Florida High Stakes Testing and Graduation Success

Laurel Alva Stanley
University of North Florida

Suggested Citation
https://digitalcommons.unf.edu/etd/319
FLORIDA HIGH STAKES TESTING AND
GRADUATION SUCCESS

by
Laurel Alva Stanley

A dissertation submitted to the Doctoral Faculty of the College of Education and Human Services in partial fulfillment of the requirements for the degree of

Doctor of Education

UNIVERSITY OF NORTH FLORIDA
COLLEGE OF EDUCATION AND HUMAN SERVICES

March, 2007

Unpublished work © Laurel Alva Stanley
The dissertation of Laurel Stanley is approved:

Signature Deleted

Joyce T. Jones, Ed.D., Chair

Signature Deleted

Candice Carter, Ph.D.

Signature Deleted

Larry G. Daniel, Ph.D.

Signature Deleted

David E.W. Fenner, Ph.D.

Accepting for the Department:

Signature Deleted

Joyce T. Jones, Ed.D., Chair
Department of Counseling and Educational Leadership

Accepting for the College:

Signature Deleted

Larry G. Daniel, Ph.D., Dean
College of Education & Human Services

Accepting for the University:

Signature Deleted

David E. W. Fenner, Ph.D.
Dean of Graduate Studies
ACKNOWLEDGEMENT

I would like to thank my family for their steadfast and loyal support throughout the demanding and rigorous dissertation process. The support of my husband, Nile, and my daughter, Jane, was essential to my finishing the project. Through their inspiration, a great deal of meaning was added to this dissertation. I would especially like to thank my brother, Mark, for helping me with the data processing of the thousands of student records that formed the basis for the statistical analysis. His computer skills are amazing and I never could have done it without him.

My dissertation committee deserves many thanks for their unwavering support throughout the dissertation process. I had the benefit of two dissertation chairs, Dr. Pritch Smith and Dr. Joyce Jones. Dr. Smith shared his tremendous expertise in multicultural education and research as he guided me on my way. Dr Jones shared her extensive knowledge of research and writing so that I could reap the harvest of a completed dissertation.

Members of the dissertation committee included Dr. Larry Daniels, who generously offered me advice and support for the statistical analysis, Dr. David Fenner who provided philosophical wisdom to inspire me, and Dr. Candace Carter, whose expertise in critical theory and social studies was especially valuable. I would also like to acknowledge the staff at the Florida Department of Education Data Warehouse for their patience and support. Finally, I would like to acknowledge Gary Orfield and other members of The Civil Rights Project at Harvard University whose work served as an inspiration for this dissertation.
DEDICATION

I dedicate this dissertation to all students and their parents, especially those in Florida, who come to school bright and eager to learn. We owe these citizens, as professional educators, a quality education. Sadly, too few children entrusted to our education system acquire the knowledge and passing high stakes test scores to fully reap the benefits that our education system has to offer. More often, those children who do not graduate have parents who are also not members of the privileged, more educated social classes in our society.

Responsibility for creating a more inclusive education system should be fully accepted by the education community. Included in this community are researchers, teachers, professors, administrators, school board officials, politicians, parents, students, and citizens. This dissertation is an attempt at responsibly producing needed information on the relationship between high stakes testing policies and the chances of student graduation from high school through the disciplined study and analysis of archival records and primary student data. My hope is that when enough quality information is produced and made available to educators and the public, then thoughtful decision making will follow and real changes will occur in our schools.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENT</td>
<td>i</td>
</tr>
<tr>
<td>DEDICATION</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td>CHAPTER 1: INTRODUCTION TO THE STUDY</td>
<td>1</td>
</tr>
<tr>
<td>- Significance of the Research</td>
<td>8</td>
</tr>
<tr>
<td>- Statement of Purpose</td>
<td>9</td>
</tr>
<tr>
<td>- Research Design</td>
<td>9</td>
</tr>
<tr>
<td>- Data Analysis</td>
<td>11</td>
</tr>
<tr>
<td>- Definition of Terms</td>
<td>12</td>
</tr>
<tr>
<td>- Assumptions</td>
<td>13</td>
</tr>
<tr>
<td>- Limitations and Delimitations</td>
<td>14</td>
</tr>
<tr>
<td>- Organization of the Study</td>
<td>15</td>
</tr>
<tr>
<td>CHAPTER 2: REVIEW OF THE LITERATURE</td>
<td>16</td>
</tr>
<tr>
<td>- A History of Florida High Stakes Testing</td>
<td>16</td>
</tr>
<tr>
<td>- Theoretical Literature</td>
<td>22</td>
</tr>
<tr>
<td>- The Appropriate Use of Tests</td>
<td>22</td>
</tr>
<tr>
<td>- The Value of High Stakes Testing and Accountability</td>
<td>24</td>
</tr>
<tr>
<td>- The Controversy of High Stakes Testing</td>
<td>25</td>
</tr>
<tr>
<td>- Eight Assumptions about High Stakes Testing</td>
<td>29</td>
</tr>
<tr>
<td>- Achievement at a High Level</td>
<td>29</td>
</tr>
<tr>
<td>- Students Not Doing Well in School</td>
<td>30</td>
</tr>
<tr>
<td>- Improved Student Learning</td>
<td>31</td>
</tr>
<tr>
<td>- Increased Motivation</td>
<td>35</td>
</tr>
<tr>
<td>- Fair and Neutral Testing</td>
<td>36</td>
</tr>
<tr>
<td>- Reliable, Research, and Valid</td>
<td>37</td>
</tr>
<tr>
<td>- Increased Equity</td>
<td>39</td>
</tr>
<tr>
<td>- Economic Productivity</td>
<td>40</td>
</tr>
<tr>
<td>- Disparate Graduation Rates: Institutional Racism</td>
<td>41</td>
</tr>
<tr>
<td>- Calculating Graduation Rates</td>
<td>44</td>
</tr>
<tr>
<td>- Conceptual Framework</td>
<td>49</td>
</tr>
<tr>
<td>- Empirical Research</td>
<td>52</td>
</tr>
<tr>
<td>- Studies of High Stakes Testing and Graduation Rates</td>
<td>54</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Number</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1</td>
<td>Conceptual Framework</td>
<td>51</td>
</tr>
<tr>
<td>Figure 2</td>
<td>Changes in Demographics from 1977 to 2005 of Percentages White Identity, Exceptional Student Education, and Limited English Proficiency Students</td>
<td>96</td>
</tr>
<tr>
<td>Figure 3</td>
<td>Raw CPI Graduation Rates in Florida 1977-2005</td>
<td>97</td>
</tr>
<tr>
<td>Number</td>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Table 1</td>
<td>Cumulative Promotion Index (CPI) in Florida: 1978-2005</td>
<td>92</td>
</tr>
<tr>
<td>Table 2</td>
<td>Periods of Florida High Stakes Testing</td>
<td>94</td>
</tr>
<tr>
<td>Table 3</td>
<td>Correlations for the Macro Study</td>
<td>99</td>
</tr>
<tr>
<td>Table 4</td>
<td>Correlations for Subgroups in Macro Study</td>
<td>100</td>
</tr>
<tr>
<td>Table 5</td>
<td>Linear Regression Effect Size for HO₁</td>
<td>100</td>
</tr>
<tr>
<td>Table 6</td>
<td>Sum of Square Breakdown for HO₁</td>
<td>101</td>
</tr>
<tr>
<td>Table 7</td>
<td>Coefficients for Linear Regression for HO₁</td>
<td>102</td>
</tr>
<tr>
<td>Table 8</td>
<td>Descriptive Statistics for Means and Standard Deviations Across Time Periods</td>
<td>104</td>
</tr>
<tr>
<td>Table 9</td>
<td>Descriptive Statistics for Mean and Standard Deviation for Time Period and Non-Standardized Residual for HO₂</td>
<td>104</td>
</tr>
<tr>
<td>Table 10</td>
<td>Test of Between-Subjects Effects for HO₂</td>
<td>105</td>
</tr>
<tr>
<td>Table 11</td>
<td>Demographics of the Random Sample of 3,000 Florida 9th Grade Student 1999-2003</td>
<td>108</td>
</tr>
<tr>
<td>Table 12</td>
<td>Graduation and Ethnic Subgroups of 1998 Retained 9th Graders</td>
<td>109</td>
</tr>
<tr>
<td>Table 13</td>
<td>Disproportionality of 1,342 Students w/o FCAT Math Scores</td>
<td>110</td>
</tr>
<tr>
<td>Table 14</td>
<td>FCAT Score Frequencies and Percentages of 3,000 Florida 9th Grade Students</td>
<td>111</td>
</tr>
<tr>
<td>Table 15</td>
<td>Measures of Central Tendency for FCAT Scores</td>
<td>112</td>
</tr>
</tbody>
</table>
Table 16 Discriminant Analysis Case Processing Summary for HO₃  
Table 17 Wilks' Lambda for HO₃  
Table 18 Standardized Canonical Coefficients and Structure Matrix HO₃  
Table 19 Cross Validation of HO₃  
Table 20 Discriminant Analysis Case Processing Summary for HO₄  
Table 21 Wilks' Lambda for HO₄  
Table 22 Standardized Canonical Coefficients and Structure Matrix HO₄  
Table 23 Cross Validation of HO₄
ABSTRACT

Florida has the longest history of high stakes testing of any state in the United States, spanning a period of 28 years, seven different testing policy periods, and five tests. In this study, analysis of statistics from the 28 years suggested that declines in graduation rates corresponded to increasingly difficult high stakes testing policy periods, new tests, and periods that did not include high stakes accountability for graduation rates. Overall, graduation rates declined through an erratic 28 years for a net loss of 4.51%. The achievement gap in graduation rates between white and black students worsened 200% from 1992 to 2003. Analysis of a random sample of 3,000 Florida 9th grade students in 1999 indicated that 42.6% of students graduated within four years with a standard diploma. School achievement variables, including grade point average, retention history, high stakes test scores, and attendance, were found to be the best predictors of individual student graduation. Implications were that to benefit the individual student for graduation, teachers and other educators must work to ensure academic success. Educators should conduct further studies to better understand the relationship between graduation success and high stakes testing policies.
CHAPTER 1

INTRODUCTION TO THE STUDY

One of the most important indicators of an education system, personal achievement, and the economic well being of individuals and society is high school graduation. The United States currently fails to graduate many children; in some school districts more than 50% of children fail to graduate. In 2000, the industrialized nations of Finland, France, Germany, Italy, Poland, the Slovak Republic, Hungary, South Korea, and Japan all graduated a higher percentage of students from upper secondary school (high school) at the typical age than the United States (Schleicher, 2002).

