

COMPARISON OF TRADITIONAL EDUCATIVE DELIVERY  
TO ONLINE EDUCATION IN UNITED STATES HISTORY  
AS MEASURED BY FLORIDA'S END-OF-COURSE EXAMINATIONS  
IN A LARGE URBAN SCHOOL DISTRICT IN CENTRAL FLORIDA

by

WILLIAM T. WILSON

A. A. Central Florida Community College, 1992  
B. A. University of Central Florida, 1998  
M. Ed. University of Central Florida, 2005

A dissertation submitted in partial fulfillment of the requirements  
for the degree of Doctor of Education  
in the School of Teaching, Learning, and Leadership  
in the College of Education and Human Performance  
at the University of Central Florida  
Orlando, Florida

Summer Term  
2014

Major Professor: Barbara A. Murray

© 2014 William Timothy Wilson

## ABSTRACT

Student participation in online courses has been growing steadily for the past decade, and the trend appears to continue the growth in this form of instructional delivery method for the foreseeable future (iNACOL, 2012). To date, little research exploring student success rates exists in the social studies. This particular study was conducted to examine what differences, if any, existed in the End-Of-Course (EOC) scores of 11<sup>th</sup> grade United States history students who took the course in a traditional, face-to-face format versus students who took the same course online through Florida Virtual School.

For this study, proper permission was received from all interested parties, and a sample of 9,339 End of Course (EOC) examinations were taken from 36 high schools in a large, urban school district in Central Florida. All identifiable data were scrubbed from the sample. Due to the extremely small sampling of online students, the One-Sample Wilcoxon test was used on four research questions to compare students in the traditional, face-to-face versus online format and based on ethnicity, gender, and free-and-reduced lunch status.

Overall, none of the One-Sample Wilcoxon tests indicated the presence of a significant difference among any subgroup—overall, White, non-White, female, male, high socioeconomic status, or low socioeconomic status. Therefore, none of the null hypotheses presented were rejected. Recommendations included replicating the study on a broader scale and conducting a qualitative study to examine the characteristics of online students, their similarities and differences, to those of students who attend class in a face-to-face format.

I dedicate this dissertation to my parents.  
Even though they're not here, they know I've made it;  
and to my children, who I hope will one day follow in my footsteps.

## ACKNOWLEDGMENTS

The number of people I would like to thank is relatively small, but nevertheless, they are a network of family, co-workers, friends and professors who throughout this journey have become dear friends, all of whom played a tremendous role. Without their assistance, I would never have gone as far as I have gone, and reached as far as I have reached. To all of you, my humble, but sincere thanks. . .

-My wife, my best friend, my support, and my love. We have walked this journey together, and words simply do not exist that can eloquently enough define the love and pride I feel. Know that without you, I simply would not be.

-Dr. Barbara Murray, my dissertation chair; thank you so very much for all your support, the experience you've shared, your wisdom, and your, at times, brutal honesty. You and Ken have both become an inspiration for my wife and me. One lesson you've taught us was to let us know that it is okay to dream big. We hope to one day "be" you two.

-Dr. Rose Taylor. Thank you so much for starting that dream.

-Many thanks to Dr. Ken Murray, Dr. Lee Baldwin and Dr. Larry Holt for providing much needed guidance, direction, and recommendations as members of my committee.

-Madelyn Thomas, thank you for the countless number of times you provided your valuable time to watch our children while we attended classes, sometimes on short notice!

-Susie Quillin and Russ Ward, thanks for being the best in-laws ever. You've provided much needed advice, relief, and, yes, the occasional shoulder to cry on.

-To David Collins, you've been that steady friend whom I could rely upon, and that has meant more to me than you'll ever know.

-To all my family and friends at work, you know who you are--thank you for everything.

## TABLE OF CONTENTS

LIST OF TABLES .....	ix
CHAPTER 1 THE PROBLEM AND ITS CLARIFYING COMPONENTS.....	1
Introduction.....	1
Statement of the Problem.....	1
Significance of the Study .....	2
Definition of the Terms.....	3
Theoretical Framework .....	6
Pioneers of Psychometrics .....	7
Two Psychometric Models Used in Assessment and Testing .....	12
The Connection of Testing and Measurement With Current Educational Policy .....	15
Research Questions.....	20
Assumptions.....	
Limitations .....	23
Delimitations.....	24
Research Methodology .....	24
Organization of the Study .....	27
CHAPTER 2 REVIEW OF LITERATURE .....	29
Introduction.....	29
Overview of Instructional Delivery Methods Leading to Online Instruction.....	30
The History of Multimedia and Online Instructional Delivery .....	39
The Nature of Education: Home-School-Home Paradigms.....	40
Correspondence Study .....	42
Distance Education Through Multimedia.....	44
Distance Education Through the Internet .....	49
An Analysis of High-stakes Testing .....	55
History of High Stakes Testing.....	56
Issues Surrounding High-Stakes Testing.....	58
The Relationship of High-Stakes Testing to Accountability and Student Achievement .....	61
State of Florida Legislative Mandates Concerning End-of-Course Examinations.....	65
Legislative Development of End-of-Course Examinations .....	65
The United States History End-Of-Course Examination.....	73
State of Florida Legislative Mandates of Online Education in Florida .....	75
Florida Virtual School (FLVS) Legislative History .....	76
K-8 Virtual School Program .....	76
Digital Learning Now Act .....	78
Overview of Florida Virtual School (FLVS) .....	80

CHAPTER 3 METHODOLOGY .....	83
Introduction.....	83
Statement of the Problem.....	83
Research Questions.....	84
Research Design.....	87
Participants.....	88
Instrumentation.....	88
Data Collection.....	91
Data Analysis.....	92
Summary.....	95
CHAPTER 4 PRESENTATION AND ANALYSIS OF DATA .....	96
Introduction.....	96
Demographic Characteristics of the Sample.....	97
Research Question 1 .....	99
Research Question 2 .....	100
Research Question 3 .....	102
Research Question 4 .....	103
Summary.....	106
CHAPTER 5 SUMMARY, DISCUSSION, AND RECOMMENDATIONS.....	107
Introduction.....	107
Summary of the Study .....	107
Summary and Discussion of Findings .....	108
Research Question 1 .....	108
Research Question 2 .....	109
Research Question 3 .....	110
Research Question 4 .....	111
Implications for Practice.....	112
Recommendations for Further Study .....	113
APPENDIX A SCHOOL DISTRICT PERMISSION TO CONDUCT THE STUDY	114
APPENDIX B INSTITUTIONAL REVIEW BOARD APPROVAL.....	1166
REFERENCES .....	1188



## LIST OF TABLES

Table 1	Variables and Statistical Tools without Results #1 .....	22
Table 2	Variables and Statistical Tools without Results #2.....	86
Table 3	Demographic Characteristics of Sample.....	98
Table 4	One-Sample Wilcoxon Test: Overall Online Performance .....	99
Table 5	One-Sample Wilcoxon Test: Online Performance by Ethnicity.....	101
Table 6	One-Sample Wilcoxon Test: Online Performance by Gender.....	102
Table 7	One-Sample Wilcoxon Test: Online Performance by Free-and-Reduced Lunch Status.....	104
Table 8	Variables, Statistical Tools, and Results .....	105

## CHAPTER 1 THE PROBLEM AND ITS CLARIFYING COMPONENTS

### Introduction

According to the International Association for K-12 Online Learning [iNACOL] (2012), online education or distance learning has gained in numbers and acceptance over the past decade. For the 2010-2011 school year, 1,816,400 students in the United States on the K-12 level were enrolled in at least one online class, up from that estimated 400,000 at the turn of the century (iNACOL, 2012). For the 2010-2011 school year, the State of Florida employed 1,500 staff to serve over 100,000 online students (Florida Department of Education [FLDOE], 2013). Two such courses taken online by Florida students have been United States history regular, and United States history honors. United States history regular or United States history honors have been included as a part of social science assessment currently mandated as part of the Next Generation Sunshine State Standards and Common Core. At the time of the study, completion of either United States history regular or United States history honors course and passage of the state of Florida 11<sup>th</sup>-grade End-of-Course examination was required for Florida high school graduation (FLDOE, 2013).

### Statement of the Problem

To date, there has been little research comparing student performance outcomes on the United States History end-of-course examinations completed by students enrolled in traditional face-to-face instruction with students enrolled in an online United States

history course. Contradictory data exist on the effects of student achievement overall in online, or distance learning. For example, a 2009 report by the United States Department of Education showed that online instruction produces similar results when compared to face-to-face learning (Groux, 2011). Yet critics of online education claim that online students suffer from high rates of dropout (Bennett, Lucchesi, & Vedder, 2010). Minnesota reported only 16% of online high school students passed the state's mathematics proficiency examination (Lemaige, 2011). Overall, students who took at least one course online in 2010 had a 34% dropout rate, versus a dropout rate of 26% for students who attended traditional face-to-face classes (Xu & Jagers, 2011). This study looks to determine if any differences in student success exist between students who take a course in a traditional, face-to-face classroom setting and those who take a similar course in an online format.

### Significance of the Study

This study focused on comparing online instructional delivery with traditional face-to-face instructional delivery of United States history regular courses and United States history honors courses in high schools as measured by the state of Florida 11<sup>th</sup>-grade End-of-Course examination. The contribution of this study may be of major importance to the design of instructional delivery in the high school setting. School district officials should be interested in any information that may serve to improve student performance outcomes on End-of-Course examinations and lead to higher school graduation rates. Legislators and department of education officials who support

mandates for online education should also be interested in the findings produced in this study. It is clear that online education is here to stay and, therefore, must receive appropriate review and be subject to quality control.

### Definition of the Terms

Online School. Either a virtual private, or public school on both the K-12 and secondary level of school where classes are taken online through the use of a computer. Credit can be earned in any one or a combination of three types of formats: asynchronous, synchronous, or hybrid. Online Education schools and programs are also sometimes referred to as “distance learning”.

Asynchronous. An online educational delivery method which allows students to work on required coursework on their own time schedules, usually has weekly deadlines.

Synchronous. An online educational delivery method that is less flexible than asynchronous, as it requires real-time commitment with conference calls, online chatting, and teleconferencing.

Hybrid. An online educational delivery method commonly used to describe courses in which some traditional face-to-face “seat time” has been replaced by online learning activities. It is also referred to as “blended learning” or “blended instruction”.

Next Generation Sunshine State Standards (NGSSS) Standards passed by the Florida State Legislature in 2010. The NGSSS have been divided using benchmarks that describe the knowledge or ability that a student should be able to demonstrate mastery of by the end of each grade level up through 12<sup>th</sup> grade. Students’ progress is measured

based on their performance on the Florida Comprehensive Assessment Test (FCAT) test and related End-of-Course (EOC) examinations. These standards are aligned with the Federal Government's Common Core standards.

End-of-Course examinations (EOCs). These are formative, criterion-based examinations given at the end of a course. They are designed to measure the level of content mastery a student has achieved. Upon successful passage of the examination, students receive credit for passing the course. The State of Florida currently uses EOCs for a number of classes that students must pass in order to receive credit in both the high school and middle school settings.

Florida Virtual School (FLVS). Part of the Florida public school system. It is also the largest state-sponsored virtual or online school in the United States. It started in 1997 as a collaboration of two county public school systems, Alachua and Orange counties, with free enrollment for Florida students. The majority of students who enroll in FLVS are enrolled part-time. They are still enrolled in their school of zone, and their virtual school enrollment often takes the form of a remedial class to make up missing credits.

Face-to-face education is the more traditional education found in a classroom setting. It is often times referred to as "traditional" schooling.

Norm-referenced. A type of test that determines if the test-taker performed better, or worse, than the others who took the same test.

Criterion-referenced. A type of test in which a person's mastery of a subject is measured based on specific standards. Most school tests are written and use pre-defined objective standards or achievement levels.

Standardized test. A test that is given and scored in a consistent, 'standardized' way. These tests are designed in such a manner that all conditions surrounding the test; format, scoring procedures and the interpretations thereof, are all the same and consistent.

Psychometrics. The field of study concerned with the theory and technique of psychological measurement, which includes the measurement of knowledge, abilities, attitudes, personality traits, and educational measurement.

No Child Left Behind (NCLB). Federal legislation signed into law by then-president George W. Bush in 2001 that enacts the theories of standards-based education reform. Pursuant to 20 USCS § 6301, NCLB ensures that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging state academic achievement standards and state academic assessments (U.S. Legal.com, 2013).

Race To The Top Initiative (RTTT). A competitive grant program designed to encourage and reward States that are creating the conditions for education innovation and reform; achieving significant improvement in student outcomes, including making substantial gains in student achievement (United States Department of Education [USDOE], 2009).

Open universities. Schools of higher learning, started in England, that have classes open to any and all people, regardless of academic background. Many are free, and are taught in a correspondence, or broadcast format to students over long distances.

Massive Open Online Courses (MOOCs). Online courses that have open enrollment with unlimited participation, and often times are free of charge to the participants. Although most classes have specific structure and learning goals, MOOCs promote a highly interactive user community where students learn through group collaboration and peer review and professors are more facilitators as opposed to dispensers of information.

### Theoretical Framework

The comparison of the End-of-Course examinations from students who have completed the state of Florida 11<sup>th</sup>-grade United States history course in a traditional face-to-face instructional delivery method to an online delivery method was central to this study. Psychometrics would be one such theory which could provide the appropriate rationalization and validation for this particular study. The following three components were identified and researched to provide the foundational theoretical underpinnings of this study: (a) a brief overview of the history of the evolution of psychometrics through the examination of early psychometricians from the late 19<sup>th</sup> century to present day, (b) an exploration of the two primary theoretical models pertinent to present day assessment and measurement, and finally, (c) the connection and impact that testing and measurement has had on current educational policy. The first two parts address the

specifics and background of statistical analysis relevant to measurement in the social studies. The third part, which is less technical, describes the complexities surrounding the practical employment of standardized testing as it relates to current educational policy.

Psychometrics were defined by Stevens (1946) as “the assignment of numerals to objects or events according to some rule” (p. 677). In the late 1800s and early 1900s, the use of psychometrics was focused primarily in the study of Eugenics and early psychological testing. Several researchers were instrumental in the development of important statistical methods that are discussed in this chapter. In modern times, however, psychometrics has been concerned with two main responsibilities: (a) the creation of instruments and procedures for measurement; and (b) the development of theoretical approaches to measurement (Collins English Dictionary, 2014). However, the origins of this definition came from varied sources and date back to the middle 19<sup>th</sup> century.

### *Pioneers of Psychometrics*

German philosopher and psychologist, Johann Friedrich Herbert, worked on the psychological study of metaphysics, and articulated methods of incorporating ways to measure observations in psychology (Miller, 2003). In addition, Herbert was also instrumental in the development of the concept of pedagogy, the science and art of education, as an academic discipline (Miller, 2003). Blyth (1981), from the British School of Education at Liverpool, England asserted that the use of pedagogy was critical



in the development of modern education. Pedagogy, he reasoned, was the inherent transfer process from the individual as the start point, to the finished product, character. It was the teacher's responsibility to use pedagogy as the process agent for that change, and, as Blyth (1981) stated, should be “. . . grounded in intellectual education” (p. 72).

Sir Francis Galton, genius and cousin to Charles Darwin, expanded upon the observations of Belgium mathematician, Lambert Quitelet, who discovered an order to individual variations in observed occurrences (Tyler, 1963). Galton was inspired to research the human condition after reading Darwin's *The Origin of Species*; specifically the chapter related to the breeding of domestic animals (Forrest, 1974). He was the first scientist to coin the term Eugenics, in 1883 (Galton, 1883). Eugenics was the study, belief, and practice of changing the human population through genetics. As a social philosopher, Galton believed that by studying results through observation, one could improve the human race by encouraging the reproduction of positive human characteristics over negative human characteristics. With enough observation and data collection and analysis, it would ultimately allow for the eventual extinction of undesirable human traits (Dowbiggin, 1997). In addition to his contributions to Eugenics, Galton noticed in his observations that often times, a pattern in the distribution of data would occur, and that pattern resembled a symmetrical, bell-shaped curve (Tyler, 1963). Galton would continue his experiments with numbers distribution as he researched heredity and environment where he used the statistical techniques of correlation, regression, and regression to the mean to explain relationships that he found in nature (Nelson, Pettersson, & Carlborg, 2013). While not the first to use the statistical

concept of regression and correlation, he was able to demonstrate how these statistical methods were applicable in the study of heredity, anthropology and psychology (Bulmer, 2003). In one particular incident at a fair, for example, Galton (1907) observed that approximately 800 people guessed the weight of a recently slain cow. To his surprise, no one guessed the exact weight, but by looking at the numbers he discovered that despite the fact that there were some wide discrepancies in the guesses by a small percentage of the people participating, when he performed the calculations he discovered that the median guessed by the people was only .08% off from the weight that was calculated by the judges. This was what he referred to as variance and led him to discover standard deviation as a way of analyzing variance (Clouser, 2007). To this end, it would be these statistical methods that would lay the groundwork for future psychometricians to develop more precise and elaborate statistical analyses (Clouser, 2007).

According to Tankard (1984), Pearson led a wide and vast career that spanned many decades and many areas, from science, to philosophy, to literature. He was a devout socialist who was inspired after meeting Galton to study Eugenics (1984). He soon became a student of Galton who became Pearson's mentor (Tankard, 1984). In his years of research, Pearson contributed greatly to the field of statistics and psychometrics. He worked with Galton in founding the journal, *Biometrika*, which focused on the development of statistical theory (Tankard, 1984). The statistical methods he developed served as the basis for present-day testing, e.g., the Correlation coefficient and its relationship with linear regression, Pearson's system of continuous curves, P-value, and Pearson's chi-squared test. He also created statistical hypothesis testing theory and

statistical decision theory, both of which remained in use in testing and data analysis at the time of this study (Pearson, 1900).

James MacKeen Cattell was an American student of Wundt in the late 1800s, and according to Butler-Bowden (2007), Cattell was considered by many as the father of modern psychology. Early in his career, he worked under Wundt and wrote his doctoral dissertation, *Psychometric Investigation*, which opened the door to the development of intelligence in measureable terms (Butler-Bowden, 2007). According to Thorne and Henley (1997), Cattell introduced the term mental intelligence as a general term in applications relating to psychological testing. During his long career, he fought hard against the belief that psychology was just simply a pseudoscience (Butler-Bowden, 2007). He also became the first American to publish in the field of psychology. Later, he became the very first president of the American Psychological Association and was credited with the establishment of psychology as a legitimate science in the United States (Woodsworth, 1944).

Charles Spearman was one of the next generation of psychometricians who would continue the work of Galton. Spearman spent 15 years early on in his career as an engineering officer in the British Army and then returned back to school and earned his doctorate in 1907. He taught at Cambridge University until his retirement in 1931 (Thomson, 1947). During his tenure, he was able to develop the statistical formulations for two-factor theory, later termed factor analysis (Lovie & Lovie, 1996). He had collected data from testing on children with dissimilar intelligence levels and ages. In the analysis of the data, he discovered a peculiar hierarchy of data that proved the empirical

and conceptual (Lovie & Lovie, 1996 ) of his theory, which created statistical validity of the data. According to Lovie and Lovie, he named the two factors, *g*, for general intelligence, and *s* for specific ability, and immediately received criticism from Thorndike, Thurstone, and especially Pearson. They disbelieved that the notion of human intelligence could be regarded in such “simplistic a notion as *g* and *s*” (Lovie & Lovie, 1996, p. 82). The debate raged on for several decades, and despite criticisms from psychometricians from several areas, Spearman, along with assistants and detractors alike, worked diligently to develop the two-factor theory into the more complex model of factor analysis in use at the time of the present study (Lovie and Lovie, 1996).

Thurstone was well known for expanding Spearman’s work, added his contributions to what would eventually become known as factor analysis (Martin, 1996). According to Horst (1955), Thurstone was responsible for the founding of the Psychometric Society. During his long career, he helped to develop new psychological measuring and scaling techniques, e.g., the development of the Thurstone scale. In addition, Horst noted that Thurstone was deeply concerned with measurement concepts such as validity, internal consistency and the “fundamental problems of measurement and identification of variables” (p. 1260), and purposed much of his work toward getting psychology out of the ivory tower and into the measurement of the attitudes of regular people in real-life situations. To that end, he was deeply involved in expanding Binet’s work when he developed the modern definition used today in both the standardized mean and deviation used in the IQ scores from the Intelligence Quotient test developed by Binet (Horst, 1955).

