TRENDS IN NAEP SCORES AMONG 17-YEAR-OLD STUDENTS IN THE ERA OF ACCOUNTABILITY

A Dissertation

Presented to

The College of Graduate and Professional Studies

Department of Educational Leadership

Indiana State University

Terre Haute, Indiana

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

© Katherine Kyler

December 2017

Keywords: school accountability, NAEP, secondary student success, graduation rates

VITA

Katherine J. Kyler

EDUCATION

2017	Indiana State University, Terre Haute, Indiana Ph.D. in Educational Leadership
2016	Indiana State University, Terre Haute, Indiana Licensure in K-12 School Administration
2012	Marian University, Indianapolis, Indiana M.A. in Teaching
2010	Marian University, Indianapolis, Indiana Licensure in Life Science, Indianapolis Teaching Fellows
2008	Purdue University, West Lafayette, Indiana B.S in Wildlife Biology

PROFESSIONAL EXPERIENCE

Present	Dean of Instruction, YES Prep Public Schools Houston, Texas
2012 – 2017	Science Teacher, Hope Academy Indianapolis, Indiana
2010 – 2011	Science Teacher, Stonegate Early College High School Indianapolis, Indiana
2008 – 2009	Biological Science Technician, National Park Service Porter, Indiana
2007 – 2008	Teacher's Assistant, Introduction to Environmental Science, Purdue University, West Lafayette, Indiana

COMMITTEE MEMBERS

Committee Chair: Bradley V. Balch, Ph.D.

Professor of Educational Leadership & Dean Emeritus

Indiana State University

Committee Member: Terry McDaniel, Ph.D.

Professor of Educational Leadership

Indiana State University

Committee Member: Mary Jo Ratterman, Ph.D.

Owner and Operator

Research and Evaluation Resources

ABSTRACT

Education has undergone a sweeping renovation throughout the last several decades as part of the school accountability movement aimed to increase student success. High school graduation rates are the highest they have been in decades. School accountability measures continue to be implemented and modified with a goal of increasing student success and closing the achievement gap (Maleyko & Gawlik, 2011). Accountability measures are in place that require data analysis and reporting of information such as graduation rates and standardized test scores (No Child Left Behind Act [NCLB], 2008). While it is important to hold schools accountable, many of the currently utilized methods to measure student success can be manipulated to improve school and district ratings (Maleyko & Gawlik, 2011). The purpose of this quantitative study was to better understand the relationship between select student demographics and low-stakes the National Assessment of Educational Progress (NAEP) scores for time periods before and after the implementation of numerous school accountability measures. Specifically, I analyzed data sets from 1990 and 1999 for the time period before NCLB (2008) and data sets from 2004 and 2012 for the time period after the implementation of NCLB. This data was examined using independent samples t tests and Cohen's d statistic. Data analysis showed that there was a significant increase in NAEP Math scores for 17-year old students in the time period before NCLB but not after. NAEP English scores did not show a significant difference before NCLB but did show a significant increase after NCLB. While NAEP scores pre and post NCLB do not demonstrate significant changes in student success, graduation rates continue to rise. This findings and conclusions of this study will benefit school districts and policy makers when

considering the effectiveness of past school accountability measures. Additionally, this study provides an example of the inconsistencies associated with high stakes measures of student success and highlights the importance of alternate indicators of success.

PREFACE

The basis of this study originally developed from my passion for access to higher education for all students. I felt as if students were less prepared for college today than they were in the past. However, the media consistently reported that high school graduation rates were higher than they had ever been. This made me wonder about the effectiveness of traditional student success indicators. The main question in my mind was, “Are we truly improving education for our students, or does it only appear that way?”

I decided I would investigate the methods used to report student success. I was also determined to investigate other, nontraditional indicators of student success to see if the data provided the similar results. I have a background in science and I consistently check sources and data validity before truly believing what appears on the surface.

ACKNOWLEDGMENTS

The path leading me to the world of education was indirect and influenced by many things. My mother is in education and although she never pushed me into it, education is where I ultimately landed. I always considered getting a Ph.D. but did not think it was something I could do at a young age. My mother pushed me to pursue further education while I was young and encouraged me to start my doctoral program. She helped me write countless essays, resumes, and cover letters along the way to get me to the place I am today and I am incredibly thankful for her support. My entire family has rallied behind me throughout my doctoral program, including my father, stepmother, brother, sister, and grandparents and I feel very fortunate to have such a solid foundation of support.

Dr. Balch was a professor I had never met but heard so many wonderful things about. Many people at Indiana State University mentioned him and thought he would be a great chair for my dissertation committee. I reached out to him with the idea for my dissertation and asked if he would be my chair and he graciously accepted. I ended up meeting Dr. Balch during my Saturday principal internship classes. I always loved hearing his wisdom as he spoke to the class. I would like to thank Dr. Balch for all the time and support he has provided me throughout the entire dissertation process. He helped me navigate the extremely difficult process of getting a restricted use data license from the National Center for Education Statistics and remained kind and patient throughout it. I have deep respect for Dr. Balch and would love to follow his career path in the future.

All of the staff members at Indiana State University have been incredibly kind and supportive throughout my entire doctoral program. Dr. McDaniel was one of the first professors I spoke with and he always made me feel important and provided kind, encouraging words of support. He joined my committee and provided many helpful improvements and much needed support. Dr. Donlan provided invaluable help when I was mapping out my dissertation idea. Dr. Monahan was very gracious and worked to get me into and through the internship program. Dr. Hampton taught me much about statistics and helped me navigate the plan for my study. Matt Jenkins provided much needed assistance in securing and safeguarding the NCES dataset for the study. Thank you all for being so helpful and kind.

I was fortunate enough to have local help throughout my doctoral process as well. Dr. Ratterman helped me when my dissertation idea was just forming and she pointed me in the right direction which ultimately led to my final dissertation. I would like to thank her for providing ideas and support, especially in the very beginning when I had an overwhelmingly vast study idea. I would also like to thank Burt Weed for helping me with formatting, time management, and proof reading.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Education has undergone a sweeping renovation throughout the last several decades as part of the school accountability movement aimed to increase student success. *A Nation at Risk* was published in 1983 and stated that the United States was losing its international edge due to low education standards (as cited in Fraser, 2014). The response to this document was astonishing and over 250 education task forces were created in the following year. Numerous educational laws were put in place that impact the current education system (Mehta, 2015). Politicians and the media continue to cite the growing high school graduation rate as evidence of improved student success in the era of school accountability (United States Department of Education [USDOE], 2015). High school graduation rates are the highest they have been in decades. According to the National Center for Education Statistics (NCES), students in the U.S. had an 81% high school graduation rate for the 2012-2013 school year, the highest rate since the cohort method for graduation rate calculations was adopted (USDOE, 2015). U.S. Secretary of Education Arne Duncan stated that

> America's students have achieved another record-setting milestone. We can take pride as a nation in knowing that we're seeing promising gains, including for students of color. This is a vital step toward readiness for success in college and careers for every student in this country, and these improvements are thanks to the hard work of teachers, principals,

students, and families. (USDOE, 2015, para. 2)

While politicians and newscasters continue to cite the high graduation rate as a gauge of the effectiveness of school accountability measures, it is important to investigate other indicators of student success. The graduation rate has been utilized as a measure of school performance and while it does serve as one kind of indicator, it is important to look into other measures (Stanley, Spradlin, & Plucker, 2008).

The National Assessment of Educational Progress (NAEP) has been investigating trends in student success since the 1970s (USDOE, 2013). The NAEP is a low stakes assessment that is used to identify educational trends. The NAEP utilizes Item Response Theory (IRT) which enables comparison of different data sets over decades (Hambleton & Jones, 1993). This quantitative study will explore NAEP scores of 17-year old students in reading and mathematics before and after the implementation of the No Child Left Behind Act (NCLB, 2008), a reauthorization of the Elementary and Secondary Education Act (ESEA).

Background of the Problem

In the past, secondary education was a luxury, not a requirement. Secondary education was not commonplace in the 1800s and the first graduation rate was not reported until 1869 (USDOE, 2014). Today, secondary education is customary and educational reforms emphasize the importance of the graduation rate. In 1869, the graduation rate, calculated as the proportion of the 17-year old population with a diploma, was only 2%. The national public high school graduation rate has peaked in recent years at 81% as noted by the USDOE (2015). While this is a substantial improvement since the 1800s it is important to examine other factors in addition to the graduation rate (Stanley et al., 2008).

The increase in secondary education came with legislation surrounding its improvement.

The original ESEA was signed into law in 1965 with the goal of improving education for the poor. In 1988 The ESEA was reauthorized by the Hawkins-Stafford Elementary and Secondary School Improvement Amendments which required the analysis of annual dropout and retention rates by the NCES (Seastrom et al., 2006). Since the reauthorization of ESEA the NCES reports have also included high school completion and graduation rates (Seastrom et al., 2006). The No Child Left Behind Act of 2001 was an impactful reauthorization of ESEA (NCLB, 2008). The goal of NCLB was to close the achievement gap and have all students proficient in math and reading by 2014 (Bell & Meinelt, 2011). The National Assessment Governing Board defines proficiency as "students reaching this level have demonstrated competency over challenging subject matter, including subject matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to subject matter" (National Assessment Governing Board; NAGB, 2015, p. 10). The reporting of high school graduation rates was required under the reauthorization of ESEA in the form of NCLB (Bell & Meinelt, 2011; Trolian & Fouts, 2011). While NCLB required the reporting of graduation rates, it did not specify how graduation rates must be calculated, making comparison meaningless (Swanson, 2003). The most recent reauthorization of ESEA was signed into law on December 10, 2015 as the Every Student Succeeds Act (ESSA). Many versions of this law were proposed and revised before ESSA passed with bipartisan support. The ESSA replaces the highly criticized NCLB (Darrow, 2016).

The NAEP "is the only nationally representative and continuing assessment of what U.S. students know and can do in various subject areas" (Hombo, 2003, p. 59). While testing was commonplace in the era of NCLB, states were able to design and implement their own standardized tests, which made meaningful comparison between states impossible (NCES, 2010). The NAEP is unique as it "provides a common measure of student achievement across

the country" (NCES, 2010, p. 1). The NAEP analyzes the educational attainment of students over time (NCES, 2009). The information gathered by the NAEP is used by policy makers and educators as a guide to student progress (NCES, 2009). NAEP results can be compared across states as the test and administration of the NAEP is the same in all states (NCES, 2010). The NAEP is designed to produce accurate and useful results that can be used to indicate educational progress (Beaton & Zwick, 1992). The NAEP consists of two kinds of assessments, the main NAEP and the long term trend NAEP (NCES, 2010). The main NAEP is given to fourth, eighth, and twelfth grade students across the nation annually and tests subjects such as math, reading, science, and writing. Long term trend NAEP is given to 9, 13, and 17-year old students across the nation every four years and tests math and reading skills. Results from the long term trend NAEP allow comparison of student progress over time and will be used for this study (NCES, 2010).

The most recent reauthorization of ESEA was signed into law on December 10, 2015 as the Every Student Succeeds Act ESSA. Many versions of this law were proposed and revised before ESSA (2015) passed with bipartisan support. The ESSA (2015) replaces the highly criticized NCLB (Darrow, 2016). ESSA provides more power to states and districts and decreases federal government's role in education. Many of the proponents of NCLB will remain, such as annual testing and accountability, however states and districts will be responsible for this (Darrow, 2016). As ESSA (2015) was just passed into law, the research in this paper will focus on the impacts of NCLB.

Statement of the Problem

School accountability measures continue to be implemented and modified with a goal of increasing student success and closing the achievement gap (Maleyko & Gawlik, 2011).

Accountability measures are in place that require data analysis and reporting of information such as graduation rates and standardized test scores (NCLB, 2008). While it is important to hold schools accountable, many of the currently utilized methods to measure student success can be manipulated to improve school and district ratings (Maleyko & Gawlik, 2011). As school accountability measures persist and evolve over time, it is important to determine any impacts the NCLB (2008) has had. If teachers and schools are to improve, there needs to be further research to determine the effectiveness of accountability measures that have been put in place in order to improve future methods (Stanley et al., 2008).

Purpose of the Study

The purpose of this quantitative study was to better understand the relationship between select student demographics and NAEP scores for time periods before and after the implementation of numerous school accountability measures. Specifically, this study will analyze data sets from 1990 and 1999 for the time period before NCLB and data sets from 2004 and 2012 for the time period after the implementation of NCLB. The independent variables in this study were year and race/ethnicity. The dependent variables in this study were long term trend NAEP scores in Math and Reading.

Significance of the Study

This study is important because it expands upon traditional indicators of student success utilized for accountability. This study investigated other indicators of success that do not allow flexibility by states or districts. While there have been many studies evaluating NCLB and NAEP independently, there is a paucity of information examining NAEP scores before and after NCLB's implementation. Future accountability measures should expand upon success or failure of past accountability measures. If past accountability measures are shown to be effective across

multiple indicators of student success then moving forward researchers can continue to implement the effective strategies. If accountability measures are shown to be ineffective then moving forward researchers can modify them to meet the goal of increasing student success. School accountability measures continue to be at the center of education and should be effective if they are going to be used in the future or modified so that they are effective at increasing student success. The impacts of past accountability measures could be important to legislators and others involved in the development of future accountability measures.

Research Questions and Hypotheses

The core research question leading this study is, "In the current era of school accountability, are there indicators of student success other than graduation rates among 17-year old test takers on the long term trend National Assessment of Educational Progress?" Additional research questions include:

R1. Is there a change among 17-year old test takers in 1990-1999 NAEP Math and Reading scores.

R2. Is there a change among 17-year old test takers in 2004-2012 NAEP Math and Reading scores.

R3. Are there significant differences among 17-year old test takers by race/ethnicity in 1990-1999 NAEP Math and Reading scores.

R4. Are there significant differences among 17-year old test takers by race/ethnicity in 2004-2012 NAEP Math and Reading scores.

R5. Is the mean score change between 1990 and 1999 comparable to the mean score change between 2004 and 2012, as measured by Cohen's d statistic, for both reading and math?

To investigate this, long term trend NAEP scores for 17-year-old students were analyzed. To better understand NAEP scores in the era of school accountability, the following null hypotheses were examined to fully address the core research question:

$H_0$1. There was not a change among 17-year old test takers in 1990-1999 NAEP Math and Reading scores.

$H_0$2. There was not a change among 17-year old test takers in 2004-2012 NAEP Math and Reading scores.

$H_0$3. There were not significant differences among 17-year old test takers by race/ethnicity in 1990-1999 NAEP Math and Reading scores.

$H_0$4. There were not significant differences among 17-year old test takers by race/ethnicity in 2004-2012 NAEP Math and Reading scores.

Additionally, the research question "Is the mean score change between 1990 and 1999 comparable to the mean score change between 2004 and 2012, as measured by Cohen's d statistic, for both reading and math?" was addressed.

Research Design

For this quantitative research study, data regarding the scores of 17-year-old students on mathematics and reading NAEP assessments were analyzed. These data were collected through the National Center for Education Statistics and was analyzed using the Statistical Package for the Social Sciences (SPSS) software program. The data analysis in this study included analysis of descriptive statistics, independent samples t tests, and Cohen's d statistic.

Delimitations

1. The study population is restricted to 17-year-old students.
2. Long term trend data include only Math and Reading scores.

3. Long term trend data are collected at the national level only and cannot be broken down by state.
4. NAEP does not assess every student in the nation, but uses a representative sample based on the racial, geographical, ethnic, and socioeconomic diversity of students in the United States.

Definition of Terms

The following definitions are provided to enhance understanding:

Achievement gap is defined as "the gap between the academic achievement of students from traditionally underrepresented or minority backgrounds and white students" (Span & Rivers, 2012, p. 2).

Elementary and Secondary Education Act (ESEA) is defined as the law signed by President Johnson in 1965 with the goal of improving education for the poor (Seastrom et al., 2006).

