinely missed an excessive number of questions and the degree of time a teenager might be willing to spend on each question if facing several hours' worth of corrective measures each week.

Another unanticipated finding was of the researcher's over-confidence that all of the students would be excited and willing to do the extra work to learn more and increase their grades. Several of the lower performing students were only interested in achieving the minimal grade required for passing the subject. The students stated that they were not interested in learning more or getting any end-of-course grade higher than they needed to pass the class and graduate. The students would calculate their grades and know the least number of corrections they had to complete to pass the class, and then only turn in the minimal amount needed. While ideally the corrective measures were meant to reinforce the concepts for all students, there was no way to convince some of the learners of the need for intrinsic improvement. As long as they had a passing grade and figured they knew enough to pass the end-of-course test, some of the students would do nothing further to help themselves.

Finally, perhaps the most surprising part of the unanticipated findings was that in neither the two research questions nor the null hypothesis did there appear to be statistical

significance supporting student self-grading and corrective measures over teacher-only grading. Based on the early works of Vygotsky, Kolb, and Bloom the research tended to support the more actively involved student. Later research, such as from Sadler and Good, also supported self-grading as a positive and likely strategy to improve learning and test scores. The researcher was surprised that in all three instances there was no statistical difference in methods.

# Limitations

While there were careful considerations during this research study to conduct it in a setting as unbiased and valid as possible, inevitably there were limiting boundaries and ways in which the findings may lack generalizability. For instance, the nature and size of the sample could have easily changed the outcome of the study. This research study was conducted in two high schools in a central North Carolina rural community. The uniqueness of this particular setting could have produced very different results if the sample sizes were larger, more varied in students' cultural or economic backgrounds and the study encompassed other high schools, including those in an urban setting.

This research study was conducted during the spring semester of the 2009 school year. At these participating schools, final semester exams for the fall were administered near the end of January because of inclement weather, which is late for a semester to conclude. In both schools, students complained about the changing schedule, continued weather problems, and seemingly grew tired of participating with corrective measures shortly after they started. Perhaps beginning this research study at the start of a new school year when students were excited and fresh from their vacation would have made a difference in the results.

This research study was also limited to the subject of 11<sup>th</sup> grade United States History. The study may have produced different results in a subject such as math or science where answers were much less subjective. The corrective measures might have been easier for students to do and quicker to complete where answers are more specific and less open to multiple interpretations, such as with a history class. Results could have been different for any of the research questions if different ages, grade levels, or subjects had also been chosen for participation.

# Suggestions for Further Research

While these study results provided unlikely statistical evidence to support student self-grading and corrective measures over teacher-only grading, additional research may be needed to generalize results to other populations or groups outside the area where the study was conducted. The population studied were 11<sup>th</sup> grade United States History students in two rural high schools. While every effort was made to include as random a sample as possible, results might have been different in a larger school or in a different geographic area. Sadler and Good's (2006) study was conducted on middle school students, but besides that particular study little research has been done on students and self-grading below the college level.

Further research might provide positive results in support of self-grading and corrective measures if additional studies were carried out at lower grade levels. Perhaps students would be more excited about making the connections between an incorrect response and the reasons why their assumptions were erroneous. High school students, as reported earlier, grew bored of the corrections process quickly. Additionally, it might prove beneficial to conduct this study in a different subject area. Any concept from an

historical perspective naturally leads itself to numerous interpretations, whereas a science or mathematics test would be much more objective in answer possibilities. Students may be more comfortable working with corrective measures when there is not as much room for argument, either from them or from the teacher, in terms of what would be an acceptable way to write a correction.

While this study focused mainly on multiple-choice tests, there might be different results if the weekly tests were in an alternative format, essay for instance. The method used in this study where students used a pre-coded answer key might have produced results more statistically significant if the students had coded the keys or if they had been taught to assess essay tests, which was done in other studies (Sadler and Good, 2006). A more complicated assessment with essay tests, and students who have been taught how to assess questions with a rubric and training, could provide a higher level of research in the high school classroom.