### *Two Psychometric Models Used in Assessment and Testing*

There are two theoretical psychometric models which underpin testing and evaluation that were pertinent to this theoretical framework. They are classical test theory and item response theory.

According to MacDonald (1999), classical test theory, also referred to as true score theory, provided a way to estimate how exact the measurement of a particular test score was. Classical test theory was most often associated with empirical applications that required a high level of reliability in psychological tests (Borsboom, 2005). This theory of reliability in test administration did not come about until the last of three necessary ingredients were discovered in the early to mid-1900's (Traub, 1997). First were the discoveries of error in measurements, as defined by researchers like Thurstone and Spearman (Traub, 1997). Spearman helped to figure out how to correct a “correlation coefficient for attenuation due to measurement error” (p. 8) and create an index of reliability in order to make that correction. This, according to Traub (1997), marked the beginning of classical test theory. Finally, in 1937, the last ingredient was added when Kuder-Richardson published their formulas that presented the ideas of lower bounds of reliability (Traub, 1997). According to Traub (1997), classical test theory was fully realized as a workable test theory when Novick used previous work from other researchers such as Yule, Kelley, and Guttman, and applied all these principals to develop the systemic treatment which remained in use at the time of the present study (Traub, 1997).

According to MacDonald (1999), classical test theory focuses on three distinct variables, X, T, and E of a population in a simple equation of  $X = T + E$  where X was simply the observed score, T was considered the True score, and E was regarded as the Error (1999). Reliability, according to Streiner (2003), was impossible to estimate directly because the True score can never be known. It could, however, be estimated in an index of .0 to 1.0, where the closer to 1.0 the more reliable the test question, and therefore, the index also could be spoken to the quality of the test question (Streiner, 2003).

Classical test theory, however, was found to have several shortcomings (MacDonald, 1999). First, it does not provide information necessary in the evaluation of single test items (Traub, 2005). Item analysis had to rely on two statistics, the P-value and the item total correlation. P-value was defined as the proportion of test takers who responded within the keyed direction, typically referred to as item difficulty (Traub, 2005). The item-total correlation provided an index of the differentiated power of the test item and was most often used in diagnosing possible test item issues such as confusing detractors in multiple choice items (Traub, 2005). Secondly, according to Traub (2005), there is a separation of test characteristics and test taker characteristics, and they can only be interpreted in context of one another. The third shortcoming is found in the very definition of what reliability means, “. . . the correlation between tests scores on parallel forms of a test” (Traub, 2005, p. 8), where different opinions could exist on just what parallel tests are. Fourth, classical test theory assumes that the standard error of measurement is consistent between all test takers. According to Hambleton,

Swaminathan, and Rogers (1991), however, to make this assumption would be impossible as scores on tests are unequal in how they measure due to differences in test takers abilities. Finally, classical test theory is test oriented, not item oriented, so predictions made concerning a particular test taker, or group of test takers would be impossible to determine (Traub, 2005).

Item response theory (IRT), according to Embretson and Reise (2000), was the next major model after classical test theory in the evolution of psychometrics and was regarded to be far superior (2000). It was a major psychometric theory that was developed principally from researchers, e.g., Lord, Rasch, and Lazarsfeld, in the 1950s and 1960s (Hambleton et al., 1991). The purpose of item response theory was to provide a framework to estimate how well assessments work, and more specifically, how well the individual items on assessments work (Hambleton et al., 1991). Many standardized tests, including 21st century computerized adaptive tests were developed from the IRT and was used as the mainstream theoretical basis for measurement, including the field of education (Embretson & Reise, 2000). The reason for this reliance on item test response was because it treated the difficulty of each test item as information to be incorporated in scaling. In other words, it modeled the response of each test taker's given ability to each item in the test (Embretson & Reise, 2000).

According to Bock and Aitkin (1981), item response theory had three basic assumptions; a unidimensional trait denoted by  $\Theta$ , test item independence, and the response of the test taker to an item. The test taker's response is defined on a scale determined by the test taker's ability, and was referred to as the item response function

(IRF). The lower the test taker's ability, the lower the chance one would answer the question correctly, and the higher the test taker's ability, the higher the chance of responding correctly (Bock and Aitkin, 1981). In addition, test items were assumed to be dichotomous. This meant that even with multiple choice test items, the answers were assumed to be either correct or incorrect (Bock and Aitkin, 1981).

According to Yu (2010), IRT was descriptive in nature because the goal was to fit the model to the data as opposed to fitting the data to the model (2010). In other words, the premise of IRT was predicated on the assumption that the probability of a correct answer to a test item was a function of the student's ability, which in modern test-taking terms, is critical (Thompson & Weiss, 2009).

#### *The Connection of Testing and Measurement With Current Educational Policy*

The connection that standardized testing has on Educational reform and policy is a complex one, and should be addressed from a few perspectives such as instructional delivery, student perspective, validity of test purpose, and the evolution of instruction and learning. Finally, the various perspectives examined in this section, along with the current policy trends which rely heavily on data, will show that an impact on standardized testing exists in terms of accountability and student achievement.

From the perspective of instructional delivery in the domain of the social studies, Gaudelli (2002) stated that history presented the most problems with regard to the use of standardized tests because of the different vantage points held by stakeholders. Those viewpoints were innumerable and "disjointed" in the search for common ground. Many



different viewpoints are held and could be influenced by stakeholders, e.g., instructors, district policy makers, curriculum writers, testing companies, parental groups, lobbyists and legislative groups. Gaudelli cautioned that standardized tests could be portrayed from any one historical viewpoint. An instructor could teach the material correctly, but from a viewpoint that did not match the test. Gaudelli asserted that a universal curriculum was an option, but the study of history constantly changes. Even if the study of history could be universalized, the trends of how history is taught continuously change (Gaudelli, 2002).

The perspective in which history is presented as instruction to the student has had an impact on testing. Similarly, the general perspective of the student has had an impact on the authenticity of standardized testing as well. According to McCoog (2005), students perform poorly on tests in the social studies discipline, especially in the realm of history. McCroog stated that students had a hard time relating to the past from their modern perspectives. In his article, he acknowledged two nearly opposing viewpoints concerning specifically the discipline of history and standardized testing. First, in 1994, Stern observed that standardized testing did not account for the ability to understand the intricacies and complexities of American history. Stern stated that the issue lay in how history was presented to students from the perspectives of the norms and values of present day. Though not viewed by Stern as the goal of instruction, he posited that national standards and standardized testing would force teachers to present it as such. Kornblith and Lasser (2004), however, had a different perspective. They stated that for

standardized testing to remain a good assessment both the tests and curricula must be reformatted every year.

The perspective of test validity itself has also played an important role in past and present standardized testing practice, impacting 21st century educational reform. According to Shepard (2013), achievement tests in public education schools a century ago were assumed to be valid due to the content reviews performed by experts. She stated that testing evolved as a more mature unified theory of validity which required all tests to have both some form of measureable construct and some form of empirical content. One of the main problems with validity, she asserted, was that tests do not always test what they are supposed to actually test or accurately measure what they are supposed to measure. For example, she cited a Texas study that showed conclusively that there were not only gains in student achievement in reading, but that gaps between white students and minority students had also narrowed significantly. In reality, in 2013, the National Assessment of Educational Progress (NAEP) testing data on reading for the same time period showed exactly the opposite. Shepard's stance was that differing validity evidence was needed in order to explain differing uses. In other words, the United States has become increasingly convinced of using one particular type of test in order to determine multiple policy decisions "predominantly for accountability purposes" (Shepard, 2013, p. 2). Tests should be framed around a specific purpose in order for decisions to be made on the most current and relevant authoritative summary of validity evidence (Shepard, 2013).

The concept of instruction and learning has changed over time, as has the understanding that individuals have when it comes to testing and measurement. According to Welner (2013), in the past, testing was purely a measurement tool designed to help teachers objectively measure the extent of student learning and achievement, and a summative tool for admission or for future class placement (2013). For those purposes, however, according to Baker (2013), in the past, normal distributions were sufficient for test design and formed the basis for test interpretations with the major goal designed to segregate test takers into labels or classifications. Unfortunately, at that time, however, objective tests missed elements that should have been measured, e.g., abstract and mechanical intelligence, social intelligence, and even interpersonal relations (Baker, 2013). In addition, the need for a more sophisticated test measurement proved a shift from an empirical approach that had a more simplistic view of validity to one that addressed the more “complex concepts of the validity argument” (p. 5). Baker asserted that a large group of educational theorists in the latter part of the 20<sup>th</sup> century, including Bloom, Skinner, Lumsdaine and Glaser, added to the body of knowledge that changed the understanding of teaching and learning to be focused on complete alignment and integration of instruction with measureable performance that had a predictable set of outcomes and with criterion-based formative and summative tests that were divided neatly into domains and standards where goals were accessible and relevant.

As stated by Shepard (2013), testing and measurement in the 21st century has changed from a focus of testing only the student to being used to assess teachers and administrators. The data have come to be used to drive state and district policy. In

addition, Henig (2013) concurred in his work that the current emphasis on high-stakes testing was reflective of the nation's shift towards an evidence based educational policy in which political leaders and stake holders rely heavily on data. According to Henig, data are sterile and affect different stake holders differently. To teachers, data are used as a part of job evaluations through such statistical algorithmic computations like the value-added model. To interest groups, data are used as a political weapon in order to place pressure on policy makers. To a political leader, data provide a comfortable safe haven in which to anchor educational policy that can be used to "absolve politicians of the responsibility of outcomes" (p. 6) that may be unfavorable to constituents (Henig, 2013). Henig, in quoting then Mayor Michael Bloomberg, noted that data can be viewed as a powerful tool that state and school districts use in order to make policy decisions: "We have a saying that in God we trust--everybody else has to bring data. . . this business of teaching to the test is exactly what we should do, as long as the test reflects what we want them to learn" (2013, p. 2).

The impact that testing has on educational policy is indeed complicated. According to Welner (2013), "meaning is created by use" (p. 1) and meaning has indeed changed as the nation steers towards more accountability through testing. Shepard (2013) cautioned, however, that ". . . formative or learning purposes are subverted when combined with summative or accountability-oriented testing" (p. 10) which can leave the possibility wide open for the realization of the law of unintended consequences in terms of educational policy. Testing, it seems, impacts educational policy not so much from the basis of fact, but from the nuances of interpretation (Henig, 2013).

### Research Questions

The following research questions and hypotheses were used to guide the study:

1. What difference, if any, exists between the End-Of-Course (EOC) examination scores of students who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus those students who complete the same course in an online format through Florida Virtual School (FLVS)?

$H_{01}$  There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.

2. What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on ethnicity, who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus those students who complete the same course in an online format through Florida Virtual School (FLVS)?

$H_{02}$  There is no significant difference between students, based on ethnicity, who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method versus the students

who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.

3. What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on gender, who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus students who complete the same course in an online format through Florida Virtual School (FLVS)?

$H_{03}$  There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method based on gender versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.

4. What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on free-and-reduced lunch, who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus students who complete the same course in an online format through Florida Virtual School (FLVS)?

$H_{04}$  There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method based on free-and-reduced

lunch status versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.

The following table presents both independent and dependent variables and which statistical tool was used.

Table 1

*Research Questions, Independent Variable, Dependent Variable and Statistical Tool.*

Research Question	Independent Variable	Dependent Variable	Statistical Tool
Research Question 1	Overall, Face-to-Face (Population)	Overall, Online (Sample)	One-Sample Wilcoxon test
Research Question 2	White, Face-to-Face (Population)	White, Online (Sample)	One-Sample Wilcoxon test
	Non-White, Face-to-Face (Population)	Non-White, Online (Sample)	One-Sample Wilcoxon test
Research Question 3	Female, Face-to-Face (Population)	Female, Online (Sample)	One-Sample Wilcoxon test
	Male, Face-to-Face (Population)	Male, Online (Sample)	One-Sample Wilcoxon test
Research Question 4	No Lunch Assistance Face-to-Face (Population)	No Lunch Assistance Face-to-Face Online (Sample)	One-Sample Wilcoxon test
	Free or Reduced Lunch Face-to-Face (Population)	Free or Reduced Lunch, Face-to-Face Online (Sample)	One-Sample Wilcoxon test

### Assumptions

1. It was assumed that students completed the United States history course successfully with fidelity in order to take the EOC Assessment.
2. It was assumed that students answered the instrument questions truthfully and to the best of their ability.
3. It was assumed that students were comfortable using computers in both formats.
4. It was assumed that the students who completed the United States history course online took the EOC examination in the school of their zoning.

### Limitations

The study had the following limitations.

1. The sample of students was drawn from a large urban school district in Central Florida; therefore, the results may not be generalizable to schools or school districts in the rest of the state or other states.
2. In addition to the results not being generalizable to schools within the state, these results from the United States history EOC examination scores should not be generalized to students in other online educational programs.
3. Other variables exist which were outside the researcher's control. Some of these variables were: uncontrollable variations in teacher quality and teacher effectiveness, student motivation, and student access to technology.



4. The numbers of online students for this study were few in number and are not representative of the student population, thus inferences were difficult to make. Any trends that were detected in the analysis of this study should not be made as generalizations.

### Delimitations

1. The study was delimited to students from a large urban school district in Central Florida who have completed Florida's United States history course regular (course code 2100310) and honors (course code 2100320).
2. The study was delimited to the United States history EOC assessment as provided by the Florida Department of Education (FLDOE).

### Research Methodology

To determine what difference, if any, existed between the End-of-Course (EOC) examination scores of students who completed the Florida United States history in the traditional, face-to-face instructional delivery format versus those students who completed the same course in an online instructional delivery format through the Florida Virtual School (FLVS), a total of 9,339 students completed the course in either the traditional or online format and had a scale score for the EOC examination. Of these students, only 10 completed the course online; the remaining students completed it in the face-to-face modality. Because the face-to-face results represent the performance in the established, status quo modality, the results of the 9,329 in the face-to-face group were

considered to represent the population. The 10 students in the online group served as the sample for comparison. The data were collected with the assistance of the school district's Office of Accountability and Assessment.

The data subsets for research questions two through four do not fit a normal distribution for the students who comprised the online cohort of this study. Due to the very small size of the sample data subsets, caution was necessary when making inferences in order to avoid the possibility of error. In order to minimize the possibility of such an error, a conservative test, the Wilcoxon signed-rank test was used. According to Kiess (1989), the Wilcoxon signed-rank test was also referred to as the One-Sample Wilcoxon test. The One-Sample Wilcoxon test was the non-parametric equivalent to the One-Sample T-Test. A non-parametric test can be advantageous when "the statistical tool requires no assumptions about the population parameters" (Lapin, 1973, p.514). According to Kiess (1989), non-parametric tests can also be referred as distribution-free tests.

Additionally, the One-Sample Wilcoxon test was used "for within-subjects designs with two levels of an independent variable" (Kiess, 1989, p.478). The One-Sample Wilcoxon test was a test that considered "both the direction and the magnitude of the differences between matched sample pairs" (Lapin, 1973, p.531). In other words, the One-Sample Wilcoxon test was used when comparing two related or matched samples to assess whether their population-mean ranks differ (Lapin, 1973).

For Research Question 1, a One-Sample Wilcoxon test was run to determine if a statistically significant difference existed between the EOC examination scores of

students from the traditional, face-to-face format and students who completed the course online. The One-Sample Wilcoxon test, the nonparametric alternative to the one-sample  $t$ -test, was chosen due to the extremely small sample size of 10 students in the online cohort; in general, a one-sample test was chosen because the performance of students in the traditionally-formatted class represents the established population value rather than another sample. The performance of online students served as the comparison sample. For this test, the EOC examination scores represented the continuous dependent variable. The median value of the face-to-face scores was calculated and used as the value to which the median value of the online scores was compared.

Research Question 2 intended to measure differences in performance between online and the established face-to-face population when taking gender into consideration. Because comparisons still needed to be made relative to the population performance of the face-to-face group, two separate One-Sample Wilcoxon tests were run, one within each of the two gender categories (female and male) using the same continuous dependent variable of EOC examination scores. Therefore, differences could be detected between the online sample and the face-to-face population with respect to each gender.

Research Question 3 followed the same pattern as Research Question 2, but instead measured differences with respect to ethnicity. Detailed ethnicity information was provided (White, Black, Hispanic, Asian, and Multiracial); however, with such a small sample size among online students, the ethnic categories were combined into White and non-White. The One-Sample Wilcoxon test was then run within these two groups to

determine differences in the median performance of online students in each of these categories as compared to the face-to-face population values.

Research Question 4 also followed the same pattern as Research Questions 2 and 3, but instead measured differences with respect to socioeconomic status. Students were separated into groups reflecting those who did not receive any free or reduced-price lunch assistance (high SES) and those who did receive such assistance (low SES). The One-Sample Wilcoxon test was then run within each of these groups to determine differences in the median performance of online students in each of these categories as compared to the face-to-face population values.

Data were collected from 36 high schools in a large urban district in Central Florida's office of Accountability and Assessment. As most students enrolled in the online class were part-time, they were required physically to take the EOC examination in the school of their zoning and not online. Thus, their scores were included with the data from the Office of Accountability and Assessment. The instrument used was the Florida End-of-Course (EOC) examination.

### Organization of the Study

The report of this research study is presented in five chapters. Chapter 1 includes the introduction and background of the study, statement of the problem, the purpose of the study, the significance of the study, the definition of terms, and the theoretical framework. The methodology used to conduct the study was outlined, including research question, limitations, delimitations, assumptions, and the organization of the study.

Chapter 2 presents the review of the literature in six parts: (a) an overview of the methodologies of instructional delivery systems leading to the development and incorporation of online or distance learning in the United States; (b) a close examination of the history of multimedia and online instructional delivery; (c) an analysis of high-stakes testing; (d) State of Florida legislative mandates concerning End Of Course examinations; (e) State of Florida legislative mandates concerning online education; and (f) an overview of the Florida Virtual School (FLVS). Chapter 3 describes the methodology used in the research study. It includes a restatement of the problem and the research questions and the methods and procedures used to conduct the study. The participants, instrumentation, data collection, and data analysis are detailed. Chapter 4 contains a report of the analysis of the data organized around each of the research questions that guided the study. Chapter 5 provides a summary of the research, a discussion of the findings, implications for practice and recommendations for future research.

## CHAPTER 2 REVIEW OF LITERATURE

### Introduction

This study looked at the End Of Course (EOC) examinations of students who completed the 11<sup>th</sup>-grade state of Florida United States history course, regular, and honors in both the online and face-to-face formats. The following six topics were identified as components which were relevant to this study; Instructional delivery methods, distance learning, End of Course Examinations, high-stakes testing, state of Florida legislative mandates for online education and End Of Course examinations and the Florida Virtual School (FLVS) and all were topics which created the underpinnings on which this study was based.

As such, in order to provide a clear rationale for this study, a better understanding of the components was necessary. The review of literature for this chapter is presented in six separate sections: (a) an overview of the methodologies of instructional delivery systems leading up to the development and incorporation of online or distance learning in the United States, (b) a close examination of the history of multimedia instructional delivery and online instructional delivery, (c) an analysis of End-of-Course examinations and high-stakes testing, (d) a summary of the legislative mandates of online education in the State of Florida, (e) a summary of the state of Florida legislative mandates of End-of-Course examinations; and (fi) an overview of the FLVS.