The dominant educational paradigm of the “new millennium” (Wikipedia, the free encyclopedia, 2005) in the United States is standardization and accountability as measured by a set of high stakes tests. Proponents of high stakes testing policies, backed up by most recent decisions in courts of law (Mulvenon, Murry, Jr., & Ritter, 2001), assert that the government has a legitimate interest in determining the quality of education and how well students and schools are achieving as measured by such assessments. The test results purportedly indicate differences in individual and group academic abilities and capabilities that are relevant. Requirements for student promotion and graduation are derived from test scores. Without the tests and associated public disclosures, differences in educational outcomes might be unrecognized or ignored.
Mandatory accountability for differences among students raises questions of equality of opportunity in education. A renewed emphasis on the more advantaged students seems to be supported by a statement of Alan Greenspan, who called for an exclusive focus in high schools on essentially a standardized honors program, supported by accountability in the form of testing, which he claimed, is needed for economic productivity in a global economy (Greenspan, 2004). The business model for education became a leading force in public education of the United States. The model is based on claims of inadequate academic achievement of students, widespread assumptions about the wisdom of business methods to improve schools, and the belief in economic benefits related to high-performance schooling (National Commission on Excellence in Education, 1983).

Emerging globalization, characterized by the intensification of social, economic, and transnational relationships around the world (Torres, 2002), supports and encompasses the business model, standardization, and high stakes testing. Global businesses create school curricula, conduct assessments, and access the for profit 2 trillion dollar international education system, as they change the definition of education from a social right to a commercial commodity (Siqueria, 2005). Privatization, for example, of education has resulted in enormous amounts of profit for a multibillion dollar industry composed of a small number of testing and publishing companies, such as McGraw-Hill, Houghton-Mifflin, and Harcourt General (Metcalf, 2002).

The massive government and business effort to standardize schools has been touted as a way to spread economic growth in the world. Caution may be warranted, however, because increased prosperity for all from globalization has not been the result,
as evidenced by the increased poverty for people in China and many other developing countries (UNESCO, 2005). Challenges to globalization from widespread social movements that offer progressive alternatives to the regressive economic policies, imperialism, and wars associated with globalization have become increasingly common and powerful (Fisher, 2001). Challenges to the high stakes testing paradigm have emerged as well from social action groups such as FairTest: The National Center for Fair and Open Testing (2004, October).

Scholars have argued that the new standardization, accountability, and high stakes testing paradigm is anti-democratic and acts in oppressive ways by normalizing inappropriate surveillance, punishment, and obedience to authority, while reducing criticism, social solidarities, and collective action (Broscio, 2003; Lipman, 2005; Vinson, 1999). Such initiatives as the standardization movement have been long contested by scholars such as Paulo Freire (1970), who referred to standardization schemes as “banking” education, where students are filled with deposits of information by the teacher for an accountable return. Progressive educators in the tradition of John Dewey have contested standardization as not accommodating the multifaceted cultural experiences of students that are known to enhance their learning (Vinson, Gibson, & Ross, 2001).

Freire (1970) associated standardization schemes with a basic, fundamental relationship between oppressor and oppressed known as prescription, where one individual or group imposes their consciousness upon another, transforming the dominated consciousness into a prescribed one in accordance with the oppressor’s thinking. Politics pervades commercial media with dissemination of information in a
political drive to influence public perception of high stakes testing, as evidenced by the payment of $241,000 in 2004 by the government to media commentator Armstrong Williams in order to make President Bush's high stakes testing policies look attractive (Kurtz, 2005). Partially as a result of such presentations and political speeches, the public perception of high stakes testing is predominantly positive (Bracey, 2000).

The testing paradigm is not new but has developed steadily for over 100 years, with the early history of 20th century standardized testing in the United States marked by the Stanford-Binet IQ test and Thorndike's standardized tests in school subjects (Grant, 2004). Government interest in mental and scholastic testing grew before and during World War II, when testing was used in the United States as a sorting tool for social and military classification purposes (Grant, 2004). When the Russians launched Sputnik in 1958, before its United States counterpart, a massive new effort in education for the purpose of obtaining a more competitive stature in mathematics and sciences occurred in the U.S.A.

About the same time, in the 1960s and 1970s, schools in the United States experienced racial desegregation. Some schools used progressive ideas, such as pedagogical approaches valuing psychology and the experience of the student (Dewey, 1956), democracy and interaction (Dewey, 1966), reflective inquiry, cooperation, and multiculturalism (Vinson, Gibson, & Ross, 2001). Many of these progressive ideas in the schools ended literally and philosophically when President Ronald Reagan claimed the education system was in a state of "sorry repair" and that reform through standardization, accountability, testing, and privatization was needed (Finn, Jr., 2004). Responsively, Secretary of Education William Bennett spearheaded a back-to-basics movement that
eliminated the supporting budgets for many vocational programs, the arts, and self-esteem programs, while concentrating on student achievement in reading and mathematics (Tellez, 2003).

The Reagan Administration appointed a National Commission on Excellence in Education, comprised of prominent businessmen and politicians that published the landmark report, *A Nation at Risk: The Imperative for Educational Reform* (1983). The report called for increased accountability, standardization, and emphasis on economic competitiveness for U.S. schools. Politically and socially, a liberal-conservative alliance supporting the school reforms of standardization and high stakes testing made the widespread adoption of the new educational paradigm inevitable (Vinson, et al., 2001). In the 1990s, President Clinton called for testing every child in the 4th and 8th grade in reading and math to insure accountability in meeting the standards (Grant, 2004).

The signature piece of legislation of President George W. Bush, the *No Child Left Behind Act* (NCLB) of 2001, included the provision that all students pass a number of high stakes tests, defined as assessments that are used to make significant educational decisions about students, schools, and educators. The high stakes tests and the required consequences for schools, students, and educators serve as the basis for a powerful accountability system that determines students' and educator's success or failure. The testing influences the professional behavior of educators, the curriculum, school funding, and public perceptions. Some research has indicated that the tests and related policies are effective in providing motivation to students and educators to obtain higher academic achievement, as evidenced by higher test scores and a narrowing of the achievement gap.
between white students and students of color (Carnoy & Loeb, 2002; Florida Department of Education, 2004a; Hanushek & Raymond, 2003; Rosenshine, 2003).

However, scholars find that the assumptions about students' motivation are faulty; the validity of claims about improvements in academic achievement are questionable; the inequity between racial/ethnic and socio-economic groups is unacceptable; the elite position of privilege of the upper classes is undemocratic; and the unintended consequences such as reduced graduation rates bring into question the merit of the high stakes testing policies (Haney, 2000; Orfield, Losen, Wald, & Swanson, 2004; Sleeter, 2003; Valenzuela, 2002). Controversy and problems with the high stakes testing are evident when students fail; then sometimes, students fail repeatedly, and they are denied graduation.

Helpful insight may be gleaned from looking at perspectives of proponents and opponents of the testing policies. Proponents claim that identifying and helping failing students with additional resources and “scientifically based” teaching methods offers these students their best chance for participation in the education and economic systems of today (U. S. Department of Education, 2001). Proponents further claim that social promotion, or the practice of passing failing students through the system, does nothing for the student or society, except perpetuate weakness. Proponents further state that the economic weaknesses of the United States in global and domestic markets are traced to educational weaknesses such as “soft curriculums” that must be overcome by increased emphasis on a rigorous program of English, mathematics, science, computer science, and foreign language (National Commission on Excellence in Education, 1983).
Opponents offer other explanations instead for low ability or lack of attention to studies for poor performance on the tests, including differing cultures, backgrounds, and experiences that might be especially valid for low-income children, ethnic minorities, and children of color (Debra P. v. Turlington, 1979; Gopaul-McNicol & Armour-Thomas, 2002; Tellez, 2003). There are also questions about performance motivation for low socio-economic and historically low performing racial/ethnic students, who may perceive themselves as marginalized from a testing system designed for predominately white, middle class, dominant culture students (Sefa Dei, Mazzuca, McIsaac, & Zine, 1997; Stephens, Sadler, & Moss, 2002). From a critical perspective, deserving students can be unjustly denied their life-altering high school diplomas when schooling, aided by different amounts of individual social capital, reproduces socio-economic classes (Bourdieu, 1986; Bourdieu & Passeron, 1977).

Analysis of power relations between the proponents of the tests and the opponents of the tests is relevant when determining who is advantaged by them. High stakes testing policies provide wealthy and powerful government officials and business people greater influence on schooling, methods, and its outcomes. Those policy makers who recommend a curricula or pedagogical practices, which are enforced by a testing regime functioning as a sorting mechanism, have the power to create the production of citizens prepared for dominant or subordinate positions in society (Grant, 2004; McLaren, 1998). Fairness is questionable because other interests besides the policy makers' are critically at stake. All citizens of this nation are affected by high stakes testing policy, especially those who may suffer the consequences of failing to graduate.
Although the courts have found that elitist results of the high stakes testing are unintentional, it is believed that evidence of the implicit support for elitism may be found in the disparate passing and graduation rates between educated, middle class, predominately white students and lower class students, particularly students of color (Orfield, Losen, Wald, & Swanson, 2004; Swanson, 2003). Scholars have claimed that standardization and high stakes testing in the school reduces the depth, scope, quality, and quantity of the curriculum, and, over the long term, standardization creates inequalities and widens the achievement gap between poor and minority students and more privileged students (Darling-Hammond, 2004; Gordon, Piana, & Keleher, 2000; Grant, 2004; McNeil, 2000; Sleeter, 2003; Stephens, et al., 2002). Due to the serious consequences of failure to graduate from high school for individuals, families, and society, critical examination of the relevant circumstances and relationships between the high stakes testing and graduation is appropriate.