Every Student Succeeds Act (ESSA) is defined as the law signed by President Obama on December 5, 2015 and is a reauthorization of the Elementary and Secondary Education Act of 1965. The goal of ESSA is to focus on complete preparation for success in college and careers for all students (USDOE, 2016).

Graduation rate is defined as the proportion of individuals completing high school through one of the methods discussed in this research.

Human capital is defined as encompassing the notion that there are "investments in people (e.g., education, training, health) and that these investments increase an individual's productivity" (Goldin, 2014, p. 1).

National Assessment of Educational Progress (NAEP) is defined as a "continuing and nationally representative assessment of what our nation's students know and can do. NAEP has often been called the 'gold standard' of assessments because it is developed using the best thinking of assessment and content specialists, education experts, and teachers from around the nation" (USDOE, 2010, p. 2). The National Assessment Governing Board is responsible for NAEP policy setting (USDOE, 2010).

National Center for Education Statistics (NCES) is defined as the agency responsible for the development of test questions, administration of the assessment, scoring of responses, analysis of data, and reporting of the results for the NAEP. NCES is a division of the Institute of Education Science within the U.S. Department of Education (USDOE, 2010).

No Child Left Behind Act (NCLB) is defined as a reauthorization of ESEA signed into law by George W. Bush in 2002 with goals of having all students proficient in reading and math by 2014 and closing the achievement gap between socioeconomic and racial groups (NCLB, 2008; Bell & Meinelt, 2011).

Proficiency is defined as "students reaching this level have demonstrated competency over challenging subject matter, including subject matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to subject matter" (National Assessment Governing Board; NAGB, 2015, p. 10).

Socioeconomic status (SES) is defined as "one's access to financial, social, cultural, and human capital resources. Traditionally, a student's SES has included, as components, parental educational attainment, parental occupational status, and household or family income, with appropriate adjustment for household or family composition" (Cowan et al., 2012, p. 4)

School accountability is defined as "using administrative data-based mechanisms aimed at increasing student achievement" (Loeb & Figlio, 2011, p. 384).

Student success is defined as the ability to master and apply specific content knowledge and skills.

Summary and Organization of the Study

This quantitative study focused on student success in the era of school accountability. The purpose of this study was to better understand the relationship between select student demographics and NAEP scores for time periods before and after the implementation of numerous school accountability measures. This study consists of five chapters. Chapter 1 provided an introduction, background of the problem, statement of the problem, purpose of the study, significance of the study, research questions, research design, delimitations, definition of terms, and summary. Chapter 2 provides a review of the literature regarding the history of the high school graduation rate, the achievement gap, the Elementary and Secondary Education Act, the No Child Left Behind Act, the National Assessment of Educational Progress, and the theoretical framework. Chapter 3 describes the research methodology and procedures that were utilized in this study. Chapter 4 explains the data analysis and Chapter 5 provides a review of the findings, conclusions, limitations, and recommendations for future research.

CHAPTER 2

LITERATURE REVIEW

This review of the literature and study findings is intended to analyze traditional student achievement indicators and their reliability, particularly at the high school level of Grades 9 through 12. The literature review, in part, answers the research question, "In the current era of school accountability, are there indicators of student success other than graduation rates among 17-year old test takers on the long term trend National Assessment of Educational Progress?" To analyze this grand tour research question, this chapter consists of six sections, (a) the history of the high school graduation rate, (b) the achievement gap,(c) the Elementary and Secondary Education Act, (d) the No Child Left Behind Act, (e) the National Assessment of Educational Progress, and (f) the theoretical framework.

The History of the High School Graduation Rate

The national public high school graduation rate has peaked in recent years at 81% as noted by the United States Department of Education (USDOE; 2015). The first graduation statistic was not reported until 1869 and at that time was determined as the proportion of the 17-year-old population with a diploma at 2% (USDOE, 2014). Secondary education, specifically Grades 9 through 12, has now become customary and many educational reforms have emphasized the high school graduation rate as a highly important factor for measuring student and school success (Fraser, 2014; USDOE, 2015). While it is important to know the proportion

of high school students who ultimately graduate, it is necessary to explore the disparities and similarities with which graduation rates are calculated before using it as a point of reference or comparison (Stanley et al., 2008).

The first American high school opened in Boston in 1821 (Fraser, 2014). Attending high school at the time was not a requirement, but a luxury for students. As noted, in 1870, only 2% of 17-year-old citizens earned a high school diploma but in the years following, the number slowly rose to 6.4% by 1900 (USDOE, 2014). The graduation rate did not have a significant increase until after World War I when it increased to 16.8% in 1920 (Fraser, 2014; USDOE, 2014). It was not until World War II ended that most American students received a high school diploma (Fraser, 2014).

Secondary education quickly became a priority in the United States over the next 50 years (Fraser, 2014). The rise in the proportion of the 17-year-old graduates is illustrated in Table 1 (NCES, 2014). Prior to 1929, high school graduation rates were only reported every decade, but beginning in 1930 high school graduation rates were reported annually (Snyder, 1993). Beginning in 1970, the Average Freshman Graduation Rate (AFGR) was reported as well, which will be discussed in further detail later in this section (NCES, 2014).

Table 1

High school graduates 1869-2012

School year	Population		17-year-old population graduate ratio
	Total	Population 17 years old	
1869-1870	16,000	815,000	2.0
1879-1880	23,634	946,026	2.5
1889-1890	43,731	1,259,177	3.5
1899-1900	94,883	1,489,146	6.4
1909-1910	156,429	1,786,240	8.8
1919-1920	311,266	1,855,173	16.8
1929-1930	666,904	2,295,822	29.0
1939-1940	1,221,475	2,403,074	50.8
1949-1950	1,199,700	2,034,450	59.0
1959-1960	1,858,023	2,672,000	69.5
1969-1970	2,888,639	3,757,000	76.9
1979-1980	3,042,214	4,262,000	71.4
1989-1990*	2,574,162	3,505,000	73.4
1998-1999	2,758,655	3,917,885	70.4
1999-2000	2,832,844	4,056,639	69.8
2000-2001	2,847,973	4,023,686	70.8
2001-2002	2,906,534	4,023,968	72.2
2002-2003	3,015,735	4,125,087	73.1
2003-2004*	3,054,438	4,113,074	74.3
2004-2005	3,106,499	4,120,073	75.4
2005-2006*	3,122,544	4,200,554	74.3
2006-2007	3,199,650	4,297,239	74.5
2007-2008	3,312,337	4,436,955	74.7
2008-2009	3,347,828	4,336,950	77.2
2009-2010	3,440,185	4,311,831	79.8
2010-2011	3,449,719	4,366,292	79.0
2011-2012	3,452,470	4,291,741	80.4

Note. Adapted from U.S. Department of Education, National Center for Education Statistics, Digest of Education Statistics, "Table 219.10, High School Graduates, by Sex and Control of School: Selected Years 1869-70 through 2023-24".

* Includes imputations for non-reporting states

In more recent decades secondary education has become vital for success and much emphasis has been placed on increasing high school graduation rates for all students, not only the affluent (USDOE, 2015). The Elementary and Secondary Education Act (ESEA) was signed into law in 1965 by President Johnson with the goal of improving education for the poor. In 1988 the ESEA was reauthorized using the Hawkins-Stafford Elementary and Secondary School Improvement Amendments which required the NCES to analyze school dropout and retention rates each year (Seastrom et al., 2006). Since the 2001 reauthorization of ESEA the NCES reports have also included high school completion and graduation rates (Seastrom et al., 2006). The 2001 reauthorization of ESEA was the No Child Left Behind Act (NCLB, 2008).

In January of 2002, President George W. Bush signed the NCLB into law with goals of having all students proficient in reading and math by 2014 and closing the achievement gap between socioeconomic and racial groups (Bell & Meinelt, 2011). The National Assessment Governing Board defines proficiency as "students reaching this level have demonstrated competency over challenging subject matter, including subject matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to subject matter" (NAGB, 2015, p. 10). The reporting of high school graduation rates was required under the reauthorization of ESEA to NCLB (Bell & Meinelt, 2011; Trolian & Fouts, 2011). Various reauthorizations of ESEA and more specifically, NCLB (2008) will be discussed in more detail later in this section.

While NCLB (2008) required state officials to report graduation rates, it did not specify how the rates must be calculated (as cited in Swanson, 2003). Swanson (2003) reported that a review of NCLB accountability plans revealed that states were using varying methods for graduation rate calculations and that

> The most common approach, pursued by 30 states (including the District), adopts a method developed by the National Center for Education Statistics, the U.S. Department of Education's statistical agency. In the initial stages of implementing their NCLB accountability systems, only 10 states intend to use a true longitudinal graduation rate calculated using data from individual students tracked over time. The accountability workbooks for the remaining 11 states include a *diverse* array of strategies for meeting compliance with the law. These approaches range from using a dropout rate rather than a graduation rate per se (2 states), to calculating completion ratios (4 states), to employing other methods including grade-to-grade promotion ratios (5 states in all). (p. 2)

The most common calculation methods used will be explained in detail. According to Swanson (2003), the longitudinal rate is defined as the

> percent of students from an entering 9th grade cohort who graduate with a regular diploma in four years. Adjustments to the original cohort may be made for students who join or leave the school system at grade-level during that four-year period. (p. 4)

The rate defined by the NCES is "regular diploma recipients as a percent of students leaving high school over a four-year period (estimated as the sum of diploma recipients and dropouts during the past four years in grades 9 through 12 respectively)" (as cited in Swanson, 2003, p. 4). The leaver rate is designated as the "proportion of the students leaving high school who have received a high school diploma for completing a public secondary education program" (Swanson, 2004, p. 45). The basic completion ratio is described as the

> number of graduates in a particular year divided by the number of entering students at some earlier point in time. The most basic form of this indicator divides graduates by the number of 9th graders four school years earlier. Adjustments to this basic indicator may

> be made to accommodate more detailed information such as mobility in and out of a school system. (Swanson, 2003, p. 4)

The dropout rate is described as the "percent of public school students enrolled in Grades 9-12 who drop out during a given school year" (Swanson, 2004, p. 53) and is measured over the course of one year. The AFGR is calculated by dividing the number of students graduating by the number of first time freshmen four years prior (Seastrom et al., 2006). With the varying methodologies used to calculate graduation rates, it is difficult to make any kind of meaningful comparison between them. Many of these methods fail to take into account many dynamics of students, such as transfer students and years to completion (Seastrom et al., 2006). As the National Governor's Association (NGA; 2005) states, "the quality of state high school graduation and dropout data is such that most states cannot accurately account for their students as they progress through high school" (p. 7).

The use of varying calculation methods results in incomparable graduation rates that may not be very accurate (Seastrom et al., 2006). According to Heckman and LaFontaine (2007), "Depending on the data sources, definitions, and methods used, the U.S. graduation rate is estimated to be anywhere from 66% to 88% in recent years-an astonishingly wide range for such a basic statistic" (p. 3). Members of the Education Trust investigated the disparities between state reported data and a common measure and found that North Carolina documented a 97% graduation rate while The Education Trust calculated a rate closer to 64% using the Cumulative Promotion Index (Stanley et al., 2008). The Cumulative Promotion Index evaluates the "proportion of students who progress from one grade to the next at the end of the school year for Grades 9, 10, and 11 multiplied by the proportion of seniors who graduate at the end of the school year" (Seastrom et al., 2006, p. 12) and was developed by Christopher Swanson, a

recognized expert on high school graduation rates. The Cumulative Promotion Index was utilized by the Education Trust because it is regarded as a more accurate measure than most as it evaluates enrollments in lieu of dropouts (Stanley et al., 2008). Regardless of which method is the most accurate, this example illustrates the disparities which may occur in the reporting of graduation data. The NGA summarized the present issues by stating, "Because the state and federal graduation and dropout data are variable and of low quality, policymakers and practitioners cannot reasonably target resources to improving high school graduation rates and too many students fail to have their education needs met" (NGA, 2005, p. 11).

To address the issue of low quality graduation data, NGA (2005) officials determined a need for a unified calculation for graduation rates nationwide. The recommendation for a unified calculation method was the cohort model (NGA, 2005). Recommendations from NGA officials stated that schools should "adopt a standard formula for calculating a four-year, cohort-based high school graduation rate, as well as the additional indicators of high school performance" (NGA, 2005, p. 11). Through the cohort model an identification number is attached to each student so that they can be tracked even if they transfer schools. The cohort method enables schools to report much more reliable data and additional information including the time to graduation and other types of certificates obtained (Stanley et al., 2008). While this may not be the paramount method for calculating graduation rates, at minimum it provides a way to compare graduation rates using common methodology.

The cohort model for graduation rate calculation was implemented in 2008 and all schools were required to report graduation rates using this method starting with the 2010-2011 school year (Balfanz, Bridgeland, Bruce, & Fox, 2013). According to the USDOE Press Release (2012), U.S. Secretary of Education Arne Duncan stated, "By using this new measure,

states will be more honest in holding schools accountable and ensuring that students succeed. Ultimately, these data will help states target support to ensure more students graduate on time, college and career ready" (para. 3). The legalization of the cohort model encourages further communication and stewardship on educational initiatives between districts, states, and the nation. Administrative professionals will now be able to discuss the benefits and shortcomings of different methods of increasing student success by having a common method for comparison (Balfanz et al., 2013). While the cohort model is beneficial as a universal calculation method, there is a need for further collaboration to ensure the dynamics of students are accurately captured and portrayed to ensure future success for all.

The most recent reauthorization of ESEA was signed into law on December 10, 2015 as the Every Student Succeeds Act (ESSA). Many versions of this law were proposed and revised before ESSA passed with bipartisan support. The ESSA (2015) replaces the highly criticized NCLB (Darrow, 2016). ESSA (2015) provides more power to states and districts and decreases the role of the federal government in education. Many of the proponents of NCLB will remain, such as annual testing and accountability, however states and districts will be responsible for this (Darrow, 2016). The ESSA (2015) will go into full effect in the 2017-2018 school year (USDOE, 2016).

Over time the importance of secondary education has increased astronomically to the point that it is nearly universal in American schooling. Many educational reforms have taken place to focus on improving high school completion for all students, not just the wealthy (NGA, 2005). With increasing importance came increasing graduation rates until they remained reasonably steady from 1970 to 2000. Using the proportion of the 17-year old population with a high school diploma, the highest observed rate occurred in 2012 at 80.4% (USDOE, 2014).

However, this is only one method of calculating high school completion and is used solely as a basis of comparison to the rates reported in the 1800s. Graduation rates are calculated using numerous methods that all have their own limitations (Seastrom et al., 2006). There is no denying the importance of a high school diploma in the twenty-first century and increasing accountability demands aim to provide this for all (Balfanz et al., 2013).

The Achievement Gap

While the achievement gap is a term that arises frequently in everyday educational language, the term was only developed recently (Meyers, 2012). As school accountability receives more focus, discussion of the achievement gap increases. The presence of an achievement gap played a large part in the initiation of school reform (Meyers, 2012). The history of the achievement gap and its relationship to school reform today provide insight into its importance.

The achievement gap is a term that is commonplace in today's era of education. However, this term was developed rather recently (Meyers, 2012). The Civil Rights era launched a journey into school reform that has not been seen before. After the legalization of the ESEA in 1965, President Johnson used a group of researchers led by James Coleman to examine the racial disparities in student achievement (Meyers, 2012). The report by Coleman's team made the presence of an achievement gap known, but that term was not used directly (Meyers, 2012). In 1966 the media presented Coleman's teams' findings using the term achievement gap for what is likely the first time. Meyers (2012) tracked the use of the achievement gap in the media from 1851 to 2004. The first discussion of such a gap occurred in 1966 and increased slowly into the 1990s. As school accountability came to the forefront of the nation's agenda in the 2000s, the use of the achievement gap in media such as the *New York Times* and *Education*

Week skyrocketed (Meyers, 2012).