## Overview of Instructional Delivery Methods Leading to Online Instruction

The comparison of the online instructional delivery with traditional face-to-face instructional delivery in high school United States history courses was an important component to this study. Thus, literature was reviewed relating to the various methodologies of instructional delivery. According to Barlow (1985), no one instructional methodology covers all teaching and learning situations. There are, however, several approaches of instructional delivery commonly used in United States education which include expository, inquiry/guided discovery and individualized instruction. Individualized instructional methods can be further categorized by programmed instruction, learning centers, contracting, mastery learning, and computer assisted instruction. Expository teaching, the basis of Ausubel's (2000) assimilation theory, is an educational method that employs lecturing as so to effectively communicate information that is relevant to the lesson at hand. Ausubel emphasized meaningful learning and retention is helped by anchoring the lesson's ideas through the use of advanced organizers. Ausubel argued that learning does not necessarily need to be rote memorization, though he did acknowledge certain inadequacies exist. Inadequacies included the instructor's use of words or language which caused students to suffer due to either the lack of vocabulary or background knowledge; the inundation of unrelated and unimportant trivia and facts; and the use of assessment which focused on rote memorization or minutiae (Ausubel, 2000). An example for a social studies class would be a teacher in a high school setting using Ausubel's methods of assimilation theory, e.g., using interactive lecture with advanced organizers to provide background information on

the concept of GDP in different countries. This method would be appropriate due to the large quantities of information to efficiently disseminate to a class (Wilensky, 2004).

Guided discovery is a method of instruction where the teacher is more a facilitator in a lesson and students learn by combining and categorizing information, emphasizing the learned contents into relationships (Bruner, 1966). In this instructional format, students are not presented with the lesson's end result. Rather, it is the students, themselves, who reach the conclusion of the lesson through the discovery of relationships. There is typically an increased amount of classroom activity (Lefrancois, 1972). According to Bruner (1966), when discovery learning is used in problem solving lessons, student retention and mastery of the material is more likely, especially when activities include stimulating audio-visual aids. One of the criticisms of the guided discovery methodology has been that this particular method cannot be used for all subjects and in all areas. For example, when students already have a large amount of background information and can extrapolate the purpose and outcome of the lesson with abstract thought, an expository approach would be more meaningful (Lefrancois, 1972). An example of guided discovery in the social sciences would be a primary school teacher teaching the concept of a map to first graders. The students, guided by the teacher, are shown examples of different types of maps and allowed to examine symbols, keys, and attributes found in the maps. Students are then shown non-examples of maps, and, again, with questions guided by the teacher, the students compare differences between the two. At the end of the lesson, students are given a set of maps and non-maps and are asked to create a group with only maps in it (Johnson, 2006).



Individualized instruction is another instructional delivery method commonly used in the United States. Lefrancois (1972) argued that each child is unique, and education and learning are dependent upon several factors. In regard to student learning and success, much of it, according to Lefrancois (1972) was due to motivational factors and students with differing levels of ability. Wu (2005) observed that classroom students were not uniform in skills, knowledge, ability, and no lesson can guarantee success for all students in all settings. Individualized instruction has been viewed as a way for teachers to customize the learning experience and has been defined as “a system that individualizes instruction by designing and programming specific learning tasks so individual learners can progress through the program at their own level of readiness and own learning rate” (Gutlek, 1983, p. 12). An effective individualized lesson has the following characteristics: (a) a predetermined lesson with specific objectives, (b) learning activities with a pre-written learning guide, (c) an evaluation of student performance, (d) instructor involvement, and (e) student responsibilities (Barlow, 1985). The customization of learning is found in other instructional delivery methodologies as well. These include mastery learning, learning centers, contracting, programmed instruction, and computer assisted instruction.

Mastery learning is an instructional delivery model developed by Bloom (1981) which specializes in producing customized lesson objectives that are “broken down” into small, individual units of study giving students the opportunity to master the concept being taught (Woolfolk, 1980). Bloom also acknowledged that all children are different. For example, children who are athletically gifted are inherently different from those that

are intellectually or musically inclined, and so the approach to learning should be different in order to help each child achieve his or her full potential (Bloom & Sosniak 1985). According to Bloom (1981), a proper mastery learning model needs (a) clear mastery objectives which are arranged such that previous and current knowledge can help with future learning and lessons, (b) a pre-established high passing standard, and (c) grading that is criterion referenced (Bloom, 1981). An example of mastery learning in the social studies would be introducing a lesson on the impact the Treaty of Versailles had on the origins of World War II. The teacher would have a pre-set number of standards students will need to show in order achieve success, or mastery. The instructor might use the following series of activities: (a) a pre-assessment exercise to establish a base of what level of background knowledge the students may have, if any; (b) a combination of didactic instructional techniques like streaming video clips, lecture, and discussion; (c) a drill of students concerning new vocabulary terms, and review of the material; and at an appropriate time (d) some kind of formative assessment, solely diagnostic in nature. Those students who reach a level of diagnostic success could continue on with an enrichment activity. Students who were unsuccessful would pursue a level of remediation, and an additional formative evaluation to see if an acceptable level of success is reached. At the end of the unit lesson, a summative evaluation would be administered in order to determine a grade (Esler & Sciortino, 1991).

Learning centers provide for another individualized instructional approach where the learning is more self-paced, either by the individual or a small group of students (Esler & Sciortino, 1991). The uses of this instructional approach can be wide and

varied, such as drill and practice, problem solving, or to encourage independent self-interest. Typically, however, the centers are focused on course content, and teachers often times will carry on with whole class activities while mixing in this form of independent, self-paced learning (Esler & Sciortino, 1991). An example of this style of instructional delivery in social studies would be appropriate for a teacher who wishes to introduce a variety of civilizations to the class. Students would have choice centers, as they would collect information about each of the civilizations from different sources, presenting their discoveries to the rest of the class. The teacher should have already prepared explicit and detailed instructions, complete with rubrics, for the students to follow for this lesson (Wilensky, 2004).

Contract learning is an instructional delivery model that enjoys a high level of adaptability to virtually any subject area and grade level. In contract learning, students have input into not only what they learn, but how, and to what extent they learn about a subject and what grade they should receive (Esler & Sciortino, 1991). The instructor makes clear and stipulates what tasks are necessary in order to receive a particular grade. Students are given ample time to consider what is being required of them and also are encouraged to provide input in the contract itself (Esler & Sciortino, 1991). One of the advantages of contract learning is that it has the tendency to motivate children to complete a lesson in subject areas in which they may not have a lot of motivation on their own (Biehler & Snowman, 1990). It increases the motivation by having distinct, attainable steps in which the instructor can give positive feedback and reward (Biehler & Snowman, 1990). Contract learning works best with students who are in higher levels of

education and does not work well with younger children (Esler & Sciortino, 1991). Critics of this method of learning, many of whom were teachers themselves, have indicated that students often times aimed for a grade which allowed them to barely pass, not the grade that they could, and should, have attained, and that this results in a poor quality of work (Esler & Sciortino, 1991). An example would be the teacher who creates a contract for the completion of assignments for the semester in a United States history class. The teacher and student would work together in creating a chart that contains achievable goals, indicating what assignments should be completed, and by what date. Each time an assignment is turned in, the date of the completion is recorded on the chart. The chart would show a positive progression, and each point of completion would reward the student with whatever external reward was previously agreed upon in the contract. The danger is that inappropriate external rewards can undermine students' motivation to learn (Woolfolk, 2001).

Programmed instruction (or learning) is an instructional delivery technique pioneered by Skinner, a behaviorist and one of the greatest psychologists of the 20<sup>th</sup> century (Biehler & Snowman, 1990). Skinner (1986) recognized that each child was unique, and learned at varying speeds and levels, thereby dictating that instruction should match the learner which would increase motivation and retention. Programmed learning was the instructional process by which subject content was divided into smaller pieces, allowing for immediate feedback and reinforcement as the student successfully completed each step of the lesson (Barlow, 1985). It was delivered either through written material or by machine (Biehler & Snowman, 1990). According to Skinner (1968), a

well-developed programmed instruction delivery should contain the following steps: (a) immediate reinforcement for a correct or incorrect answer; (b) close instructor monitoring; (c) student learning at own rate; (d) high rates motivation due to student success; and (e) multiple stop and start points. The advantage to programmed learning, when properly managed, was that it was self-paced and appealed to both slow and fast learners (Biehler & Snowman, 1990). However, critics contended that researchers had only been able to show slightly higher scores for students who learned in this fashion than those who learned using nonprogrammed methods (Biehler & Snowman, 1990).

Two reasons have been given for this weak effect: programs designed poorly and students who became frustrated and lost motivation as the constant response-feedback system become tedious to many students (Biehler & Snowman, 1990). An example of programmed instruction in the social studies could be the use of a computer program used for remediation of a failed course in civics that could be taken by students who attend summer school. One such program is called compass learning. It is a self-paced program that is divided into small, easily attainable steps for the student. Upon completion of the program, students would be permitted to retake the EOC examination for civics, and earn the credit needed to go onto high school. Though programmed instruction has not retained its universal popularity as a delivery method, researchers and educators still use it occasionally, carefully applying it where it is most valuable (Biehler & Snowman, 1990). Programmed learning has evolved, giving way in the 1980s to what may be perceived as the next step in self-paced learning, computer-assisted instruction (Biehler & Snowman, 1990).

Computer-assisted instruction is an instructional delivery method defined as the “use of computers to present programs or otherwise facilitate or evaluate learning” (Biehler & Snowman, 1990). It can be traced back to Skinner’s work in programmed instruction as a way to deliver instruction through the use of simple teaching machines, all of which have since been replaced by the computer (Biehler & Snowman, 1990). The advantages of computer-assisted instruction were numerous; most programs were self-paced and interactive, were highly personalized to fit any subject area and any student, especially students with disabilities. Additionally, computer-aided instruction provided one-on-one interaction with the student and provided immediate feedback to answers given. Finally, computer-aided instruction was used as a way of tracking and monitoring progress. It provided a level of privacy for students who were embarrassed about possibly giving an incorrect answer in a regular classroom setting (Biehler & Snowman, 1991). Though the use of computers as a method of instructional delivery has grown immensely, Skinner cautioned that the computer’s effectiveness as a teaching tool was dependent on its programming: “A good program of instruction guarantees a great deal of successful action. . . and in a well-designed instructional program, students gobble up their assignments” (Skinner, 1984, p. 125). Similar to programmed instruction, a teacher may use computer-assisted instruction to help in the social studies. For example, in a classroom with full inclusion, teachers may have students who are limited in their comprehension of English, yet may be proficient in their home language. These students could use the computer which covers the same content as the rest of the class. McGraw-Hill provides such a computer program in both Spanish and English for all middle-grades

social studies courses. Students can master the content at their own pace; at the same time they can make connections and improve comprehension in English as well (Jacobsen, Eggen, & Kauchak,2008).

Although the advantages of computer-assisted instruction were numerous as a tool in helping to facilitate learning, students lacked a level of interaction with instructors. In order to broaden and enrich the online or e-learning experience, a need existed for students to interact with facilitators. This was solved with the development of both hardware and software which made the transition from simple, computer-assisted instruction to e-learning possible (Jethro, Grace, & Thomas, 2012). With the invention of the internet and the modern sophistication of computer technology, the popularity of online education grew immensely in the first decade of the 21st century (iNACOL, 2012). At the time of the present study, online education was no longer just a simple computer program which tutored students who needed extra help with basic skills with computer-assisted instruction. It had evolved into entire self-paced courses that provide credit recovery programs for students who have failed courses and/or dropped out of school by allowing them to get back on track to graduate through blended learning with qualified instructional teachers (Lips, 2010). Over the past few years, growth in online education, or e-learning, has expanded the range of courses available to students, especially in small rural or inner-city schools and provided highly qualified teachers in subjects where qualified teachers were unavailable. In addition, online education has provided flexibility for students who face schedule conflicts, dropouts, migrant youths, pregnant students, students that were in jail and students who were homebound. Online

education allowed these students to continue their studies outside the classroom (Lips, 2010). According to iNACOL,

Gone are the days when it was assumed online learning was only for gifted students... Why? In a word: flexibility. Online learning allows students to choose when, where, and at what pace they want to learn, so personalization is possible in ways that, before few educators or students could imagine. (iNACOL, 2009, p. 4).

### The History of Multimedia and Online Instructional Delivery

The examination of distance learning was an important component in this study. Distance learning, though not new, has had an impact on how many stakeholders view the current trends in education. A somewhat curious analogy that has some bearing on the development of online education can be found in some educational trends, The concept of learning at home to learning in brick-and-mortar schools and the current popularity of home learning through online education has mirrored work trends of home/factory/home during the roughly same time period. In this section, the nature of the movement in education and work in the United States is reviewed. Also included is a chronological description of the development of correspondence study, the use of multimedia, and the Internet in distance learning.



*The Nature of Education: Home-School-Home Paradigms*

Similar trends to the past and current educational experience in the United States can be found in an analogy regarding the three phases of work trends in both the industrial revolution and information age. According to De Vries (1994), the first phase of work trends found in the industrial revolution of the 18<sup>th</sup> century began with home-based or community-based cottage industries where the work and production took place largely in people's homes. Prior to the Industrial Revolution, most products were custom-made by hand, and often times producers only worked part-time. Likewise, according to Cremin (1970), education for much of the Colonial era was provided largely for the purpose of reading the bible and was taught privately in homes and select schools such as church Sunday schools. Only in New England was there an attempt to establish public schools during the colonial era, but they were relatively few in number.

Additionally, in the early days, schools were not universal or compulsory. There were very few facilities for girls, and curriculum was limited. In the middle and southern colonies, very few public schools existed, as the plantation system was more suited for the hiring of tutors. If wealthy enough, parents would send colonial children to private schools. In larger cities, Sunday schools were used as platforms for rudimentary education, but were often times not originated for purely religious reasons, "the rudiments of reading and writing to the children who worked during the week with the added benefit of keeping them off the streets for the Sabbath (Cremin, 1970, p. 62).

The next phase found workers leaving the home during the second industrial revolution, as new production and managerial methods, e.g., scientific management, were

used to increase production and efficiency by creating factories and workplaces outside the home (DeVries, 1994).

Similar characteristics can be seen in the educational history of the United States as a whole in terms of student enrollment. With the adoption of the U. S. Constitution, education became a reserved power of the states as tax-payer funded public institutions. Students who were previously taught at home, were now physically attending a school for their learning (Cremin, 1970). According to Mondale and Patton (2001), much of the impetus for the creation of state-run common schools was be summed up by Horace Mann (2001), an American educational reformist and advocate of state-sponsored universal public schools. To Mann, many children grew up illiterate, uneducated, and ill-equipped to participate in the ongoing industrial revolution, which required a more technically savvy work force (Mondale & Patton, 2001). The purpose of universal public schools, according to Mann, was to create a public that was no longer ignorant and to pay for that schooling with the use of tax dollars. According to Mondale and Patton, Mann believed that this education should take place in non-sectarian public schools and be taught by well-trained professional teachers so that all children from all classes could share the same common learning experience and prepare them for a lifetime of future employment. With the passage of time, schools across the nation grew in size and scope to meet the needs of society in as efficient a manner as possible; as traditional brick-and-mortar schools with students attending classes in a face-to-face setting with instructor-driven curriculum (Herbst, 1997).

At the time of the present study, the third phase of work in the United States was being illustrated by people leaving the offices and factories in favor of working at home, largely due to the impact of the internet. Kaufmann (2012), a writer for *Forbes*, an internet entrepreneur and expert on the globalization of the economy, has expressed the belief that conventionally held ideas of both workplace and work-times have been changing. The American workforce has been undergoing yet another change, increasing production and efficiency in an ever-evolving global economy; but this time both work and production are being performed by workers from their homes. Correspondingly, in the past 20 years, public school students have been increasingly leaving classrooms in favor of learning which takes place once again in the home. For example, a decade ago, fewer than 400,000 students were enrolled in online courses, but in the 2010/2011 school year; nearly two million students were enrolled in at least one online course (iNACOL, 2012).

### *Correspondence Study*

The history of distance learning, or distance education, can be traced as far back as ancient times, though in the United States, the first documented correspondence-style coursework was from a 1728 advertisement printed in the *Boston Globe* when a teacher, promised to teach shorthand to students via weekly lessons by mail (Holmberg, Hrsg, & Busch, 2005). In the middle 1800s, teachers and even private schools began to advertise educational courses through mail on a regular basis. For example, in 1837 Sir Isaac Pitman developed an entirely mail-based curriculum to teach the Pitman Shorthand to

anyone that would pay for the materials, postage, and a small fee of one penny (Moore & Kearsley, 2011). According to Pittman (2003), William Harper, president of the University of Chicago was largely credited for helping to establish the first correspondence school from a regular school institution in 1892. The United States was not the only country during this time period to develop a correspondence-based educational system. Australia and Canada, both countries that had remote populations separated by long distances, developed state-sponsored educational correspondence courses that taught elementary, middle, and high school students, and even trained teachers. The government of Australia went so far as to develop the Department of Correspondence Studies in 1911 (Sumner, 2000).

According to Kett (1994), education as a whole expanded greatly during the progressive era of the early 1900s with the explosion of high schools and colleges all throughout the United States. As stated by Kett, correspondence schools were no exception, especially when it came to addressing the needs of American students in largely rural areas, where spending much-needed money to build physical schools was not practical. Stakeholders looked to correspondence schools for answers in meeting the needs of (a) students who were far removed from population centers, (b) fathers who looked for ways to advance or change careers, or (c) dropouts looking for ways to gain an education to become more marketable in a rapid and technologically changing world (Kett, 1994).

According to Clark (1906), an educational historian, correspondence schools offered flexibility insofar as time was concerned, but that flexibility came with a price.

Coursework was very focused, as opposed to the broader, more liberal education one could expect to receive at a regular school or college (Clark, 1906). In addition, many students struggled due to various inadequacies such as the lack of positive feedback and encouragement, a lack of reading comprehension, or a lack of self-discipline required which led to dropping out because they underestimated the level of difficulty. As evidence, less than 3% of the students who began a course actually finished it (Clark, 1906).

New technologies that were developed in the first half of the 20th century led to the demand for more technologically savvy workers and the need for more schooling options, most of which were found at the secondary education level and targeted mostly adults. Correspondence schools, both public and private, began to experiment with new audio-visual delivery systems such as radio and television to reach a broader audience and enhance the learning experience.

### *Distance Education Through Multimedia*

According to Lee (2008), an educational historian, incorporating multimedia education became popular with the advent of new technology in the early part of the 20th century. At first, correspondence education was largely viewed as a way to enhance the individualistic aspect of education during the progressive era. An increase in the numbers of correspondence students, however, led to additional demand to make distance learning individualized on a much larger scale and what was referred to as “mass-individualism on a massive scale” (Lee, 2008, p. 249) while at the same time improving

existing instructional delivery designs in order to provide a more complete learning experience.

Lee (2008) asserted that the evolution in distance education came in the form of multimedia as a response to the increased demand of correspondence students in the early part of the 20th century. Lessons and materials in printed format were not forgotten but were integrated into the educational experience. In 1922, Pennsylvania State University was the first school to offer courses through the radio, and just three years later by 1925, over 200 universities and colleges were granted broadcasting licenses (Reid & Day, 1942). The increase in the use of radio as a way of disseminating information was an example of the trend of automation through scientific measurement, yet problems remained in individualizing education for each learner as progressives advocated. One of the answers to the problem of differentiation and individualized educational experiences, according to Lee, was to create a myriad of choices of coursework with a vast spread of offered classes ranging from “courses in horticulture and home economics, to textile fabrication, refrigeration and the history of European civilization. . . aimed at professionals, workers, school children, university students, military men, house wives, or seamen” (Lee, 2008, p. 248). Wilhelm, in closing a 1950 conference session focused on correspondence courses, stated the problem clearly: “Every lesson should be adapted to meet precisely the interests, talents, and needs of the individual pupil” (Lee, 2008, p. 247). The harsh reality about distance learning was that much of the formatting followed an already established plan of presenting previously prepared lesson plans with students demonstrating competency by the successful completion of a test (Lee, 2008).

Problems, however, remained in how to address the characteristics of learners and the needs of students. One distance learning issue related to the problems in communicating via distance that were not present in face-to-face dialogues. According to Bizhan Nasseh (1997), the concept of education by radio was a major reason for development of educational television by the mid-20th century. One of the events that furthered the acceptance of television as a medium of distance education was a 1956 study funded by the Ford Foundation which focused on the application of television instruction in conjunction with correspondence study. Researchers concluded that television instruction was not a method. Rather, it was a means by which instruction could be transmitted from one location and reach an unlimited number of students without having to be in a physical, brick-and-mortar school. Additionally, according to Nasseh (1997), no significant difference was found in the performance of regular students who were taught by means of television transmission or by a combination of correspondence study and television.