Significance of the Research

Relevant insights into high stakes testing policies and their consequences were gleaned from studying Florida, because Florida has the longest and most comprehensive history of high-stakes testing in the United States and low graduation rates. Too many students in Florida, especially those in racial/ethnic minority groups, the poor, and immigrant students are not receiving an adequate free public education, as evidenced by low Florida graduation rates of 53% overall, 41% for African Americans, and 52% for Hispanics (Orfield et al., 2004). The study of graduation success and high stakes testing in Florida contributed to the knowledge base by providing statistical analysis of the relationship between 28 years of high stakes assessment testing and graduation success.
Through exploring relationships between success in graduation from high school and demographic variables such as grades, Florida Comprehensive Achievement Test (FCAT) scores, retention history, free/reduced lunch status, and gender, insights into the mechanics of success were gained that may be useful to teachers, educators, students, citizens, and scholars. Identification of predictive demographic and achievement variables may help stakeholders work together towards a time when every child graduates from high school in the United States.

Statement of Purpose

One purpose of the study was to produce information and analysis concerning the relationship between high-stakes testing policies and the likelihood of students graduating from high school in Florida. Educational leaders, policymakers, researchers, citizens, students, and families may find the information produced by this study of broad relationships between the powerful high stakes testing policies and high school graduation useful.

Another purpose of the study was to produce descriptive and analytical information about how achievement and demographic variables influence the likelihood for an individual student to graduate from high school in Florida. Important variables influencing a student’s graduation prospects were identified. The information produced by this study may be used by students, families, teachers, educators, researchers, and policymakers focusing on the individual student level.

Research Design

The study was conducted in a macro phase and a micro phase. The macro phase investigated the State of Florida graduation rate of students from high school as related to
the seven time periods when five different high stakes testing programs were used: (a) 1978, during the basic skills test, SSAT-1, but before any graduation test; (b) 1979, the functional literacy test, SSAT-11 was required for graduation; (c) 1980 - 1982, during the suspension of withholding high school diplomas due to the Debra P. v. Turlington (1979) court case; (d) 1983 – 1990, implementation of a revised Secondary School Admission Test, SSAT-11 with increased standards based rigor, mandatory retention, and graduation consequences; (e) 1991 -1997, during the High School Competency Test (HSCT) of the 1990s; (f) 1998 – 2002, during the FCAT testing from 1998 to 2002; and (g) 2003 - 2005 during the FCAT testing and the inclusion of AYP for school grades. The time periods were adjusted for the variation of demographics of racial/ethnic identity, exceptional student education, and limited English proficient students. Data consisted of Florida Department of Education aggregated high school enrollment counts, graduation counts, and demographic data. Analysis consisted of linear regression, univariate analysis of variance, and descriptive statistics.

The micro phase examined whether high stakes test scores, history of retention, attendance records, grades, racial/ethnic identity, free or reduced lunch status, limited English proficiency, and special education status could be used to predict individual student graduation success. Using data from school records of a random sample of 3,000 public high school students in Florida, the independent variables of high stakes test scores (FCAT), history of retention, attendance records, grades, racial/ethnic identity, free or reduced lunch status, limited English proficient, gender, and exceptional student education status and the dependent variable graduation success were examined with
descriptive statistics and discriminant analysis to help understand the factors predictive of student graduation success.

Data Analysis

The macro study analysis consisted of descriptive and inferential statistical analysis of aggregated data from the Florida Department of Education covering a period of 28 years. The independent variables were the percentages of white identity, exceptional student education, and limited English proficiency, and seven different high stakes testing policy periods. The dependent variables were annual graduate rates each year for 28 years and a non-standardized residual, a variable created by removing the variations from the percentages of white identity, exceptional student education, and limited English proficiency from graduation rate. Descriptive statistics included the Cumulative Promotion Index (CPI) for 28 years, demographics of percentages of white identity, exceptional student education, and limited English proficiency, description of seven different policy periods, and correlations between the independent and dependent variables. Inferential statistics consisted of linear regression.

The micro phase consisted of descriptive and inferential analysis; however, the data were at the individual student level. The independent variables included gender, racial/ethnic identity, free/reduced lunch status, limited English proficiency, exceptional student education, grades, retention history, attendance, and FCAT scores for the 8th and 10th grade. The dependent variable was graduation from high school. Descriptive statistics furnished information on frequencies and percentages. Discriminant analysis was used and then cross validated with three random samples to find out how good a
model of graduation success could be derived from the variables. The strength of independent variables in predicting graduation success was determined.

Definition of Terms

In this study, the following terms will be defined as follows.

- Graduation will refer to a student's award of a standard high school diploma, obtained within four years of entering the ninth grade.

- Meritocracy is a system where good student work is justly rewarded and bad student work is not rewarded or is punished.

- Racial/ethnic identity describes assigned labels for student descriptors that government provides in required demographic forms for school attendance.

- Standards refer to delineated frameworks of field specific knowledge that all students are expected to know and schools are expected to teach.

- Standardization refers to the educational reform that aligns school curriculum, instruction, and assessment with state determined standards, typically conforming to nationally determined standards set by professional or political organizations.

- The No Child Left Behind (NCLB) Act of 2001 was signed into law by President George W. Bush for the purpose of mandating educational reform in the U. S.

- Accountability is an education reform that holds students, teachers, administrators, and schools accountable for performance on measures such as high stakes tests, Adequate Yearly Progress (AYP) of schools, and adherence to a required standardized educational program.
• States must identify concrete annual performance targets, known as Adequate Yearly Progress (AYP) that include content standards, standards for student and school academic achievement, and graduation rates.

• High stakes tests are mandatory standardized tests that are used to make significant decisions about students, teachers, and schools.

• High stakes testing policies refers to a comprehensive set of related policies, including required adherence to a standardized curriculum and prescribed pedagogical methods, mandatory retention in grade for poor performance, graduation test requirements, school and student grading with prescribed consequences for students, teachers, administrators, and schools, and economic consequences for communities.

• A macro perspective looks at the workings of broad policies over whole populations. For this study, the macro perspective examines the workings of state and national high stakes testing policies over Florida aggregated graduation rates for a period of 28 years.

• A micro perspective looks at the detailed functioning of people and phenomena. For this study, the micro perspective examines data from individual students and their demographic characteristics, academic achievements, and school records as related to graduation over a four year period.

Assumptions

One assumption of the study was that accurate archival records were published by the State of Florida Department of Education publications, including the *Florida Profiles of Staff and Student Data* (1986-2005) and others. Another was that graduation and
enrollment numbers and the resulting graduation rates in Florida measure the same phenomena over the 28 year period. A third was that different periods involving high stakes testing are comparable in a meaningful way in the context of the variables chosen for this quantitative analysis. A fourth was that the migration within the two year period for Cumulative Promotion Index calculations was not a threat to the validity of the study.

Limitations and Delimitations

The study used aggregated and disaggregated State of Florida Department of Education data; therefore, results were limited to those derived from this data. Statistical tests cannot be used to indicate causation, but they may point the way for further meaningful studies. Illumination of the relationship between high stakes testing policies and graduation may be achieved through descriptive statistics, multiple regression analysis, univariate analysis of variance, and discriminant analysis.

Analysis of data in the form of enrollment counts, graduation counts, demographic data, and historic periods of high stakes testing focused on relationships between graduation rate trends and high stakes testing for the last 28 years. In addition, data from a random sample of 3,000 individual high school students' records from 1999-2003 in Florida provided insight into relationships between graduation success and student demographic and achievement data. Use of a random sample of Florida public school data rather than creating data through surveys, collecting data from individual schools, or creating data from experiments provided a high degree of contextual relevance for this examination of high school graduation in Florida.
Organization of the Study

The dissertation contains five chapters. Chapter One acquaints the reader with the study. Included in this chapter are the background of the study, the significance of the research, a statement of purpose, an overview of the research design, delimitations and limitations, and the organization of the study.

Chapter Two contains a review of relevant theoretical and empirical literature that highlights and synthesizes knowledge about relevant facets of the paradigm of high stakes testing policies and graduation success. A conceptual framework for the study was synthesized and is graphically displayed.

Chapter Three contains a description of the research design, methods, and procedures of the study. Research question, hypotheses, and data analysis procedures are found in this chapter.

Chapter Four presents research data and analysis. A summary of the research questions related to the data analysis completes the chapter.

Chapter Five elaborates, discusses, and draws conclusions from a synthesis of research data, findings, and the conceptual framework discussed in the review of literature. Recommendations for future practice and research in educational leadership conclude the study.
CHAPTER 2
REVIEW OF THE LITERATURE

The disciplined pursuit of truth supported by evidence is the basis of theory that in turn guides practice. Building on the scholarly work of other researchers and conducting empirical research provides stimulation and information for students, researchers, and citizens and helps them in the pursuit of truth and to focus future studies. This review of literature is designed to further these goals by presenting, analyzing, and synthesizing relevant literature.

The first section in the review of literature presents the history of Florida high stakes testing with the purpose of building understanding for the context of this study. The second section on theoretical literature offers a review of expert analysis of high stakes testing rationales, related concepts, and different perspectives of prominent scholars in educational research, educational leadership, education theory, and related disciplines. A conceptual framework for this study is developed. The empirical research section presents current relevant research. The knowledge base of high stakes testing policies related to graduation from high school is advanced through the analysis in the review of literature.