Finally, further research might prove a valid use for student self-grading and corrective measures if the strategy was tried on weaker students or students who have difficulty with reading comprehension. In North Carolina, every student who is on a standard diploma tract must score the minimal prerequisite Level III on the end-of-course test to pass the course and be eligible for graduation. Currently there is no provision made to accommodate exceptionalities if a student is going to graduate with a regular diploma. The result is many students who receive special education services have to be retested or have to repeat the whole course and go through the end-of-course testing procedure again. It might prove beneficial to use self-grading and corrective measures for at-risk and weaker students that traditionally feel intimidated or left behind in the

normal classroom. Perhaps the corrections format could be altered to better accommodate their needs, while the resulting benefits cited earlier, such as increased selfefficacy, motivation, and excitement to learn would return for those students.

While increased learning and higher standardized test scores continue to be the focal point of today's schools, educators must continue to remember that tests are merely tools. From the works of early learning theorists such as Dewey and Yvgotsky, to today's researchers like Sadler and Good, the one thing that has remained constant is the idea of increased student involvement and active participation in the learning process. While this research study on student self-grading and corrective measures found that this particular strategy was not statistically significant over teacher-only grading for promoting more learning and higher test scores, there are still sufficient studies from a wealth of sources supporting the benefits of students playing an active part in their learning. There is merit, both academically and emotionally with using classroom corrective measures and perhaps this study could be the starting point for a different direction utilizing the corrections process.

APPENDIXES

# APPENDIX A

Dear Beth,

We are pleased to inform you that your above study has been approved by the Liberty IRB. This approval is extended to you for one year. If data collection proceeds past one year, or if you make changes in the methodology as it pertains to human subjects, you must submit an appropriate update form to the IRB. Attached you'll find the forms for those cases.

Thank you for your cooperation with the IRB and we wish you well with your research project. We will be glad to send you a written memo from the Liberty IRB, as needed, upon request.

Sincerely,

Fernando Garzon, Psy.D. IRB Chair, Liberty University Center for Counseling and Family Studies Liberty University 1971 University Boulevard Lynchburg, VA 24502-2269 (434) 592-4054 Fax: (434) 522-0476

### APPENDIX B

U.S. History: Goal 1 Test Name 1. Which economic activity was of primary importance to the South during the early years of the United States? a) plantation agriculture c) commercial fishing b) mining and manufacturing d) trade and shipbuilding 2. During the Federalist period, political participation in most states was only open to which group? a) all adult males who had reached the age of twenty-one c)white male landowners b) white men and women over the age of twenty-one d) all native-born citizens 3. Which region of the United States had an economy that depended on shipbuilding, trade, and manufacturing? a. the Northwest c) the South b) New England d) the Southwest 4. The political status of women in the early years of the United States could be best summarized by which statement? a) few women held jobs outside the home b) women could not own property c) women formed societies for moral improvement d) women were not eligible to vote 5. What was the main reason that Native Americans were not part of the political process during the early years of the united States? a) they were not citizens b) they wanted to maintain their tribal customs c) language barriers d) the passage of the Bill of Rights 6. what act of Congress created the federal court system? a) the Land Ordinance of 1785 c) the Northwest Ordinance of 1787 b) the Judiciary Act of 1789 d) the passage of the Bill of Rights 7. Which constitutional change was made to guarantee the rights and liberties of American citizens? a.) the Bill of Rights c) the Virginia and Kentucky Resolves b) the two-party system d) the Twelfth Amendment

8. Which political view was shared by the Federalists?

a) nullification	c) loose interpretation of the Constitution
b) states' rights	d) judicial review by the Supreme Court

9. What was the main reason that Thomas Jefferson and the Democratic-Republicans opposed Hamilton's plan to create a national bank?

a) they believed that it was unconstitutional

b) they believed that the bank would unfairly aid the northern states

c) they opposed central banks in general

d) they thought that a national bank would only benefit big business

10. Which group supported the Federalists?a) western farmersc) Southern plantation owners

b) Northern businessmen d) landless wage earners

11. What was the result of the political disagreements between Alexander Hamilton and Thomas Jefferson?

a) the two-party system	c) the Bill of Rights
b) Northern businessmen	d) the Judiciary Act