Nasseh (1997) reported that in 1960, the Correspondence Education Research Project was developed with the hope of conducting further research and better defining the status of distance education in the United States. In 1968, the name of the organization was changed to the Division of Independent Study, and its purpose was to create new ways for the delivery of instruction through media like video, television, and telephone (Nasseh, 1997).

King (2008) defined educational television using the following words of Diamond: “. . . broad term usually applied to cultural and community broadcasting which

may include some programs for instructional use” (p. 59). Specifically, television instruction was defined as “television used in the formal classroom context on any educational level” (King, 2008, p. 59). This definition went hand-in-hand with the research performed by Seibert, a professor in the field of instructional television at several universities. Seibert, according to King, studied the effects of education by broadcast and instructional television, i.e., *Studies in the Utilization of Television in the Schools*, and *International Seminar on Instructional Television*.

King (2008) also reviewed the work of Lane, another pioneer in the use of instructional television, who began the Stoddard Plan experiment in 1958. Lane used televisions in the classroom as a way to meet teacher shortage needs in large classroom settings. During her tenure with the state of Kentucky, she went on to produce 300 televised lessons and trained 200 teachers in educational television through a Ford Foundation grant. Because of her work, the United States Congress called upon her expertise to streamline and modernize the American Samoa educational system, upgrading it to a televised-based curriculum (King, 2008). On the state level, Crabbe worked to produce and develop educational television in school districts and in state-wide colleges and post-secondary level courses. Much of the work by instructional television pioneers like Seibert, Sikes, and Crabbe resulted in creating an environment where television was utilized not only in the classroom but specifically in distance education (King, 2008).

On the national level, the National Education Association established its first multimedia department, the Department of Visual Instruction (DVI) with the intent to



become the pre-eminent voice for the national audio-visual movement (King, 2008). During World War II, the DVI developed and produced audio-visual materials for military training. Duncan (2005), a researcher for East Carolina University, reported that the United States military had used correspondence courses for decades for skill training, cross training, and for receiving credit for promotion within the military (2005).

In 1968, the Carnegie Foundation helped to establish the Articulated Instructional Media Project (AIM), which laid the groundwork for the Open University concept, whereby institutions of higher learning would deliver formal, accredited coursework via broadcast media such as radio or television (Daniel, 1999). Great Britain was the first country to utilize the Open University concept, offering televised courses in 1971. This was thought to be a way to offer post-secondary education to a much larger audience without the constraints of classrooms, timeframes, and national boundaries. The Open University was designed to make degrees more accessible to learners who were unable to attend traditional programs, campus-based courses (Lee, 2008). The United States quickly followed with New York's Empire State College starting master's and doctoral degree programs in the latter months of 1971. As programs in both Great Britain and the United States prospered, countries such as Spain, Australia, Germany, and Canada also started their own Open Universities, with many being labeled Mega Universities. Daniel wrote in 1999 that enrollments in some of these mega-universities had an excess of over 100,000 students. According to Daniel, the connection between the multimedia delivery methods found in Open Universities and the use of computers started with Luskin, a pioneer in e-learning and an expert in global learning. Luskin helped to launch the

Coastline Community College which blended Computer Aided Instruction (CAI) with tele-courses in 1976 as a college without walls (Daniel, 1999).

### *Distance Education through the Internet*

The delivery of online distance education has many overlapping terms, e-learning, technology-enhanced learning (TEL), computer based training (CBT), internet-based training (IBT), web-based training (WBT), virtual learning environments (VLE), and cyber-learning. Although each designation has had some specific connotations associated with it, the term, e-learning, has been used in this review unless a particular reference called for a more specific term. This section addresses several aspects of e-learning and its impact on distance education including the following: (a) the definition of the three generations of distance education, (b) the three types of instruction, and (c) the history of online education.

#### Three Generations

According to Bates (2007), CEO of Skype and expert and author on the history of internet technologies, distance education has gone through three generational evolutions. The first generation, print-based correspondence education, was correspondence schools in print with a one-directional format whereby teachers presented information and students responded by completing tests sent through the mail. Making solid inferences of student learning was difficult with this format of distance education because student-teacher interaction was limited to correspondence sent through the mail. The second

generation of distance education included the use of multimedia and added the dimension of audiovisual technologies. This generation, the industrial mode, according to Bates, included the use of a specialized division of labor by either radio or television transmission or a combination of both to present materials to a vast and wide audience through the Open University system. Bates noted that this generation was marked by large numbers of students in the post-secondary level and a heavy involvement of state-sponsored initiatives. Additionally, this style of distance education retained the previous one-directional format of correspondence schools whereby communication occurred between teacher and student largely through letters in the mail. The addition of the internet marked the next evolution, or third generation, of distance education. According to Bates, the internet changed distance education from a one-directional delivery system to one that added the dimension of true student-teacher interaction and allowed for immediate correction and feedback. The level of interaction between teacher and student was enhanced through the development first of email followed by instant-messaging concepts found in most modern online educational computer program platforms. According to Bates (2007), teacher-student interaction achieved levels of interconnectedness never before thought possible with the advent of social media programs like Twitter, Facebook and Youtube.

## Types of Instruction

Online education has typically taken one of three styles of instruction: synchronous, asynchronous, and blended or hybrid. According to Johnson (2007), a University of Otago professor, asynchronous is an online educational delivery method which allows students to work on required coursework on their own time schedules and usually has weekly deadlines. Asynchronous has expanded in the past few years to utilize newer online technologies such as blogs and discussion boards. Johnson stated that this style of online learning was advantageous for students who could not or did not wish to attend school in a traditional setting, e.g., children with health issues and students desiring to work on their own time, and proceed at their own pace in a “low-stress” environment. Another advantage of asynchronous is the ability to review a lecture without the possible fear of embarrassment from other students if there is still a question. This style of online learning relies on students who have a high level of self-motivation, and also have the ability to communicate effectively through writing (Johnson, 2007).

Johnson (2007) defined synchronous learning as an online educational delivery method that is less flexible than asynchronous, as it requires real-time commitment with conference calls, online chatting, and teleconferencing. The bonus is that the real-time interaction allows for the exchange of thoughts and ideas, and therefore, a higher level of interconnectedness where the teacher can provide immediate feedback (Johnson, 2007). This form of online learning also requires someone who is highly motivated and has the ability to commit to a less flexible schedule than an asynchronous format. An example of

a synchronous face-to-face discussion in modern terms could be the use of Skype, a chat room, or a virtual classroom where real-time collaboration is possible.

The third type of online learning is called blended or hybrid. The International Association for K-12 Online Learning [iNACOL] (2012) defined blended or hybrid learning as an online educational delivery method commonly used to describe courses in which some traditional face-to-face seat time has been replaced by online learning activities. iNACOL estimated that nearly two-thirds of all school districts currently offer some form of online or blended programs. Hybrid learning can be expensive on the K-12 level, as this style of learning requires having enough computer access for all students to be able to complete their work. An example of a hybrid format in the K-12 setting that seemed to be working was found at the Carpe-Diem Collegiate High and Middle School in Yuma, Arizona (Baker, 2013). At this school, students spent four days a week, spending half of the day working with teachers in small-group settings and the second half of the day working online. Baker indicated that the format allowed students greater flexibility to target areas of weakness or, for those who wish to do so, to move ahead at their own pace.

Though there have been plans to duplicate the model in Indiana and Nebraska, hybrid learning is not without its critics. Baker (2013) wrote of Safier's fears that states may look towards this model as an excuse to increase the ratio of teachers to students despite the lack of hard evidence and data from this new school.

## History of E-learning

The history of e-learning goes back over a half century to 1960, where the University of Illinois implemented coursework that utilized computers that communicated with each other. Woolley (1994), a pioneer in online and social communities asserted that Bitzer, along with some other U of I staff and students, wrote the Programmed Logic for Automatic Teaching Operations (PLATO) software, the first computer-assisted instruction system. Additionally, Woolley claimed that PLATO was networked to a dozen computer mainframes all over the world which allowed for a level of multi-user computing. Many of the modern concepts of computer-level communication like forums, bulletin board systems, online testing, email, chat rooms, picture languages, instant messages, remote screen sharing and multiplayer games were originated and developed on PLATO platforms (Woolley, 1994).

In 1963, Luskin, an internet pioneer and important advocate for computers in higher education, worked to develop a computer-assisted instruction consortium of all community and state colleges in California. This allowed a level of student interaction that set the stage for the next evolution of e-learning, the blending of computers with post-secondary education (Harasim, 1990). According to Harasim (1990), this blending of computer-assisted instruction of the 1970s with new programming capabilities which allowed for an early form of e-learning, computer-mediated communication, pioneered by Turoff at the New Jersey Institute of Technology. At first, e-learning was aimed mostly towards higher education with schools like the University Of Phoenix and the

New York Institute of Technology launching early forms of virtual campuses in the first half of the 1980s (Harasim, 1990).

The drawback of these forms of early online distance education schools was the fact that the courses offered were very similar to those found in typical correspondence schools. The curriculum was largely one-directional, and there was little opportunity for either student or instructor to communicate or receive feedback (Harasim, 1996). The computer and specifically internet technology continued to advance in the 1990s with the release of the World Wide Web to the public. The popularity of e-learning, e.g., online education, grew considerably with the introduction of a new way to offer courses online. In 1996, the United Kingdom's Open University revolutionized the way courses were taught by starting an online learning environment system called WebCT where users interfaced with each other; teachers communicated with students, and students could communicate with each other either through email or by instant messaging. WebCT, purchased by Blackboard Learning, became the foundational basis for most online schools and courses available at the time of the present study. In 2013, it was used by over 20 million students worldwide in over 30 different countries (Blackboard Learning, 2013).

The newest trend in online distance education came in the form of massive open online courses (MOOCs). Developed in 2008, MOOCs became the latest addition to online education, with the *New York Times* naming 2012 as the Year of the MOOC (*New York Times*, 2013). The purpose of MOOCs was to allow for interactive participation of online students on a massive scale where forum members interactively develop an online

learning community. These used traditional methods of online instruction but also incorporated open licensed video and blended in access through social media sites like Facebook, YouTube, Instagram and Twitter, all of which were free of charge (Daniel, 2012). Several large accredited universities, e.g., Harvard, Stanford, the University of California, Berkley, and the Massachusetts Institute of Technology, have joined educational technology companies like Udacity and Coursera to offer free undergraduate courses which had enrollments in excess of 160,000 per class (Wikipedia, 2013). As of the writing of this review, EdX was poised to announce a partnership with Google in order to create a large scale platform, MOOC.org, in which to help universities, institutions, businesses, governments and teachers to host their courses for an audience that can span the entire world to be launched sometime in the first half of 2014 (EdX, 2013).

### An Analysis of High-stakes Testing

High-stakes testing and End-of-Course examinations were two important dimensions of the rationale for this particular study. According to Meisels, he defined high-stakes testing as “The use of readiness or achievement tests for the classification, retention, or promotion of students” (Meisel, 1989, p. 17). Additionally, Seifert and Sutton (2009) wrote that the test poses more than just important consequences for the test taker. The current direction of national trends includes much broader implications, most notably how high-stakes tests are not only used to assess student achievement but also for teacher accountability purposes.



### *History of High Stakes Testing*

This section of the review contains a brief account of the history of and current issues surrounding high-stakes testing as it pertains to public education in the United States. Also presented are some examples of the growing body of research, both past and present, which explain the relationship high-stakes testing has with accountability and student achievement from multiple perspectives. Also in this section of the review is an analysis of End Of Course (EOC) examinations for the state of Florida, its history, what subjects that EOC examinations cover, and all pertinent information regarding the state of Florida United States history EOC examination in particular as detailed by the Florida Department of Education (FLDOE).

According to Nichols and Berliner (2007), high-stakes testing can be traced as far back as ancient China, where students would study for years to pass tests in order to be eligible for employment as civil servants. For the past century, tests like the Iowa Test of Basic Skills (ITBS), the Stanford Achievement Test (SAT) and the Metropolitan Achievement Test (MAT) have been developed to assess characteristics such as aptitude, intelligence, and achievement (Nichols & Berliner, 2007). Nichols and Berliner noted, however, that the modern emphasis on testing has changed from one that simply assesses students for minimum competency to an emphasis of evaluating teachers, administrators, schools, and districts. These researchers also observed that testing was being used as a lever to drive educational policy.

According to Nichols and Berliner (2007), this current trend can be traced back to the authorization and implementation of the 1965 Elementary and Secondary Education

Act (ESEA). The implementation of this law was a reaction to the concern America had over the threat that the Soviet Union and the spread of communism across the globe, highlighted with the sensationalized launch of Sputnik. Concern for the United States educational system did not stop there. Nichols and Berliner indicated that the 1983 *Nation at Risk* report startled the nation into a renewed sense of urgency to fix America's failing schools and resulted in a flurry of legislation based on "mistaken factual claims" (p. 4) made in *A Nation at Risk*. Among broad educational policy changes in the 1980s and 1990s, as a result of the report, was an increase in the demand for more tests (Nichols & Berliner, 2007).

With the passage of No Child Left Behind Act of 2001, a system of accountability was established by the federal government that required states to adopt a system to annually evaluate students, teachers, administrators, and even schools (NCLB, 2001) that was based on students' performance rankings on standardized tests. Schools were either rewarded or punished in accordance with annual gains and achievements based on test score performances placed into some form of accountability matrix that were to be determined by each state (Nichols & Berliner, 2007).

According to the Partnership for Assessment of Readiness for College and Careers [PARCC] (2014) website, additional funding was to be awarded to states that used assessments based on Common Core standards with the passage of the Race to the Top initiative that was signed into law in 2009. At the time of the present study, such standardized assessments were currently being readied for full implementation in the 2014-2015 school year by the PARCC consortium. These assessments were computer-

based assessments for mathematics and language arts focused on preparing students for readiness to either enter college or directly into a career in the workforce (PARCC, 2014).

### *Issues Surrounding High-Stakes Testing*

Ravitch (2011), former assistant secretary of education and an authority on high-stakes testing, acknowledged the heated debate and numerous issues surrounding high-stakes testing. She claimed that the tests hold major consequences for not only test takers but also stakeholders, e.g., parents, teachers, schools, districts, and even state and federal departments of education. Similarly, the American Psychological Association [APA] (2014) noted that the concept of high stakes was not part of the characteristics of a test but rather the consequences that are placed on the results of the test.

In addition to Ravitch's assertion, Au (2007), a California State University professor, stated high-stakes testing does not always correctly measure the individual's knowledge or skills and the results could be misused. For example, a test might claim to be a general reading test, but the test might be designed such that it actually tests if an individual had read a specific book (Au, 2007). Test misuse can also arise when data from tests are incorrectly or inaccurately used to compare student performance. Similarly, as earlier determined in *Debra P. v. Turlington*, 1981, the state of Florida was prohibited from using a particular high school graduation test because black students, who were segregated and forced to attend inferior schools, were held accountable for material that was not taught yet was covered in the test. Although the use of tests that

include material not taught to students may be appropriate if used to assess if schools were doing their job, holding students accountable in such a case is inappropriate (Au, 2007).

According to the National Research Council (1999), public perception on American schools are based on “personal and anecdotal” observations and from tests results that are published, such as the National Assessment of Educational Progress (NAEP) and Trends in International Mathematics and Science Study (TIMSS). NAEP and TIMSS are tests designed to provide an overview, and not to measure the performance of individual students, as testing populations change from year to year and state to state. Additionally, according to the National Research Council, tests like the SAT and ACT, which have been used for college entrance, are decision-making tools. They do not provide accurate information about overall levels of academic achievement, as the sample of test takers is far from representative of a district, a state, or the nation as a whole. The danger in these forms of test misuse is that stakeholders and policy makers can make poor educational policy decisions based on bad or inaccurate data. History provided an example of this when in 1917 the president of Columbia University used the *Thorndike Tests for Mental Alertness* to refuse Jewish students admittance (Crouse & Trusheim, 1988). Another example can be seen in a 1997 lawsuit *People Who Care v. Rockford Board of Education* where testing was used to move black students to lower tracks even when there were white students with lower test scores who were assigned to higher tracks.

Another issue was discussed by the APA in a 2000 summary report which claimed high-stakes tests were often presented in long, single examinations which could cause test anxiety and stress for some students. Indeed, in a similar report issued in 1999 by the National Research Council's Committee on Appropriate Test Use, it was stated that it is difficult if not counter-productive for a child to take a test which could be hours long and who cannot sit still for 10 minutes and expect an accurate and authentic assessment of said child's progress.

High-stakes tests have also been blamed for the "narrowing" of curriculum (Au, 2007). As published by the National Research Council (1999), one of the unintended consequences of high-stakes testing has been the school's responsibility for remediation, and preparation of students in order to pass more complex and complicated curricular concepts in areas like reading and math, at the detriment of other, non-essential courses like art and sports. In addition to adjustments to educational programs on the school level, the National Research Council found that if the accountability stakes were high enough, teachers would change the curriculum and instruction to reflect the curriculum content that was assessed by the test.

Assessments like high-stakes tests are used for a variety of reasons. Tests can be used to measure student achievement, provide feedback, identify problems in learning, inform administrators and the public about the overall state of learning and use of tax dollars, and they can be used for placement and promotion. Such tests, however, are being used for teacher, school, and district accountability and are also being used to drive educational policy (Au, 2007; Nichols, Glass, & Berliner, 2012; Ravitch, 2011). Instead,

APA (2000) recommended that “Any decision about a student’s continued education, such as retention, tracking, or graduation, should not be based on the results of a single test, but should include other relevant and valid information” (p. 2).

*The Relationship of High-Stakes Testing to Accountability and Student Achievement*

The topic of high-stakes testing and the relationship it has on student achievement and accountability has been a highly controversial one. According to a 2003 study by Greene, a senior fellow at the Manhattan Institute for Policy Research who studied the relationship between scores of students who took both low-stakes tests and high-stakes tests, results were similar nationally. Greene defined a high-stakes test as one that was used for accountability purposes; low-stakes tests were not used for accountability purposes. He found that there was a “very strong” adjusted average correlation (0.88) in student achievement, and a moderate average correlation (0.45) between year-to-year score gains of high-stakes tests compared to low-stakes tests. The state of Florida, specifically, had the strongest correlation of 0.96 between the high-stakes and low-stakes test scores and a moderately strong 0.71 correlation between the year-to-year gains of high-stakes and low-stakes test scores. Greene concluded that if low-stakes tests had a similar outcome to that of high-stakes tests, the stakes attached to high-stakes tests were not distorting the test outcomes. Therefore, in his opinion, high-stakes tests accurately measured students’ true levels of achievement and could and should be used for reliable policy tools (Greene, 2003).

Evidence that high-stakes testing might not equate to real learning gains was shared in the late 1980s in Cannell's highly criticized 1987 study. Cannell claimed that throughout the 1970s and 1980s, districts were reporting steady gains in norm-referenced test scores. In 1988, all 50 states reported testing that was above publishers' national norms. Cannell dubbed this phenomenon the "Lake Wobegon effect" where somehow, all 50 states were above average (Cannell, 1989; Shepard, 1990). Much of this effect, according to Cannell, was due to factors like inaccurate initial norms, and teachers that affected test results by teaching to the test. All of which, he claimed, did not really reflect true learning gains, just the more savvy test-taking skills of those being tested (Cannell, 1987). According to a 1990 report by Shepard, however, educational policymakers and lawmakers dismissed the report, claiming that his data were wrong, and that he did not understand the statistics that he used, thus invalidating his conclusions as to why the test scores were high. She believed that the high achievement found in all 50 states was authentic and realistic (Shepard, 1990).

Linn, in a 1990 study, examined the relationship between standardized testing and student achievement, revealing findings mirroring those of Cannell. Linn also determined that all 50 states claimed that they scored above the national average on standardized tests. In addition, Linn also questioned many of the state's performance claims. It was discovered that some of the test performance increases could be accounted for as a function of both preparation performed by the teachers prior to the test and the level of familiarity that students had with the style and type of tests used (Linn, 1990). Though Linn confirmed Cannell's basic conclusions that test scores were exaggerated,

these findings, according to Shepard (1990), were largely ignored by stakeholders and more importantly, policymakers.