The History of Florida High Stakes Testing

Florida began assessment of basic skills testing in 1972, although the tests were mainly used for research purposes (FL DOE, 2003a). In 1977, Florida tested basic skills...
of all students in public schools in grades 3, 5, 8, and 11 with the SSAT-1. Graduation testing in the form of a functional literacy test began in 1979, when Florida became the first state in the nation to require 11th grade students to pass both the SSAT-1 and a functional literacy graduation test (SSAT-11) in order to graduate.

Problems with disparate graduation rates surfaced immediately because the high school graduation test resulted in a racial disparity with 20% of black students failing compared to 2% of white students (Abrams, 2004). Recognition of this racial disparity resulted in an unsuccessful lawsuit over the claim that students were being denied their rights to graduate on the basis of graduation tests (Debra P. v. Turlington, 1979). For the period from 1980 through much of 1982, students were not refused graduation on the basis of the SSAT-11.

The courts consistently held in this case and similar cases that while high stakes tests have negative effects on minority children of failure and increased retentions or loss of graduation, these effects are unintentional (Heubert & Hauser, 1999; Mulvenon, Murry, Jr., & Ritter, 2001). The states were deemed to have a legitimate interest in education, and the tests were a way to hold the students accountable for their academic performance.

Encouraged by the 1983 National Commission of Excellence report, A Nation at Risk, the Florida testing trend changed in the 1980s to support a more rigorous and demanding curriculum with strict retention policies. The business model for education became a leading force in education based on claims of inadequate academic achievement of students, widespread assumptions about the wisdom of business methods to improve schools, and the belief in economic benefits.
Starting in 1983, the revised SSAT-11 (Secondary School Admission Test) was given at grades 3rd, 5th, 8th, and 10th in Florida and required for graduation. The tests relied on retention for those failing the tests to raise standards; however, these retention policies were unsuccessful in stimulating higher achievement (Morris, 2001). Morris conducted an aggregate analysis of schools in Florida during the 1980s and found racial disparities for retention. Most improvements in academic achievement from retention occurred at the more affluent schools, with little demonstrable improvement following remediation at schools serving mostly poor children and children of color. The state began to avoid retention and promote grade passing from 1987 to 1990 and ended basic skills testing in 1989. This trend was short lived.

The increasingly difficult High School Competency Test (HSCT) was administered from 1990 to 1997. This was a required graduation test with enforced retention policies in the lower grades. Some schools began to be labeled as “low performing.” The model was reinforced in 1996 by a standards based (Sunshine State Standards) curriculum and testing that rewarded and/or punished students, teachers, and schools on the basis of their performance (Florida DOE, 2003a).

The Florida Comprehensive Assessment Test (FCAT), first administered in 1998, was designed to measure student achievement of the Sunshine State Standards, in writing, reading, and math. Florida Governor Jeb Bush signed the A+ legislation mandating a standard based accountability system that assigned schools grades of A-F. Testing in grades 3-10 was phased in, with 2001-2002 the first year all grades were tested.
There are three different types of FCAT in Florida. One of three types of FCAT is a criterion-referenced assessment, where the interpretation of student scores is by comparison with a set standard or cut score on the test (FL DOE, 2002a). Initial selection of test items is made from a pool of test items created by content area specialists following the Sunshine State Standards and final selection of test items is according to criterion referenced procedures. The second is a norm-referenced FCAT that tests students in grades 3-10 and ranks the students according to a national sample of students. The third FCAT test is a rubric scored writing test given in grades 4, 8, and 10. Students receive their FCAT scores according to five achievement levels and a corresponding score ranging from 100 to 500. Achievement at level 2 or above is required for promotion in the third grade, while graduation from high school in 2004/2005 requires scores at or above the cut scores of 300 in reading and mathematics (Florida DOE, 2005a).

The state monitors and enforces compliance on FCAT testing through the Florida A+ Accountability plan that primarily uses student performance on the FCAT to assign grades for schools (A-F). However, other factors are considered, such as performance by similar schools, poverty, mobility, out-of-school suspension, absenteeism, retention, dropout, and graduation rates. Low school grades may prompt social pressure, the removal of funds from schools in the form of vouchers to students for attendance at private schools, and reward or punitive measures for teachers and administrators. Monetary rewards are given to high performing or improving schools, bonuses are given to teachers, and the state has the authority to replace principals, teachers, or to close schools because of low test scores.
Florida high stakes testing is supported by federal law, the No Child Left Behind (NCLB) Act of 2001, and adherence to NCLB is required for states to receive federal funds. NCLB is an extension of the Elementary and Secondary Education Act (ESEA), which was first enacted in 1965. Title I of ESEA directs most of the funds towards provisions to meet the educational needs of poor children and continues in this mode with NCLB, albeit by high stakes testing requirements.

One purpose of the NCLB is to raise the achievement level of all students and to reduce the achievement gap or differences between the average achievement levels between different groups of students. Assessment and accountability are required for all states, with states retaking the authority to determine how students will be held responsible for test performance. All states must administer annual reading and math tests to all students in grades 3-8 and in one year between grades 9-12, starting in 2005-2006. Science testing will begin in 2005.

In order to comply with NCLB, state plans must also identify concrete annual performance targets, known as Adequate Yearly Progress (AYP). Florida began counting AYP in assigning school grades in 2001-2002. Included in AYP must be content standards, standards for student and school academic achievement, and graduation rates. Graduation rate for the NCLB is defined as

The percentage of students, measured from the beginning of high school, who graduate from high school with a regular diploma (not including an alternative degree that is not fully aligned with the States academic standards, such as a certificate or a GED) in the standard number of years. (U. S. Dept. of Education, 2001, 34 C. F. R. sec. 200)
The state of Florida DOE method used to calculate graduation rates starts with the number of graduates of a cohort, including recipients of standard diplomas, special diplomas, and GED diplomas, but excluding certificate recipients (Florida DOE, 2002b). The number of graduates is divided by the number of original cohort members plus incoming transfers on the same schedule minus students who transferred out, left for private school, home education, adult education, or deceased students. How the status of students who transferred out, or left for home or adult education, is determined or verified is not clear; neither is the difference in the formula for the GED diploma inclusion and the adult education exclusion. Alternative schools are included with regular high schools.

Although graduation rates have not been emphasized for school accountability as much as high-stakes test scores in the past, graduation rates are an important component of NCLB. Inclusion of graduation rates as an accountability measure may counter any pressure to push students out of school in order to cause high stakes testing scores to rise.

Research into the relationship between high stakes testing policies and graduation over the 28 year period may help formulate policy decisions in the future. In this study, three research questions are posited for investigation: a) To what extent are high-stakes testing policies related to graduation rates from high school in Florida?, b) Can changes in high-stakes testing policies be useful in predicting a change in the number of students graduating from high school?, and c) What relationships may be found between graduation from high school and students’ high-stakes test scores, history of retention in
grade, attendance, grades, membership in racial groups, gender, socio-economic class, limited English proficiency, and exceptional student education?

Theoretical Literature

Graduation from high school is an important personal accomplishment and an indication of the success of the educational process. Many scholars view universal high school completion as a key to a high quality life, informed voting potential in our democracy, and attainment of a solid position in society as a productive wage earner (Hauser, 2002). High school graduates earn a higher income, find it easier to find and keep jobs, are less likely to require public assistance or have increased early childbearing, poor health, incarceration, or become victims of crime than people who do not graduate from high school (Rumberger, 2001; Swanson, 2004).

The Appropriate Use of Tests

Recommendations of the National Council on Education Standards, a bipartisan group established by President George W. Bush, have supported the use of standards and testing. The belief that standards contribute to learning and provide teachers with a coherent guide for instructional practice guides supporters of standards (Sandholtz, Ogawa, & Scribner, 2004). Standards improve student learning and provide equality of educational opportunity by holding all students accountable for the same body of knowledge. Testing ensures that students receive an appropriate education and demonstrate competency in the learning necessary to do well in life (Florida DOE, 2004a).

Standards and testing have been supported by many business leaders so that graduates may be employed in the technologically advanced business environment.
Achieve (2004) is an organization created by business leaders and governors that supports raising academic standards, testing, and academic achievement in schools for the goals of postsecondary education, work, and citizenship. Achieve (2004) has as part of its mission creating additional value for the high school education with benefits to the student and society.

Standards for appropriate test use have been set by several organizations, including the congressionally mandated committee of the National Academy of Sciences (Heubert & Hauser, 1999), the American Educational Research Association (AERA, 2000), the American Psychological Association (APA, 1999), the National Association of School Psychologists (NASP, 2003), the American Evaluation Association (AEA, 2002), and the National Council on Measurement in Education (AERA, APA, & NCME, 1999). The focus in these organizations concerning high stakes test use is that all children should have an opportunity to achieve a high level of educational attainment. Both intended and unintended consequences of the high stakes testing policies are of concern. These organizations oppose the inappropriate use of tests that does not improve teaching or learning. Serious problems with the testing recommended by these professional organizations include the creation of a narrowly constrained education system that fails to provide the review, meta-evaluation, and validation processes (AEA, 2002).

A common theme is that high-stakes promotion and graduation tests should cover only the material and abilities for which students have had an opportunity to learn and adequate resources to learn (Heubert & Hauser, 1999). Exactly how children obtain an opportunity to learn material is unclear, for in many cases, socio-cultural factors, such as level of parent education and quality of the teacher, play a major role in providing
opportunities to learn. Another issue is that the tests concentrate on school knowledge valued by the dominant social classes, which provides an unfair advantage to the members of these classes.