Meyers (2012) found three sharp increases in the use of the achievement gap in the media which appear to overlap with key educational movements. The first escalation in use occurred from 1997 to 1999 when states began initiating their own accountability systems. The next inflation of the term occurred from 2000-2003 when presidential debates and the reauthorization of ESEA as NCLB (2008) took place (Meyers, 2012). The most recent surge occurred between 2004 and 2007 as serious sanctions from NCLB (2008) were imposed on school districts. To confirm the rise in use of the achievement gap, Meyers (2012) examined its use in the academic journal *Phi Delta Kappan* and found similar results. The achievement gap is now included in many pieces of educational legislation, initially NCLB (Meyers, 2012). The rise in the discussion of the achievement gap comes with increased focus on what it means and how to close it.

The achievement gap remains at the forefront of education policy. Today the achievement gap is defined as "the gap between the academic achievement of students from traditionally underrepresented or minority backgrounds and white students" (Span & Rivers, 2012, p. 2). The closure of the achievement gap was a top priority in the era of NCLB. Numerous systems have been put in place to narrow the achievement gap (Span & Rivers, 2012). The National Assessment of Educational Progress (NAEP) is used for comparison of trends in the achievement gap over time as it provides long term trend data. NAEP assessments will be discussed in further detail later in this section. The first long term trend assessment was given in 1973 for mathematics and in 1971 for reading (USDOE, 2013). The scores on the long term trend NAEP over time have increased since 1971 for all ages and groups. While there is still an average 21 point gap between White students and Black and Hispanic students' scores, the

achievement gap has decreased by almost half since the initial assessment in all age groups (USDOE, 2013). The achievement gap in long term trend NAEP scores has also narrowed for 17-year old students. As the USDOE (2013) reports, "Larger gains for Black and Hispanic students than for White students narrowed the White – Black and White – Hispanic gaps to about half of what they were in the 1970s" (p. 18). While the evaluation of NAEP data can provide insight into trends in the achievement gap, no explanation is given for the cause of these changes (USDOE, 2013).

The Elementary and Secondary Education Act

President Lyndon B. Johnson signed The Elementary and Secondary Education Act (ESEA) into law in 1965 (Fraser, 2014). Johnson referred to the law as "the most sweeping educational bill ever to come before Congress. It represents a major new commitment of the Federal Government to quality and equality in the schooling that we offer our young people" (as cited in Fraser, 2014, p. 310). By signing the ESEA bill into law educational reform became a priority as never before in the United States. The years that follow ESEA only increased the prominence of school reform in the country (Meyers, 2012).

The ESEA provides financial support to schools that serve traditionally underprivileged students. As written, the purpose of the ESEA (1965) is to "provide financial assistance to local educational agencies serving areas with concentrations of children from low-income families to expand and improve their educational programs by various means" (p. 27). The ESEA is the main source of federal funding for vulnerable students. The largest financial component of the ESEA is Title I (Thomas & Brady, 2005). According to the original ESEA (1965), Title I provides "financial assistance to local educational agencies for the education of children of low-income families" (p. 27). In 1965 the ESEA provided approximately $1 billion to school

districts in need (Thomas & Brady, 2005).

ESEA experienced numerous reauthorizations over the last 40 years, starting in the 1980s. When President Ronald Reagan was elected in 1980, the funding provided through Title I decreased drastically (Thomas & Brady, 2005). Reagan restricted the role of the federal government in education while emphasizing the lack of performance in America's schools. *A Nation at Risk*, a report by the National Commission on Excellence in Education, was published in 1983 and this critique had a big impact on the nation. The authors stated that educational standards in the United States were too low and that other nations were beginning to surpass the U.S. due to this (as cited in Fraser, 2014). Reagan used this report as support for his position on education (as cited in McDonnell, 2005). Reagan passed the Education Consolidation and Improvement Act of 1981 which drastically reduced federal funding in education (McDonnell, 2005). Through the Education Consolidation and Improvement Act of 1981 Title I was converted to Chapter I which came with decreased funding and served fewer students (Thomas & Brady, 2005). ESEA was reauthorized in 1988 and required states to set and document achievement levels for the first time. Documentation of achievement levels by states took the form of scores on standardized tests and ESEA funds were based on these results (Thomas & Brady, 2005). Standardized testing as a way to allocate funds will play a vast role in further reauthorizations of the ESEA (Thomas & Brady, 2005).

President George H.W. Bush took office in 1989 and he continued to focus on changing education policy. An educational summit was held in Virginia in 1989 to create academic achievement goals to increase student achievement (Thomas & Brady, 2005). At the summit it was determined that ESEA needed accountability based on more difficult academic standards and national testing (Thomas & Brady, 2005). President Bush initiated the America 2000

legislation in 1991 that aimed for national academic standards and testing. While America 2000 passed through both houses of Congress, it failed in the Senate due to a filibuster (McDonnell, 2005). America 2000 was important nonetheless as it opened the door to school reform based on academic standards. Although President Bush was the only president after 1965 whose term did not have a reauthorization of the ESEA, he still significantly influenced educational policy (McDonnell, 2005).

Throughout the 1990s ESEA continued to transform. In 1992 President Bill Clinton was elected to office. Clinton continued to focus on the standards based school reform from the Bush administration (Thomas & Brady, 2005). Clinton's educational initiative took the form of Goals 2000: Educate America Act. Goals 2000 encouraged states to implement performance and content standards and provided grants to states that adopted these standards, developed education improvement plans, and assessed student achievement on the standards (McDonnell, McLaughlin, & Morrison, 1997). Another reauthorization of ESEA occurred in 1994 under President Clinton. The ESEA was reauthorized as the Improving America's Schools Act (IASA) (McDonnell, 2005). Public Law 103-382 states the purpose of IASA "is to enable schools to provide opportunities for children served to acquire the knowledge and skills contained in the challenging State performance standards developed for all children (1994, Sec, 1001, d). In order to receive Title I funds under IASA, states were required to develop, teach, and assess content standards. States were also required to set goals for adequate yearly progress (AYP) and report schools that failed to meet AYP under IASA (Thomas & Brady, 2005; McDonnel, 2005). Schools that did not meet AYP were required to make improvements (Thomas & Brady, 2005). Many changes transformed the role of the ESEA in the 1990s and even more came in the 2000s (Thomas & Brady, 2005).

The ESEA was the first step towards increased federal involvement in education and its initial implementation started a major transformation of federal education policy (Thomas & Brady, 2005). Throughout the ESEAs history many educational initiatives and reauthorizations occurred that shifted the role federal government has in education. While the 1994 reauthorization of ESEA, entitled IASA, provided a new framework of standards based reform, the next reauthorization would greatly expand that initiative (Thomas & Brady, 2005).

The No Child Left Behind Act

A reauthorization of the ESEA occurred in 2001 when the No Child Left Behind Act (NCLB; 2008) was signed into law by George W. Bush (Bell & Meinelt, 2011). NCLB was an extension of the ESEA but had more stringent regulations and sanctions associated with it than previous acts. NCLB expanded upon previous ideas and had unique plans that required much more from states (Dee & Jacob, 2010).

NCLB was passed in Congress with bipartisan support in the House and a vote of 381-41(DeBray-Pelot & McGuinn, 2009). The Senate had equally strong bi-partisan support with a final vote of 87-10 (DeBray-Pelot & McGuinn, 2009). This is a remarkable achievement considering the differing perspectives of Democrats and Republicans on educational policy. Since the ESEA was passed in 1965, Democratic and Republican leaders have had very different priorities in educational reform as demonstrated by the educational policy changes that occurred under different presidents (DeBray-Pelot & McGuinn, 2009).

While IASA suggested standards-based reform, NCLB (2008) required all states, the Capitol, and Puerto Rico to design accountability systems based on performance and to have them operational by 2006. The goal of NCLB (2008) was to have 100% of all students in all subgroups proficient in math and reading by 2014 (Wei, 2009). As stated in the legislation, “The

purpose of this title is to ensure 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" (P.L. 107-110, 2002, Sec. 1001). NCLB specifically aimed to improve low-performing schools (Maleyko & Gawlik, 2011). The accountability systems of NCLB intended to eliminate the achievement gap between minority and majority groups (Maleyko & Gawlik, 2011).

Under NCLB (2008), states were required to follow specific mandates in order to receive Title I funds. Wei (2009) noted these mandates included annually testing all students in third through eighth grade, developing yearly measurable objectives in reading and math, and making Adequate Yearly Progress (AYP). To meet these goals under NCLB, rigorous standards were adopted (Maleyko & Gawlik, 2011). Many things can happen to Title I schools when they fail to meet AYP. Title I schools that don't meet AYP for two consecutive years are labeled in need of improvement and are required to implement a school improvement plan, dedicate 10% of their Title I dollars to professional development, and offer all students school choice (Porter, Linn, & Trimble, 2005). Title I schools that don't meet AYP for three consecutive years are required to provide additional services to students from low-income families. Schools that fail to meet AYP for four consecutive years are labeled in need of corrective action and must implement at least one corrective action, such as developing a new curriculum (Porter et al., 2005). Title I schools that fail to meet AYP for five years in a row are required to design a school restructuring plan. Schools that fail to meet AYP for six consecutive years must implement the restructuring plan (Porter et al., 2005). NCLB had greater federal sanctions associated with it than previous school reform efforts (Meyers, 2012).

While the numerous requirements under NCLB (2008) seemed very stringent, there was

quite a bit of flexibility allowed. States are able to develop their own accountability plans, choose their own assessments, trajectories, and subgroup sizes. States are also allowed to choose whether or not they use confidence intervals, performance indexing, or allow retesting (Wei, 2009). According to Maleyko and Gawlik (2011), "States have the ability to statistically manipulate their AYP implementation, which may give a false impression to the public that AYP is a consistent measure of school effectiveness across the country" (p. 601). The way states choose to measure AYP has a major influence on whether or not their schools make AYP (Maleyko & Gawlik, 2011).

The method states choose for their line of trajectory impacts whether or not schools make AYP. The line of trajectory shows how states plan to meet 100% proficiency by 2014 (Maleyko & Gawlik, 2011). States may use a stair-step trajectory with an initial performance increase in 2005 followed by increases every three years (Porter et al., 2005). States may also choose a straight-line approach with consistent increases each year. Another method is the front-loaded trajectory where larger gains are made in the beginning. A back-loaded trajectory is another option where larger gains are made in the final years (Porter et al., 2005). Only three of the trajectory options were utilized by states. Four states chose to use the straight-line method, 19 states chose the stair-step trajectory, 24 states chose to use the back-loaded approach, and three states did not specify their choice (Porter et al., 2005). The choice states make to reach proficiency levels greatly impacts whether or not their schools meet AYP (Porter et al., 2005).

Subgroup requirements differ by state and states are able to choose the least amount of students for a subgroup to count towards AYP (Maleyko & Gawlik, 2011). Manipulating subgroup size can greatly impact the number of schools making AYP. Two states did not identify their minimum subgroup sizes and the other 48 states had minimum subgroup sizes from

five to 100 students. The smaller the minimum number of students required per subgroup, the greater the standard (Porter et al., 2005). Schools that have more racial diversity and more subgroups have a smaller chance of making AYP (Maleyko & Gawlik, 2011).

The use of confidence intervals greatly impacts schools making AYP. When states utilize confidence intervals, any student that has a score within that boundary is regarded as proficient in regards to AYP. School success may be perceived as higher than it is according to the number of schools meeting AYP due to the use of confidence intervals (Maleyko & Gawlik, 2011).

> For example, if 35% of the students in a given subgroup were proficient or above in a given subject when the objective for that student and year was 40% or more proficient, the school would still meet its AYP requirement for that subject if the confidence interval had a width of 12 percentage points. (Porter et al., 2005, p. 34)

The highest level of accountability comes without the use of confidence intervals and 11 states chose this method. Two states chose a confidence interval smaller than a one-tailed 95% confidence interval and 16 states chose a 99% confidence interval. The chosen confidence interval creates variance in state schools making AYP (Porter et al., 2005).

Porter et al. (2005) analyzed Kentucky's design choices to demonstrate the variability afforded under the previously described state choices. Kentucky used a minimum subgroup size of 60, a two-tailed 99% confidence interval, and a delayed AYP increase trajectory until 2005. Based on these choices, 90% of Kentucky's schools reached AYP in 2003 and 94% met AYP in 2004. In contrast to these choices, if Kentucky had utilized a minimum subgroup size of 30, the proportion of schools making AYP would have declined to 84% in 2003 and 89% in 2004 (Porter et al., 2005). Had Kentucky chosen not to use a confidence interval, the percentage of

schools making AYP would have been 61% for 2003 and 72% for 2004. If Kentucky had not used a confidence interval and chose a straight-line trajectory starting in 2003, only 45% of schools would have met AYP in 2003 and 59% in 2004 (Porter et al., 2005). Had Kentucky implemented the smaller minimum subgroup, not used a confidence interval, and used a straight-line trajectory, only 31% of schools would have met AYP in 2003 and 44% in 2004 (Porter et al., 2005). The variability of schools meeting AYP illustrated in these choices can lead to misleading results between states.

There are many critics of NCLB that believe it provided the federal government with too much control and that NCLB's implementation led to unintended consequences (Maleyko & Gawlik, 2011). One of these unintended consequences is educational triage. Educational triage consists of prioritizing students that are close to the proficiency level over those who are well below or above it (Maleyko & Gawlik, 2011). Focusing on these students enables schools to increase their test scores. Students who are not near the proficiency cutoff suffer because they may not receive extra time and resources (Maleyko & Gawlik, 2011). Other criticisms of NCLB include narrowing the curriculum to focus on tested subjects, teaching to the test, and a host of other measures focused on improving ratings (Maleyko & Gawlik, 2011).

The National Assessment of Educational Progress

"The National Assessment of Educational Progress (NAEP) is the only nationally representative and continuing assessment of what U.S. students know and can do in various subject areas" (Hombo, 2003, p. 59). The NAEP was first used in the 1969-1970 school year (Beaton & Zwick, 1992). While testing was commonplace in the era of NCLB, states were able to design and implement their own standardized tests, which made meaningful comparison between states impossible (NCES, 2010). States were allowed to determine their own

curriculum and achievement tests, which led to inconsistent testing policies between states (Hombo, 2003). The NAEP is incredibly useful as "NAEP is the only regularly administered, Congressionally mandated national assessment program" (Beaton & Zwick, 1992, p. 95). The NAEP is unique as it "provides a common measure of student achievement across the country" (NCES, 2010, p. 1). The NAEP analyzes the educational attainment of students over time (NCES, 2009). The information gathered by the NAEP is used by policy makers and educators as a guide to student progress (NCES, 2009). NAEP results can be compared across states as the test and administration of the NAEP is the same in all states (NCES, 2010).

The NAEP is a federally required assessment by the National Center for Education Statistics (NCES) and must be reauthorized by Congress periodically (Hombo, 2003). The National Assessment Governing Board oversees policy for the NAEP. The Governing Board is comprised of 26 members that are chosen by the U.S. Secretary of Education. These individuals include state legislators, governors, business representatives, local and state school officials, educators, and the public (NCES, 2010). The Commissioner of Education Statistics carries out the assessment. The Commissioner serves as head of the NCES which is part of the USDOE's Institute of Education Sciences (NCES, 2009). Quality control of the assessment is monitored by the Associate Commissioner for Assessment in the NCES. The Associate Commissioner directs the creation, administration, scoring, and analysis of the NAEP (NCES, 2009).

Participation in the NAEP is incredibly important because "it informs us how student performance has changed over time, and allows states to compare their progress with that of other states and the nation as a whole" (NCES, 2010, p. 9). Before NCLB state participation in NAEP was voluntary. Schools receiving Title I funds are now required to participate in reading and math NAEP for students in Grades 4 and 8 biannually, if they are selected, however the

NAEP is still voluntary in all other subjects (Hombo, 2003). More than a million students take the NAEP annually (NCES, 2009). "NCES uses a sampling procedure to ensure that those selected to participate in NAEP will be representative of the geographical, racial, ethnic, and socioeconomic diversity of schools and students across the nation" (NCES, 2010, p. 8).