12. Which actions were parts of Hamilton's economic plan?

a) foreign competition, taxes, and private loans

b) tax-revenue, selling public lands, and federal funding

c) free trade, free silver, and state banks

d) assumption of states' debts, tariffs, national bank

13. What measures did Hamilton propose to pay off the nation's debts?

a) a protective tariff and excise taxes on whiskey

b) a federal income tax

c) the free and unlimited coinage of silver currency

d) reducing federal funding

14. What part of Hamilton's economic plan had the goal of protecting American manufacturers from foreign competition?

a) the Coinage Act of 1792	c)tariffs
b) payment of states' debts	d) taxes

15. what was the significance of the Whiskey Rebellion?

a) it demonstrated that the new government under Washington could not prevent anti-tax rebellions

b) Washington quickly stopped the rebellion to demonstrate the effectiveness of the new government

c) Hamilton proposed a negotiated settlement to move the new capital to the Maryland/Virginia border

d) support for the Federalists increased amongst farmers in the west and south

16. Alexander Hamilton's vision for the future of the United States supported which economic goal? a) territorial expansion on the western frontier c) the growth of agriculture b) increased trade, business, and manufacturing d) establishing new colonies 17. Which of the following best describes how Democratic-Republicans, such as Thomas Jefferson, interpreted the Constitution? a) a strict interpretation of the Constitution b) a weak interpretation of the Constitution c) an activist interpretation of the Constitution d) a loose interpretation of the Constitution 18. Supporters of the Democratic-Republican party were mostly: a) farmers in the South and the West c) landless wage earners b) free blacks and Native Americans d) bankers and businessmen 19. What were the main features of the Alien and Sedition Acts? a) restriction of foreign immigration and penalties for criticism of government officials b) the expansion of a secret federal police force that would spy on American citizens c) the creation of a federal agency to regulate the content of books and newspapers d) to increase the number of legal immigrants allowed into the United States each year 20. Why were Republicans opposed to the Alien and Sedition Acts? a) the acts harmed the war effort c) the acts encouraged immigration b) the acts threatened civil liberties d) the acts helped big business 21. Which were written to protest the Alien and Sedition Acts? a) Letters to an American Farmer c) Virginia and Kentucky Resolves b) the South Carolina Exposition d) Washington's Farewell Address 22. What was promoted by "nullification"? a) the right of states to cancel federal laws that are unconstitutional b) the national government's powers over the states are supreme c) the Supreme Court may not strike down laws passed by Congress d) the President may be impeached for "high crimes and misdemeanors" 23. What document listed below upholds the principle of states' rights? a) the Federalist Papers c) the Albany Plan of Union b) the Olive Branch Petition d) the Virginia and Kentucky Resolves 24. What did the Supreme Court do in the case of *Marbury v. Madison* (1803)? a) affirm the constitutionality of nullification theory b) establish the principle of judicial review c) weaken the power of the federal judiciary d) struck down the constitutionality of the B.U.S.

25. What was established in the case of Marbury v. Madison?

a) presidential power to over-rule the federal courts

b) the Supreme Court's authority to determine what is constitutional

c) the right of the states to nullify federal laws

d) congressional power to oversee federal court decisions

26. which best explains the principle of judicial review?

a) the Supreme Court has the power to remove federal district judges from the bench

b) the Supreme Court serves as the primary federal trial court in the United States

c) the Supreme Court decides whether laws are constitutional or unconstitutional

d) the Supreme Court serves as the chief prosecutor in cases involving federal law

27. What was the primary goal of American foreign policy during the early years of the United States?

a) avoiding war and foreign alliances

c) increased naval powerd) opening new markets

b) establishing new colonies

28. What was the intent of the Embargo Act and why did it fail?

a) It was meant to help Great Britain in its war with France without requiring the US to commit military personnel to the conflict, but it failed because US troops inevitably became involved in the fighting.

b) It was a response to the insult of the "XYZ Affair" but it failed because the US did not have the military strength to back up its actions.

c) It was meant to avoid war by forbidding trade between the US and foreign nations, thus preventing the impressments of US sailors. It failed, however, because it had little effect on Great Britain and hurt the US economy by damaging business.

d) It was meant to keep the French and British from establishing future colonies in the Western Hemisphere, but it failed because Great Britain's navy was too powerful for the US to resist.