Recommendations of professional organizations, such as the American Education Research Association (AERA), the American Psychological Association (APA), the National Council on Measurements in Education (NCME), the National Association of School Psychologists (NASP), the American Evaluation Association (AEA, 2002), and the National Academy of Sciences state that additional methods of evaluating students should be included in promotion and graduation decisions beyond high stakes test scores (AERA, APA, & NCME, 1999; NASP, 2003). Unfortunately, many decisions about students' futures are currently made, by legislative mandate, by one high-stakes test. In addition, recommendations of these professional organizations are that the tests should not be used if the usage leads to placements that are not beneficial to the students, such as commonly occurs with tracking, retention, or remedial placements (Heubert & Hauser, 1999). Examination of the relationship between graduation and high stakes testing policies is useful in providing information on whether the policies are beneficial to students.

The Value of High Stakes Testing and Accountability

High-stakes tests are used to facilitate: (a) accountability of teachers, administrators, schools, districts, and states; (b) decisions about students; (c) program evaluation; (d) tracking of long-term trends; and (e) diagnosis (Heubert & Hauser, 2000). Many people need to have information about the quality of their schools in an easy to use format or rating, which is what high-stakes testing provides. Testing is a popular
instrument of accountability and reform for several reasons (Linn, 1998). The tests are less expensive than other educational reforms, such as increasing instructional time, reduced class size, and investments in better quality teachers. Testing is quickly implemented and attracts media attention that is corporate-funded and stimulating to the economic interests of government in schooling.

Changes in the curriculum result from the testing that would be difficult to require without testing mandates to drive them. Such changes focus the curriculum on a common culture that is predominately represented by the standardized testing paradigm that has dominated testing throughout the 20th Century. Not all education researchers are satisfied with standardized constructs as the focus of education. Eisner (2001) described the testing movement as a rational endeavor with specified, quantifiable outcomes that can be used for accountability. He expressed concern that over-rationalization of the goals of education through standardization will miss many of the qualitative values of education, while not attending to relevant real life educational needs. Instead of serving to transmit cultural knowledge, increase self-actualization for citizens, and increase economic equity, education has become a commodity or standardized product designed to serve a global, capitalist economy and to maintain the power of dominant cultures.

The Controversy of High Stakes Tests

Proponents claim that the testing and associated policies are research based (Scheurich, Skria, & Johnson, 2004; Skria & Scheurich, 2004) and benefit all students, including historically disadvantaged groups. According to the Florida Department of Education, in 2004, the largest number of 3rd graders in the state's history, 66%, were reading on grade level (Florida DOE, 2004a). Of high school seniors who failed the
FCAT in 2003, 58% were enrolled in some type of post-secondary education in 2004, including compensatory programs, vocational programs, technical programs, or community college. The achievement gap had shrunk with African Americans improving by 35% and Hispanics by 25%. The testing is popular with the public who perceives improvements in the quality of education and increases in equity of resource allocation as a result of the testing policies.

Proponents believe that the high stakes character of the policies is important to obtaining good results, because otherwise students and educators may not take the tests seriously. States with strong consequences of rewards and punishment have produced higher achievement gains than states with weak rewards and punishments, as measured from data from the state tests and the National Assessment of Educational Progress (NAEP) tests (Westchester Institute for Human Services Research, 2003). Evidence was provided by Stanford University economists Carnoy and Loeb (2002), who through statistical analysis found that in states with strong accountability systems, the achievement gap had shrunk with gains on state tests for African American and Hispanic 8th graders that were 1 to 1.75 standard deviations larger than the .5 standard deviation gains made by white students.

Critics believe that the high stakes testing policies do not improve the quality of education, may worsen the achievement gap, and lack fundamental fairness to all groups of students. A Joint Organization Statement on No Child Left Behind (NCLB) Act, October 21, 2004 (Fairtest: The National Center for Fair & Open Testing, 2004) was issued by 50 prominent education organizations, including the Association for Supervision and Curriculum Development, Council for Exceptional Children, National
Alliance of Black School Educators, National Education Association, National School Boards Association, and others. These organizations endorsed a statement that criticized high stakes testing because of an over emphasis on standardized testing; teaching to the test; too much emphasis on sanctions for low performing schools; the exclusion of low-scoring children; and inadequate funding. Alternative assessments, better teacher professional development, and less stressful testing regimes were supported by these 50 organizations.

Some educational leaders have voiced concerns that the oppression from the tests goes beyond unfair and inadequate assessment. The tests are deemed as oppressive to all students because the tests undermine quality teaching and expose all students to a narrow curriculum that is taught to the test, instead of engaging students in relevant learning and critical examination of our world (Apple, 2001; Grant, 2004; Kozol, 2005; Ladson-Billings, 1994; McNeil, 2000; Orphal, 2000; Sleeter, 2003). Standardized curriculum and high stakes tests often negatively impact students through excessive exercises designed to raise test scores (Kozol, 2005; McLaren, Martin, Farahmandpur, & Jaramilo, 2004; Valenzuela, 2002). Parents, students, and educators have reported feeling oppression from the tests in the form of mental anxiety and physical threats to future life chances. Grant (2004) noted that students who do not do well enough on the tests will not graduate or be accepted into college. With high-stakes testing, those who enjoy the privileges of success on the tests, graduation, and college are usually white, middle and upper class, or of the educated social classes (Apple, 2001; Grant, 2004; Madaus & Clarke, 2001; McLaren et al., 2004; McNeil, 2000; Orfield et al.; 2004; Sheets, 2005; Sleeter, 2003; Swanson, 2004; Tellez, 2003; Valenzuela, 2002).
High stakes testing policies were designed to provide all students with quality education through teaching a standards-based curriculum with the goal that every child can achieve at a high level. However, different political groups developed compromises, which were used to help formulate high stakes testing policies. Thus, the policies often consist of contradictory messages and results. According to Apple (2001), high stakes testing policies were formulated by political liberals (such as Senator Edward Kennedy who drafted NCLB), neo-liberals (capitalists), neo-conservatists (stressing a core culture that must be defended from corrupting influences such as multiculturalism), and national professional organizations, who actually developed the standards often in constructivist ways. Despite the stated intention of the NCLB law, the result of the emphasis on standards and testing has led to many students being retained and eventually failing to graduate.

In addition, the role of government in creating, mandating, and enforcing the high stakes testing regime has been questioned by education professionals who have devoted lifetimes to the study of education, educational psychology, and educational leadership. NCLB reflects a change from the historic role of limited federal involvement in education to comprehensive federal involvement via enforced standardized tested mandates from the government. The change from the conservative position before the Reagan years, which favored limited government involvement in education, towards the current position of a dominant role for the government in education, has occurred. As Anderson (2005) has noted, NCLB has not yet been fully implemented; thus, it is too early to tell if this change will result in excesses in federal control over education. However, criticism of NCLB and other high stakes testing programs already exists concerning unintended
negative consequences including over-regimentation of classes, over-control of the
curriculum, over-control of teaching, and racial and social disparities in funding (Kozol,
2005).

**Eight Assumptions about High Stakes Testing Policies**

If the decried goals of improved educational attainments for all children are not
achieved, then insight into the causes may be gained from examining and criticizing the
accompanying assumptions. It is, therefore, valuable to examine eight of the common
assumptions supporting high stakes testing.

*Achievement at a High Level*

The first assumption is that all students can achieve at a high level. Many
scholars, both proponents and critics of testing, support the assumption that all students
can achieve at a high level; however, widespread systematic success for all students has
not been achieved as evidenced by the still wide achievement gap (Apple, 2001;

In cases where schools have gone beyond high stakes testing and practiced
culturally relevant (Ladson-Billings, 1994) teaching with knowledgeable, caring teachers
who are willing to be student advocates and sustain a vision of success for students, then
high achievement has resulted. Examples of critically engaged schools where high
academic achievement occurred with respect for the cultures, histories, experiences, and
vocational needs of their students are Central Park East School in New York and Rindge
School of Technical Arts located in the Boston area (Apple, 2001).

If success with all students is possible, the problem lies with the system rather
than individual students. Relying on high stakes testing is an attempt to treat the problem
of school failure by treating the system. However, the treatment is limited to the standards-based curriculum, the testing, and school grading without attending to broad based structural inequalities, learning not included on the tests, or specific, relevant needs of students and schools. To illustrate, the Stranahan and Borg (2005b) study used a multilevel mode of student achievement to analyze the impact of school level versus student level characteristics on achievement. Results indicated that student background variables are more important predictors than other significant school variables, such as high quality teachers, class size, and low teacher turnover. In particular, student poverty was such a powerful predictor that even if poor students attended high quality schools, the achievement gap between economically disadvantaged groups and the economically advantaged persisted.

Students Not Doing Well in School

The second assumption is that students are not doing well in school. From the perspective of A Nation at Risk (1983), the United States needs high-stakes standardized testing because students are lacking in academic achievement and the necessary knowledge to compete in a technological global economy. However, Berliner and Biddle (1995) studied aggregated and disaggregated national scores of the Scholastic Aptitude Test, from 1971 to 1993, and the National Assessment of Educational Progress testing program, from 1973 to 1986, which showed no recent drop in standardized test scores. Minority students had made the most gains in standardized achievement and IQ scores, although the achievement gap between minority and white students remained substantial (Berliner & Biddle, 1995).
**Improved Student Learning**

A third assumption is that high stakes testing will improve student learning. Improved accountability systems, increased resources in schools where needed, and improved "scientifically based" teaching result in improved learning for students (Cizek, 2001). To advance these efforts, educators frequently experience professional development programs that are standards based, data driven, and more accommodating towards special populations with the purpose of leaving no child behind. Students receive tutoring as needed, community support, and rewards for achievement on the tests. Schools and teachers receive monetary rewards, public recognition, and career advancement for high achievement.