For national NAEP tests, anywhere from 6,000 to 20,000 students are tested for each subject from each grade. For state reported assessments approximately 3,000 students from 100 schools in every state are tested in each subject and grade level (NCES, 2010). According to Hombo (2003), "Minority students and students attending private schools are oversampled in NAEP to provide a sufficiently large sample for the subgroup analysis and reporting mandated by the U.S. Congress" (p. 61). In the years prior to 2002 national and state NAEP samples were independent of each other. After 2002 the design changed to the national sample being a subsample of pooled state samples (Hombo, 2003). Administration of the NAEP takes about 90 minutes and is done by a federal contractor (NCES, 2009; Hombo, 2003). The federal government covers all testing costs associated with the NAEP and "having a single contactor responsible for all NAEP field data collection provides continuity and consistency throughout the assessment" (Hombo, 2003, p. 61). Since the NAEP does not test all students in the nation there is a margin of error associated with sample size. Only statistically significant changes from year to year are discussed in the annual report, The Nation's Report Card (NCES, 2010).

The NAEP is designed to produce accurate and useful results that can be used to indicate educational progress (Beaton & Zwick, 1992). The NCES designs test questions and NAEP utilizes many question types including written response, multiple choice, and reading aloud (Beaton & Zwick, 1992; NCES, 2010). The NAEP has continued to progress throughout the years as new technology and information is made available (Beaton & Zwick, 1992). Every

NAEP test is “built from a content framework that specifies what students should know and be able to do in a given grade. The National Assessment Governing Board, which sets NAEP policy, oversees the creation of the NAEP frameworks” (NCES, 2010, p. 4). The NAEP frameworks are developed based on contributions from teachers, subject experts, school administrators, policymakers, teachers, parents, and others (NCES, 2010). After the framework is completed, the questions are developed by over 300 people from the framework committee, curriculum specialists, state and district representatives, state assessment directors, university educators, state content specialists, and subject experts (NCES, 2010). The questions are then reviewed and updated before undergoing pilot testing (NCES, 2010).

The NAEP consists of two kinds of assessments, the main NAEP and the long term trend NAEP (NCES, 2010). The main NAEP is given to students in Grades 4, 8, and 12 across the nation annually and tests subjects such as math, reading, science, and writing. National results are available for all tests and state and some urban results are available for fourth and eighth grade assessments (NCES, 2010). Long term trend NAEP is given to 9-, 13-, and 17-year-old students across the nation every four years and tests math and reading skills. Results from the long term trend NAEP allow comparison of student progress over time (NCES, 2010). The main NAEP is given annually between January and March. The long term trend NAEP is given every four years in the fall for 13-year-old students, the winter for 9-year-old students, and the spring for 17-year-old students (NCES, 2010).

The NAEP continues to evolve its reporting methods over time. The first results in 1969 were reported item by item for the whole nation, specific regions, and demographic groups and also included trend data (Lane et al., 2009). This method shifted in 1984 when the NAEP began using summarized scale scores for each content area. In 1986 the NAEP implemented The

Nation's Report Card reporting method, which is still used today (Lane et al., 2009). As school accountability started to become an increasingly important policy issue, NAEP began to report achievement levels by state in The Nation's Report Card in 1990 (Lane et al., 2009). Focused reporting of performance by subgroup also started during this time (Lane et al., 2009). The current reporting in The Nation's Report Card consists of nationwide results and some state and urban results. Results provided represent student performance in public and private schools, Department of Defense schools, and Bureau of Indian Education schools while results from individual students and schools are not reported (NCES, 2010). The Nation's Report Card provides results by gender, socioeconomic status, demographic groups, and race/ethnicity (NCES, 2010). Scale scores are provided for the NAEP and are 0-300 or 0-500 point scales. Percentile scores are also reported for each grade level. There are five percentiles to demonstrate student progress at lower levels from the 10^{th} to 25^{th} percentiles, middle level at the 50^{th} percentile, and higher levels from the 75^{th} to 90^{th} percentiles (NCES, 2010). Achievement levels are also reported for each grade taking the NAEP assessment. Achievement levels consist of basic, proficient, and advanced and are knowledge students should have in certain grade levels (NCES, 2010). "The percentage of students at or above these levels, labeled *Basic*, *Proficient*, and *Advanced*, adds to the interpretation of student performance on NAEP" (Hombo, 2003, p. 60). In addition to providing assessment results, the NAEP collects background information for analysis from the schools, teachers, and students (Hombo, 2003). Prior to the school accountability era the NAEP mainly served as a survey of student knowledge over time. In the NCLB era, the NAEP now has another important purpose which is to analyze state assessment results and to compare student progress between states (Hombo, 2003).

Theoretical Framework

In 1983 *A Nation at Risk* was published by the National Commission on Excellence in Education (NCEE). The authors of the report stated that the educational standards in the United States were too low and that other nations were beginning to surpass the U.S. due to this (as cited in Fraser, 2014). Since *A Nation at Risk* (NCEE, 1983) was published education policy has been an issue of rising importance (Mehta, 2015).

When the National Commission on Excellence in Education was formed, the Reagan administration wanted to abolish the Department of Education, not expand it. *A Nation at Risk* (NCEE, 1983) was initially reported in a White House ceremony and Reagan disregarded the findings of the report (Mehta, 2015). However, the day after the ceremony the U.S. Government Printing Office received over 400 requests for the report in one hour and more than six million requests that year. In 1984 over 250 task forces were established to examine education and recommend changes. The overwhelming media coverage of *A Nation at Risk* (NCEE, 1983) played a huge role on its impact in education (Mehta, 2015).

A Nation at Risk (NCEE, 1983) "invoked a crisis so far-reaching in its impact that it still governs the way we think about public education 30 years later" (Mehta, 2015, p. 20). Drafts of the report indicate "that the report's inflammatory rhetoric about a system in crisis was a conscious choice made by some on the commission in order to increase its impact" (Mehta, 2015, p. 22). The rhetoric in the report, the prominence of its authors, and the timing of the report likely played a role in the overwhelming response. *A Nation at Risk* (NCEE, 1983) was written during a recession when international competitors were becoming a threat. "Contrasting this declining educational picture with the centrality of skills and human capital in the knowledge-based, postindustrial economy, the report linked the future of the nation's

international economic competitiveness with the reform of its educational system" (Mehta, 2015, p. 21).

Concern for the status of the United States in the world economy mounted after the publishing of *A Nation at Risk* (NCEE, 1983). Education is considered one of the main determinants of labor market success and the method used to measure returns to education is the human capital theory (Free, 2010). The basis of human capital theory is

> Status attainment, or the level of success achieved by an individual in society, is a direct result of educational levels, personal values and skills, and other individual characteristics and abilities. Education is seen as an investment in human capital, not unlike the investment a business might make in machinery or new technology. The greater the investment in a person's human capital, the higher the probability of success. (Healey & Stepnick, 2016, p. 49)

Modern human capital theory is used to highlight the importance of training and education as a key piece of the new world economy as education has been discussed as primarily a monetary device in Western nations. Human capital theory has been the most significant economic theory in Western education and has set the basis for numerous governmental policies (Fitzsimons, 1999).

Human capital theory is more than a theory that further education yields higher income. The explanation for this correlation is that education increases skills that increase productivity and increased productivity then leads to greater earnings (Strober, 1990). Human capital theory in education also includes skills, dispositions, knowledge, and resources of adults in schools that promote student learning. In order for schools to maximize effectiveness, teachers need to be adaptable to changes in learning needs, characteristics, and conditions of students. This requires

the use of standards and assessments for goal-setting and monitoring (Smylie, 1996). However, standards and assessments should be used to guide action, not dictate it. Human capital theory in education also incorporates social accountability through professional skills and outcomes, not bureaucratic rules and procedures (Smylie, 1996).

The use of standards and assessments can be beneficial when it is used to guide action, not dictate it as mentioned above. NCLB involved many sanctions and incentives through the use of high stakes standardized testing. Dee and Jacob (2010) stated

> The hallmark features of this legislation compelled states to conduct annual student assessments linked to state standards to identify schools failing to make 'adequate yearly progress' (AYP) toward the stated goal of having all students achieve proficiency in reading and math by 2013-2014 and to institute sanctions and rewards based on each school's AYP status. A fundamental motivation for this reform is the notion that publicizing detailed information on school-specific performance and linking that 'high-stakes' test performance to the possibility of meaningful sanctions can improve the focus and productivity of public schools. (p. 149)

Incentives for performance measures often lead to activities that improve measured performance, but not the value of the work (Hout & Elliott, 2011). This phenomenon is referred to as Campbell's law, which is described as

> The more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor. (Campbell, 1975, p. 49)

According to Campbell (1975), when test scores are the goal of teaching, their value for educational progress is lost and the educational process becomes distorted.

NAEP is a low stakes test that does not have any incentives or sanctions associated with it. Therefore, NAEP scores are not as subject to corruptibility under Campbell's Law as indicators associated with school accountability such as high stakes state tests or graduation rates. The NAEP utilizes Itemized Response Theory (IRT), which is a statistical theory on how performance is related to measured abilities by items in a test (Hambleton & Jones, 1993). IRT originated in the 1940s as a way to differentiate examinee proficiency from the tasks they were given and to differentiate tasks independently of examinees who took them. IRT aims to "provide an estimate of the probability of a correct response on a test item as a function of the characteristics of the item (e.g. difficulty, discrimination)" (De Champlain, 2010, p. 113). Based on this, IRT models try to "explain observed (actual) item performance as a function of an underlying ability (unobserved) or latent trait" (De Champlain, 2010, p. 113).

IRT is used in NAEP to create scales that condense major findings from the vast amount of student performance demonstrations (Beaton & Johnson, 1992). These scales provide useful achievement descriptions without having the problems associated with average percent correct statistics such as dependency on selected items, limits to meaningful trend measurements, and lack of proficiency distribution estimates in a population. Not all students taking NAEP will have the same items or all items from the pool (Beaton & Johnson, 1992). However, NAEP scales enable all students to be on a common scale which allows analysis of trends in a consistent metric. The scale score for a student represents proficiency and can be used to approximate, or even reproduce, the response for any question from the item pool (Beaton & Johnson, 1992). NAEP reports number-right scores for a hypothetical 500 item test and scores typically range from 100 to 400. Students are not given one score on NAEP, but rather five possible values. This process "assigns a student to a group five times, each time using a different plausible value

and one fifth of the student's sampling weight" (Beaton & Johnson, 1992). The use of IRT in NAEP allows measurement of trends over time using a reliable metric even though item pools vary with time (Beaton & Johnson, 1992).

The graduation rate is frequently cited as evidence of improvement through NCLB. However, this is susceptible to corruption under Campbell's Law. According to the NCES, students in the U.S. had a high school graduation rate of 81% for the 2012-2013 school year, the highest rate since the cohort method for calculating graduation rates was adopted (USDOE, 2015). U.S. Secretary of Education Arne Duncan stated

> America's students have achieved another record-setting milestone. We can take pride as a nation in knowing that we're seeing promising gains, including for students of color. This is a vital step toward readiness for success in college and careers for every student in this country, and these improvements are thanks to the hard work of teachers, principals, students, and families. (USDOE, 2015, para. 2)

John Gomperts, leader of the GradNation campaign and president and CEO of America's Promise Alliance stated, "When schools are held accountable and students are given support to help them stay in school and on track, real progress is possible" (as cited in USDOE, 2015, para. 4). While the high graduation rate is commendable, graduation rates may not be the most reliable source of student achievement data as previously discussed.

Evaluating school and student progress has become a priority in recent years as political leaders seek to demonstrate the effectiveness of school reform efforts (Fuller, Wright, Gesicki, & Kang, 2007). Political leaders highlight growth factors such as state test scores and a high graduation rate. However, many of these factors are arbitrary in nature and susceptible to corruption (Fuller et al., 2007). As discussed, state test scores are not useful as a means for

comparison for many reasons, such as lack of uniformity between states and changing proficiency standards.

Fuller et al. (2007) examined student performance from the NAEP and state testing data from 1992-2006. The researchers concluded that "We should not rely on state testing programs and the jagged trend lines that stem from their results. Instead, it is important to focus on historical patterns informed by the NAEP" (Fuller et al., 2007, p. 275). Fuller et al. further stated, "state trend lines often look like a jagged mountain range, erratically moving up and down as tests are changed and proficiency bars are moved" (p. 268). Analysis of student achievement trends on state tests in regards to the effects of federal legislation can be difficult as it is hard to ascertain if the reform efforts caused the improvement or if improvements began before that (Fuller et al., 2007).

Dee and Jacob (2010) examined NCLB's impact on student achievement. The researchers chose to use NAEP data because it is a low-stakes assessment. The NAEP is not associated with state assessments and because of this NAEP data should not be impacted by "accountability-driven test-score inflation, such as may result from 'teaching to the test" (Dee & Jacob, 2010, p. 58). Dee and Jacob also discussed the possibility of state test scores being inflated as states decrease proficiency standards in order to increase AYP. There are many discrepancies between state test results and NAEP results and it is important not to rely on only one test as a source of data (Lee, 2007).

High stakes tests results, or tests that have meaningful consequences attached to them, can be misleading because as testing stakes rise, schools and teachers are more likely to prepare students specifically for tested content (Cole & Osterlind, 2008; Shepard, 1990). This kind of intense preparation for high stakes assessments is what many call teaching to the test and can

compromise the legitimacy of the results (Shepard, 1990). An examination of high stakes testing by John Cannell in 1988 demonstrated these phenomena even before high stakes testing was at the intensity it is today (as cited in Nichols, 2007). Cannell (1988) examined state and district results from the Iowa Test of Basic Skills and found that states reported findings where over 50% of students performed above average. Based on the reports, Cannell believed that public reporting pressure led to data manipulation to make the results look more satisfactory. The findings in this study are referred to as the *Lake Wobegon effect* (Nichols, 2007). The Lake Wobegon effect demonstrated score inflation had taken place, but did not explain its cause. A possible explanation is "Under conditions of pressure (i.e., being evaluated publicly by student test results), teachers and principals changed their behavior to focus instruction more intently on the test" (Nichols, 2007, p. 51).

To examine the theory of teaching to the test, Linn, Graue, and Sanders (1990) paralleled the Iowa Test of Basic Skills scores to NAEP scores and found that the Iowa Test of Basic Skills scores increased more rapidly and to greater levels than NAEP scores, even after accounting for testing differences. According to Nichols (2007), "the importance of looking at a comparable no stakes test (such as the NAEP) for evidence that high-stakes policies are working to increase student learning was an important outcome of Lake Wobegon studies" (p. 51).

Another example of score inflation in high stakes testing comes from Texas. Texas was at the head of the accountability movement in the 1990s as the state consistently increased testing sanctions (Nichols, 2007). Texas portrayed high success rates from their accountability methods to the public. When the scores from the Texas Assessment of Academic Skills (TAAS) were compared with NAEP scores, researchers found that student achievement was not much different in Texas than any other state (Nichols, 2007). Texas NAEP scores ranked in the 46th

percentile, putting Texas near the median (Camilli, 2000). The inflated state test scores most likely resulted from score manipulation and teaching to the test, similar to the example from the Lake Wobegon studies (Nichols, 2007). According to Haney (2000), "In the opinion of educators in Texas, schools are devoting a huge amount of time and energy preparing students specifically for TAAS" (p. 1). In addition to score inflation, Haney found other issues including missing students and exclusions. The examples above help illustrate corruptibility according to Campbell's Law.