In a 2001 study of high-stakes tests from 18 states, two researchers from Arizona State University's College of Education found that data from high-stakes testing programs appeared distorted and corrupted, and no discernable evidence proving that high-stakes tests reflect true student knowledge and learning (Amrein & Berliner, 2002). In their study, they looked specifically at a number of standardized tests: the ACT, SAT, NAEP and AP tests.

Nichols et al. (2012) repeated a study they originally conducted in 2006 which used an Accountability Pressure Index (API) to look at the relationship of high-stakes testing and student achievement as measured by the National Assessment for Education Progress (NAEP) of 25 states in both reading and mathematics. In both the 2006 and 2012 studies, Nichols et al. were consistently able to find a strong positive correlation between the pressure index and the NAEP scores in mathematics, and a weaker correlation between the pressure index and the NAEP scores in reading (Nichols et al., 2012). Ultimately, their research showed that the impact of accountability-based policies from high-stakes testing reflected in student achievement was, in their words, "varied, limited, and relatively inconclusive" (p. 26). They continued with a possible explanation that though it was possible that teachers had become more efficient in training students for the test, it was difficult to isolate any sort of causal effect between policy implementation and student achievement (Nichols et al., 2012). They summarized their findings as follows: ". . . it becomes more difficult to disentangle the effects of pressure



on student ability to take tests from pressure that genuinely affects student learning” (Nichols, et al., 2012, p. 27).

According to Nichols and Berliner (2007), another perspective concerning the impact high-stakes testing had on not only student achievement, but educational policy as a whole, was reflected in a social science law commonly referred to as Campbell’s Law. Campbell’s Law was a social science principle developed in the 1970s by Donald T. Campbell. According to Nichols and Berliner (2007), Campbell’s Law stated that achievement tests were useful when they were used for what they were originally designed for: the normal conditions of teaching and general competence. The problem arose, they stated, when used improperly to assess teachers, administrators, and districts. When they were also used as levers of educational policy, however, the data were distorted in undesirable ways. They stated that the pressures of high-stakes testing eroded the validity of said test scores, which in turn destroyed the validity of the very things (teacher, administrator, and school assessment) that tests were being used to assess and measure (Nichols & Berliner, 2007).

Nichols et al. (2012) found the purpose of high-stakes testing was to improve student and school performance and achievement by using the tactic of attaching negative consequences through penalties to standardized test performances. Nichols et al. stated that this logic was predicated on the hopes that in the face of negative consequences, low performing schools and students would work harder and more effectively and efficiently to increase scores, thereby avoiding negative consequences.

## State of Florida Legislative Mandates Concerning End-of-Course Examinations

End-of-Course (EOC) examinations were central to the theme of this study, and a close examination of Florida legislative mandates concerning EOC examinations was appropriate. This section provides an account of the chronological development of End-of-Course examinations in general as a construct of the Next Generation Sunshine State Standards (NGSSS). Also addressed are the specific guidelines that were used in the design and construction of the United States history EOC examination.

### *Legislative Development of End-of-Course Examinations*

According to the Florida Department of Education (FLDOE), there were two realities which put pressure on overhauling and reevaluating Florida's Sunshine State Standards for United States History. First, in 2005, outside consultants were hired to review the original standards passed in 1996. Their findings suggested a greater amount of specificity was needed in the benchmarks to clearly guide teachers as to what they were supposed to teach. Second, the federal legislation of the No Child Left Behind Act of 2001 held that both schools and districts were accountable for what, and how well each child was learning.

In the spring of 2008, pursuant to F. S. A. 1003.41, the legislature adopted the next generation of educational standards, commonly referred to as the Next Generation Sunshine State Standards (NGSSS). Specifically, F. S. A. 1003.41(4)(b) focused on the social studies, detailing what courses were to be taken in the k-12 curriculum, and how they were to be organized. The following courses were mandated; geography, United

States and world history, government, civics, economics, and humanities (F. S. A. 1003.41, 2008). Those courses, were to be “rigorous and relevant and provide for the logical, sequential progression of core curricular content that incrementally increases a student’s core content knowledge and skills over time” (F. S. A. 1003.41{4}{d}). In addition, F. S. A. 1003.41(4)(e) stated that these courses were to integrate “critical thinking and problem solving skills, communication, reading and writing skills. . . collaboration skills; contextual and applied learning skills; technology-literacy skills; information and media-literacy skills; and civic engagement skills”. It was determined that the State Board of Education was to devise these new standards in time to be adopted by the legislature by December 31 of 2011 (F. S. A. 1003.41{2}).

In 2008, the Florida legislature also adopted the Student Assessment Program for Public Schools. In this statute, the Florida legislature mandated that the purpose of an assessment program was to “provide information needed to improve the public schools by enhancing the learning gains of all students and to inform parents of the educational progress of their public school children” (F. S. A. 1008.22). To accomplish this, the statute continued by stating that the state was to, “. . . assess the annual learning gains of each student toward achieving the Sunshine State Standards appropriate for the student’s grade level (F. S. A. 1008.22{b}). There were two primary reasons for this. First, the state wanted to be able to identify the educational strengths and needs in order to determine promotion and graduation status (F. S. A. 1008.22{c}); the second purpose was to assess “. . . how well educational goals and curriculum standards are met at the school, district and state levels” (F. S. A. 1008.22{d}). The reasoning for the measuring

of goals and curriculum standards was to improve the development of “educational programs and policies” (F. S. A. 1008.22{e}) and the purpose of assessment was also to determine the performance of Florida students as compared to the rest of the country (F. S. A. 1008.22{f}). To that end of assessment and accountability, the state of Florida Board of Education was charged with developing and implementing a student achievement testing program called the Florida Comprehensive Assessment Test (FCAT) to “measure a student’s content knowledge”, and this test was to be administered as a comprehensive exam first in the reading and mathematics “annually in grade 3 through 10” (F. S. A. 1008.22{c}).

2008 was also the first year that the Florida legislature mentioned an End-of-Course (EOC) examination. In statute F. S. A. 1008.22(c), EOC examinations may be administered in addition to the assessments given annually in the FCAT. EOC examinations were to be aligned with Common Core content established in the NGSSS. The commissioner of education was charged with establishing EOC examinations that “measured student skills and competencies” adopted by the NGSSS (F. S. A. 1008.22{c}{1}). Testing for EOC examinations was to be developed by contract either through approved private or public vendors, postsecondary institutions, school districts, or a collaboration thereof. Tests were instructed to be criterion-based, aligned with Common Core standards, and worded in such a way that the assessment(s) could be measureable (F. S. A. 1008.22{c}).

In 2009, F. S. A. 1008.22 (c) stated that “A statewide standardized end-of-course assessment is to be administered within the last 2 weeks of the course”. Additionally, in

2009, the Florida Commissioner of Education was charged with developing and designing, based on collaboration and input from school districts, student testing programs for end-of-course examinations based on a “field testing process and psychometric analyses. . . and an evaluation or determination of the effect of test items on such. . .” (F. S. A. 1008.22{c}).

In 2010, the FLDOE, in accordance with the passage of the Next Generation Sunshine State Standards (NGSSS) found under Title XLVIII in F. S. A. 1008.22(2)(a) that students were mandated beginning in the 2010-2011 school year to take a certain number of EOC examinations in order to achieve promotion to the next grade level or graduate high school (2010). In addition, EOC examinations were to be aligned with Common Core standards as set by NGSSS, and were to be comprehensive, rigorous and standardized (F. S. A. 1008.22{2}{a}). According to F. S. A. 1008.22(2)(a), these EOC assessments were to be administered statewide in addition to FCAT assessments. A passing grade in the Algebra I EOC was necessary for course credit. In Geometry, the EOC was to account for no more than 30% of the overall course grade. Additionally, the 2010 statute instructed the Florida Commissioner of Education to “Provide technical assistance to school districts in the implementation of state and district testing programs and the use of the data produced pursuant to such programs” (F. S. A. 1022{14}{e}), to “Study the cost and student achievement impact of secondary end-of-course assessments, including web-based and performance formats, and report to the Legislature prior to implementation” (F. S. A. 1022.14{f}).

Beginning in 2011, the Florida legislature added another provision to statute 1022.2(a)(2). An EOC examination for Biology I was mandated to be implemented starting the spring of the 2011-2012 school year. Students enrolled in Biology I starting the 2012 school year would be required to earn a passing score on the Biology I EOC examination in order to receive course credit for completion (F. S. A. 1022.2{a}{2}). Secondly, the legislature also passed a similar mandate for civics. During the 2012-2013 school year, an EOC assessment shall be given to students in the middle school level (F. S. A. 1022.2{a}{2}). At first, the field test was to account for only 30% of the overall grade, with a passing score on the EOC examination starting with the 2014-2015 year required for course completion, and promotion onto high school (F. S. A. 1022.2{2}). Future legislation has been expected to rescind the pass/fail for course credit component of the civics EOC examination, and return the seventh-grade civics EOC examination to no more than 30% of the overall grade for the course. Additionally, in 2011, subsection “d.” of Florida statute 1022.2 stated that “. . . Contingent upon funding provided in the General Appropriations Act. . . the Commissioner of Education shall establish an implementation schedule for the development and administration of additional statewide, standardized end-of-course assessments in English/Language Arts II, Algebra II, chemistry, physics, earth/space science, United States history and world history” (F. S. A. 1022.2{d}). The previous wording of the statute was amended in 2012 to include the following additional subsection, “Beginning with the 2014-2015 school year, all statewide and end-of-course assessments shall be administered online” (F. S. A. 1022.14{g}).

The primary purpose of EOCs was to increase student achievement and improve college and career readiness (FLDOE, 2014). The state of Florida EOC examinations were designed to provide student achievement and learning gains data to stakeholders like students, parents, teachers, school administrators and district personnel (FLDOE, 2014). The intention was to use the data for multiple reasons: (a) to drive better instruction, (b) to compare state and national norms, and (c) to allow the public to assess the cost benefit to the expenditure of taxpayer dollars. According to the Florida legislature, data from EOC examinations that assess their achievement levels provides for measuring students' strengths, weaknesses, and needs. Data from the EOC examinations are also used in making decisions regarding school accountability by assessing if educational goals and curriculum standards have been met by the school and the district at state and national levels.

At the time of the study, NCS Pearson, Inc. was the contractor responsible for the development, administration and scoring of the EOC examinations (FLDOE, 2014). According to the FLDOE EOC Test Administration Manual (2014), PearsonAccess.com was the website where all test preparation, setup, and administration occurs. It is also the website where reporting tasks are developed. In addition to computer access, students taking EOC examinations are required to use their student IDs and passwords for access to the examination itself. At the writing of this study, for the beginning of spring 2013, private school students may also participate in the Algebra I, Biology I, and Geometry EOC assessments. No provisions yet are available for the 7<sup>th</sup>-grade Civics, and the 11<sup>th</sup>-

grade United States history EOC assessments for private schools (FLDOE EOC Test Administration Manual, 2014).

The Next Generation Sunshine State Standards were adopted by the Florida State Board of Education on July 27, 2010 (FLDOE, 2014). The new standards for Language Arts, History, Social Studies, Science, and other subjects like reading and writing were designed to align state standards with Common Core standards. It was these standards upon which the state of Florida EOC examinations have been based (FLDOE, 2014).

EOCs in the state of Florida were designed to measure student achievement and determine competency for specific courses. Spring, 2011 marked the first administration of EOC examinations for Algebra I, Geometry, and Biology I. Spring 2012 was the field test of the 11<sup>th</sup>-grade United States history EOC in select schools. Spring 2013 saw the addition of both 11<sup>th</sup>-grade United States history full roll-out state-wide of the EOC examination and the 7<sup>th</sup>-grade Civics field test in select schools. Spring 2014 has been slated to introduce the full state-wide roll-out of the 7<sup>th</sup>-grade Civics EOC examination (FLDOE, 2014).

All state-wide standardized EOC assessments were designed to use scale scores that match other standardized tests given throughout the state, e.g., FCAT reading, writing, and science. The scale scores ranged from a low of 1 to a high of 5, with the state board determining what constituted a passing score. The FLDOE used the following definitions to differentiate student success in regard to the content of the NGSSS. Level 1 students have demonstrated an inadequate level of success of the content for the NGSSS. Level 2 students have demonstrated a below satisfactory level of



success of the content for the NGSSS. Level 3 students have demonstrated a satisfactory level of success for the content of the NGSSS. Level 4 students have demonstrated an above satisfactory level of success for the content of the NGSSS. Level 5 students have demonstrated mastery of the most challenging content of the NGSSS. For the United States history EOC examination, specifically, achievement level 3 is considered passing. The United States history EOC examination cut scores are as follows: achievement level 1 has a scale score from 325-369, achievement level 2 has a scale score of 370-378, achievement level 3 has a scale score of 379-397, achievement level 4 has a scale score of 398-417, and achievement level 5 has a scale score of 418-425 (FLDOE, 2014).

All state of Florida EOC examination assessments are computer-based; however, paper-based versions are provided to students who require such a format if an accommodation is necessary to comply with a student's individual education plan (IEP). Although test specifications state that a computer mouse is to be made accessible for every computer, touchpads will also be made for students who are comfortable using them. In addition, according to the Florida Department of Education, training has been made available for school computer technicians to configure school computers for optimum settings. Computer-based testing platform protections have been put in place to (a) ensure that no other applications can be used that would disturb the continuity of the test, and (b) keep accidental logging out of any computer-based high-stakes testing to a minimum (FLDOE, 2014).

According to the state of Florida EOC Test Administration Manual, electronic Practice Assessment Tests (ePATs) are available to all students to practice on their own time. These can be found online through the FLAssessments.com website (2014).

### *The United States History End-Of-Course Examination*

Concerning the state of Florida United States history EOC examination specifically, the first state-wide rollout test was administered between April 22, 2013 and May 3, 2013 where it was given in a 160 minute session with a single, 10-minute break after the first 80 minutes. The FLDOE allowed additional time if necessary but stated that testing must be completed within the same school day. The state reported three primary categories of scores for the state of Florida United States history EOC examination: (a) the percentage each reporting category comprises, (b) the state mean for each reporting category, and (c) the content area scores for each reporting category (FLDOE, 2014).

There are four different test forms for the state of Florida United States history EOC examination, coded A, B, C, and D, with questions common to all four forms. Test questions have been organized in a selected-response format, with four possible answer choices. In addition, test forms were developed and constructed by content and psychometric testing experts who created an equal level of difficulty by using operational and field-test statistics (FLDOE, 2014). Test forms were reviewed by committee members and were determined to be fair, and test items were aligned with state standards (FLDOE, 2014).

The test was divided into three reporting categories, or time periods: category 1, the late 19<sup>th</sup> and early 20<sup>th</sup> centuries (1860-1910) which contained 33% of the test item score points and focused on the issues related to the United States Civil War, Reconstruction, the end of the frontier, industrialization, and changes to American society at the beginning of the twentieth century; category 2, Global Military, Political, and Economic Challenges (1890-1940) which contained 34% of the test item score points and focused on the issues related to the rise of United States as a world military power, its increased involvement in world affairs including its participation in both world wars, and the changing social, political, and economic forces affecting the United States at home during the 1920s and 1930s; category 3, the United States and the Defense of the International Peace (1940-2010) which contained 33% of the test item score points and focused on the issues related to World War II, the Cold War, and the domestic social revolutions of the late 20<sup>th</sup> century, and the challenges face the United States at the dawn of the 21<sup>st</sup> century (FLDOE, 2014). It should be noted that the state mean score for each reporting category may be different and cannot be used to compare students' achievement levels between test forms (FLDOE, 2014).

In order to compare student achievement levels, the state must first determine raw scores, also known as content area scores, which are the actual number of questions that are correctly answered on the United States history EOC examination. Due to the fact that there are four different forms of the test, raw scores are equated. Equating means taking the raw score, and if necessary, adjusting the score to maintain a continuity of the difficulty level across all four forms. In order to yield a valid interpretation of student

performance, the equating process ensures that the scale scores on the different test forms have the same meaning and are comparable (FLDOE, 2014).

Scores have been reported in the *U.S. History EOC Assessment Student Report* (2013), a two-page report. The first page provides the reader with an explanation of the assessment and shows the scale score. It provides a graph which compares the student's score with those of other students across the state, and also shows the state mean. The graph also indicates the student's ranking within the lowest, middle, or highest third of test takers. The second page provides the student's content area score, with information translated into both Spanish and Haitian/Creole. The top part provides the explanation of the content area score. The rest of the page is divided into three columns. The first column shows the actual number of points in each of the content areas; the second column displays the possible number of points in each of the content areas; and the third column shows the state mean for each of the content areas to allow the reader the opportunity for comparison. The state mean for each of the content areas reveals the points earned by the students across the state for each reporting category of that particular test form (FLDOE, 2014).

#### State of Florida Legislative Mandates of Online Education in Florida

Legislative mandates were important considerations in reviewing the literature and providing a rationale for the present study. Many of the legislative mandates and actions reviewed in this section were concerned with and had direct impact upon the Florida Virtual School (FLVS). This section contains a brief legislative history of FLVS

and a more detailed discussion of two of the more current legislative mandates: the Virtual Instruction Program Law of 2008 and the Digital Learning Act of 2013.

Title XLVIII, Chapter 1002, Section 37 paragraph 1(a) of the Florida statutes established the “development and delivery of online and distance learning education”. Florida’s Commissioner of Education was assigned the responsibility for monitoring the progress of Florida’s Virtual School (FLVS). Paragraph 1(b) explained the mission of the virtual school as one that provided technology-based opportunities to a variety of students, such as students seeking accelerated access to graduate from high school early, home-schooled, or rural students who did not have access to higher-level courses.

#### *Florida Virtual School (FLVS) Legislative History*

In 1997, a startup funding grant of \$200,000 was awarded to Alachua and Orange counties to develop an online virtual high school (FLVS, 2014). In 2000, the legislature named FLVS as an “independent educational entity” in legislative statute 1002.37, giving it the same rights as the other 67 counties in the state of Florida (FLVS, 2014). In 2003, the legislature changed the funding of FLVS, allowing it to take part of the Florida Educational Funding Program (FEFP) with section 1004.04 of the Florida Statutes (FLDOE, 2014).

#### *K-8 Virtual School Program*

In 2005, the state of Florida legislature passed statute 1002.415, or the K-8 Virtual School Program (Florida Virtual School, 2013). A virtual school program was created

within the FLDOE for the purpose of creating academic instruction and to make it available to full-time students from kindergarten through grade 8 using online technology (Florida Virtual School, 2013). A selection process was used to determine which schools were allowed to deliver the program instruction. The schools were pilot programs which, in order to receive funding, had to comply with the following: be nonsectarian in its hiring practices, admission policies, operations and programs; comply with statutory antidiscrimination provisions; participate in the state's school accountability system; charge no tuition or registration fee; provide all related coursework materials, computer hardware, and software in each household for enrolled students; have its administrative office within the state and have its administration and instructional staff members be state residents (F. S. A 1002.415{1}). If all procedures were met and approved, the state would grant the schools an initial three year contract receiving full-time eligibility funds from the General Provisions Fund updated annually for each full-time K-8 enrolled student. Upon proper documentation of enrollment and proof of attendance, payments would be made to the school four times a year. Students were subject to compulsory attendance requirements, and were required to take statewide standardized assessments in the school of the students zoning. If a virtual school received a performance category grade of D or F, it was required to file and receive approval of a school improvement plan. The contract with the State of Florida's Department of Education was to be terminated if a performance grade of D or F was received for two of any four consecutive years. Additionally, the Department of Education could choose to not renew a contract if the school: failed to completely participate in the state's educational accountability

system; failed to receive a “C” or higher of the school grading system; failed to meet generally accepted standards of fiscal management; violated the law; or if the state legislature failed to fund the program (F. S. A. 1002.415{5}).

The program was fully implemented in 2008, named the “Virtual Instruction Program” (FLDOE, 2014). It defined a Virtual Instruction Program (VIP) as a program which occurs in an online interactive environment in which both the teacher and student are separated either by time or space (FLDOE, 2014). This law required all school districts in the state of Florida to provide online and distance learning instruction available to full-time virtual students in grades k through 8 by the 2009-2010 school year. Teachers who were hired by an online or blended charter school were required to be certified or have an endorsement in the area in which they were teaching. Districts were required to provide either their own VIP, establish a contract with FLVS, enter into a contractual agreement with an approved provider, or pair up with another district, Florida college, or virtual charter school (FLDOE, 2014). In addition, the statute was amended to require full-time online programs to expand their coverage to all grades, K-12 (FLDOE, 2014).