How well high stakes tests measure student learning is debatable. Student learning, when measured by standardized achievement tests, usually the state high stakes test in question, creates a logically circular pattern where the definition of learning is also the test to measure learning, and teaching to the test constitutes the curriculum. Learning that takes place outside of this circular pattern may not be adequately assessed and knowledge outside the pattern may not be taught. Some scholars (Grant, 2004; McNeil, 2000; McNeil & Valenzuela, 2001) have voiced criticism about undesirable consequences of student learning, such as reduced conceptual learning as a result of narrowing of the curriculum.

In addition, interpretation of raises in test scores should be made with caution because many factors can affect test scores besides academic achievement. Research on test scores, conducted by Linn, Graue, and Sanders (1991), has shown that increases in
scores are typically realized after introducing a new test after an initial drop. The inference is that test score gains probably are not the result of improved academic achievement, but are the result of proficiency in specific test taking skills.

Assuming that standardized tests accurately measure academic achievement, learning in the era of high stakes testing has improved. For example, there are trends in achievement on the National Assessment of Educational Progress (NAEP) test that show student reading and mathematics achievement have moderately improved from the early 1970s to 1999 (Chatterji, 2004). Carnoy, Loeb, and Smith (2000) found statistical evidence that suggested that Texas had largely achieved its goals of improved academic achievement during the 1990s period with the TAAS high stakes test. Carnoy and Loeb (2002) found strong evidence that academic achievement in high stakes testing states, as measured by the grade 4 and 8 NAEP mathematics exam between 1996 and 2000, increased by more than one half of a standard deviation for white students, black students, and Hispanic students.

Skrla, Scheurich, Johnson, and Koschoreck (2004) documented the rise in student achievement in Texas from 1994 to 2000. As measured by the high stakes test, TAAS, overall student achievement rose 23.4%, for African American students, 34.7%, and for Hispanic students, 30.7%. However, criticism of the Skrla et al. study included that Texas had changed the pass cut off level from 70% in 1991 to 50% in 2000, which made claims of gains in students' scores invalid (Haney, 2004).

Other research, however, has indicated less than exceptional gains overall in Texas during the 1990s TAAS period. Linton and Kester (2003) found that using NAEP data, there was no significant change in student achievement in Texas for black or white
students, although there was a significant rise in student achievement for Hispanic students. Smaller differences in reported gains in student achievement from 1994 to 1998 were found using NAEP data compared to the more impressive gains using TAAS data (Camilli, 2000; Klein, Hamilton, McCaffrey, & Stecher, 2000).

Furthermore, other research studies (Amrein & Berliner, 2002b, 2003; Kornhaber & Orfield, 2001; Madeus & Clarke, 2001) do not support the assumption that high-stakes testing improves student learning. Amrein and Berliner (2002) conducted archival time-series analysis for the SAT, ACT, AP, and NAEP tests that compared high stakes testing states to national averages for 1999-2000. Results indicated that high stakes testing had the following effects: (a) no effect on SAT scores, (b) evidence of decreased ACT scores in 67% of states with high stakes testing, and (c) no evidence of increased learning in AP tests. In addition, NAEP data showed high stakes testing states increased only in reading, and changes in test scores fluctuated with the exclusion rates for eligible test takers.

Some have criticized Amrein and Berliner's (2002) study for comparing states with high stakes tests to the national average, rather than using a comparison group (Rosenshine, 2003). Rosenshine used the same data as Amrein and Berliner in the 2002 study, but he used a control group and cohort analysis to find that students in high stakes testing states had comparatively higher achievement in 4th and 8th grade math NAEP scores than low stakes states. Amrein-Beardsley and Berliner (2003) then re-analyzed the data by comparing high stakes states to low stakes states with essentially the same results as in their original study, except that 4th grade mathematics showed significant gains for high stakes testing states.
In response to Amrein and Berliner (2002a), Braun (2004) conducted a cross-sectional analysis of the same NAEP data. The results of Braun's re-analysis of 1999-2000 data strongly favored the high-stakes testing states over states without high stakes testing for having increased NAEP standardized test scores in grades 4 and 8. However, when Braun's analysis (2004) was conducted with a pseudo-longitudinal design, involving following two different cohorts from grade 4 to grade 8, the comparisons slightly favored the low-stakes testing states for student achievement, noting it was impossible to conclude that high stakes testing does not improve achievement. Because of the large number of other educational initiatives and data collection improvements introduced at the same time, Braun found it difficult to attribute changes in achievement exclusively to testing.

In Florida, student academic achievement as measured by the Florida Comprehensive Assessment Test (FCAT) from 1999 to 2003 showed differences in scores between elementary and secondary students, with elementary students achieving at a higher level on the FCAT than secondary students (Chatterji, 2004). In the criterion referenced FCAT reading, 60-70% of 10th grade students scored below the passing level in 2003, which was a reduction of scores from previous years. In the norm referenced FCAT reading, 10th grade students have scored on average below the national median in the last few years.

Thus, conclusions about whether high stakes testing increases student learning are difficult to make. Different studies have reached conclusions on this matter. Design of the studies and circumstances surrounding the environmental data collected can result in different conclusions.
Increased Motivation

The fourth assumption is that the tests will increase motivation. Related to this assumption is the conclusion that students believe they do not need to be high achievers in high school in systems without high stakes testing, but the tests help correct that perception and help students to achieve academically in school and after graduation (Bishop, 2004). The sanctions and rewards associated with high-stakes tests provide motivation for increased academic achievement of students. Gaylor, Chudowsky, Hamilton, Kober, & Yeager (2004) concluded from a review of research that studies, overall, show that students are taking the tests more seriously than ever, with students who believed that the tests counted for graduation reporting greater efforts on the tests. Research by Kornhaber and Orfield (2001) and Madaus and Clarke (2001) found weak statistical evidence that high-stakes testing sanctions and rewards provided strong motivation for students.

High stakes testing policies are complex because more than just the test is involved, such as retention in grade. A study conducted by Roderick and Engel (2001) on whether retention provided motivation for retained students to work harder and achieve more than non-retained students indicated that students were motivated when supported by teachers in an efficacious environment, but not in unsupported environments. Numerous other studies have found that retention policies do not have a widespread history of success in motivating students (Grissom & Shepard, 1989; Hauser, 2002; Holmes, 1989; Rumsberger, 1995; Smith & Shepard, 1989; Valenzuela, 2002).

Some research attests to the negative motivational effect of the high stakes testing policies. Wheelock, Bellel, and Haney (2000) examined the self portraits of tested
students, where students depicted themselves as angry, anxious, and bored by the tests, finding that older students were more disillusioned than younger students. Other questions of the effectiveness of the motivation from standardized testing arise because students of different historically subjugated minorities are on average one standard deviation below their counterparts on standardized tests in some countries, such as Japan, India, and United States.

According to John Ogbu (1978), the group differences shown on standardized tests are not attributable to racial or ethnic differences; instead, the differences are attributable to sociological differences or differences in cultural group motivations. Histories of subordinate, oppressed groups, such as African Americans and Mexican Americans, have been found to support an identity system that may be opposed to the dominant white class and associated standardized testing (McLaren, 1998; Valenzuela, 2000). Through resistance to schooling, subordinate students retain at least some control over their production of knowledge.

_Fair and Neutral Testing_

A fifth assumption is that tests are a fair and neutral way to evaluate student academic achievement and decide whether students deserve promotion or graduation. The value of high stakes testing resonates with the principle that the tests justly award students who work hard in a fair and impartial way and punish students who do not work hard (Natriello & Pallas, 1998). The test and education accountability system employ logic based on the premise of meritocracy, that all children have a fair chance to succeed in school rather than recognize structural and symbolic inequalities (Ladson-Billings, 1998).
However, testing all children with the same test does not recognize socio-economic inequalities and the necessity to adjust education to meet the needs of children, including the poor, thereby serving to reinforce inequalities in public education (Kozol, 1991). From this perspective, the tests measure unequal backgrounds and opportunities in a mode favoring the dominant, white, middle class culture (Madaus & Clarke, 2001).

In addition, Kellow and Willson (2001) reported that potential measurement error rates in scoring the tests are about 2%. These errors in test scoring may involve approximately 1.1 million students misclassified as failing the test each year when these students should have passed. For these students, chances for a prosperous life may be unjustly destroyed.

Reliable, Research based, and Valid

The sixth assumption is the tests are reliable, research based, and valid. Reliability refers to the consistency of results of the tests and is not a characteristic of the test itself. Statistical analysis by Greene, Winters, and Forster (2004) found that the Florida A+ plan has produced FCAT scores that correlate at 0.96 with the Stanford-9 standardized test across all grade levels and subjects. However, reliability does not attest to the validity of the tests to measure educational achievement.

High stakes testing concentrates on achievement according to the “common culture” of the United States, as a fair and desirable mode of evaluating all students. Test creators claim that the tests enable valid comparisons among social groups because the test items are written in a culturally unbiased way that uses different cultural contexts within the test items. In order for a test to yield valid results, however, the assumption
that the tests are not culturally biased must be credible to those taking the test (Kane, 2002; Tellez, 2003).

How credible the tests are to diverse cultural groups may be questioned. For example, the tests are given in English and may not be credible, reliable, or valid for students with limited English skills (Abedi, 2004; Venenzuela, 2000). While test developers utilize concepts of racial/ethnic identity, ethnicity, and socioeconomic status when developing assessment instruments, these concepts are used as variables that are supposedly interchangeable with culture (Gopaul-McNicol & Armour-Thomas, 2002). Culture tends to be inappropriate as an independent variable because of the meaningful differences within and among cultural groups. The dimensions of culture that have psychological implications fail to be operationally defined sufficiently to permit measurement.