29. Which of the following statements might have been heard from a "War Hawk" prior to the War of 1812?

a) "We must go to war! Great Britain has violated our right to open trade on the seas by impressing our sailors into their own service. Even more, they encourage the Indians on the frontier to oppose and resist our westward expansion."

b) "We must not rush to war. Great Britain has a powerful navy and we are in no position to resist her.

c) "It is my contention that this convention here in Hartford send ambassadors to Washington to express our disappointment with the government's waging of this war."d) "It is not our desire to possess new lands or take any territory from Great Britain. We merely want to show our enemy that we will not be intimidated on the high seas."

30. What document replaced the Articles of Confederation and have greater powers to the new United States government?

a) the Virginia and Kentucky Resolutions

b) the Bill of Rights

c) the US Constitution

d) the Declaration of Independence

31. How did the US respond to the "XYZ Affair," and how did it affect relations between the US and France?

a) With outrage/ it ended relations between the two nations for a time.

b) With pleasure/ it began a new era in positive US-French relations.

c) Irritated/ the US went into debt paying money to France.

d) With disappointment/ it meant that France and Great Britain would be allies against the US.

32. "If, after careful consideration, the legislature of the great state of Virginia comes to the conclusion that the federal government has overstepped its bounds in passing this law- if we find it to be unconstitutional in its very nature- then we will, as a state, refuse to subject ourselves to it." The quote is advocating what?

a) Federalism

c) Democratic-Republicanism

b) the doctrine of nullification

d) impressments and nationalism

33. "....remember the ladies, and be more generous and favorable to them than your ancestors. Do not put such unlimited power into the hands of the Husbands. Remember, all men would be tyrants if they could. If particular care and attention is not paid to the ladies, we are determined to foment a rebellion, and will not hold ourselves hound by any laws in which we have no voice or representation."

The above quote comes from whom?

a) the wife of a southern plantation owner demanding the right to equal pay

b) Martha Washington demanding that women be allowed to run for public office

c) Abigail Adams demanding that women be granted suffrage

d) Dolly Madison demanding that women be granted the right to free speech

34. Which of the following was considered a "necessary evil" and was not abolished despite the fact that it seemed to contradict the principals of the Declaration of Independence?

a) the formation of political parties

c) the institution of slavery

b) attacks on Native Americans on the frontier

d) the Embargo Act

35. What was the primary significance of Pickney's Treaty?

a) it kept America out of was with Great Britain

b) it allowed western farmers to deposit their goods in New Orleans

c) it ended the U.S. alliance with France

d) it resulted in the purchase of the Louisiana Territory from Spain

36. Which party was elected to power in the election of 1800?

a) the Know-Nothings	c) the Federalists
b) the Democratic-Republicans	d) the Whigs

, I ,

37. Which president was elected in 1800?

a) John Adams	c) Thomas Jefferson
b) James Madison	d) George Washington

38. From what nation did the United States purchase the Louisiana Territory?

- a) France c) Spain
- b) England d) Russia

39. What was the significance of the Louisiana Purchase (1803)?

a) it doubled the territorial size of the United States

b) it re-established the 1778 alliance with France

c) it halted American expansion at the Mississippi

d) it renewed Indian attacks on the western frontier

40. Thomas Jefferson originally hesitated to purchase the Louisiana Territory mostly because?

a) he doubted that Congress would approve of the funds needed for the purchase

b) he knew that most of the territory was too dry and unsuitable for agriculture

c) he was opposed to white settlement on lands belonging to Native Americans

d) it conflicted with his political belief in a strict interpretation of the Constitution