Other issues associated with the use of high stakes test scores as indicators of student success arise from the rate of change associated with the tests. States quickly adopt and change policies regarding high stakes testing which makes it challenging to determine their effects (Nichols, 2007). Nichols (2007) aimed to study the influence high stakes testing had on student achievement and found that

> The most rigorous studies on high-stakes testing do not provide convincing evidence that high-stakes testing has the intended effect of increasing student achievement. Moreover, the modest gains found in some studies should be viewed with caution since the findings indicate that increases in achievement could be the result of teaching to the test. (p. 57)

Finally, Nichols stated that "studies of high-stakes testing policy must turn to 'audit' tests such as the NAEP for an indication that learning has occurred (p. 58). As such, this report will examine NAEP scores of 17-year-old test takers as opposed to high school graduation rates or standardized test scores to investigate trends in student success in the era of school accountability.

Summary

This chapter discussed an abundance of research regarding secondary education in the United States as well as historically associated indicators of student success. School accountability measures have increased over time and are at the forefront of educational policy today. The use of the high school graduation rate as an indicator of the success of school accountability measures may be effective, but it is important to determine if this is supported in other areas.

The first section of this chapter examined the historic use of the high school graduation rate in the United States. The high school graduation rate has been reported since 1869 when it was at 2%. During this time period the high school graduation rate was reported as the proportion of the 17-year-old population with a diploma. Utilizing this method for reporting graduation rates, the graduation rate was reported at an all-time high of 81% in recent years. Over the years, the way the calculation of the high school graduation rate transformed and there was a common method nationwide. The introduction of the cohort rate was the first unified nationwide calculation rate but this does not encapsulate all of the dynamics of secondary students.

The next section of this chapter explained the Elementary and Secondary Education Act of 1965. The signing of this bill represented a major shift in educational reform from the past. The ESEA aimed to close the achievement gap through numerous school reform measures. The ESEA has been reauthorized numerous times since its original implementation. Each reauthorization came with increased school accountability measures.

The third section of this chapter discussed the achievement gap. The achievement gap was identified during the Civil Rights era after the legalization of the ESEA in 1965. President

Johnson had a group of researchers examine racial inequalities in student achievement and their report was published. Since then, the media has used the term achievement gap to describe the presence of the racial disparities in student achievement.

The proceeding section of this chapter investigated the most recent reauthorization of ESEA: the No Child Left Behind Act. President George W. Bush signed NCLB into law in 2001. NCLB was an extension of ESEA with much stricter regulations and sanctions than previous school reform efforts. NCLB required all states to design and implement performance-based accountability systems with the goal of having 100% of students proficient in math and reading by 2014. In order to receive Title I funds, states were also required to perform annual testing, develop annual measurable objectives, and make AYP. There were numerous sanctions for schools that failed to make AYP including school improvement plans, offering school choice, and designing and implementing a restructuring plan.

The fifth section of this chapter examined the National Assessment of Educational Progress. The NAEP has been administered since the 1969-1970 school year and allows comparison between states. This is helpful to measure growth as states are allowed to design their own curriculum and accountability tests. NAEP is the only test that has been regularly administered across the United States in the same format. The NAEP is designed to provide results to indicate educational progress.

CHAPTER 3

METHODOLOGY

This chapter discusses the research methodology that was utilized including the null and alternative hypotheses, basic method, sample population, and data analysis. The purpose of this quantitative study was to better understand the connection between select student demographics and NAEP scores for time periods before and after the implementation of numerous school accountability measures. The independent variable in this study was school year and race/ethnicity. The dependent variables in this study were long term trend NAEP scores in Math and Reading.

Research Questions and Null Hypotheses

The core research question leading this study was, "In the current era of school accountability, are there indicators of student success other than graduation rates among 17-year old test takers on the long term trend National Assessment of Educational Progress?" To investigate this, long term trend NAEP scores for 17-year old students were analyzed. Additional research questions include

R1. Is there a change among 17-year old test takers in 1990-1999 NAEP Math and Reading scores.

R2. Is there a change among 17-year old test takers in 2004-2012 NAEP Math and Reading scores.

R3. Is there significant differences among 17-year old test takers by race/ethnicity in 1990-1999 NAEP Math and Reading scores.

R4. Is there significant differences among 17-year old test takers by race/ethnicity in 2004-2012 NAEP Math and Reading scores.

R5. Is the mean score change between 1990 and 1999 comparable to the mean score change between 2004 and 2012, as measured by Cohen's d statistic, for both reading and math?

To better understand NAEP scores in the era of school accountability, the following null hypotheses were examined:

$H_0$1. There was not a change among 17-year old test takers in 1990-1999 NAEP Math and Reading scores.

$H_0$2. There was not a change among 17-year old test takers in 2004-2012 NAEP Math and Reading scores.

$H_0$3. There were not significant differences among 17-year old test takers by race/ethnicity in 1990-1999 NAEP Math and Reading scores.

$H_0$4. There were not significant differences among 17-year old test takers by race/ethnicity in 2004-2012 NAEP Math and Reading scores.

Additionally, the research question "Is the mean score change between 1990 and 1999 comparable to the mean score change between 2004 and 2012, as measured by Cohen's d statistic, for both reading and math?" was addressed.

Basic Methodology

Quantitative research is defined as "a means for testing objective theories by examining the relationship among variables. These variables, in turn, can be measured, typically on instruments, so that numbered data can be analyzed using statistical procedures" (Creswell,

2009, p. 4). Data for this study was extracted from the NAEP database. A restricted use data license was acquired for each data set (see Appendix). There was one large extraction for each year, 1990, 1999, 2004, and 2008. Within each year, data for 17-year-old test takers was extracted targeting Math scores, Reading scores, and race/ethnicity. This data was examined using independent samples t-tests and Cohen's d statistic to support or refute the previous hypotheses.

Description of the Sample

Long-term trend NAEP is a national test for public and private schools and is intended to measure progress of students nationally, not individually (Perie, Moran, & Lutkus, 2005). The focus of this research was the Math and Reading scores of 17-year old students in public schools taking the long-term trend NAEP. Long-term trend NAEP is administered every four years to a national sample to measure student performance in Math and Reading and can be compared with student performance since the 1970s due to the nature of the test (USDOE, 2010). Data for long-term trend NAEP is collected in the spring for 17-year-old students (Rampey, Dion, & Donahue, 2009).

For long-term trend NAEP, "a nationally representative sample of students is selected, and their results are generalized to the nation as a whole" (Perie, Moran, & Lutkus, 2005, p. 4). To make certain the sample was representative of students across the nation, "a sampling plan was created to randomly select schools and students to participate" (Perie, Moran, & Lutkus, 2005, p. 5). In 1990, the 17-year-old sample population for long-term trend reading included 4,383 students. For the math assessment the same year, the sample population included 4,411 students. In 1996, the math assessment sample size was 3,539 students and the reading assessment sample size was 4,669 students (Campbell, Voelkl, & Donahue, 1997). The 1999

long-term trend math assessment was comprised of 3,795 in the sample population. For the reading assessment in the same year, 5,288 students were included in the sample population (Campbell, Hombo, & Mazzeo, 2000).

In 2004, the sample population included over 8,300 for the reading assessment and 7,600 for the math assessment (Perie, Moran, & Lutkus, 2005). The 2008 sample population included 9,600 students for each assessment (Rampey, Dion, & Donahue, 2009). For the 2012 NAEP long-term trend reading assessment there were 9,000 participants in the sample with a target population of 3,468,000. The 2012 NAEP long-term trend math assessment included 8,600 students with a target population of 3,472,000. (USDOE, 2012). All data used for this study is existing data and was requested following Indiana State University's Institutional Review Board (IRB) approval. A restricted use data license was granted for the restricted use data. Public records data was also utilized for this study.

Statistical Software and Hardware

Analysis of complex data samples such as the long term trend NAEP requires specific statistical software. *Statistical Package for the Social Sciences, SPSS,* was utilized for this study after the data was extracted from the restricted use data disc. The data was extracted utilizing the *NAEP Data Toolkit* provided with the restricted use data CD-ROM. *NAEPEX* is the data extraction program in the toolkit that was utilized for choosing variables, extracting data, and generating SPSS command statements (USDOE, 2011).

There were many security requirements associated with the restricted use data set. Following security procedures, data was stored on a stand-alone computer in a secured location. Only authorized users of the data had access to this location and when the data was not being used it was under lock and key according to IES requirements. Additional security requirements

included the appointment of security personnel that met IES qualifications, password restrictions, read-only access, encryption codes, and overwriting of data. After the data analysis was complete, the data disc was returned, the computer was decommissioned, and the hard drive was formatted.

Statistical Analysis

Independent samples t-tests were utilized to address the null hypotheses for this research. For the first null hypothesis, "There was not a change among 17-year old test takers in 1990-1999 NAEP Math and Reading scores", the year was the independent variable and the NAEP math score was the dependent variable. An independent samples t-test was completed and analyzed for statistical significance using an alpha level of .05. This t-test was manipulated again with the NAEP reading score as the dependent variable. The independent samples t-test for NAEP Math scores from 1990-1999 showed significant results. Based on this, the effect size was examined using Cohen's d. The second null hypothesis, "There was not a change among 17-year-old test takers in 2004-2012 NAEP Math and Reading scores" was also tested for significance using independent samples t-tests with an alpha level of .05. The independent and dependent variables were year and math score, respectively. This test was also manipulated with reading score as the dependent variable utilizing the independent samples t-test. Significance was found for NAEP Reading scores from 2004-2012. As a result, Cohen's d was utilized to examine effect size. The third null hypothesis, "There were not significant differences among 17-year-old test takers by race/ethnicity in 1990-1999 NAEP Math and Reading scores" was tested for significance using independent samples t-tests. The t-tests were analyzed for significance by each subgroup; White, Black, Hispanic, and Other. The fourth null hypothesis, "There were not significant differences among 17-year old test takers by race/ethnicity in 2004-

2012 NAEP Math and Reading scores" was tested for significance using independent samples t-tests. The t-tests were analyzed for significance by each subgroup; White, Black, Hispanic, and Other.

The research question, "Is the mean score change between 1990 and 1999 comparable to the mean score change between 2004 and 2012, as measured by Cohen's d statistic, for both reading and math?" could not be addressed for these years. The significance that was found before the implementation of NCLB was for a different content area than the significance that was found post NCLB. Due to these findings, it was not meaningful to compare effect sizes of different subjects.

Data Storage and Confidentiality

NAEP public records data and NAEP restricted use data were both utilized in this study. The public records NAEP data that was accessed for this study is based on public records data, and therefore are not confidential. Restricted use data licenses were granted through the IES following a lengthy process involving security protocols for hardware and personnel. A Principal Project Officer (PPO), Senior Official (SO), and Systems Security Officer (SSO) were appointed to ensure data confidentiality procedures were followed according to IES protocol. The PPO was responsible for daily operations involving the data usage. The SO had the legal authority to sign for the restricted use data on behalf of Indiana State University. The SSO managed the security of the data. All data was stored in an Indiana State University owned and password protected computer with appropriate security protocols according to the IES. Following the use of the restricted use data, the data disc was returned, the computer was decommissioned, and the hard drive was formatted.

Summary

The purpose of this chapter was to provide a framework of the research methodology that was utilized in this study. Research questions and null hypotheses were explained and the basic method and description of the sample were explained. In addition to this, the statistical software and the statistical analyses that were utilized for this quantitative research study were described. The main purpose of this study was to better understand the relationship between select student demographics and NAEP scores for time periods before and after the implementation of numerous school accountability measures.

CHAPTER 4

DATA ANALYSIS

The purpose of this quantitative study was to better understand the connection between select student demographics and NAEP scores for time periods before and after the implementation of numerous school accountability measures. The independent variables in this study were school year and race/ethnicity. The dependent variables in this study were long term trend NAEP scores in Math and Reading.

Restricted-use data licenses were applied for and granted to receive NAEP data from the years 1990, 1999, 2004, and 2008. Data licenses were received and NAEP Math and reading scores from 1990, 1999, 2004, and 2008 were analyzed. However, restricted-use data from 2012 was not available at the time of this study. In order to effectively analyze data from the time period after the implementation of NCLB, 2012 NAEP data was needed. NAEP Data Explorer (NDE) is a tool provided by the National Center for Education Statistics and enables custom data analysis including significance testing, regression analysis, and gap analysis (USDOE, 2011). In order to provide a larger timeframe for data analysis post NCLB, NDE was utilized for further data analysis. NDE provided data for descriptive statistics and significance testing.

All t-tests were independent samples t-tests using an alpha level of .05 (USDOE, 2011). Independent samples t-tests were chosen for this study in order to compare separate samples of students on the mean NAEP Math and Reading scores over a time period prior to the

implementation of NCLB and a time period post NCLB. The independent variable was categorical with two levels and the dependent variables were continuous. Additionally, participants appeared in only one group. Due to the nature of the data, this study met the design requirements for an independent measures t-test as discussed above.

Assumptions for the independent samples t-test involve independence, normal distribution, and homogeneity of variance. As previously mentioned, long-term trend NAEP, "a nationally representative sample of students is selected, and their results are generalized to the nation as a whole" (Perie, Moran, & Lutkus, 2005, p. 4). To make certain the sample was representative of students across the nation, "a sampling plan was created to randomly select schools and students to participate" (Perie, Moran, & Lutkus, 2005, p. 5). Additionally, sample sizes for this study were in the thousands and did not violate the assumption of homogeneity of variance.

Research Questions and Null Hypotheses

The core research question leading this study was, "In the current era of school accountability, are there indicators of student success other than graduation rates among 17-year old test takers on the long term trend National Assessment of Educational Progress?" To investigate this, long term trend NAEP scores for 17-year-old students were analyzed. Additional research questions include:

R1. Is there a change among 17-year old test takers in 1990-1999 NAEP Math and Reading scores.

R2. Is there a change among 17-year old test takers in 2004-2012 NAEP Math and Reading scores.

R3. Are there significant differences among 17-year old test takers by race/ethnicity in 1990-1999 NAEP Math and Reading scores.

R4. Are there significant differences among 17-year old test takers by race/ethnicity in 2004-2012 NAEP Math and Reading scores.

R5. Is the mean score change between 1990 and 1999 comparable to the mean score change between 2004 and 2012, as measured by Cohen's d statistic, for both reading and math?

To better understand NAEP scores in the era of school accountability, the following null hypotheses were examined:

$H_0$1. There was not a change among 17-year old test takers in 1990-1999 NAEP Math and Reading scores.

$H_0$2. There was not a change among 17-year old test takers in 2004-2012 NAEP Math and Reading scores.

$H_0$3. There were not significant differences among 17-year old test takers by race/ethnicity in 1990-1999 NAEP Math and Reading scores.

$H_0$4. There were not significant differences among 17-year old test takers by race/ethnicity in 2004-2012 NAEP Math and Reading scores.

Additionally, the research question "Is the mean score change between 1990 and 1999 comparable to the mean score change between 2004 and 2012, as measured by Cohen's d statistic, for both reading and math?" was addressed.

Null Hypothesis 1

$H_0$1. There was not a change among 17-year old test takers in 1990-1999 NAEP Math and Reading scores. Data analyzed for this hypothesis included average scale score, standard deviation, and t-tests. Table 2 below provides the selected descriptive statistics. An independent

samples t-test revealed that on average there was a statistically significant difference in NAEP Math scores from 1990 (*M* = 304.56, *SD* = 31.10) to 1999 (*M* = 308.20, *SD* = 30.77), *t*(8193) = -4.37, *p* = .0077 (*Cohen's d* = .10). Cohen's d indicated a small effect size. The null hypothesis for NAEP Math scores from 1990-1999 was rejected. An independent samples t-test revealed that on average there was not a statistically significant difference in NAEP Reading scores from 1990 (*M* = 290.21, *SD* = 41.27) to 1999 (*M* = 287.76, *SD* = 41.79), *t*(8509) = 2.22, *p* = .16. The null hypothesis for NAEP reading scores from 1990-1999 failed to be rejected.