### *Digital Learning Now Act*

On June 2, 2011, Florida’s governor signed House Bill 7197, coined the Digital Learning Now Act (FLVS, 2014). The legislation, titled F. S. A. 1002.321 (2011) mandated that all students should have elements of “high quality, digital learning.” Education, according to statute, needed to be customized to the needs of the students

using digital content and contain an infrastructure that supported digital learning. To that end, in 2011 school districts were required to operate either part-time or full-time instructional programs or enter into a contract with “Florida Virtual School, or a blended learning institution, or full-time virtual charter school of instruction”

(F. S. A. 1002.45{4}{a}). All students who entered high school in the 2011-2012 year were required to take at least one course online in order to graduate with a diploma (FLDOE, 2014). The law stated that an online high school course taken in the middle grades six through eight also fulfilled the requirement for graduation (FLVS, 2014). In addition to the online requirement for graduation, the legislation made it easier for the creation of more private online and blended charter schools in Florida. Also, for the first time, the Integrity Of Online Courses component of the statute made it unlawful for any person to take a course online or to take an exam for any other person for any reason, such as for money (FLDOE, 2014), making it punishable as a misdemeanor in the second degree. Maxwell (2011) criticized the Act in a *Tampa Bay Times* column. He expressed his belief that the mandates forced smaller districts around the state to divert limited funds in order to develop new computer labs for students who may not have access to the internet at home to comply with the online requirement for graduation, and that the state is doing very little to provide funding to low populated districts for computers and lab space (Maxwell, 2011).



## Overview of Florida Virtual School (FLVS)

The Florida Virtual School website, (FLVS) was founded in 1996 in a grant awarded jointly to both Alachua and Orange counties (FLVS, 2013). In 2000, legislation established FLVS as an independent education entity with the same legal status as the other 67 school districts in Florida and became the first statewide k-12 public charter school that was completely online, as pursuant to statute 1002.37. During the same year, the Committee on International and Trans-regional Accreditation (CITA) and the Southern Association of Colleges and Schools (SACS) sent letters of accreditation to FLVS. At the time of the study, Julie Young was the current CEO, and there were six members on the FLVS board of trustees: Michael Olenick, Board Chairman; Lady Dhyana Ziegler, Vice Chair; Marva Johnson, Tammie Nemecek, Brian Cunningham, and Linda Pellegrini, board members (FLVS, 2014). According to F. S. A. 1002.37(2), the board of trustees shall be appointed by the governor to four-year staggered terms (FLVS, 2014).

According to F. S. A. 1002.37(2), board of trustees shall be responsible for the following: meeting at least four times a year; developing state-of-the-art technology-based educational systems that is cost-effective, educationally sound, marketable, and capable of sustaining a self-sufficient delivery system through the Florida Education Finance Program (FEFP); seek different avenues to raise and generate revenue and enter into agreements with distance learning vendors and providers; acquire, enjoy, use and dispose of patents, copyrights, trademarks and licenses; be responsible for the administration and control of school funds; accrue supplemental revenue from all

activities or sources; “administer and maintain personnel programs for all employees by adopting rules, policies, and procedures related to the appointment, employment and removal of personnel” (F. S. A. 1002.37{2}); determine the compensation, salaries, and fringe benefits of all employees; establish rules and procedures for admission of students; establish and distribute school procedures; enter into franchise agreements with other Florida school districts; enable employees to be eligible to participate in the Florida Retirement System (FRS); establish performance and accountability measures for both employees and students; submit to the Board of Education both forecasted and actual enrollment and credit completions; provide for content and custody of school records; maintain financial records and report said records under the “prescribed uniform system of financial records and accounts for the schools of the state” (F. S. A. 1002.37{a}).

In 2003, FLVS changed its funding methods from the state legislature. At present, FLVS was the only public school in Florida where the funding that is received is tied directly to performance. Funding was changed to the Florida Education Finance Program (FEFP) where the schools could receive full-time equivalency funding based on course completion and performance, as opposed to seat time from a traditional brick-and-mortar school (FLVS, 2014). Full-time equivalency is defined as a student who was enrolled and received direct instruction and completion of the course in order for FLVS to receive funds from FEFP (FLDOE, 2014). Each half credit course successfully completed generates 0.0834 un-weighted full-time equivalency (FTE). Six courses per semester are required to generate full-time funding. Supporters and advocates of FLVS

have stressed that the savings to the state have been significant, averaging a savings of \$1,048 per student over regular public school weighted FTE (iNacol, 2014).

In addition, FLVS receives no funding for exceptional student education, capital outlay, transportation, or any other significant funding stream that provides substantial funding for brick-and-mortar schools. The only funding stream from which FLVS benefits is for instructional materials. These funds are dedicated to online course creation and maintenance, teacher training, and class size reduction (iNacol, 2014).

FLVS has serviced full-time and part-time K-12 students, and all credits and diplomas earned are generated from the enrolled students' home school location. At the time of the present study, FLVS was associated with all 67 school districts in the state and serviced students from all over the world. As of the 2012/2013 school year, it had 411,000 successful half-credit semester completions (FLVS, 2014).

## CHAPTER 3

### METHODOLOGY

#### Introduction

According to iNACOL, online education or distance learning has gained in numbers and acceptance over the past decade. For the 2010-2011 school year, 1,816,400 students in the United States on the K-12 level were enrolled in at least one online class, up from the estimated 400,000 of 10 years ago (iNACOL, 2012). For the 2010-2011 school year, the State of Florida employed 1,500 staff to serve over 100,000 online students (FLDOE, 2013). One course taken by Florida students is United States history. This course has been included as a part of the social science assessment now mandated as part of both the Next Generation Sunshine State Standards and Common Core. At the time of the study, United States history was also currently required for Florida high school graduation (FLDOE, 2013).

#### Statement of the Problem

To date, little research has compared student performance outcomes on the United States history End-of-Course examinations completed by students enrolled in traditional face-to-face instruction with the outcomes for students enrolled in online United States history courses. Contradictory data exist on the effects of student achievement overall in online or distance learning. For example, a 2009 report by the United States Department of Education showed that online instruction produced similar results when compared to

face-to-face learning (Groux, 2011). Yet, critics of online education claim that online students suffer from high rates of dropout (Bennett et al., 2010). Minnesota reported only 16% of online high school students passed the state's mathematics proficiency examinations (Lemaige, 2011). Overall, students who took at least one course online in 2010 had a 34% dropout rate versus 26% for students who attended traditional face-to-face classes (Xu & Jagers, 2011).

### Research Questions

The following research questions and hypotheses were used to guide the study:

1. What difference, if any, exists between the End-Of-Course (EOC) examination scores of students who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus those students who complete the same course in an online format through Florida Virtual School (FLVS)?

H<sub>01</sub> There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.

2. What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on ethnicity, who complete the Florida United States

history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus those students who complete the same course in an online format through Florida Virtual School (FLVS)?

H<sub>02</sub> There is no significant difference between students, based on ethnicity, who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.

3. What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on gender, who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus students who complete the same course in an online format through Florida Virtual School (FLVS)?

H<sub>03</sub> There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method based on gender versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.

4. What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on free-and-reduced lunch, who complete the Florida

United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus students who complete the same course in an online format through Florida Virtual School (FLVS)?

H<sub>04</sub> There is no significant difference between students who take Florida’s United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method based on free-and-reduced lunch status versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.

The following table presents the variables, statistical tools, and if the Null Hypothesis for each was accepted or rejected.

Table 2

*Research Questions, Independent Variable, Dependent Variable, and Statistical Tool.*

Research Question	Independent Variable	Dependent Variable	Statistical Tool
Research Question 1	Overall, Face-to-Face (Population)	Overall, Online (Sample)	One-Sample Wilcoxon test
Research Question 2	White, Face-to-Face (Population)	White, Online (Sample)	One-Sample Wilcoxon test
	Non-White, Face-to-Face (Population)	Non-White, Online (Sample)	One-Sample Wilcoxon test
Research Question 3	Female, Face-to-Face (Population)	Female, Online (Sample)	One-Sample Wilcoxon test
	Male, Face-to-Face (Population)	Male, Online (Sample)	One-Sample Wilcoxon test

Research Question	Independent Variable	Dependent Variable	Statistical Tool
Research Question 4	No Lunch Assistance Face-to-Face (Population)	No Lunch Assistance Face-to-Face Online (Sample)	One-Sample Wilcoxon test
	Free or Reduced Lunch Face-to-Face (Population)	Free or Reduced Lunch, Face-to-Face Online (Sample)	One-Sample Wilcoxon test

### Research Design

This quantitative, non-experimental, ex-post facto comparison research study was designed to test the extent to which there was a relationship between the End-of-Course examinations of students who completed the state of Florida 11<sup>th</sup>-grade United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face instructional delivery format versus those students who complete the same course in an online instructional delivery format through Florida Virtual School (FLVS) and was not to suggest a causal inference.

This study was based on the premise of studying to what extent differences may or may not exist. This study was non-experimental, as variables were pre-identified and not manipulated. This study was ex-post facto in that the events had already occurred, and no independent variable was manipulated.



### Participants

The population for this study was drawn from one Central Florida school district that had 36 high, charter, and alternative schools which offered United States history and United States history honors classes in a traditional, face-to-face format. For this study, the target population consisted of all students who were enrolled in either United States history (course code 2100310) or United States history, honors (course code 2100320) and entered grade 9 or beyond during the 2012-2013 school year in this district.

A total of 9,339 students completed the course in either the traditional or online format and had a scale score for the EOC examination. Of these students, only 10 completed the course online; the remaining students completed it in the face-to-face modality. Because the face-to-face results represent the performance in the established, status quo modality, the results of the 9,329 in the face-to-face group were considered to represent the population. The 10 students in the online group served as the sample for comparison. The data were collected with the assistance of the school district's Office of Accountability and Assessment.

### Instrumentation

Historically speaking, and according to Carmines and Zeller (1979), there has been little regard for the science of measurement in the social sciences. Measurement in the social sciences was more of an abstract concept rather than a focused part of the social sciences. As an indication of the truthfulness of Carmines' and Zeller's statements, United States history was never added as part of the Florida Comprehensive

Assessment Test (FCAT) or FCAT 2.0. However, in the present era of accountability, this too has changed. Measurement in the social sciences, specifically United States history, is indeed real and is an integral part of not only student success but also teacher, school, and district accountability. Any simple measurement of United States history will not, however, meet the state's expectations of accountability under NGSSS (FLDOE, 2013). The measurement used in the present study to investigate online vs. traditional face-to-face delivery modes was the state of Florida United States history End-Of-Course Assessment (EOC). The EOC measures a student's level of achievement in accordance with the NGSSS (FLDOE, 2013).

The United States history EOC Assessment was determined to be a reliable and valid instrument prior to its use. Reliability is the extent in which a test can yield the same consistent results after repeated trials. Validity is achieved if what a test does what it actually is supposed to do (Carmines, 1979). In 2012, 30 sample test items from the Florida United States history EOC examination were field tested with 50,000 students representing 55 school districts and 243 schools in Florida. The test items went through a careful analysis for both reliability and validity and were revised as needed prior to the first full implementation of the assessment in the 2012-2013 school year (FLDOE, 2013).

The Florida United States history EOC Assessment was a 60-item standardized, criterion-referenced, multiple-form examination that was given to each student during a one-week window from the end of April to the first week of June, 2013. The EOC was administered to all students who were enrolled in either regular or honors United States history who entered Grade 9 or beyond during the 2012-2013 school year; the assessment

comprised 30% of students' final grades. The examination was administered to most students via a computer-based test (CBT) platform. Students with disabilities and those who required accommodations in accordance with their Individual Education Plans (IEPs) or their Section 504 plans, however, received a paper-based test. The test was administered during a single 160-minute session and allowed for a 10-minute break after 80 minutes. Despite the time restriction, students were still given additional time if needed, so long as the test was completed the same school day. Subsequent assessments in the future were scheduled to be graded on a pass/fail basis with a passing grade required in order to achieve course credit (FLDOE, 2013).

According to the FLDOE (2013), the contextual framework of the EOC examination included the following characteristics of the test:

- The test includes 60 questions, with multiple versions.
- Approximately 6 to 10 items are experimental in nature, being field tested for future use and are not included in the student's test score.
- The T-score scale ranges from 20 to 80, with 50 being the statewide average.
- 33% of the questions concern late 19th and early 20th century, 1860-1910.
- 34% of the questions are based on global military, political, and economic challenges, 1890-1940.
- 33% of the questions address The United States and the defense of the international peace, 1940-present.
- EOC assessment items were categorized using a model based from the Webb's Depth of Knowledge (DoK) and the cognitive classification system

used for the National Assessment of Educational Progress (NAEP) of low, moderate, and high cognitive questions.

- 20-30% of the questions are to be of a low cognitive complexity level.
- 45-65% of the questions are to be a moderate level of cognitive complexity.
- 15-25% are of a high level of cognitive complexity.

### Data Collection

Quantitative methodologies were used in data collection for this particular study.

Contact was initiated with the Florida Virtual School's (FLVS) Instructional Programs Manager by phone and subsequently through email (Appendix A). It was determined that in order for the researcher to continue, proper applications needed to be completed and submitted. Temporary access to a demonstration mode for both 7th-grade United States Civics and 11th-grade United States history was granted, which permitted the researcher access to research and study samples of online coursework for analysis and later evaluation (FLVS, 2013).

Contact was next initiated with the County Senior Executive Director for the Accountability and Assessment Office in the target district for permission to access United States history EOC data for both traditional school students and students who had completed the course through FLVS. The vast majority of students who were enrolled in FLVS were enrolled only part-time. These students were also still enrolled in the school for which they were zoned. As a result, those students who completed the course online were required to take the EOC examination at their zoned school of record. Written

confirmation of approval to collect data was received from the Senior Executive Director for the Accountability and Assessment Office of the school district (Appendix A). After the proposal for the dissertation was approved by the researcher's committee, approval to conduct the study was also sought and received from the University of Central Florida's Institutional Review Board (Appendix B).

### Data Analysis

During the 2013-2014 school years, United States history was delivered to students in two formats: a traditional face-to-face classroom setting and an online format through the Florida Virtual School (FLVS). United States history is a year-long course; its completion is necessary in order to receive credit towards graduation (FLDOE, 2013). Upon acquisition of data, students' EOC scores and demographic information were analyzed using SPSS (v. 21).

The data subsets for research questions two through four do not fit a normal distribution for the students who comprised the online cohort of this study. Due to the very small size of the sample data subsets, caution was necessary when making inferences in order to avoid the possibility of error. In order to minimize the possibility of such an error, a conservative test, the Wilcoxon signed-rank test was used. According to Kiess (1989), the Wilcoxon signed-rank test was also referred to as the One-Sample Wilcoxon test. The One-Sample Wilcoxon test was the non-parametric equivalent to the One-Sample T-Test. A non-parametric test can be advantageous when "the statistical tool requires no assumptions about the population parameters" (Lapin, 1973, p.514).

According to Kiess (1989), non-parametric tests can also be referred as distribution-free tests.

Additionally, the One-Sample Wilcoxon test was used “for within-subjects designs with two levels of an independent variable” (Kiess, 1989, p.478). The One-Sample Wilcoxon test was a test that considered “both the direction and the magnitude of the differences between matched sample pairs” (Lapin, 1973, p.531). In other words, the One-Sample Wilcoxon test was used when comparing two related or matched samples to assess whether their population mean ranks differ (Lapin, 1973).

For Research Question 1, a nonparametric One-Sample Wilcoxon test was run to determine if a statistically significant difference existed between the EOC examination scores of students from the traditional, face-to-face format and students who completed the course online. The One-Sample Wilcoxon test, the nonparametric alternative to the one-sample *t*-test was chosen because the performance of students in the traditionally-formatted class represented the established population value rather than another sample. The performance of online students served as the comparison sample. For this test, the EOC examination scores represented the continuous dependent variable. The median value of the face-to-face scores was calculated and used as the value to which the median value of the online scores was compared.

Research Question 2 was intended to measure differences in performance between online and the established face-to-face population when taking gender into consideration. Because comparisons still needed to be made relative to the population performance of the face-to-face group, two separate One-Sample Wilcoxon tests were run, one within

each of the two gender categories (female and male) using the same continuous dependent variable of EOC examination scores. Therefore, differences could be detected between the online sample and the face-to-face population with respect to each gender.

Research Question 3 followed the same pattern as Research Question 2, but instead measured differences with respect to ethnicity. Detailed ethnicity information was provided (White, Black, Hispanic, Asian, and Multiracial); however, with such a small sample size among online students, the ethnic categories were combined into White and non-White. The One-Sample Wilcoxon test was then run within these two groups to determine differences in the median performance of online students in each of these categories as compared to the face-to-face population values for both the white and non-white samples.

Research Question 4 also followed the same pattern as Research Questions 2 and 3, but instead measured differences with respect to socioeconomic status. Students were separated into groups reflecting those who did not receive any free or reduced-price lunch assistance (high SES) and those who did receive such assistance (low SES). The One-Sample Wilcoxon test was then run within each of these groups to determine differences in the median performance of online students in each of these categories as compared to the face-to-face population values for both the free-and-reduced lunch status students, and those that did not receive free-and reduced lunch.

## Summary

This chapter restated the purpose of the research and the research questions and hypotheses. All available data were used from the students' EOC scores for this study from students who completed the United States history course in both the traditional face-to-face and online formats. The instrument, Florida's United States history EOC examination was discussed at length, including the particulars of the field test administered during the 2012-2013 school year. Data collection procedures and description of the data analysis were shared as well. Chapter 4 contains a report of the results of the analysis of the data.



## CHAPTER 4 PRESENTATION AND ANALYSIS OF DATA

### Introduction

This study was conducted to examine if any differences in student achievement could be identified between students who completed the 11<sup>th</sup>-grade regular Florida United States history course (course code 2100310) or honors United States history course (course code 2100320) in a traditional, face-to-face format and those students who completed the same course in an online format through the Florida Virtual School (FLVS). The data for the study consisted of the 9,339 students who completed the EOC examination in a large, urban school district in Central Florida. The EOC examination scale scores were used as a measure of student achievement.

Analyses aligned with the four research questions in addressing differences between the face-to-face and virtual performance for the overall population as well as for three other demographic factors: (a) gender, (b) ethnicity, and (c) free-and-reduced lunch status. Although little research to date has compared student achievement on 11<sup>th</sup>-grade Florida United States history EOC examinations, the goal of this study was to contribute to the design of more effective instructional delivery methods which could improve student achievement. This chapter presents the results of the data analysis for the four stated research questions.

### Demographic Characteristics of the Sample

The demographic characteristics of the sample are presented in Table 3. Of the population, nearly all (99.9%) took the course in the face-to-face modality. Regarding the other demographics, both the large face-to-face population and the smaller online sample were nearly evenly divided between male and female students. With respect to ethnicity and to free-and-reduced lunch status, however, the very small online sample was not as equivalent. Various ethnic subgroups were not represented at all in the online population, as 70% were White and 30% were of non-White ethnicities (Hispanic and Asian). On the other hand, only 29.8% of the face-to-face population were White. Likewise, only 20% of the online sample received lunch assistance, as compared to 59.8% for the face-to-face population. Despite these discrepancies, it is important to be mindful of the fact that the extremely small online sample could easily lead to this mismatched representation as compared to the larger face-to-face sample.