The research base utilizes criterion referenced test item construction theory and classic test construction theory that emphasize maximizing differences among students through norm referencing procedures (Florida DOE, 2002a). The norm referenced test item selection policy guarantees that the lower portion of the student body will fail and suffer the punishing consequences of retention and failure to graduate (Haney, 2002). The norming group is limited and unlikely to afford equal representation to all groups of students who take the tests, thus making the test inappropriate for groups not adequately included in the norming procedure (Fenner, 1999). Students of diverse cultures, especially if they are of lower socio-economic status, often are underrepresented in the norming process. They also achieve at a far lower rate on the tests and unfairly suffer the negative educational, social, and economic consequences.
Psychometrically valid methods of setting the cutoff scores for passing and the margin of error in scoring, programming, and reporting test scores are important for quality tests. However, studies by Haney (2000, 2001) indicated that the passing scores set on TAAS, the Texas state accountability test, were arbitrary, discriminatory, and failed to take measurement error into account. The 1991 passing scores were set arbitrarily at 70% correct without evidence that such cutoff scores differentiated among students by any criterion other than performance on the TAAS. Additional arbitrary setting of cutoff scores occurred in 2000 with the passing rate lowered to 50%. The vast improvement of pass rates on the TAAS can be partially explained by this change in cutoff scores. In addition, Haney (2000) suggested the actual standard error of measurement for TAAS tests was 20 – 40% greater than that reported in the TAAS 1996-97 Technical Manual.

Increased Equity

The seventh assumption is that the tests will increase equity. High stakes testing policies pursue the strategy that the achievement gap can be overcome as a result of the scientifically based testing, strict retention policies, timely remediation, and extra resources applied in a skillful way. Critics believe that the policies aim at overcoming weaknesses, failures, and other deficits in children, teaching, or schools. Deficit thinking from a scientific based perspective historically has involved racism and claims of genetic inferiority, cultural inferiority, or lack of socialization (Menchaca, 1997). The history of deficit policies in education chronicles a dismal record of success (Valencia, 1997).

In spite of criticism, high stakes testing policies provide greater resources to poorly performing schools, which has helped schools reach students who previously had
been ignored (Florida DOE, 2003b). Different reports exist as to the success of these efforts or whether the achievement gap is growing larger or smaller. For example, Scheurich, Skria, and Johnson (2004) reported that four large Texas districts had made dramatic progress in teaching and learning, and these districts had produced equitable educational success for literally all children in the districts. Other research (Chatterji, 2004; Klein, Hamilton, McCaffrey, & Stecher, 2000; Lee, 2004), using long term NAEP data, did not support the hypothesis that the achievement gap had narrowed between blacks and whites in the 1990s.

A study of whether the Massachusetts Comprehensive Assessment Test (MCAT) mitigates or exacerbates inequalities compared teacher assigned grades and high stakes test grades (Brennan, Kim, Wenz-Gross, & Siperstein, 2001). The study used data from a sample of eighth grade students from suburban schools to statistically analyze MCAT and teacher assigned grades for differences between racial/ethnic minority students and white students and differences between girls and boys. Results indicated that the MCAT reduces the average competitive position of African American students in math and of girls in math and science. Racial/ethnic comparisons indicated that the African American-white achievement gap and the Latino/Latina-white gap on MCAT math scores was larger by about .50 standard deviations than when using teacher assigned grades.

Economic Productivity

The eighth assumption is that high stakes testing enhances economic productivity by better preparing the workforce for the demands of the modern world. Economic benefits in the form of profits for testing companies, politicians, and the privileged
classes have been substantial (Grant, 2004). The assumption that high-stakes testing supports economic gains for everyone, however, has been criticized because the ownership gap between the wealthy classes and the working classes has widened, and the middle class has grown smaller. Studies of traditional tests that show only nominal statistical links to measures of worker productivity indicate that testing may not increase working class economic well-being (Levin, 2001).

Business interests have had a major influence on creating standards and testing, and those business interests have been well served by education in ensuring a supply of able technical workers and cheap unskilled labor (Bartlett, Frederick, Gulbransen, & Murillo, 2002). Unfortunately, cheap, unskilled labor is the only work usually available for people without high school diplomas. The industrial work force once provided work to under-educated people, but these jobs are becoming scarce as evidenced by the 2.3 million jobs lost since 1991 (Orfield et al., 2004). High school dropouts earn only 70% of what graduates earn, and the economic impact is more severe for some minority groups, such as Hispanic dropouts who only earn 57% of what graduates earn. In 2003, more than half of African American males in Chicago without a high school diploma were unemployed.

Disparate Graduation Rates: Institutional Racism

Many people of color are not well served by the current education system. Kozol (2005) reported in that in some areas of New York City, only 20% of black children are expected to graduate from high school. The large numbers of ethnic minority students who are retained in grade and who fail to graduate suggest that schools and teachers do not know how to teach these populations (Delpit, 1995). Curricula and instruction
representing the knowledge of the traditionally dominant classes (white, middle or upper classes), rather than the knowledge of traditionally subordinate populations (people of color, poor people, many minorities), may not be the most effective way to teach or test children of color, poor people, and many minorities (Gay, 2000; Irvine & York, 1995; Sheets, 2005).

Divergent cultural learning styles, such as African American preferences for verbal and kinesthetic learning rather than visual learning (Gay, 2000; Irvine & York, 1995; Ladson-Billings, 1994; Shade, 1997; Sheets, 2005), are not accounted for by the current paradigm of high stakes testing. The English-only nature of high stakes tests, such as the Texas Assessment of Academic Skills (TAAS), has been found to discourage students from graduating from high school in the following ways: (a) limited-English proficient students, who are often immigrants, have difficulty passing English-only tests and (b) English-only tests and courses negate Mexican students’ culture and language (Valenzuela, 2002).

Valenzuela (2002) noted that narrowing of the curriculum caused by excessive focus on tested material may create academic vulnerability for students, especially children of color and poor children who must also contend with cultural insensitivity in schools, segregation, unqualified teachers, and parents who may not possess high levels of education. Lack of a culturally relevant education is manifested through inappropriate teaching and curriculum that can cause students to resist engagement in education with destructive consequences, including lower achievement or failure to graduate (Cammorota, 2004; Ladson-Billings, 1994).
Sometimes failure to recognize and value multicultural students' background and experiences leads to blaming failure on deficits in children rather than weaknesses in the education system, schools or teachers. Valencia (1997) noted that common explanations for children failing standardized tests are consistent with the deficit perspective, such as genetic inferiority, lack of socialization, cultural dysfunction, language impairment, or lack of effort or motivation of students. Cultural shortcomings, such as lack of family structure, lack of valuing of education, and lack of education in the family have been blamed for the achievement gap. McLaren (1998) described deficit thinking as equivalent to educators' avoidance of responsibility or any critique of themselves, the schools, or society. In deficit thinking, society blames the students, who have virtually no power, while protecting those people in positions of power from scrutiny.

Some current variations of the deficit perspective have qualified that the problematic deficits are deficits in academic skills (Torgesen, 2004), rather than cultural deficits; however, "soft" deficits in oral skills and background knowledge are often diagnosed in children of color and poor children. Results of low scores using standardized testing instruments often mean special education or remedial placements patterned after the medical model of disease treatment, designed to "cure" the problem or deficit. Unfortunately, special education and remedial track placements have a dismal record, especially for minorities, for lacking educational opportunities, obtainment of a standard high school diploma, or prospects for post graduation employment (Oswald, Coutinho, & Best, 2002).

One effect of the policies for schools that serve lower income students who are also minorities is for test scores to act as gatekeepers for many children that prevent them
from entering colleges, qualifying for scholarships, and entering careers that offer a more prosperous life (Cammarota, 2004; Delpit, 1995; Grant, 2004; Kozol, 2005). For example, the average yearly earnings of a high school dropout are $12,400 compared to $21,000 for a high school graduate and $41,000 for a college graduate (Cammarota, 2004). Socially, this disparity contributes to racial prejudices, stereotypes, crime, and poor health of members of non-graduating classes, with hegemony for the wealthy and dominance for elite classes.

School policies, funding, curriculum, pedagogy, and student composition occur within a context of power. High stakes testing policies assist in reproducing social class consistent with cultural reproduction theory (Apple, 2001; Kozol, 2005; Tanaka, 2002). Schools and teachers, acting in a context of power, actively push students into tracks according to their socio-economic status, thus preserving positions of economic and social privilege for the upper, usually white, classes. Lower class students, more often non-white, are denied a chance for advancement in their position through the avenue of education by the active reproduction on social class (Bourdieu & Passeron, 1977). Efforts to change the context of power are refuted, such as when efforts to broaden the knowledge base of teaching by recruiting more racially and ethnically diverse teachers are thwarted by institutionalizing standardized testing for teacher competency, resulting in fewer teachers of color, a narrower curriculum, and increased ethnocentric teaching (Smith, 2000).

Calculating Graduate Rates

The NCLB law requires that states must include graduation rates as an academic accountability indicator at the high school level. Many states have not reported
graduation rates in the past or instead have reported a dropout rate (Swanson, 2003). There is no reliable longitudinal system in the United States that tracks individual students through their schooling to see if they graduate (Miao & Haney, 2004). Other questions that are inconsistently answered in most state graduate rates include: who should count as a graduate, what is a regular number of years to graduation, and how should the population base be adjusted for transfers between schools, grade retention, and migration into or out of the state. In the 14 states of the southeastern United States, nine different methods of calculating graduation rates make it impossible to make comparisons regarding how many students are graduating using the “official” rates. The difficulty in calculating accurate historic graduation rates lies in the partial lack of complete and consistent data available.

Problems have occurred because of the use of GED certificates to count as regular diplomas, under-coverage of certain population segments, such as incarcerated students, and reporting bias that schools may use to improve their graduation rates (Greene & Winters, 2002; Swanson, 2003, 2004). For example, although students who leave school to pursue adult education (GED) are removed from the graduation formula in Florida, students who pursue and obtain a GED while in school are considered as graduating. This is possible because Florida has a “GED Exit Option” where students who can read at the 7th grade level and are in jeopardy of not graduating can pursue a GED in alternative classes while staying in school (Florida DOE, 2004b). Students who pass the GED but not the FCAT are awarded a State of Florida High School Diploma; whereas students who pass the FCAT are awarded a Standard Diploma. For example, South Plantation High School in Broward County, Florida, was documented as telling low performing
students to enroll in GED classes one month before the FCAT exam was given (Gaylor et al., 2004).