Table 2

NAEP Descriptive Statistics 1990-1999

Year	Jurisdiction	Subject	Average Scale Score	Standard Deviation
1990	National	Math	304.56	31.10
1999	National	Math	308.20*	30.77
1990	National	Reading	290.21	41.27
1999	National	Reading	287.76	41.79

Note. Created using data from U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, National Assessment of Educational Progress, 1990 and 1999 Long-Term Trend Assessments.
*Indicates statistical significance using an alpha level of .05

Null Hypothesis 2

$H_0$2. There was not a change among 17-year old test takers in 2004-2012 NAEP Math and Reading scores. Data analyzed for this hypothesis included average scale score, standard deviation, and t-tests. Table 3 below provides the selected descriptive statistics. An independent samples t-test revealed that on average there was a statistically significant difference in NAEP Reading scores from 2004 (*M* = 282.70, *SD* = 44.64) to 2012 (*M* = 286.89, *SD* = 42.25), *t*(17065) = -6.01, *p* = .0045 (*Cohen's d* = .09). Cohen's d indicated a small effect size. The null

hypothesis for NAEP reading scores from 2004-2012 was rejected. An independent samples t-test revealed that on average there was not a statistically significant difference in NAEP Math scores from 2004 (M = 305.30, SD = 30.47) to 2012 (M = 306.32, SD = 30.78), t(16059) = -2.07, p = .35. The null hypothesis for NAEP Math scores from 2004-2012 failed to be rejected.

Table 3

NAEP Descriptive Statistics 2004-2012

Year	Jurisdiction	Subject	Average Scale Score	Standard Deviation
2004	National	Math	305.30	30.47
2012	National	Math	306.32	30.78
2004	National	Reading	282.70	44.64
2012	National	Reading	286.89*	42.25

Note. Created using data from U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, National Assessment of Educational Progress, 2004 and 2012 Long-Term Trend Assessments.
*Indicates statistical significance using an alpha level of .05

Null Hypothesis 3

$H_0$3. There were not significant differences among 17-year old test takers by race/ethnicity in 1990-1999 NAEP Math and Reading scores. Data analyzed for this hypothesis included average scale score, standard deviation, and t-tests. Table 4 below provides the selected descriptive statistics by race/ethnicity. An independent samples t-test revealed that on average there was a statistically significant difference in NAEP Math scores for White students from 1990 (M = 309.49, SD = 29.47) to 1999 (M = 314.76, SD = 28.18), p = .001 Additionally, an independent samples t-test revealed that on average there was a statistically significant difference in NAEP Math scores for Hispanic students from 1990 (M = 283.50, SD = 31.85) to 1999 (M =

292.69, *SD* = 29.66), p = .02 There were no other statistically significant differences in Math or Reading from 1990 to 1999 in any other group.

Table 4

NAEP Average Scale Scores 1990-1999

	Student Ethnicity							
	White		Black		Hispanic		Other	
Year	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
				Math				
1990	309.49	(29.47)	288.55	(27.87)	283.50	(31.85)	312.48	(33.81)
1999	314.76*	(28.18)	283.31	(27.57)	292.69*	(29.66)	319.73	(28.70)
				Reading				
1990	296.64	(39.58)	267.34	(39.18)	274.80	(40.73)	290.12	(41.23)
1999	294.63	(40.42)	263.95	(37.95)	270.70	(39.67)	289.79	(41.23)

Note. Created using data from U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, National Assessment of Educational Progress, 1990 and 1999 Long-Term Trend Assessments.
*Indicates statistical significance using an alpha level of .05

Null Hypothesis 4

$H_0$4. There were not significant differences among 17-year old test takers by race/ethnicity in 2004-2012 NAEP Math and Reading scores. Data analyzed for this hypothesis included average scale score, standard deviation, and t-tests. Table 5 below provides the selected descriptive statistics by race/ethnicity. An independent samples t-test revealed that on average there was a statistically significant difference in NAEP Reading scores for White students from 2004 (*M* = 288.58, *SD* = 43.72) to 2012 (*M* = 295.13, *SD* = 40.85), *p* = .0002 . Additionally, an independent samples t-test revealed that on average there was a statistically significant difference in NAEP Reading scores for Black students from 2004 (*M* = 261.51, *SD* = 39.92) to 2012 (*M* = 269.08, *SD* = 39.94), p = .004 Lastly, an independent samples t-test revealed that on average

there was a statistically significant difference in NAEP Reading scores for Hispanic students from 2004 (*M* = 266.85, *SD* = 44.54) to 2012 (*M* = 273.84, *SD* = 40.77), p = .03 There were no other statistically significant differences in Math or Reading from 2004 to 2012 in any group.

Table 5

NAEP Average Scale Scores 2004-2012

	Student Ethnicity							
	White		Black		Hispanic		Other	
Year	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*	*M*	*(SD)*
				Math				
2004	311.15	(28.73)	284.23	(27.59)	291.84	(28.76)	312.87	(30.27)
2012	313.69	(28.18)	287.91	(28.96)	294.44	(29.57)	317.71	(30.47)
				Reading				
2004	288.58	(43.72)	261.51	(39.92)	266.85	(44.54)	291.03	(42.69)
2012	295.13*	(40.85)	269.08*	(39.94)	273.84*	(40.77)	295.94	(41.12)

Note. Created using data from U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, National Assessment of Educational Progress, 2004 and 2012 Long-Term Trend Assessments.
*Indicates statistical significance using an alpha level of .05

Additional Research Question

The research question, "Is the mean score change between 1990 and 1999 comparable to the mean score change between 2004 and 2012, as measured by Cohen's d statistic, for both reading and math?" could not be addressed for these years. The statistical significance found from 1990 to 1999 revealed a significant increase in Math scores during that time period. The statistical significance found from 2004 to 2012 revealed a significant increase in Reading scores for that time period. As statistical significance was not found in the same subject across time periods, the Cohen's d statistic cannot be reasonably compared. In both instances, Cohen's d revealed a small effect size.

Summary

This chapter examined the core research question, "In the current era of school accountability, are there indicators of student success other than graduation rates among 17-year old test takers on the long term trend National Assessment of Educational Progress?" To investigate this, long term trend NAEP scores for 17-year-old students were analyzed. Four null hypotheses and one additional research question were addressed.

From 1990 to 1999 before the implementation of NCLB, 17-year old test takers' average NAEP scale scores showed a statistically significant increase in Math from 304.56 to 308.20 respectively. When the data was broken down by race/ethnicity, White students showed a statistically significant increase in Math scores from 309.49 to 314.76 and Hispanic students showed a statistically significant increase from 283.50 to 292.69. Black students showed a decrease in Math scores from 288.55 to 283.31, although it was not statistically significant. Students identifying as Other exhibited an increase that was not statistically significant from 312.48 in 1990 to 319.73 in 1999.

After the implementation of NCLB, from 2004 to 2012, 17-year old NAEP test takers' average scale scores in Math did not exhibit a statistically significant change as a whole or by race/ethnicity. Overall, NAEP Math scores displayed a slight increase from 305.30 in 2004 to 306.32 in 2012. Students identifying as Other showed the largest gain in Math scores from 2004 to 2012, increasing from 312.87 to 317.71 respectively. Black students exhibited an increase in Math scores from 284.23 in 2004 to 287.91 in 2012 while Hispanic students' scores increased from 291.84 to 294.44 correspondingly. White students showed a slight increase in Math scores from 311.15 in 2004 to 313.69 in 2012. While there was a slight increase in Math scores overall and by race/ethnicity, none of the changes were identified as statistically significant.

NAEP Reading scores from 1990 to 1999 showed a slight decrease as a whole and by race/ethnicity. In 1990 the average scale score for 17-year old test takers on NAEP Reading was 290.21 and then decreased to 287.76 in 1999. White students saw a slight decline in scores from 296.64 in 1990 to 294.63 in 1999 while Black students saw a decline from 267.34 to 263.95 respectively. Hispanic students' scores decreased from 274.80 in 1990 to 270.70 in 1999 and students identifying as Other displayed a change from 290.12 to 289.79 correspondingly. While slight declines were displayed across the board in NAEP reading scores from 1990-1999 none of the changes were statistically significant.

From 2004 to 2012 17-year old NAEP Reading test takers exhibited a statistically significant increase overall. As a whole, NAEP Reading scores increased from 282.70 in 2004 to 286.89 in 2012. Black, White, and Hispanic students all displayed a statistically significant increase in NAEP Reading scores from 2004 to 2012. Black students' scores increased from 261.51 in 2004 to 269.08 in 2012. White students NAEP Reading scores increased from 288.58 in 2004 to 295.13 in 2012 while Hispanic students' scores increased from 266.85 to 273.84 respectively. Students identifying as Other exhibited an increase from 291.03 in 2004 to 295.94 in 2012, although it was not statistically significant.

There were many changes in NAEP scores pre and post NCLB as displayed in Figure 1 below. However, most score changes were not statistically significant. Pre NCLB Math scores increased significantly while post NCLB NAEP Math scores increased by only one point. Before the implementation of NCLB, NAEP Reading scores were in decline then showed significant increases after NCLB. As one statistically significant increase was pre NCLB in Math scores and the other was post NCLB in Reading scores, the effect sizes according to Cohen's d could not be meaningfully compared. The effect size according to Cohen's d was

small for each. While the results did not indicate drastic changes in student success as a whole post NCLB, graduation rates continue to climb in the post NCLB era as displayed in Figure 2 below.

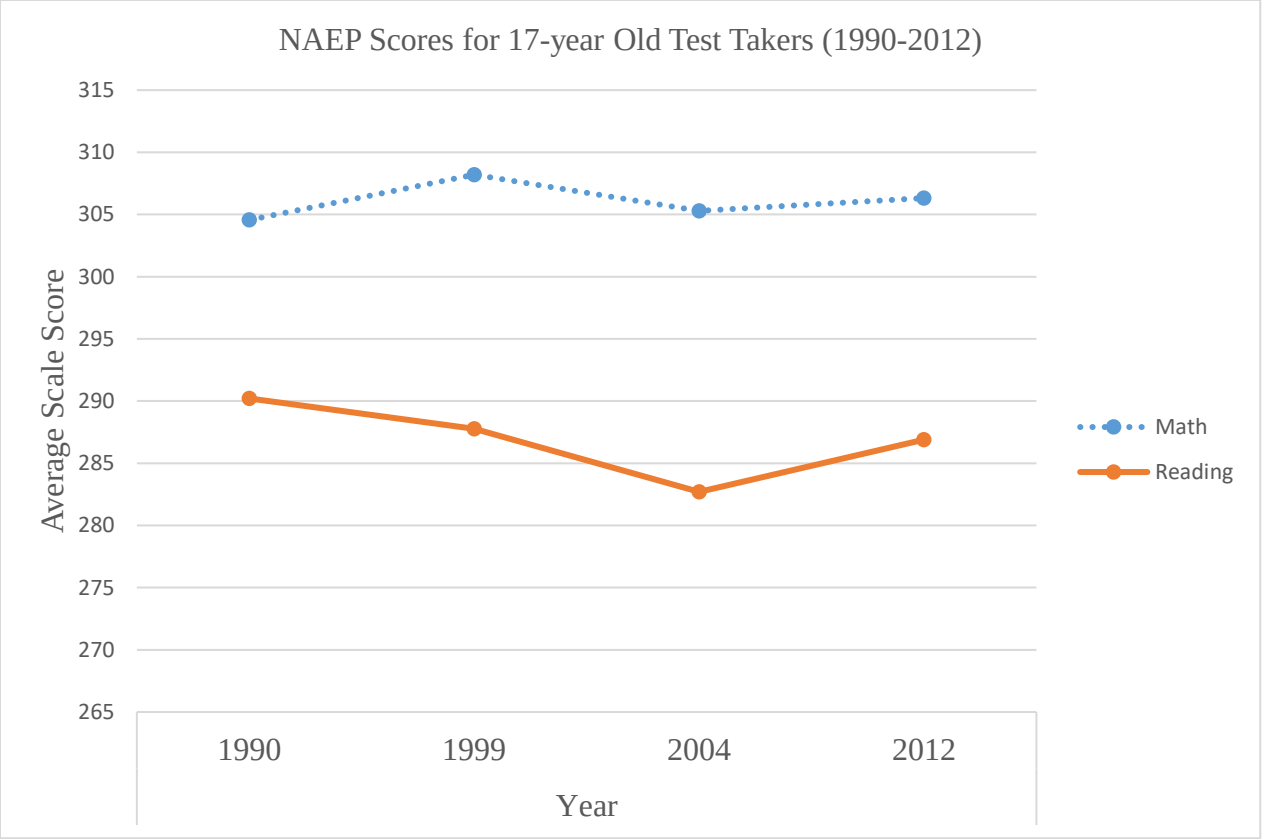

Figure 1.

NAEP Average Scale Scores 1990-2012

Note. Created using data from U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences, National Assessment of Educational Progress, 1900, 1999, 2004, and 2012 Long-Term Trend Assessments.

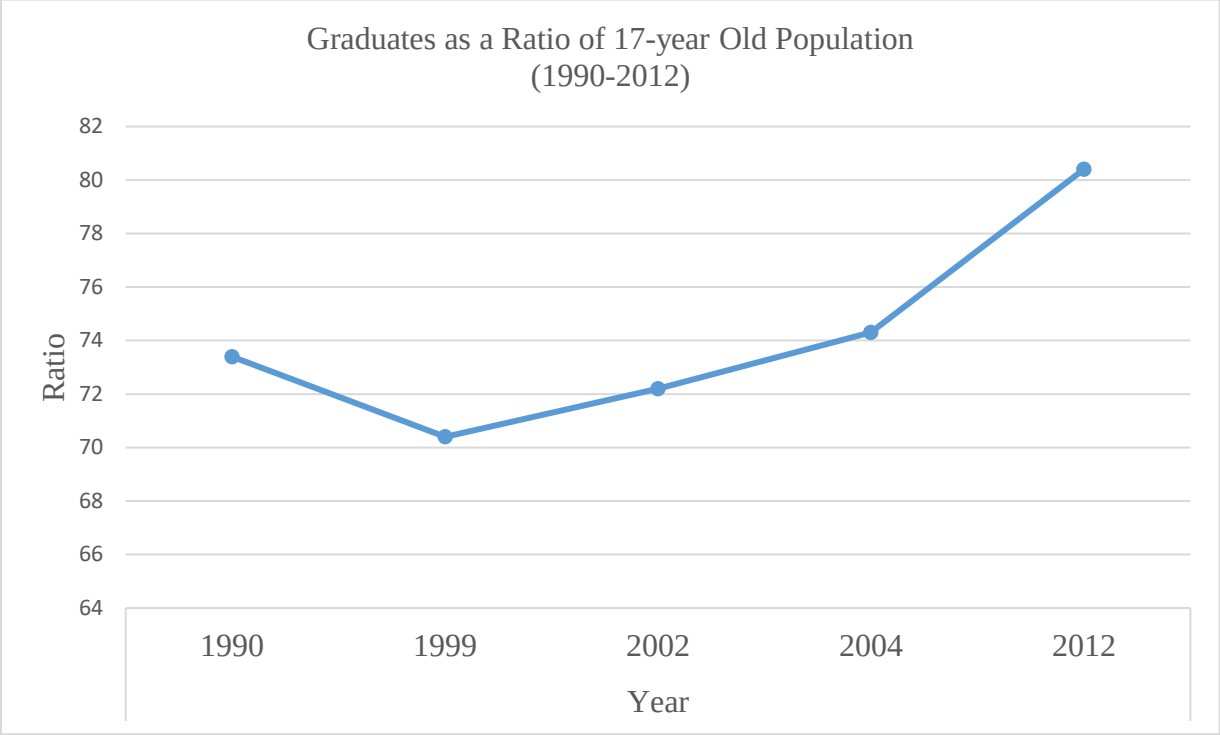

Figure 2.