41. What was the purpose of the Lewis and Clark Expedition?

a) to map and explore the Louisiana Territory

b) to negotiate with France over the purchase of Louisiana

c) to stop Native American attacks on the frontier

d) to conduct scientific experiments on agricultural techniques

42. Whose aid was essential to the success of the Lewis and Clark expedition?

a) Sacajawea	c) Tecumseh
b) Henry Clay	d) Jebediah Smith

43. What was the primary purpose of the Embargo Act of 1807?

a) to	encourage foreign trade	c) to strengthen the U.S. Navy

b) to avoid war with Britain d) to encourage domestic industry

44. Who encour	ged the declaration of war against Britain in 1812?
a) Federalists	c) northern businessmen
b) war hawks	d) Native American tribes

45. What was a primary cause of the Wa	ar of 1812?
a) French seizure of American ships	c) trade embargos
b) the decline of American trade	d) impressment

46. What were the goals of the "war hawks" in Congress?

a) stop French attacks on shipping, acquire Louisiana Territory

b) stop British impressments, end Indian attacks, acquire more territory

c) establish naval bases, acquire colonies in the Caribbean

d) free the slaves, increase protective tariffs, expand federal power
47. Which best explains Tecumseh's reason for siding with the British in the War of 1812?

a) the U.S. government had continually lied to the Indians

b) Tecumseh knew that the British were going to win the war

c) the U.S. had forced Indians to march on the Trail of Tears

d) Tecumseh hoped to stop American settlement in the west

48. Which event is an example of the conflict between states' rights and the federal government during the War of 1812?

a) the XYZ affair	c) the Treaty of San Lorenzo
b) Jay's Treaty	d) the Hartford Convention

49. Which American victory occurred after the Treaty of Ghent ended the War of 1812?a) the burning of Washington, DCb) the Battle of Horseshoe Bendc) the Battle of New Orleansd) the invasion of British Canada

50. How did the result of the War of 1812 impact the United States?

a) it resulted in the loss of American territory

b) it resulted in increased national pride and confidence

c) it resulted in a new alliance with France

d) it resulted in the gain of new U.S. territories in Canada

## APPENDIX C

## Test Correction Criteria

Each test correction must address the following requirements. Please write in complete sentences.

- 1. Who or what is the subject of this question?
- 2. What is this question specifically asking?
- 3. Where have the actions in this question taken place?
- 4. When is the event in question taking place?
- 5. What was your initial belief the question was asking and how was it different from what the question was really asking?

## References

- Alber, S.R., & Walshe, S.E. (2004). When to self-correct spelling words: A systematic replication. *Journal of Behavioral Education*, *13*(1), 57-66.
- Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. Assessment & Evaluation in Higher Education, 32(2), 159-181.
- Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through selfassessment. *Theory in to Practice*, 48 (1), 12-19.
- Arrendondo, D.E., & Block, J.H. (1990). Recognizing the connections between thinking skills and master learning. *Educational Leadership*, 47(5), 4-10.
- Ary, D., Jacobs, L., Razavieh, A., & Sorensen, C. (2006). Introduction to research in education (7<sup>th</sup> ed.). California: Thomson.
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Education Researcher*, 36(5), 258-267.
- Bergstrom, J., & O'Brien, L. (2001). Themes of discovery. *Educational Leadership*, 58(7), 5.
- Black, P., & William, B. (1998). Assessment and learning. *Assessment in Education*, 5(1), 7-74.
- Block, J., Efthim, H., & Burns, R. (1989). Creating effective mastery learning schools. New York: Longman.
- Bloom, B. (1968). Learning for mastery. *Evaluation Comment*, 1(2), 1-12.
- California State Postsecondary Education Commission. (2001). Assessment and Change: The role of student testing in California education. Higher education update.