Table 3

*Demographic Characteristics of Sample, number and percentage of Face-to-Face and Online students as broken down by Gender, Ethnicity, and Free-and-reduced lunch status subgroups.*

Descriptor	Face-to-Face		Online	
	#	%	#	%
Overall	9,329	99.9	10	0.1
Gender				
Male	4,719	50.6	5	50.0
Female	4,610	49.4	5	50.0
Ethnicity				
Asian	382	4.1	1	10.0
Black	2,535	27.2	0	0.0
Hispanic	3,386	36.3	2	20.0
American Indian/Pacific Islander	32	0.3	0	0.0
Multiracial	214	2.3	0	0.0
White	2,780	29.8	7	70.0
Free-and-reduced lunch				
No	3,713	39.8	8	80.0
Yes	5,575	59.8	2	20.0
Unknown	41	0.4	0	0.0

### Research Question 1

*What difference, if any, exists between the End-of-Course (EOC) examination scores of students who complete the regular Florida United States history course (course code 2100310), or honors United States history course (course code 2100320), in a traditional, face-to-face format, versus those students who complete the same course in an online format through Florida Virtual School (FLVS)?*

*H<sub>01</sub> There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.*

Florida United States history EOC examination scale scores were used to measure and analyze student achievement for Research Question 1. Due to the inherent setup involving the population-level face-to-face results serving as the point of comparison for the online sample, a one-sample test was selected; furthermore, due to the very small sample size of the online group, the nonparametric One-sample Wilcoxon test was utilized. Results from the One-Sample Wilcoxon test are presented in Table 4.

Table 4

*One-Sample Wilcoxon Test: Overall Online Performance*

Group	<i>n</i>	Median	<i>Z</i>	<i>p</i>
Face-to-Face (Population)	9,329	52.0	1.22	.22
Online (Sample)	10	55.5		

\**p* < .05. \*\**p* < .01.

The median scale score for the face-to-face population was 52.0, and the median for the online sample was 55.5. The One-Sample Wilcoxon test indicated that though the online sample median was higher than that of the face-to-face population, the difference

was not statistically significant ( $Z = 1.22, p = .22$ ). Therefore, insufficient information existed to reject the null hypothesis that there is no difference between overall online and face-to-face performance.

### Research Question 2

*What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on ethnicity, who complete the regular Florida United States history course (course code 2100310), or honors United States history course (course code 2100320) in a traditional, face-to-face format versus those students who complete the same course in an online format through Florida Virtual School (FLVS)?*

*H<sub>02</sub> There is no significant difference between students, based on ethnicity, who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.*

Florida United States history EOC examination scale scores were used to measure and analyze student achievement for Research Question 2. As in Research Question 1, a One-Sample Wilcoxon test was utilized to determine the presence of any significant differences in the median score of the online sample as compared to the face-to-face population. Due to the added factor of ethnicity, two separate One-Sample Wilcoxon tests were utilized: one for White students and one for non-White students. The determination to reduce the various ethnicities into the comparison groups of White and non-White originated from the fact that so few non-White students completed the course online; comparison groups comprising the individual non-White ethnicities would be extremely small. Results from the One-sample Wilcoxon tests are presented in Table 5.

Table 5

*One-Sample Wilcoxon Test: Online Performance by Ethnicity*

Group	<i>n</i>	Median	<i>Z</i>	<i>p</i>
White				
Face-to-Face (Population)	2,780	57.0	-1.44	.15
Online (Sample)	7	50.0		
Non-White				
Face-to-Face (Population)	6,549	50.0	1.60	.11
Online (Sample)	3	65.0		

\* $p < .05$ . \*\* $p < .01$ .

Among White students, the median scale score for the face-to-face population was 57.0, and the median for the online sample was 50.5. The One-Sample Wilcoxon test indicated that though the online sample median was lower than that of the face-to-face population, the difference was not statistically significant ( $Z = -1.44$ ,  $p = .15$ ). Therefore, insufficient information existed to reject the null hypothesis that there is no difference between overall online and face-to-face performance among White students.

The reverse pattern held true for non-White students. The median scale score for the face-to-face population was 50.0, and the median for the online sample was 65.0. However, although the online sample median was notably higher, the difference was not significant ( $Z = 1.60$ ,  $p = .11$ ). It is important to remember the very small sample sizes, particularly in this non-White group ( $n = 3$ ). Thus, the applicability of the statistical inference to general online performance was not particularly strong. Insufficient information existed to reject the null hypothesis that there is no difference between overall online and face-to-face performance among non-White students.

### Research Question 3

*What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on gender, who complete the regular Florida United States history course (course code 2100310), or honors United States history course (course code 2100320) in a traditional, face-to-face format, versus students who complete the same course in an online format through Florida Virtual School (FLVS)?*

*H<sub>03</sub> There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method based on gender versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.*

Florida United States history EOC examination scale scores were used to measure and analyze student achievement for Research Question 3. As in Research Questions 1 and 2, a One-Sample Wilcoxon test was utilized to determine the presence of any significant differences in the median score of the online sample as compared to the face-to-face population. Due to the added factor of gender, two separate One-Sample Wilcoxon tests were utilized: one for female students and one for male students. Results from the One-Sample Wilcoxon tests are presented in Table 6.

Table 6

*One-Sample Wilcoxon Test: Online Performance by Gender*

Group	<i>n</i>	Median	<i>Z</i>	<i>p</i>
Female				
Face-to-Face (Population)	4,719	50.0	0.41	.69
Online (Sample)	5	48.0		
Male				
Face-to-Face (Population)	4,610	53.0	1.75	.08
Online (Sample)	5	64.0		

\**p* < .05. \*\**p* < .01.

Among female students, the median scale score for the face-to-face population was 50.0, and the median for the online sample was 48.0. The One-Sample Wilcoxon test indicated that though the online sample median was lower than that of the face-to-face population, the difference was not statistically significant ( $Z = 0.41, p = .69$ ). Therefore, insufficient information existed to reject the null hypothesis that there is no difference between overall online and face-to-face performance among female students.

The reverse pattern held true for male students. The median scale score for the face-to-face population was 53.0, and the median for the online sample was 64.0. However, though the online sample median was notably higher, the difference was not significant ( $Z = 1.75, p = .08$ ). As among female students, insufficient information existed to reject the null hypothesis that there is no difference between overall online and face-to-face performance among male students.

#### Research Question 4

*What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on free-and-reduced lunch, who complete the regular Florida United States history course (course code 2100310), or honors United States history course (course code 2100320) in a traditional, face-to-face format, versus students who complete the same course in an online format through Florida Virtual School (FLVS)?*

*H<sub>04</sub> There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method based on free-and-reduced lunch status versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.*

Florida United States history EOC examination scale scores were used to measure and analyze student achievement for Research Question 4. As in prior research



questions, a One-Sample Wilcoxon test was utilized to determine the presence of any significant differences in the median score of the online sample as compared to the face-to-face population. Due to the added factor of socioeconomic status, two separate One-Sample Wilcoxon tests were utilized: one for students who received no lunch assistance and one for students receiving free or reduced lunch. Results from the One-Sample Wilcoxon tests are presented in Table 7.

Table 7

*One-Sample Wilcoxon Test: Online Performance by Free-and-Reduced Lunch Status*

Group	<i>n</i>	Median	<i>Z</i>	<i>p</i>
No Lunch Assistance				
Face-to-Face (Population)	3,713	55.0	0.77	.44
Online (Sample)	8	58.0		
Free or Reduced Lunch				
Face-to-Face (Population)	5,575	49.0	1.34	.18
Online (Sample)	2	52.0		

\**p* < .05. \*\**p* < .01.

Among students of a higher socioeconomic status, the median scale score for the face-to-face population was 55.0, and the median for the online sample was 58.0. The One-Sample Wilcoxon test indicated that though the online sample median was higher than that of the face-to-face population, the difference was not statistically significant ( $Z = 0.77, p = .44$ ). Therefore, insufficient information existed to reject the null hypothesis that there is no difference between overall online and face-to-face performance among students who do not receive free or reduced lunch.

A similar pattern held true for students of a lower socioeconomic status. The median scale score for the face-to-face population was 49.0, and the median for the online sample was 52.0. However, though the online sample median was higher, the difference was not significant ( $Z = 1.34, p = .18$ ). As was found with students' peers of higher socioeconomic status, insufficient information existed to reject the null hypothesis that there is no difference between overall online and face-to-face performance among students who received free or reduced lunch.

Table 8 presents the variables, statistical tools, and the results of the Null Hypothesis for each research question was either accepted or rejected.

Table 8

*Research Questions, Variable, Statistical Tool, and if Null Hypotheses were rejected.*

Research Question	Variable	Statistical Tool	Null Hypothesis rejected or not
Research Question 1	Overall Face-to-Face (Population) Online (Sample)	One-Sample Wilcoxon test	Not rejected
Research Question 2	White Face-to-Face (Population) Online (Sample)	One-Sample Wilcoxon test	Not rejected
	Non-White Face-to-Face (Population) Online (Sample)	One-Sample Wilcoxon test	Not rejected

Research Question	Variable	Statistical Tool	Accepted/Rejected
Research Question 3	Female Face-to-Face (Population) Online (Sample)	One-Sample Wilcoxon test	Not rejected
	Male Face-to-Face (Population) Online (Sample)	One-Sample Wilcoxon test	Not rejected
Research Question 4	No Lunch Assistance Face-to-Face (Population) Online (Sample)	One-Sample Wilcoxon test	Not rejected
	Free or Reduced Lunch Face-to-Face (Population) Online (Sample)	One-Sample Wilcoxon test	Not rejected

### Summary

This chapter presented the findings associated with the analyses conducted to address the major research questions of the study. After providing some initial descriptive information regarding the demographic composition of the face-to-face population and the online sample, the inferential statistical analyses were presented. None of the One-Sample Wilcoxon tests indicated the presence of a significant difference among any subgroup--overall, White, non-White, female, male, high socioeconomic status, or low socioeconomic status. Therefore, none of the null hypotheses presented were rejected. In the next chapter, the findings are summarized and discussed. Implications for practice and recommendations for further research are also offered.

## CHAPTER 5 SUMMARY, DISCUSSION, AND RECOMMENDATIONS

### Introduction

This chapter provides an overall summary of the study followed by a summary and discussion of findings organized around the four research questions. Potential implications for practice and recommendations for future research are offered in hopes that this research may be useful to policy makers on the state level and those school district officials responsible for the design of instructional delivery in high school settings.

### Summary of the Study

The problem in this study was twofold. Little research had been completed that compared the performance outcomes of students who took the state of Florida 11<sup>th</sup>-grade United States history course in the traditional face-to-face versus online formats. Secondly, the existing data from the research were contradictory as to the value of one format over the other (Bennett & Vedder, 2010; Groux, 2011; Lemaige, 2011, Xu and Jagers, 2011). According to iNACOL (2012), online education will continue to grow in the foreseeable future. The study may be of some significance to school district officials and policy makers on the state level as the findings of this study will add more to the body of knowledge concerning online education.

## Summary and Discussion of Findings

### *Research Question 1*

*What difference, if any, exists between the End-Of-Course (EOC) examination scores of students who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus those students who complete the same course in an online format through Florida Virtual School (FLVS)?*

*H<sub>01</sub> There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.*

The findings resulting from the analysis of data to answer the first research question indicated no statistically significant difference existed between the traditional face-to-face group, and the online group. The online sample performed better with a median score of 55.5 than did the face-to-face group with a median score of 52.0. Although according to the One-Sample Wilcoxon test, this was not statistically significant ( $Z = 1.22, p = .22$ ), it remains to be seen if this is actually educationally relevant for two reasons. First, of the 9,339 students who completed the EOC examination in a large, urban school district in Central Florida, there was a very small sample size of the online group with which to make a comparison. United States history is still primarily taught traditionally, using face-to-face instruction. It is possible that the relatively small number indicates that the online course is relatively new. Thus, lack of public knowledge may impact the number of students who could take the course. Secondly, many students take courses through Florida Virtual School for remediation

purposes, i.e., to gain credit for courses that they have previously failed. Due to the nature of the data, it was impossible to tell if the online students were taking this course for the first time or the second time for remediation and credit.

### *Research Question 2*

*What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on ethnicity, who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus those students who complete the same course in an online format through Florida Virtual School (FLVS)?*

*H<sub>02</sub> There is no significant difference between students, based on ethnicity, who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.*

The findings resulting from the analysis of the second research question indicated that there was no statistically significant difference in the EOC examination scores of the traditional group versus the online group, based on ethnicity. Two separate One-Sample Wilcoxon tests were utilized: one for White students and one for non-White students. For the White students, the median score of the face-to-face population was 57.0, and the median score for the online sample was 50.5. For the non-White students, the median score for the face-to-face population was 50.0, yet the median score for the online sample was 65.0.

### *Research Question 3*

*What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on gender, who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus students who complete the same course in an online format through Florida Virtual School (FLVS)?*

*H<sub>03</sub> There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method based on gender versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.*

The findings resulting from the analysis of the third research question indicated that there was no statistically significant difference in the EOC examination scores of the traditional group versus the online group, based on gender. As in the previous research question, two separate One-Sample Wilcoxon tests were utilized: one for female students and one for male students. For the female students, the median score of the face-to-face population was 50.0, and the median score for the online sample was 48.0. For the male students, the median score for the face-to-face population was 53.0, yet the median score for the online sample was 64.0. Insufficient information existed to reject the null hypothesis. Thus, it can be stated that there was no difference between the online and face-to-face students, either male, or female.

#### *Research Question 4*

*What difference, if any, exists between the End-of-Course (EOC) examination scores of students, based on free-and-reduced lunch, who complete the Florida United States history course, regular (course code 2100310), or honors (course code 2100320) United States history in a traditional, face-to-face format, versus students who complete the same course in an online format through Florida Virtual School (FLVS)?*

*H<sub>04</sub> There is no significant difference between students who take Florida's United States history course regular (course code 2100310) and honors (course code 2100320) in a traditional method based on free-and-reduced lunch status versus the students who take the same course online through Florida Virtual School (FLVS) as measured by the End-of-Course (EOC) examination scores.*

The findings resulting from the analysis of the fourth research question indicated that there was no statistically significant difference in the EOC examination scores of the traditional group versus the online group, based on free-and-reduced lunch status.

Similar to the previous two research questions, two separate One-Sample Wilcoxon tests were utilized: one for students who received no lunch assistance and one for students who received lunch assistance. For the non-free-and-reduced lunch status students, the median score of the face-to-face population was 55.0, and the median score for the online sample was 58.0. For the free-and-reduced lunch status students, the median score for the face-to-face population was 49.0, yet the median score for the online sample was 52.0. Again, the small sample size proved problematic, as insufficient information existed to reject the null hypothesis. Analysis showed some of the results did begin to approach significance. While there are trends in the data, the findings show there was no difference between the online and face-to-face students, based on socio-economic status.



### Implications for Practice

This study was conducted to compare the success of students who completed the 11th-grade United States history course in a face-to-face setting with that of students who completed the same course in an online format as measured by performance on the EOC examination. A factor impacting this study was the relatively small number of students who completed the online course which limited the possibility of any definitive conclusions due to the lack of sufficient data.

An issue that might have implications for practice is that the 11th-grade United States history course has traditionally been taught exclusively in a face-to-face format and understanding the complexities surrounding the motivations of students who take coursework online would be beneficial. Additionally, district and state officials should note that the 2013 school year was the inaugural administration for the 11th-grade United States history EOC examination. Therefore, officials should continue to closely monitor student EOC examination performance for both the online and face-to-face groups in the future.

District and state officials should be aware that there were multiple limitations in this study. These limitations, if addressed and further explored, could one day help develop more sound policy for state-mandated EOC examinations.

### Recommendations for Further Study

Online education is a growing trend and will likely continue to expand for the foreseeable future. Online charter schools, private and post-secondary colleges, and universities will continue to grow and compete for students in a world that is becoming increasingly smaller, interconnected and technologically more advanced and complex. Further research into the impact that online learning has on education as a whole is recommended.

Due to little research concerning online education in the social studies the lack of literature on this specific discipline indicates that a need for further research exists. One such possibility would be a replication of the present study expanded to a broader geographic area, perhaps central Florida and ultimately the entire state. Replicating this study to include a larger sample size would also allow for the use of statistical analysis that might produce more robust findings. As such, a state-wide comparison might be helpful in improving the design of instructional delivery in high school settings. Additionally, waiting a few years for this End Of Course assessment to mature before replicating this particular study on a wider scope would be advantageous.

Another possibility could be a qualitative study to explore the nature of online students, their motivation and rationale for taking courses online. School district officials should be interested in any information that may serve to improve student performance outcomes on End-of-Course examinations which lead to higher school graduation rates.

APPENDIX A  
SCHOOL DISTRICT PERMISSION TO CONDUCT THE STUDY

Accountability, Research, and Assessment  
P.O. Box 271  
Orlando, FL 32802-0271

RESEARCH REQUEST FORM

- Project Title
- Purpose and Research Problem
- Instruments
- Procedures and Proposed Data Analysis

RECEIVED JUN 25 2013

Requester's Name William T. Wilson Date 6/25/13

E-mail \_\_\_\_\_  
Address **Redacted Personal Information**

Institutional Affiliation University of Central Florida

Project Director or Advisor Dr. Barbara Murray Phone \_\_\_\_\_

Degree Sought: (check one)  Associate  Doctorate  Bachelor's  Not Applicable  Master's  Specialist

Project Title: A parallel comparison of Traditional versus Online Education in U.S. History as measured by Florida's end of course exam

ESTIMATED INVOLVEMENT

PERSONNEL/CENTERS	NUMBER	AMOUNT OF TIME (DAYS, HOURS, ETC.)	SPECIFY SCHOOLS BY NAME AND NUMBER OF TEACHERS, ADMINISTRATORS, ETC.
Students			
Teachers			
Administrators			
Schools/Centers			
Others (specify)	<u>Annual Data</u>		

Specify possible benefits to students/school system: A better understanding of the impact instructional methods have on both traditional + online education.

ASSURANCE

Using the proposed procedures and instrument, I hereby agree to conduct research in accordance with the policies of the Orange County Public Schools. Deviations from the approved procedures shall be cleared through the Senior Director of Accountability, Research, and Assessment. Reports and materials shall be supplied as specified.

Requester's Signature William T. Wilson

Approval Granted:  Yes  No Date: 6-25-13

Signature of the Senior Director for Accountability, Research, and Assessment [Signature] *Remove all references to GAPS. Non-identifiable data will be released.*

NOTE TO REQUESTER: When seeking approval at the school level, a copy of this form, signed by the Senior Director, Accountability, Research, and Assessment, should be shown to the school principal who has the option to refuse participation depending upon any school circumstance or condition. The original Research Request Form is preferable to a faxed document.

APPENDIX B

INSTITUTIONAL PERMISSION TO CONDUCT THE STUDY



University of Central Florida Institutional Review Board  
Office of Research & Commercialization  
12201 Research Parkway, Suite 501  
Orlando, Florida 32826-3246  
Telephone: 407-823-2901, 407-882-2012 or 407-882-2276  
[www.research.ucf.edu/compliance/irb.html](http://www.research.ucf.edu/compliance/irb.html)

From : UCF Institutional Review Board #1  
FWA00000351, IRB00001138  
To : William Wilson  
Date : January 23, 2014

Dear Researcher:

On 1/23/2014 the IRB determined that the following proposed activity is not human research as defined by DHHS regulations at 45 CFR 46 or FDA regulations at 21 CFR 50/56:

Type of Review: Not Human Research Determination  
Project Title: A COMPARISON OF TRADITIONAL VERSUS  
ONLINE EDUCATION IN UNITED STATES HISTORY  
AS MEASURED BY FLORIDA'S END OF COURSE  
EXAM  
Investigator: William Wilson  
IRB ID: SBE-14-09988  
Funding Agency:  
Grant Title:  
Research ID: N/A

University of Central Florida IRB review and approval is not required. This determination applies only to the activities described in the IRB submission and does not apply should any changes be made. If changes are to be made and there are questions about whether these activities are research involving human subjects, please contact the IRB office to discuss the proposed changes.

On behalf of Sophia Dziegielewski, Ph.D., L.C.S.W., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 01/23/2014 09:59:50 AM EST

A handwritten signature in black ink that reads "Joanne Muratori".