Graduation Ratio 1990-2012

Note. Created using data from U.S. Department of Education, National Center for Education Statistics, Digest of Education Statistics, "Table 219.10, High School Graduates, by Sex and Control of School: Selected Years 1869-70 through 2023-24".

CHAPTER 5

CONCLUSIONS AND IMPLICATIONS

This quantitative study aimed to examine the relationship between select student demographics and NAEP scores for time periods before and after the implementation of NCLB. Specifically, this study analyzed data sets from 1990 and 1999 for the time period before NCLB and data sets from 2004 and 2012 for the time period after the implementation of NCLB. The independent variables in this study were year and race/ethnicity. The dependent variables in this study were long term trend NAEP scores in Math and Reading.

While there have been many studies evaluating NCLB and NAEP independently, there is a scarcity of information examining NAEP scores before and after NCLB's implementation. This study expanded upon traditional indicators of student success utilized for accountability by considering other indicators of success that do not allow flexibility by states or districts. Future accountability measures should expand upon success or failure of past accountability measures. School accountability measures remain at the center of education and should be effective if they are going to be used in the future or modified so that they are effective at improving student achievement. The impact of past accountability measures could be important to legislators and others involved in the development of future accountability measures.

This chapter is divided into four sections. The first part of this chapter is the discussion of findings which includes a summary of the quantitative data analysis. The second section of

this chapter provides implications of the quantitative data analysis. The next piece of this chapter examines the limitations of this study. The final section of this chapter contains possible areas for future research.

Findings

The fundamental research question leading this study was, "In the current era of school accountability, are there indicators of student success other than graduation rates among 17-year old test takers on the long term trend National Assessment of Educational Progress?" To investigate this grand tour question, long term trend NAEP scores for 17-year-old students were analyzed. Five research questions and four supporting null hypotheses were developed. Each of the null hypotheses will be detailed in the proceeding subsections.

Null Hypothesis 1

$H_0$1. There was not a change among 17-year old test takers in 1990-1999 NAEP Math and Reading scores. An independent samples t-test revealed that on average there was a statistically significant difference in NAEP Math scores from 1990 to 1999. Cohen's d indicated a small effect size. The null hypothesis for NAEP Math scores from 1990-1999 was rejected. An independent samples t-test revealed that on average there was not a statistically significant difference in NAEP Reading scores from 1990 to 1999. The null hypothesis for NAEP reading scores from 1990-1999 failed to be rejected. The data analysis in this study revealed that there was a statistically significant improvement in NAEP Math scores before the implementation of NCLB and there was not a statistically significant change in NAEP Reading scores before the implementation of NCLB (2008).

Null Hypothesis 2

$H_0$2. There was not a change among 17-year old test takers in 2004-2012 NAEP Math and Reading scores. An independent samples t-test revealed that on average there was a statistically significant difference in NAEP Reading scores from 2004 to 2012. Cohen's d indicated a small effect size. The null hypothesis for NAEP reading scores from 2004-2012 was rejected. An independent samples t-test revealed that on average there was not a statistically significant difference in NAEP Math scores from 2004 to 2012. The null hypothesis for NAEP Math scores from 2004-2012 failed to be rejected. While the data analysis from this study indicated that pre NCLB (2008) NAEP Math scores showed statistically significant improvement, post NCLB NAEP Math scores failed to show statistically significant changes. Prior to NCLB, NAEP reading scores did not show statistically significant changes overall. Post NCLB data analysis revealed that NAEP Reading scores improved significantly.

Null Hypothesis 3

$H_0$3. There were not significant differences among 17-year old test takers by race/ethnicity in 1990-1999 NAEP Math and Reading scores. An independent samples t-test revealed that on average there was a statistically significant difference in NAEP Math scores for White students from 1990 to 1999. Additionally, there was a statistically significant difference in NAEP Math scores for Hispanic students from 1990 to 1999. There were no other statistically significant differences in Math or Reading from 1990 to 1999 in any other group.

Null Hypothesis 4

$H_0$4. There were not significant differences among 17-year old test takers by race/ethnicity in 2004-2012 NAEP Math and Reading scores. An independent samples t-test revealed that on average there was a statistically significant difference in NAEP Reading scores

for White, Black, and Hispanic students from 2004 to 2012. There were no other statistically significant differences in Math or Reading from 2004 to 2012 in any group. Data analysis from this study revealed that prior to NCLB (2008), White and Hispanic students showed statistically significant gains in NAEP Math scores and post NCLB White, Black, and Hispanic students showed statistically significant improvements in NAEP Reading scores.

Additional Research Question

The research question, "Is the mean score change between 1990 and 1999 comparable to the mean score change between 2004 and 2012, as measured by Cohen's d statistic, for both reading and math?" could not be addressed for these years. The statistical significance found from 1990 to 1999 revealed a significant increase in Math scores during that time period. The statistical significance found from 2004 to 2012 revealed a significant increase in Reading scores for that time period. As statistical significance was not found in the same subject across time periods, the Cohen's d statistic cannot be reasonably compared. Additionally, in both instances, Cohen's d revealed a small effect size.

Implications

The implications of this study may prove beneficial to a variety of stakeholders. Students, parents, teachers, school leaders, and policy makers can all utilize this information when working to improve education for all. This study did not find any evidence that indicated the increasing high school graduation rate is indicative of increased secondary student success. Statistically significant gains were found in Math before NCLB (2008) was implemented but not after the implementation of NCLB. While there were no statistically significant gains in NAEP Reading pre NCLB, there were improvements post NCLB. While Reading scores saw slight improvement post NCLB, Math scores stagnated. Although there was no evidence of

disproportionate student improvement post NCLB compared to pre NCLB, secondary graduation rates continue to rise.

Implications for Policy Development

School accountability measures are in place that require data analysis and reporting of information such as graduation rates and standardized test scores (NCLB, 2008). While it is important to hold schools accountable, many of the currently utilized methods to measure student success can be manipulated to improve school and district ratings (Maleyko & Gawlik, 2011).

The NAEP "is the only nationally representative and continuing assessment of what U.S. students know and can do in various subject areas" (Hombo, 2003, p. 59). While testing was commonplace in the era of NCLB (2008), states were able to design and implement their own standardized tests, which made meaningful comparison between states impossible (NCES, 2010). The NAEP is unique as it "provides a common measure of student achievement across the country" (NCES, 2010, p. 1). The NAEP analyzes the educational attainment of students over time (NCES, 2009). The information gathered by the NAEP is used by policy makers and educators as a guide to student progress (NCES, 2009). NAEP results can be compared across states as the test and administration of the NAEP is the same in all states (NCES, 2010). The NAEP is designed to produce accurate and useful results that can be used to indicate educational progress (Beaton & Zwick, 1992).

Data manipulation has been discussed as an issue of concern regarding student success indicators associated with NCLB (2008). However, many politicians continue to reference indicators such as the graduation rate as evidence of student success. As this study has shown, NAEP score do not provide any evidence to support the demonstrated rise in graduation rates as evidence of increased student success. NCLB has been replaced with ESSA (2015) which

decreases some of the federal control over education. It is important to examine the lack of evidence supporting the positive impact of NCLB as school accountability measures move forward. Utilizing high stakes indicators of student success tied to school funding provided too much incentive to focus on or manipulate specific indicators of student success.

When working on school accountability initiatives policy makers should consider the impact of NCLB (2008) when moving forward. While high stakes student success indicators were rising, low stakes indicators such as the NAEP did not show evidence to support the disproportionate level of improvement associated with those indicators. Policy makers need to consider a greater variety of student success indicators when moving forward with school accountability initiatives in order to ensure that students are truly making significant gains.

Implications for State Education Agencies

Many state education agencies utilize high stakes test results to inform educational policy decisions. As presented throughout the study, solely using high stakes indicators of secondary student success as a measure of student improvement leaves much open for manipulation. State education agencies should utilize results from other indicators of student success in their region.

The NAEP is one example of a low stakes assessment that state educational agencies could evaluate in addition to high stakes assessments when making educational decisions. States could not only utilize Long-Term Trend NAEP, but could also utilize State NAEP assessments. State NAEP assessments began as Trial State Assessments (TSA) and were conducted in 1990, 1992, and 1994. In 1996, State NAEP was no longer used as a trial assessment. When the ESEA (1965) was reauthorized in 2001, all states that received Title I funds were required to take the State NAEP in Reading and Math. The State NAEP assessments follow the National Assessment Governing Board frameworks and use current assessment methodology (NCES,

2016). Since 2003, all 50 states, including the District of Columbia, have participated in NAEP State assessments in Math and Reading in the fourth and eighth grades (NCES, 2016). Additionally, numerous states have participated in other State NAEP assessments as shown below (Table 6).

Table 6

History of State Participation in NAEP Science, 2000-2015: Public Schools

Jurisdiction	Grade 4				Grade 8					
	2000	2005	2009	2015	1996	2000	2005	2009	2011	2015
Alabama	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Alaska	—	—	—	—	Y	—	—	—	Y	—
Arizona	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Arkansas	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
California	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Colorado	—	Y	Y	—	Y	—	Y	Y	Y	—
Connecticut	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Delaware	—	Y	Y	Y	Y	—	Y	Y	Y	Y
District of Columbia	—	—	—	—	Y	—	—	—	Y	—
Florida	—	Y	Y	Y	Y	—	Y	Y	Y	Y
Georgia	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Hawaii	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Idaho	Y	Y	Y	Y	—	Y	Y	Y	Y	Y
Illinois	Y	Y	Y	Y	—	Y	Y	Y	Y	Y
Indiana	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Iowa	Y	—	Y	Y	Y	—	—	Y	Y	Y
Kansas	—	—	—	Y	—	—	—	—	Y	Y
Kentucky	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Louisiana	Y	Y	Y	—	Y	Y	Y	Y	Y	—
Maine	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Maryland	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Massachusetts	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Michigan	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Minnesota	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Mississippi	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Missouri	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Montana	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Nebraska	Y	—	—	Y	Y	Y	—	—	Y	Y
Nevada	Y	Y	Y	Y	‡	Y	Y	Y	Y	Y
New Hampshire	—	Y	Y	Y	‡	—	Y	Y	Y	Y
New Jersey	—	Y	Y	Y	‡	—	Y	Y	Y	Y
New Mexico	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
New York	Y	—	Y	Y	Y	Y	—	Y	Y	Y
North Carolina	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
North Dakota	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Ohio	Y	Y	Y	Y	—	Y	Y	Y	Y	Y
Oklahoma	Y	Y	Y	Y	—	Y	Y	Y	Y	Y
Oregon	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Pennsylvania	—	—	Y	—	—	—	—	Y	Y	—
Rhode Island	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
South Carolina	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
South Dakota	—	Y	Y	Y	—	—	Y	Y	Y	Y
Tennessee	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Texas	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Utah	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Vermont	Y	Y	—	Y	Y	Y	Y	—	Y	Y
Virginia	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Washington	—	Y	Y	Y	Y	—	Y	Y	Y	Y
West Virginia	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Wisconsin	‡	Y	Y	Y	Y	‡	Y	Y	Y	Y
Wyoming	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y

Note. Created using data from U.S. Department of Education, National Center for Education Statistics, National Assessment of Educational Progress, "History of Participation in Public Schools".

All states have a wealth of information regarding student success at a state level in Math and Reading since 2003 or earlier, and many states have student success data in Science as well. States should utilize this information, along with other regional data, to monitor student success at the state level and target areas for improvement.

Other regional assessments that measure growth, such as the Northwest Evaluation Association (NWEA) should also be utilized to pinpoint areas of concern. The NWEA has been in place for close to 40 years and is a not-for-profit educational organization (NWEA, 2017). The main assessment associated with NWEA is Measures of Academic Progress (MAP). Over 7,400 educational agencies utilize NWEA for assessments that measure student growth, provide professional development to increase student success, and conduct research involving assessment validity and data interpretation (NWEA, 2017).

Implications for Parents and Students

Parents and students have a wealth of information at their hands in addition to results from high stakes standardized assessments. Parents and students can find information at a local level by investigating results of the State NAEP assessments. Student performance can be analyzed at a district level and compared to results from districts around the nation. Parents and students can "view trends, download snapshot reports, and compare results for each district/jurisdiction to the nation and to other districts" (NCES, 2015, para. 1).

The NWEA (2017) is another assessment that parents and students can utilized to explore student success. Parents and students receive NWEA reports and resources for the MAP assessment. This report includes a Student Profile which displays student progress from all MAP assessments through the current assessment and also includes learning paths and growth goals (NWEA, 2017). Additionally, NWEA provides a Student Goal Setting Worksheet that

diagrams assessment history and growth projections for the student. Parents and students can use this information to highlight areas of achievement and areas to focus on for growth (NWEA, 2017). This study highlighted the importance of analyzing other assessment indicators in addition to high stakes indicators and NAEP and NWEA are examples of this type of resource.

Implications for School Teachers and Leaders

School teachers and leaders have the ability to make a direct impact on student achievement for students in their schools. As this study highlighted, there are many examples of schools manipulating high stakes indicators of student success to receive additional funds associated with those indicators. School teachers and leaders can improve student success in their schools directly by not only placing emphasis on high stakes test scores, but also evaluating other indicators of student success that aren't attached to accountability measures. School teachers and leaders can evaluate long term trend NAEP results to determine how student success is changing over time at a national level. Additionally, school teachers and leaders can evaluate State NAEP results to examine student success at a local level and can also compare this to other districts and the nation as a whole.

NWEA (2017) provides school teachers and leaders with valuable information tied directly to their students. School teachers and leaders receive additional information from NWEA MAP reports that parents and students do not. School teachers and leaders can access district reports, grade breakdowns, projected proficiency reports, student growth summaries, and individual student reports. District reports provide current and historical data to help inform the district-level decision making process. Grade breakdown reports provide one spreadsheet with all student data which enables grouping and analysis of students across the school. Projected proficiency reports allow school teachers and leaders to view the projection of student

performance on state and college readiness tests. Individual student reports allow school teachers and leaders to adjust instruction and create individual student goals (NWEA, 2017).

Limitations

Limitations of this study involved the availability of specific data. Restricted use data licenses were granted for this study for the years of 1990, 1999, 2004, and 2008. However, the data needed for the 2012 analysis was not available at the time of this study. Due to this limitation, data analysis was altered to be completed through NDE. Long-term trend NAEP studies allow for data analysis over the course of multiple years but it is not possible to disaggregate the data to specific states for a different perspective of student achievement.

The study's data quality is limited to the extent that the following are true:

1. Participants answered questions to the best of their ability.
2. The sample of 17-year-old students is representative of the 17-year-old student population nationwide.
3. Long term trends NAEP scores can be compared over time.

Future Research

Recommendations for future research include the continued use of low stakes indicators of student success. As school accountability initiatives evolve over time it is imperative that policy makers utilize many indicators of student success to evaluate the impacts school accountability initiatives have on student achievement. NCLB (2008) has been replaced by ESSA (2015) and there will be a wealth of knowledge to gain from the transition. Impacts of ESSA should be evaluated using reliable, comparable data sources such as the NAEP. As new information is revealed, policy makers should continue to improve upon existing initiatives in order to provide the best possible outcomes for students.

There is room for further research during the transition into and throughout ESSA (2015). After ESSA has been in place for several years, long term trend NAEP data could be analyzed to determine if student success in Math and Reading has had any positive, significant changes since the introduction of ESSA. If significant changes are present in NAEP scores it could be an indication of ESSA having a positive impact on student achievement. If there is an absence of positive, significant change, policy makers may want to consider adjusting some of the methods utilized in ESSA.