Sacramento, CA: California State Postsecondary Education Commission. (ERIC Document Reproduction Service No. ED452218)

- Clymer, J., & William, D. (2006). Improving the way we grade science. *Educational Leadership, Best of 2006-07*, 2-12.
- Croft, M., Waltman, K., Middleton, K., & Stevenson, E. (2005, April). The impact of school-level accountability on local test preparation practices. Paper presented at the National Council on Measurement in Education (NCME), Montreal, Quebec.
- Davis, J., & Rand, D. (1980). Self-grading versus instructor grading. Journal of Educational Research, 73(4), 207-211.
- Deeter, T., & Prine, D. (1998). Standardized tests: Summary of results 1997-1998.
  Focus on standardized testing (Des Moines Public Schools IA, p. 18). Des
  Moines, IA: Des Moines Public Schools, IA, Dept. for School Improvement and
  Employee Relations.
- Dewey, J. (1983). The influence of the high school on educational methods (Vol. 4, 1896). *American Journal of Education, 19*(4), 407-18.
- Dewey, J. (1995). Science as subject matter and as method. *Science and Education*, 4(4), 391-98.
- Dweck, C.S. (2000). Self-theories: Their role in motivation, personality, and development. Philadelphia: Psychology Press.
- East Side High School (2007). [School demographics and statistics]. Unpublished raw data.
- Edwards, N. (2007). Student self-grading in social statistics. *College Teaching*, 55(2), 72-76.

- Falchikov, N., & Boud, D. (1989). Student self-assessment in higher education: A metaanalysis. *Review of Educational Research*, *59*(4), 395-430.
- Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research.* 70(3), 287.
- Feuerstein, R. (1980). Instrumental enrichment: An intervention program for cognitive modifiability. Baltimore: University Park Press.
- Foley, E. (2002). Generally accepted principles of teaching and learning and their implications for local education support systems. Providence, RI: Brown University. (ERIC Document Reproduction Service No. ED 427990) Retrieved April 2, 2008, from Annenberg Institute for School Reform, Brown University Web site: http://www.schoolcommunities.org/images/GAPTL.pdf
- Forbes, S., Popard, M. A., & McBride, M.A. (2004). To err is human; to self-correct is to learn. *Reading Teacher*, 57(6), 566-572.
- Freeman, M., (1995, Dec.) Peer assessment by groups of group work. Assessment and Evaluation in Higher Education, 20(3), 289. Retrieved September 9, 2007 from Academic Search Premier Database.
- Freeman, M. (2007). Peer assessment by groups of group work. Assessment and Evaluation in Higher Education, 20(3), 289.
- Fuller, B., Gesicki, K., Kang, E., & Wright, J. (2006). Is the No Child Left Behind Act working? The reliability of how states track achievement. (Working Paper No. 06-1). Berkley, CA: Policy Analysis for California Education, PACE. (ERIC Document Reproduction Service No. ED 492024)

- Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books, Inc.
- Gewertz, C. (2007). States mull best way to assess their students for graduation. *Education Week*, 26(37), 1.
- Grobe, C.S., & Rendle, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, *17*(6), 612-634.
- Guskey, T. (2007). Closing achievement gaps: Revisiting Benjamin S. Bloom's learning for mastery. *Journal of Advanced Academics*, 19(1), 8-31. Retrieved April 29, 2008, from Academic Search Premier Web site.
- Hall, J.K. (2007). Redressing the roles of correction and repair in research on second and foreign language learning. *Modern Language Journal*, *91*(4), 511-526.
- Hayes, W. (2006). The progressive education movement: Is it still a factor in today's schools? New York: Rowan & Littlefield Education.
- Horn, C. (2003). High-stakes testing and students: Stopping or perpetuating a cycle of failure? *Theory into Practice*, 42(1), 30-41.
- Jarimillo, J. (1996). Vygotsky's sociocultural theory and contributions to the development of constructivist curricula. *Education*, *117*(1), 133-141.
- Johns, K.R. (2001). Going around the circle again: Exploring Kolb's theory of growth and development (Report No. JI-96-042). Alexandria, VA: National Council of Juvenile and Family Court Judges. (ERIC Document Reproduction Service No. ED 456 089)