IRB Coordinator

## REFERENCES

- Alphen, A., Halfens, R., Hasman, A., & Imbos, T. (1994). Likert or Rasch? Nothing is more applicable than good theory. *Journal of Advanced Nursing*, 20(1), 196-201.
- Allen, I. E., & Seaman, J. (2008). *Staying the course: Online education in the United States*. Needham, MA: Sloan.
- Amrein, A. L., & Berliner, D. C. (2002). High-stakes testing & student learning. *education policy analysis archives*, 10, 18.
- American Psychological Association. (2013). *Appropriate Use of High-Stakes Testing in Our Nation's Schools*. Retrieved from <http://www.apa.org/pubs/info/brochures/testing.aspx?item=2>
- Anderson, T., & Dron, J. (2011). Three Generations of Distance Education Pedagogy. *International Review of Research in Open & Distance Learning*, 12(3).
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258-267.
- Ausubel, D. (2000). *The acquisition and retention of knowledge*. Boston, MA: Kluwer.
- Baker, C. (2013). *Blended learning: Teachers plus computers equal success*. Retrieved from <http://www.deseretnews.com/article/865569876/Blended-learning-teachers-plus-computers-equal-success.html?pg=all>
- Barlow, D. (1992). *Educational psychology*. Herndon, VA: TechBooks.
- Bates, A. T. (2004). *Technology, e-learning and distance education*. Routledge.
- Bedford, L.A. (2009). The professional adjunct: An emerging trend in online instruction. *Online Journal of Distance Learning Administration*, 12(3). Retrieved from <http://www.westga.edu/~distance/ojdl/browsearticles.php>
- Bennett, D., Lucchesi, A. R., & Vedder, R. K. (2010). *For profit higher education: growth, innovation and regulation*. Retrieved from [http://heartland.org/sites/all/modules/custom/heartland\\_migration/files/pdfs/29010.pdf](http://heartland.org/sites/all/modules/custom/heartland_migration/files/pdfs/29010.pdf)
- Biehler, R., & Snowman, J. (1990). *Psychology applied to teaching*. Boston, MA: Houghton Mifflin.

- Blackboard Learning. (2013). *About Blackboard*. Retrieved from:  
<http://www.blackboard.com/>
- Bloom, B. S. (1976). *Human characteristics and school learning*. New York, NY: McGraw-Hill.
- Bloom, B. S. (1981). *All our children learning: A primer for parents, teachers, and other educators*. New York, NY: McGraw-Hill.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I: Cognitive domain*. New York, NY: McGraw-Hill.
- Bloom, B. S., & Sosniak, L. A. (1985). *Developing talent in young people*. Ballantine Books.
- Blyth, A. (1981). From individuality to character: The Herbartian sociology applied to education. *British Journal of Educational Studies*, 29(1), 69-79.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: Application of an EM algorithm. *Psychometrika*, 46(4), 443-459.
- Borsboom, D. (2005). *Measuring the mind: Conceptual issues in contemporary psychometrics*. Cambridge University Press.
- Bruner, J. (1977). *The process of education*. Cambridge, MA: Harvard University Press.
- Bulmer, M. G. (2003). *Francis Galton: pioneer of heredity and biometry*. JHU Press.
- Burge, L. (2008). Crafting the future: Pioneer lessons and concerns for today. *Distance Education*, 29(1), 5-17.
- Bush, G. W. (2001). No child left behind.
- Butler-Bowdon, T. (2007). *Fifty Psychology Classics*. Nicholas Brealey Publishing.
- Campion, N. (1990). Post-Fordism and research in distance education. In T. Evans (Ed.), *Research in distance education*. Geelong, AW: Deakin University.
- Cannell, J. J. (1989). How Public Educators Cheat on Standardized Achievement Tests: The "Lake Wobegon" Report.
- Carmines, E. G., & Zeller, R. A. (Eds.). (1979). *Reliability and validity assessment* (17). New York, NY: SAGE.



- Clark, J. J. (1906). The correspondence school—its relation to technical education and some of its results. *Science*, 24(611), 327-334.
- Clauser, B. E. (2007). The life and labors of Francis Galton: A review of four recent books about the father of behavioral statistics. *Journal of Educational and Behavioral Statistics*, 32(4), 440-444.
- Computer-assisted instruction (CAI). (2013). In *Encyclopædia Britannica*. Retrieved from <http://www.britannica.com/EBchecked/topic/130589/computer-assisted-instruction-CAI>
- Cremin, L. A. (1970). *American education: The colonial experience, 1607-1783* (Vol. 2). New York, NY: Harper & Row.
- Crouse, J. and Trusheim, D. (1988). *The case against the SAT*. Chicago: University of Chicago Press.
- Daniel, J. S. (1999). *Mega-universities and knowledge media: Technology strategies for higher education*. New York, NY: Kogan Page.
- Daniel, J. S. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education*, 3.
- De Vries, J. (1994). The industrial revolution and the industrious revolution. *Journal of Economic History*, 54(2), 249-70.
- Dillon, E., & Rotherham, A. J. (2007). States' evidence: What it means to make 'Adequate Yearly Progress' under NCLB. *The Explainer. Education Sector*. Washington DC [www.educationsector.org](http://www.educationsector.org).
- Distance Education. (n.d.) In Wikipedia.com. Retrieved from: [http://en.wikipedia.org/wiki/Distance\\_education](http://en.wikipedia.org/wiki/Distance_education)
- Dowbiggin, I. R. (1997). *Keeping America sane: Psychiatry and eugenics in the United States and Canada, 1880-1940*. Cornell University Press.
- Duncan, S. (2005). The US Army's impact on the history of distance education. *Quarterly Review of Distance Education*, 6(4), 397-404.
- Embretson, S., Reise, S., & Reise, S. P. (2000). *Item Response Theory for Psychologists*. Lawrence Earlbaum associates. Inc., NJ.

- Esler, W. K., & Sciortino, P. (1991). *Methods for teaching: An overview of current practices*. Raleigh, NC: Contemporary.
- Evans, T., & Nation, D. (1993). *Reforming open and distance education*. London, UK: Kogan Page.
- Florida Department of Education. (2013). Bureau of K-12 Assessment. Retrieved from <http://fcats.fldoe.org/eoc/>
- Florida Virtual School. (2010). *Florida virtual school jobs overview*. Retrieved from <http://jobs.flvs.net/careers-flvs>
- Florida Virtual School. (2013). *Online learning solutions*. Retrieved from <http://www.flvs.net/educators/Pages/learning-solutions.aspx>
- Forrest, D. W. (1974). *Francis Galton: The life and work of a Victorian genius*. Oxford, England: Taplinger.
- Galton, F. (1883). *Inquiries into human faculty and its development*. Macmillan.
- Galton, F. (1907). *The Ballot box*. Retrieved from [http://galton.org/cgi-bin/searchImages/galton/search/essays/pages/galton-1907-ballot-box\\_1.htm](http://galton.org/cgi-bin/searchImages/galton/search/essays/pages/galton-1907-ballot-box_1.htm)
- Gaudelli, W. (2002). US kids don't know US history: The NAEP study, perspectives, and presuppositions. *The Social Studies*, 93(5), 197-201.
- Greene, J., Winters, M., & Forster, G. (2004). Testing high-stakes tests: Can we believe the results of accountability tests?. *The Teachers College Record*, 106(6), 1124-1144.
- Groux, Catherine. (2011) Online Education's Importance Confirmed by Release of Growth Data. Retrieved from [http://www.usnewsuniversitydirectory.com/articles/online-educations-importance-confirmed-by-release\\_11805.aspx](http://www.usnewsuniversitydirectory.com/articles/online-educations-importance-confirmed-by-release_11805.aspx)
- Guri-Rosenblit, S. (2009). *Digital technologies in higher education: Sweeping expectations and actual effects*. New York, NY: Nova Science.
- Gutlek, G. (1983). *Education and schooling in America*. Prentice-Hall.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Harasim, L. M. (1990). *Online education: Perspectives on a new environment*. Greenwood.

- Harasim, L. M. (1996). Online education. *Computer networking and scholarly communication in the twenty-first-century university*, 203-214.
- Henig, R. J. (2013). The politics of testing when measures 'go public'. *Teachers College Record*, 115(9).
- Herbst, J. (1997). *The once and future school: Three hundred and fifty years of American secondary education*. Psychology Press.
- Holland, J. G., & Skinner, B. F. (1961). *The analysis of behavior: A program for self-instruction*.
- Holmberg, B., Hrsg, B., & Busch, F. W. (2005). *The evolution, principles and practices of distance education* (Vol. 11). Bis.
- Horst, P. (1955). LL Thurstone and the Science of Human Behavior. *Science*, 122(3183), 1259-1260.
- International Association for K-12 Online Learning. (2012). *Fast facts about online learning*. Retrieved from [http://www.inacol.org/press/docs/nacol\\_fast\\_facts.pdf](http://www.inacol.org/press/docs/nacol_fast_facts.pdf)
- Jacobsen, D. A., Eggen, P. D., & Kauchak, D. P. (2008). *Methods for teaching: Promoting student learning in K-12 classrooms*. Allyn & Bacon.
- Jethro, O. O., Grace, A. M., & Thomas, A. K. (2012). E-learning and its effects on teaching and learning in a global age. *Indian Journal of Education and Information Management*, 1(2), 73-78.
- Johnson, A. P. (2006). *Making connections in elementary and middle school social studies*. SAGE.
- Johnson, H. (2007). Dialogue and the construction of knowledge in e-learning: Exploring students' perceptions of their learning while using Blackboard's asynchronous discussion board. *European Journal of Open, Distance and E-Learning*, 1.
- Kaufmann, M. (2012). *The Internet Revolution is the New Industrial Revolution*. Retrieved from <http://www.forbes.com/sites/michakaufman/2012/10/05/the-internet-revolution-is-the-new-industrial-revolution/>
- Kett, J. F. (1994). *The pursuit of knowledge under difficulties: From self-improvement to adult education in America 1750-1990*. Stanford University Press.
- Kiess, H. O. (1989). *Statistical concepts for the behavioral sciences*. Allyn & Bacon.

- King, K. (2008). Instructional Television and Educational Media Resources at the National Public Broadcasting Archives. *TechTrends*, 52(4), 59.
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.
- Kline, P. (1979). *Psychometrics and psychology*. London: Academic Press.
- Kornblith, G. J., & Lasser, C. (2004). "Will That Be on the Exam?" The Role of Testing in Teaching and Learning American History. *Journal Of American History-Bloomington*, 90(4), 1379-1380.
- Lapin, L. L. (1973). *Statistics of modern business decisions*. Harcourt Brace Jovanovich.
- Lee, F. (2008). Technopedagogies of mass-individualization: correspondence education in the mid twentieth century. *History and Technology*, 24(3), 239-253.
- Lefrancois, G. (1975). *Psychology for teaching*. Belmont, CA: Wadsworth Publishing.
- Lefrancois, G. (1972). *Psychological theories and human learning*. Monterey, CA: Brooks/Cole.
- Lemaige, Sarah. (2011). Audit flags dropout rates, scores at online schools. *Star Tribune*. Retrieved from <http://www.startribune.com/local/south/130145813.html>
- Linn, Robert L., M. Elizabeth Graue, and Nancy M. Sanders. "Comparing state and district test results to national norms: The validity of claims that "everyone is above average"." *Educational Measurement: Issues and Practice* 9.3 (1990): 5-14.
- Lips, D. (2010). *How online learning is revolutionizing k-12 education and benefiting students*. Retrieved from <http://www.heritage.org/research/reports/2010/01/how-online-learning-is-revolutionizing-k12-education-and-benefiting-students>
- Lewin, T. (2013). Universities abroad join partnerships on the web. *The New York Times*, 21.
- Loutchko, I., Kurbel, K., & Pakhomov, A. (2002, May). Production and delivery of multimedia courses for internet based virtual education. In *the World Congress" Networked Learning in a Global Environment: Challenges and Solutions for Virtual Education"*, Berlin, Germany (pp. 79-96).

- Lovie, P., & Lovie, A. D. (1996). Charles Edward Spearman, FRS (1863-1945). *Notes and Records of the Royal Society of London*, 75-88.
- Martin, O. (1996). [Psychological measurement from Binet to Thurstone,(1900-1930)]. *Revue de synthese/Centre international de synthese*, (4), 457-493.
- Maxwell, B. (2011). Education Mandate for Online Class Puts Pressure on Florida's Rural Districts. *Tampa Bay Times*. Retrieved from <http://www.tampabay.com/opinion/columns/education- mandate-for-online-class-puts-pressure-on-floridas-rural/1194575>
- McCoog, I. J. (2005). American History's Problem with Standardized Testing. *Online Submission*.
- McDonald, R. P. (1999). Test theory: A unified treatment. *Mahwah, NJ: Lawrence Erlbaum*.
- Meisels, S. J. (1989). High-Stakes Testing in Kindergarten. *Educational Leadership*, 46(7), 16-22.
- Miller, E.J. (2003). "Teaching methods, the Herbartian revolution and Douglas Clay Ridgley at Illinois State Normal University". *Journal of Geography* 102, 110–120.
- Miller, G. (1974). *The American Journal of Psychology* 87(1/2), 279-288
- Miller, J. (1991). *Report questioning 'crisis' in education triggers an uproar*. Retrieved from <http://www.Edweek.org>
- Mondale, S., & Patton, S. B. (Eds.). (2001). *School, the story of American public education*. Boston: Beacon Press.
- Moore, M. G. & Kearsley, G. (2005). *Distance education. A systems view* (2nd ed.). Belmont, CA: Thomson Wadsworth.
- Moore, M. G., & Kearsley, G. (2011). *Distance education: A systems view of online learning*. CengageBrain. com.
- NCLB. (n.d.) In [legaldefinitions.com](http://legaldefinitions.com). Retrieved from <http://definitions.uslegal.com/n/no-child-left-behind-act/>
- Nasseh, B. (1997). A brief history of distance education. *Adult Education in the News*.

- Nelson, R. M., Pettersson, M. E., & Carlborg, Ö. (2013). A century after Fisher: time for a new paradigm in quantitative genetics. *Trends in Genetics*, 29(12), 669-676.
- Nichols, S. L., Glass, G. V, Berliner, D.C. (2012) High-stakes Testing and Student Achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives*, 20 (20)
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high-stakes testing corrupts America's schools*. Cambridge, MA: Harvard Education Press.
- Partnership for Assessment of Readiness for College and Careers. (2014). Retrieved from <https://www.parconline.org/>
- Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 50(302), 157-175.
- Pearson, Karl (1901). *National Life from the Standpoint of Science*. Retrieved from <http://archive.org/stream/nationallifefro00peargoo#page/n6/mode/2up>
- Pittman, V. V. (2003). Correspondence study in the American university: A second historiographic perspective. *Handbook of Distance Education*, 21-35.
- Power, M., & Vaughan, N. (2010). Redesigning online learning for international graduate seminar delivery. *Journal of Distance Education/Revue de l'éducation à distance*, 24(2). Retrieved from <http://www.jofde.ca/index.php/jde/issue/view/59>
- Power, M. (2008). The emergence of blended online learning. *Journal of Online Learning & Teaching*, 4(4). Retrieved from [http://jolt.merlot.org/vol4no4/power\\_1208.htm](http://jolt.merlot.org/vol4no4/power_1208.htm)
- psychometrics. (n.d.). *Collins English Dictionary - Complete & Unabridged 10th Edition*. Retrieved February 15, 2014, from Dictionary.com website: <http://dictionary.reference.com/browse/psychometrics>
- Ravitch, D. (2002). *A brief history of testing and accountability*. Retrieved from <http://www.hoover.org/publications/hoover-digest/article/7286>
- Ravitch, D. (2007) *Is U.S. education better than ever?* Retrieved from [http://www.huffingtonpost.com/diane-ravitch/is-us-education-better-\\_b\\_75441.html](http://www.huffingtonpost.com/diane-ravitch/is-us-education-better-_b_75441.html)

- Ravitch, Diane. (2012). *For-profit online charter scandal*. Retrieved from <http://dianeravitch.net/2012/09/02/for-profit-online-charter-scandal/>
- Ravitch, D. (2011). *The death and life of the great American school system: How testing and choice are undermining education*. Basic Books.
- Reid, C. A., Kolakowsky-Hayner, S. A., Lewis, A. N., & Armstrong, A. J. (2007). Modern psychometric methodology applications of item response theory. *Rehabilitation Counseling Bulletin*, 50(3), 177-188.
- Reid, S., & Day, D. (1942). Radio and records in education. *Review of Educational Research*, 12(3), 305-322.
- Rudalevige, A. (2003). No Child Left Behind. *No child left behind*, 23-54.
- Saccuzzo, D. P. (2012). *Psychological testing: Principles, applications, & issues*. Cengage Learning.
- The Sandia Report and U.S. Achievement: An Assessment. (1994). *The Journal of Educational Research*, 87(3), 133-47.
- Sammons, M. C., & Ruth, S. (2007). The invisible professor and the future of virtual faculty. *International Journal of Instructional Technology and Distance Learning*, 3(1).
- Seifert, K and Sutton, R. (2010). *Educational Psychology*. New York. Orange Grove Texts Plus.
- Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching the test?. *Educational Measurement: Issues and Practice*, 9(3), 15-22.
- Shepard, L. A. (2013). Validity for What Purpose?. *Teachers College Record*, 115(9).
- Skinner, B. F. (1984). The shame of American education. *American Psychologist*, 39(9), 947.
- Skinner, B. F. (1986). Programmed Instruction Revisited. *Phi Delta Kappan*, 68(2), 103-10.
- Smith, S. (1946). On the Theory of Sales of Measurement. Retrieved from: [http://www.mpopa.ro/statistica\\_licenta/Stevens\\_Measurement.pdf](http://www.mpopa.ro/statistica_licenta/Stevens_Measurement.pdf)

- Stephen, B., & Morris, M. (2013) *The inevitable coming impact of online education on state universities and rational response*. Retrieved from <http://www.public.iastate.edu/~vardeman/What%20is%20Coming-Final.pdf>.
- Stern, S. (1994). Beyond the Rhetoric: An Historian's View of the National Standards for United States History. *Journal of Education*, 176(3), 61-71.
- Stevens, S. S. (1946). On the theory of scales of measurement.
- Streiner, D. L. (2003). Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, 80(1), 99-103.
- Sumner, J. (2000). Serving the system: A critical history of distance education. *Open learning*, 15(3), 267-285.
- Tankard, J. W. (1984). *The statistical pioneers*. Cambridge, MA: Schenkman Publishing Company.
- Thompson, K. (2005). *Constructing educational criticism of online courses: A model for implementation by practitioners*. Unpublished doctoral dissertation. University of Central Florida: Orlando, FL
- Thompson, N., & Wiess, D. (2009). Computerised and adaptive testing in educational assessment. *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing*, 127-133.
- Thomson, G. (1947). Charles Spearman. 1863-1945. *Obituary Notices of Fellows of the Royal Society*, 5(15), 373-385.
- Thorndike, E.L. (1903). *Educational psychology: Briefer course*. New York. Science Press.
- Thorne, B. M., & Henley, T. B. (1997). *Connections in the history and systems of psychology*. Houghton, Mifflin and Company.
- Traub, R. E. (2005). Classical test theory in historical perspective. *Educational Measurement: Issues and Practice*, 16(4), 8-14.
- Tyler, L. E. (1963). *Tests and measurements*. Englewood Cliffs, NJ: Prentice Hall.
- United States Department of Education. (2009). *Race To The Top*. Retrieved from <http://www2.ed.gov/programs/racetothetop/index.html>



- U.S. Department of Education. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies* Retrieved from: <http://www2.ed.gov/rschstat/eval/tech/evidence-based-practices/finalreport.pdf>
- Vollmer, Jamie. (2010). *Schools cannot do it alone*. Fairfield, IA: Enlightenment Press
- Welner, K. G. (2013). Consequential validity and the transformation of tests from measurement tools to policy tools. *Teachers College Record*, 115(9), 1.
- Woodworth, R. S. (1944). James McKeen Cattell: 1860-1944. *Psychological Review*, 51(4), 201.
- Woolfolk, A. (2004). *Educational psychology*. (9<sup>th</sup> Ed.). Boston, MA: Pearson Allyn & Bacon.
- Wu, J. (2005). A View from the Classroom. *Educational Leadership*, 62(4), 40-44.
- Xu, D., & Jagers, S. S. (2011). *Online and hybrid course enrollment and performance*. Retrieved from <http://ccrc.tc.columbia.edu/Publication.asp?UID=872>
- Yu, C. H., & Ds, P. (2010). A Simple Guide to the Item Response Theory (IRT) and Rasch Modeling.
- Zemsky, R., & Massy, W. F (2004). *Thwarted innovation. What happened to e-learning and why?* Retrieved from <http://www.irhe.upenn.edu/Docs/Jun2004/ThwartedInnovation.pdf>