State NAEP assessments and NWEA MAP assessments provide additional student success data. State NAEP assessments could be analyzed to evaluate student success in the fourth and eighth grades during the transition and throughout ESSA (2015). State NAEP assessments can provide a wealth of information at a district level that could be compared to other districts and the nation as a whole. As discussed earlier in this chapter, all states will have State NAEP data for Math and Reading in the fourth and eighth grades. Additionally, numerous states have data for Science in the fourth and eighth grades. NWEA MAP assessments could also provide a wealth of information on student success throughout ESSA. Analysis of student achievement data from State NAEP and NWEA MAP assessments can provide additional touchpoints for comparison of student success during and throughout future school accountability initiatives.

In addition to analyzing NAEP and NWEA scores during ESSA (2015), passing rates on state mandated accountability tests could be analyzed. While high stakes tests may not be the most reliable data source as previously discussed, they could serve as a useful point of comparison to NAEP and NWEA data. If NAEP and NWEA scores and state passing rates are improving, this would provide further evidence of the effectiveness of ESSA. If all of these

indicators are declining this would provide an indication that ESSA may need to be reevaluated by policy makers.

Progress in NAEP and NWEA scores and state passing rates could also be compared to other indicators of secondary student success. This study examined the inconsistencies of the calculation of the high school graduation rate and also showed that while graduation rates continued to rise, NAEP scores did not follow the same trajectory. The new cohort method of graduation rates provides a unified graduation calculation method across all states. The cohort graduation rate could be evaluated along with the NAEP scores before and after NCLB (2008) and also compared to the original reported graduation rates. Researchers could investigate the cohort graduation rate's relationship to NAEP scores and whether or not it provides a more accurate representation of secondary student success than previously reported graduation rates. The cohort graduation rate can continue to be investigated alongside NAEP scores throughout ESSA (2015) as well. With a new, unified method for calculating graduation rates among states and a new school accountability act, student success indicators may show a stronger association than they did in the era of NCLB.

There is room for further research during the transition into and through ESSA. A qualitative study could be completed that analyzes different methods of student success initiatives between English and Math courses. As English scores increased after the implementation of NCLB (2008) and Math scores remained stagnant, information could be gained from the methods utilized in the different courses. Teachers could be surveyed regarding methods of instruction, remediation, and retention as part of this study.

Another area for future research involves the analysis of past NAEP scores. NAEP score trajectories could be analyzed for groups of years starting with the inception of NAEP in the

1970s through today. Researchers could analyze significance and score trajectory since the 1970s and determine if there is a larger pattern of change in student success over a longer period of time. A combination of many of the previously mentioned methods could provide an interesting evaluation of secondary student success over time. After analysis of NAEP scores, state mandated test passing rates, and graduation rates, all of the data could be plotted alongside the other data and examined for consistency or lack thereof.

Further studies could be produced involving NAEP scores and school accountability measures in the era of NCLB (2008). NAEP data for other grade levels could be analyzed and compared to passing rates on state mandated tests pre and post NCLB. This could also be analyzed throughout ESSA (2015). Additionally, other NAEP assessments that provide information by state could be utilized such as the NAEP state assessments. Data from NAEP state assessments could be utilized as another metric for comparison to state mandated assessments or other state indicators of student success in primary and secondary grades.

REFERENCES

Balfanz, R., Bridgeland, J., Bruce, M., & Fox, J. H. (2013). *Building a grad nation: Progress and challenge in ending the high school dropout epidemic - 2013 annual update*. Washington, D.C.: Civic Enterprises, the Everyone Graduates Center at Johns Hopkins University School of Education, America's Promise Alliance, and the Alliance for Excellent Education. Retrieved from http://www.civicenterprises.net/MediaLibrary/Docs/Building-A-Grad-Nation-Report-2013_Full_v1.pdf

Beaton, A. E., & Zwick, R. (1992). Overview of the National Assessment of Educational Progress. *Journal of Educational Statistics, 17*(2), 95-109.

Bell, A. L., & Meinelt, K. A. (2011). A past, present, and future look at No Child Left Behind. *Human Rights, 38* (4), 11-14.

Camilli, G. (2000). Texas gains on NAEP: Points of light? *Educational Policy Analysis Archives, 8* (42). Retrieved from http://epaa.asu.edu/ojs/article/view/433

Campbell, D. T. (1975). Assessing the impact of planned social change. In G. Lyons (Ed.), *Social Research and Public Policies: The Dartmouth/OECD Conference*. Hanover, NH: Dartmouth College Public Affairs Center.

Campbell, J. R., Voelkl, K. E., & Donahue, P. L. (1997). *NAEP 1996 trends in academic progress.* (NCES 97-985r). Washington DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from http://nces.ed.gov/nationsreportcard//pdf/main1996/97985r.pdf

Campbell, J. R., Hombo, C. M., & Mazzeo, J. (2000). *NAEP 1999 trends in academic progress: Three decades of student performance.* (NCES 2000-469). Washington DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/main1999/2000469.pdf

Cowan, C. D., Hauser, R. M., Kominski, R. A., Levin, H. M., Lucas, S. R., Morgan, S. L., Spencer, M. B., Chapman, C. (2012). Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Los Angeles: Sage.

Darrow, A. (2016). The Every Student Succeeds Act (ESSA). *General Music Today, 30*(1), 41-44.

DeBray-Pelot, E., & McGuinn, P. (2009). The new politics of education. *Educational Policy, 23*(1), 15-42.

De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education*, *44,* 109-117.

Dee, T., & Jacob, B. (Summer 2010). Evaluating NCLB. *Education Next*, 54-61.

Dee, T.S., & Jacob, B. A. (Fall 2010). The impact of No Child Left Behind on students, teachers, and school. *Brookings Papers on Economic Activity,* 149-207.

Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10 (1965).

Every Student Succeeds Act of 2015, 20 U.S. C. § 6301 (2015).

Fitzsimons, P. (1999). Human capital theory and education. *The Encyclopaedia of Educational Philosophy and Theory*. M. Peters, T. Besley, A. Gibbons, B. Žarnić, P. Ghiraldelli (Eds.). Retrieved from from http://eepat.net/doku.php?id=human_capital_theory_and_education

Fraser, J. W. (Ed.). (2014). *The school in the United States: A documentary history.* New York, NY: Routledge.

Free, R. C. (Ed.) (2010). *21st century economics: A reference handbook, volume 1*. Thousand Oaks, CA: SAGE Publications, Inc.

Fuller, B., Wright, J., Gesicki, K. & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind? *Educational Researcher, 36*(5), 268-278.

Goldin, C. (2014). Human capital. *Handbook of Cliometrics*. Retrieved from http://scholar.harvard.edu/files/goldin/files/human_capital_handbook_of_cliometrics_0.pdf

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, Fall,* 38-47.

Haney, W. (2000, August). The myth of the Texas miracle in education. *Education Policy Analysis Archives, 8*(41). Retrieved from http://epaa.asu.edu/ojs/article/view/432/828

Healey, J. F., & Stepnick, A. (2016). *Diversity and society: Race, ethnicity, and gender.* Thousand Oaks, CA: SAGE Publications, Inc.

Heckman, J. J., & LaFontaine, P. A. (December 2007). *The American high school graduation rate: trends and levels* (NBER Working Paper 13670). Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w13670

Hombo, C. M. (2003). NAEP and No Child Left Behind: Technical challenges and practical solutions. *Theory Into Practice, 42*(1), 59-65.

Hout, M., & Elliott, S. W. (Eds.). (2011). *Incentives and test-based accountability in education*. Washington, D.C.: National Academies Press.

Institute of Education Sciences, National Center for Education Statistics (December 2016). National Assessment of Educational Progress. Retrieved from https://nces.ed.gov/nationsreportcard/about/state.aspx

Lane, S., Zumbo, B. D., Abedi, J., Benson, J., Dossey, J., Elliott, S. N., Kane, M., Linn, R., & Paredes-Ziker, C. (2009). Prologue: An introduction to the evaluation of NAEP. *Measurement in Education, 22*, 309-316.

Lee, J. (2007). Do national and state assessments converge for educational accountability? A meta-analytic synthesis of multiple measures in Maine and Kentucky. *Applied Measurement in Education, 20*(2), 171-203.

Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). 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), 5-14.

Loeb, S., & Figlio, D. (2011). School accountability. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education, Vol. 3* (pp. 383-423). San Diego, CA: North Holland.

Maleyko, G., & Gawlik, M. A. (2011). No Child Left Behind: What we know and what we need to know. *Education, 131*(3), 600-624.

McDonnell, L. M. (2005). No Child Left Behind and the federal role in education: Evolution or revolution? *Peabody Journal of Education, 80*(2), 19-38.

McDonnell, L. M., McLaughlin, M. J., & Morrison, P. (Eds.) (1997). *Educating one and all.* Washington, D.C.: National Academy.

Mehta, J. (2015). Escaping the shadow: A Nation at Risk and its far-reaching influence. *American Educator,* 20-44.

Meyers, C. (2012). The centralizing role of terminology: A consideration of achievement gap, NCLB, and school turnaround. *Peabody Journal of Education, 87*(4), 468-484.

National Assessment Governing Board. (2015*). Reading framework for the National Assessment of Educational Progress.* Retrieved from http://www.nagb.org/content/nagb/assets/documents/publications/frameworks/reading/2015-reading-framework-abridged.pdf

National Center for Education Statistics. (2009). *The Nation's report card: An overview of procedures for the NAEP assessment, (NCES 2009-493).* Washington DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/about/2009493.pdf

National Center for Education Statistics. (2010). *An Introduction to NAEP*. Washington DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/about/introduction_to_naep_2008.pdf.

National Center for Education Statistics. (2011, August 03). *Tools for Analyzing NAEP data.* Washington DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from https://nces.ed.gov/nationsreportcard/researchcenter/datatools2.aspx

National Center for Education Statistics. (2013). *NAEP 2012 trends in academic progress.* (NCES 2013-456). Washington DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from http://nces.ed.gov/nationsreportcard/subject/publications/main2012/pdf/2013456.pdf

National Center for Education Statistics. (2014). *High school graduates, by sex and control of school: Selected years, 1869-70 through 2023-24*[Table 219.10]. Washington DC: U. S. Department of Education, Institute of Education Sciences.

National Center for Education Statistics. (2015). *Mathematics assessment*. Washington DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from https://www.nationsreportcard.gov/profiles/districtprofile

National Commission on Excellence in Education. (1983). *A Nation at Risk.* Retrieved from http://www2.ed.gov/pubs/NatAtRisk/risk.html

National Governor's Association, Task Force on State High School Graduation Data. (2005). *Redesigning the American High School.* Retrieved from http://www.nga.org/files/live/sites/NGA/files/pdf/0507GRAD.PDF

News Briefs: Gaps in NAEP and proficiency levels on state tests. [News Briefs]. (2007). *Gifted Child Today, 30*(4), 7-9.

Nichols, S. L. (2007). High-stakes testing: Does it increase achievement? *Journal of Applied School Psychology, 23*(2), 47-64.

No Child Left Behind Act of 2001, 20 U.S.C. § 6319 (2008).

Northwest Evaluation Association, MAP Reports Reference (July 2017). Retrieved from https://teach.mapnwea.org/impl/PGM2_MAP_Reports_Reference.pdf

Perie, M., Moran, R., and Lutkus, A. D. (2005). *NAEP 2004 trends in academic progress: Three decades of student performance in reading and mathematics.* (NCES 2005-464). Washington, DC: Government Printing Office. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2005/2005464.pdf

Porter, A. C., Linn, L., & Trimble, C. S. (2005). The effects of state decisions about NCLB adequate yearly progress targets. *Educational Measurement: Issues and Practice, 24*(4), 32-39.

Rampey, B. D., Dion, G. S., & Donahue, P. L. (2009). *NAEP 2008 trends in academic progress.* (NCES 2009-479). Washington DC: U. S. Department of Education, Institute of Education Sciences. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2008/2009479.pdf

Seastrom, M., Chapman, C., Stillwell, R., McGrath, D., Peltola, P., Dinkes, R., & Xu, Z. (2006). *User's guide to computing high school graduation rates, volume 1: Review of current and proposed graduation indicators* (NCES 2006-604). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Shepard, L. A. (1990). Inflated test score gains: Is the problem old norms or teaching to the test? *Educational Measurement: Issues and Practice, 9*(3), 15-22.

Smylie, M. A. (1996). From bureaucratic control to building human capital: The importance of teacher learning in education reform. *Educational Researcher, 25*(9), 9-11.

Span, C. M., & Rivers, I. D. (2012). Reassessing the achievement gap: An intergenerational comparison of African American student achievement before and after compensatory education and the Elementary and Secondary Education Act. *Teachers College Record, 114*, 1-17.

Stanley, K. R., Spradlin, T. E. & Plucker, J. A. (2008). Calculating high school graduation rates. *Education Policy Brief, 6*(5). Indiana University, Center for Evaluation and Education Policy.

Strober, M. H. (1990). Human capital theory: Implications for HR managers. *Industrial Relations, 29*(2), 161-186.

Swanson, C. B. (2004). *High school graduation, completion, and dropout (GCD) indicators.* Washington, D.C.: The Urban Institute. Retrieved from http://webarchive.urban.org/publications/411116.html

Swanson, C. B. (2003). *NCLB implementation report: State approaches for calculating high school graduation rates*. Washington, D.C.: The Urban Institute. Retrieved from http://www.urban.org/url.cfm?ID=410848

Thomas, J. Y., & Brady, K. P. (2005). The Elementary and Secondary Education Act at 40: Equity, accountability, and the evolving federal role in public education. *Review of Research in Education, 29*, 51-67.

Trolian, T. L., & Fouts, K. S. (2011). No child left behind: Implications for college student learning. *About Campus,* 2-7.

U.S. Department of Education. (2012, November 26). *States report new high school graduation rates using more accurate common measure* (Press Release). Archives of U.S. Department of Education. Retrieved from http://www.ed.gov/news/press-releases/states-report-new-high-school-graduation-rates-using-more-accurate-common-measure

U.S. Department of Education. (2015, February 12). *U.S. high school graduation rate hits new record high.* (Press Release). Retrieved from http://www.ed.gov/news/press-releases/us-high-school-graduation-rate-hits-new-record-high

U.S. Department of Education. (2016, May 4). *Transitioning to the Every Student Succeeds Act (ESSA).* Retrieved from http://www2.ed.gov/policy/elsec/leg/essa/faq/essatransitionfaqs050316.pdf

Wei, X. (2012). Are more stringent NCLB state accountability systems associated with better student outcomes? An analysis of NAEP results across states. *Educational Policy, 26*(2), 268-308.

APPENDIX: RESTRICTED USE DATA LICENSE

License number: 17010022

Dear Brad Balch,

We have completed our review of your application. Your application has been approved. Keep a copy of this License approval email in your License file. Please reference your License number in any future correspondence.

We will be mailing the data and a copy of your License paperwork to you in the next few business days. Please retain the documents in your License file.

Once you receive the data, please contact the IES Data Security Office for the password that you will need to unencrypt the data. You must secure this password from unauthorized access.

You are licensed for the data you requested in your Formal Request. The data you will receive are on loan for the time period specified in your Formal Request commencing with the date of the NCES Commissioner's signature on the signed License document. Once your License expires, please send the data back to IES and close the License or submit an amendment to renew the License.

Please read and have other project staff read the IES Restricted-use Data Procedures Manual at: http://nces.ed.gov/statprog/rudman/toc.asp . Violations of any of the License or security requirements by any project staff could result in License cancellation.

If during the course of your research you need to add project staff, take staff off of the project, need additional restricted-use data, extend the time period of your License, or need to close your License, please read how to do any of these at:http://nces.ed.gov/statprog/instruct_mod.asp

Any draft reports or other pre-publication documents that use or contain IES restricted-use data must be reviewed by the IES Data Security office prior to their dissemination outside the licensed project staff. Please send these reports to the email address below.

If you have any questions, please contact us.

IES Data Security

Department of Education/IES/NCES

550 12th Street SW, Room 4060

Washington, DC 20202

Office: 202-245-7674

IESData.Security@ed.gov