- Kirby, N., & Downs, C. (2007). Self-assessment and the disadvantaged student: Potential for encouraging self-regulated learning? Assessment & Evaluation in Higher Education, 32, 475-494.
- Kitsantas, A., & Doster, J. (2004). Developing self-regulated learners: Goal setting, selfevaluation, and organizational signals during acquisition of procedural skills. *Journal of Experimental Education*, 72(4), 269-287.
- Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools.* Portsmouth, NH: Heinemann.
- Kolb, D.A. (1983). Experiential Learning: Experience as the Source of Learning and
- Development. New York: Prentice Hall.
- Kozulin, A. (1998). *Psychological tools: A sociocultural approach to education*.Cambridge, MA: Harvard University Press.
- Kulik, C. C., Kulik, J.A., & Bangert-Drowns, R.L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60, 265-299.
- Levykh, M.G. (2008). The affective establishment and maintenance of Vygotsky's zone of proximal development. *Educational Theory*, *58*(1), 82-101.
- Locker, J., & Cropley, M. (2004). Anxiety, depression and self-esteem in secondary school children: An investigation into the impact of standard assessment tests (SATs) and other important school examinations. *School Psychology International*, 25(3), 333-345.

- Mathan, S.A., & Koedinger, K. R. (2005). Fostering the intelligent novice: Learning from errors with metacognitive tutoring. *Educational Psychologist*, 40(4), 257-265.
- McMillan, J., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, 87(1), 40-49.
- Merriam-Webster's Online Dictionary. (2008). Retrieved December 6, 2008, from http://www.merriam-webster.com/dictionary/learning
- Michlitsch, J. F., & Sidle, M.W. (2002). Assessing student-learning outcomes: A comparative study of techniques used in business school disciplines. *Journal of Education for Business*, 77(3), 125-130.

McNeil, L.M. (2000). Contradiction of school reform. New York: Rutledge.

North Carolina Department of Public Instruction. (1996). *North Carolina end of course tests*. Raleigh, NC: North Carolina Department of Public Instruction. Retrieved December 10, 2009, from North Carolina Department of Public Instruction Website:

http://www.dpi.state.nc.us/docs/accountability/testing/NC%20End%20of%20Course%20Tests.pdf

North Carolina Department of Public Instruction. (1997). *State board of education releases results of first year under ABCs*. Raleigh, NC: North Carolina Department of Public Instruction. Retrieved January 12, 2009, from North Carolina Department of Public Instruction Website: http://www.ncpublicschools.org/abc\_results/results\_97/press\_release.html

- North Carolina Department of Public Instruction. (2004). *School-based management and accountability procedures manual*. Raleigh, NC: North Carolina Department of Public Instruction. (ERIC Document Reproduction Service No. ED486604)
- North Carolina Department of Public Instruction. (2004). Understanding the North
   Carolina end-of-course tests (Assessment Brief, Volume 9, Number 4). Raleigh,
   NC: North Carolina Department of Public Instruction. (ERIC Document
   Reproduction Service No. ED485446)
- North Carolina Department of Public Instruction. (2006). *Glossary of Terms*. Raleigh,
   NC: North Carolina Department of Public Instruction. Retrieved February 7,
   2008, from North Carolina Department of Public Instruction Web site:
   http://ncreportcards.org/src/glossary.jsp
- North Carolina Department of Public Instruction. (2006-07). Reports of disaggregated state, school system (LEA), & school performance data for 2006-07. In *NC school report cards*. Raleigh, NC: North Carolina Department of Public Instruction. Retrieved February 17, 2008, from North Carolina Department of Public Instruction Web site: http://www.ncreportcards.org/src/performance.jsp
- North Carolina Department of Public Instruction (2008a). *Mission statement*. Raleigh,
   NC: North Carolina Department of Public Instruction. Retrieved November 5,
   2008, from North Carolina Department of Public Instruction Web site:
   http://ncpublicschools.org/curriculum/
- North Carolina Department of Public Instruction (2008b). The Multiple-choice test development process. In *Accountability services division*. Raleigh, NC: North Carolina Department of Public Instruction. Retrieved February 7, 2008, from

North Carolina Department of Public Instruction Web site:

http://www.ncpublicschools.org/accountability/testing/eog/multiple?&print=true

- Olson, L. (2006). As AYP bar rises, more schools fail: Percent missing NCLB goals climbs amid greater testing. *Education Week*, *26*(4), 2.
- Orsmond, P., Merry, S., & Callaghan, A. (2004). Implementation of a formative assessment model incorporating peer and self-assessment. *Innovations in Education and Teaching International*, *41*(3), 273-290.
- Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66(4), 543-578.
- Phelps, R. (1997). The extent and character of system-wide student testing in the United States. *Educational Assessment*, *4*(2), 89.
- Phelps, R. (2000). Trends in large-scale testing outside the United States. Educational Measurement: Issues and Practice, 19(1), 11-21.

Piaget, J., (1952). The origins of intelligence in children. New York: Basic Books.

- The Princeton Review. (2003). *Testing the testers 2003: An annual ranking of state accountability systems executive summary* [Review]. New York: Author.
- Ross, J., & Starling, M. (2009). Self-assessment in a technology-supported environment: The case of grade 9 geography. Assessment in Education: Principles, Policy, & Practice, 15 (2), 183-199.
- Sadler, P., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, *11*(1), 1-31.
- Slavin, R.E. (2006). Educational Psychology: Theory and practice (8<sup>th</sup> ed.). Boston: Pearson Education, Inc. (Original work published 1986)

- Stefani, L. (1994). Peer, self and tutor assessment: Relative reliabilities. Studies in Higher Education. 19(1), 69.
- Sterling, D. (2008). Assessing student presentations from three perspectives. *Science Scope*, *31*(5), 34-37.
- Stotsky, S., (2005). The state of English standards, 2005. Washington, DC: Thomas B.Fordham Institute. (ERIC Document Reproduction Service No. ED485523)
- Strong, B., Davis, M., & Hawks, V. (2004). Self-grading in large general education classes: A class study. *College Teaching*, 52(2), 52-57.
- Struyven, L., Dochy, F., Janssens, S., Schelfhou, W., & Gielen, S. (2006). The overall effects of end-of-course assessment on student performance: A comparison between multiple choice testing, peer assessment, case-based assessment and portfolio assessment. *Studies in Educational Evaluation*, 32(3), 202-222.

Students self-grade: A successful model. (2007). Teaching Professor, 21(7), 6-8.

- Tan, K. (2008). Qualitatively different ways of experiencing student self-assessment. *Higher Education Research and Development*, 27(1), 15-29.
- Thompson, T., & Newsome, K. (2002). Is state assessment a viable tool for reflection of classroom assessment? (Report No. 143). Columbus, OH: North American Chapter of the International Group for the Psychology of Mathematics Education. (ERIC Document Reproduction Service No. ED471757)
- U.S. Census Bureau: State and County Quickfacts. (2007, May 7). Retrieved August 29, 2007, from: http://quickfacts.census.gov/qfd/states/37/37157.html

- Vermunt, J., & Yermetten, Y. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, 16(4), 359-384.
- Viel-Ruma, K., Houchins, D., & Fredrick, L. (2007). Error self-correction and spelling: Improving the spelling accuracy of secondary students with disabilities in written expression. *Journal of Behavioral Educations*, 16(3), 291-301.
- Vygotsky, L. S. (1978). *Mind in Society*. M. Cole, V. John-Steiner, S. Scribner, & E. Souberman (Eds.). Cambridge, MA: Harvard University Press.
- Walberg, H.J. (1986). *Syntheses of research on teaching* (3<sup>rd</sup> ed.) (M.C. Wittrock, Ed.). New York: Macmillan.
- West Side High School (2007). [School demographics and statistics]. Unpublished raw data.
- Whiting, B., Van Burgh, J.W., & Render, G. F. (1995). *Mastery learning in the classroom*. Paper presented at the annual meeting of the American educational Research Association, San Francisco.
- Yeh, S. (2001). Test worth teaching to: Constructing state-mandated tests that emphasize critical thinking. *Educational Researcher*, *30*(9), 12-17.
- Yoshida, R. (2008). Teachers' choice and learners' preference of corrective feedback types. *Language Awareness*, *17*(1), 78-99.