Misclassification of students who leave school, sometimes as transfers, GED, or home schooling students without any proof of such intentions, has been found to inflate graduation rates and inaccurately raise test scores in other states. In one school in Texas, the reported dropout rate was supposedly zero, despite over half of the student body leaving school between the 9th and 12th grade (Gaylor et al., 2004). The school had increased the promotion requirement to pass the 9th grade, which caused increased 9th grade retention, while increasing scores on the 10th grade TAAS. To illustrate, in 1998, the 9th grade enrollment was approximately 900, the 10th grade enrollment was 580 and the 10th grade TAAS pass rate was 68%. In contrast, in 1999, the 9th grade enrollment was 1,200, the 10th grade enrollment was 235, and the 10th grade pass rate was 86%.

Challenges to the accuracy and reliability of some graduation rate calculations have been reported. Haney (2000, 2001) argued that the Texas Education Agency (TEA) dropout rate was invalid because of the inconsistent system of classifying school leavers in dozens of different ways that are not counted as dropouts. For example, if a student met all school requirements but failed the TAAS, then the student was classified as a non-dropout leaver. Also, students who left school to pursue a GED were not classified as dropouts.

Different ways of calculating graduation rates have been developed by educational researchers. A simple way of calculating graduation rates, using reliable public enrollment counts, is the Basic Completion Ratio (BCR) (Swanson, 2003) (see Appendix 1). The calculation involves dividing the number of graduates by the number
of ninth graders three school years before. The measure is not stable on a local level, however, because of changes in student population due to migration and population growth.

The Greene Method (Miao & Haney, 2004) uses the average enrollment of graduates in grades 8, 9, and 10 to get a smoothed estimate of the cohort’s first time enrollment. Greene averages this for three different years to get a more refined estimate of the 9th grade enrollment. In the Greene formula, the numerator is the number of regular diplomas in the senior year (i+4) and the denominator is the estimate of the 9th grade (i) cohort. The Green method has been adopted by several organizations, such as the Bill and Melinda Gates Foundation and the Buckeye Institute; however, this method has been criticized for ignoring student migration, inconsistencies in student enrollment, and inconsistencies in graduation counts (Phelps, 2005).

The Cumulative Promotion Index (CPI) approximates the probability that a student entering the 9th grade will graduate with a regular diploma four years later (Swanson, 2003). The graduation rate represented by the CPI is acceptable for NCLB accountability and has the advantage of a short period of time for data gathering, using data from the start of one school year and the start of the next school year. The shorter time period minimizes the effect of migration. The CPI uses enrollment and graduation counts for calculations that make comparisons among states and districts possible. High school graduation is represented as a stepwise process composed of three grade-to-grade promotion transitions (9-10, 10-11, 11-12) and the graduation event (12-graduation). The forward-looking CPI is useful only for establishing graduation rates given the prevailing conditions at the time of data collection. The statistics are analyzed for estimated cohorts
rather than actual longitudinal cohorts based on tracking of individual students over the years. For a mathematical description of CPI calculations, see Appendix 2.

The National Center for Education Statistics calculates a graduation rate using the percent of students who graduate versus the percent who drop out (Swanson, 2003). Most school districts use this model for NCLB. Problems with the model include lack of adequate records; indeed, currently only 54% of districts nationwide have adequate data available for the calculations.

Several other methods of calculating graduation rates are currently used by educational researchers and state departments of education. The most complicated is the Warren Method (Miao & Haney, 2004) that uses both the United States Department of Education Common Core of Data (CCD) and U. S. Census data to calculate a rate that more accurately accounts for migration than most other methods. Miao and Haney (2004) found in a comparison study of methods, including the 9th grade basic completion rate, the Greene Method, the CPI, and the more complex Warren method, that the Warren method yielded no evidence of better results than other methods.

In a study conducted by Swanson (2003), differences in graduation rates from using the NCES, BCR, and CPI methods were obtained. The national average district graduation rate for the year 2000 illustrates these differences: NCES (85%), BCR (78%) and the CPI (73%) (Swanson & Chaplin, 2003). The overall national graduation rates (calculated only for BCR and CPI) were 69% and 66%, respectively. The NCES method is the only one to use state dropout rate data in the calculations; the other methods use enrollment data from the United States Department of Education Common Core of Data. The percent of students included in the analysis was lower for minority groups than white
groups; for example, it was 60% for Native Americans and 90% for whites (Swanson, 2003). This indicates that less accurate records are kept for schools with high minority populations.

**Conceptual Framework**

The conceptual framework for this study was synthesized from a critical perspective using a background of the theoretical and empirical literature presented in this chapter (see Figure 1). The time frame for the conceptual framework is the first decade of the new millennium in the United States, a period characterized by conservative political dominance, incomplete civil rights changes of the 20th century, pressing economic and social needs of many people, globalization, media penetration, and multicultural perspectives of local, national, and international individuals and groups. In this time frame, the educational reform movement supported by politicians and business has emerged with a set of specific, highly controlled, and closely monitored high-stakes testing policies.

The set of high stakes testing policies constitute the dominant educational paradigm of the decade with profound implications for students. Legislative mandates for annual testing of all students, retention in grade, required adherence to a standardized curriculum, pressures placed on students, teachers, and schools because of sanctions, and graduation exams have intended and unintended implications for students that affect the prospects for their high school graduation for students.

Intended consequences include a focus on academic achievement, attainment of high levels of achievement, a reduction of the achievement gap, and increased economic prosperity. Not all students reach these goals, and unintended consequences, such as
narrowed educational experiences, failure to graduate, reduced life chances, and weakened democratic participation, may be related to high stakes tests. Different groups are affected differently by the testing in the following areas: (a) achievement within test parameters may or may not improve, (b) there may be high rates of retention and special education referrals, (c) there may be many students who do not graduate from high school, and (d) there may be differential economic, social, and political consequences, including increased social class stratification (see Figure 1).
Figure 1. Conceptual Framework

Environment of Conservative Politics, Economic and Social Needs, Globalization, Media Penetration, and Multicultural Society

Educational Reform Movement Supported by Politicians And Business

Legislative Mandates for State and National Standardization and High Stakes Testing

Impact and Unintended Consequences of High Stakes Testing Policies:
Differential Affects on Groups
1. Achievement with test parameters may or may not improve.
2. High retention rates and special education referrals
3. Too many students do not graduate from high school.
4. Differential economic, social, and political consequences for society.
Empirical Research

Different studies have resulted in conflicting reports over whether graduation rates are stable, rising, or falling. According to Swanson (2004), national rates have been stable or slowly rising since the mid 1990's, between 65% and 66%. Haney, Madaus, Abrams, Wheelock, Miao, and Gruia (2004) found that national rates had improved slightly in the early 1980s, dipped slightly in the late 1980s, and then fell steadily beginning in the early 1990s, from 78% in the early 1990s, to 75% in the late 1990s.

According to Chatterji (2004), graduation rates for the nation and Florida are declining, with Florida 10-12% less than the national average in 2000. Using data from the National Center Education Statistics (NCES) Common Core of Data surveys, Chatterji (2004) reported that graduation rates in Florida went from 65.0% in 1992, to 57.6% in 1997, to 55.2% in 2000. Florida's graduation rate declined from 73% in 1992 to 63% in 2000 (Haney et al., 2004). According to the Florida DOE (2004), the state graduation rate for 2002-2003 was 69%, following a steady pattern of increases in graduation rates since 1998. In 2003-2004, the rate had increased to 71.6%.

Current national high school graduation rates show there is racial disparity in who graduates from high school. High school graduation rate calculations using the Cumulative Promotion Index (CPI) resulted in an average national graduation rate for 2002-2003 of 68% (Orfield et al., 2004; Swanson, 2004). Graduation rates for subgroups follow patterns, such as 76% for whites and Asians in 2002-2003, while rates for traditionally disenfranchised racial groups were far less, with 50% for black students, 51% for Native Americans, and 53% for Hispanics graduating from high school.
The rates vary by states, with the four states with the lowest graduation rates for all students in 2002-2003 being South Carolina (50.7%), Florida (53.0%), Nevada (54.7%), and Georgia (55.5%) (Orfield et al., 2004). The four states with the lowest graduation rates for blacks were New York (35.1%), Ohio (39.6%), Nevada (40.5%), and Florida (42.0%). The four states with the lowest graduation rates for Hispanics were New York (31.9%), Massachusetts (36.1%), Michigan (36.3%), and Iowa (40.5%). The four states with the lowest graduation rates for white students were Florida (57.9%), Nevada (62.0%), Georgia (62.4%), and Mississippi (63.3%). These rates were calculated using the Cumulative Promotion Index (CPI).

Graduation rates for students have been calculated for subgroups by gender, disabilities, limited English proficiency, and school demographic make-up, including region, urban/suburban, segregation, and percentage of poor students. Nationally, boys had an 8% less chance of graduating than girls (Orfield et al., 2004). Graduation rates for students with disabilities were far lower than students in other groups, only 32% in 2002-2003 (Orfield et al., 2004). In Florida, less than 25% of students with disabilities graduated in 2002-2003. Schools with a higher degree of racial segregation or where there was a majority of minority students had a rate of 56.4%, compared to the majority white rate of 74.1%. Schools with a higher degree (>38%) of poor students had a rate of 57.6%, compared to the other schools (<38%) rate of 76.0%. Central city schools had a rate of 57.5%, compared to suburban schools rate of 72.7%.

Orfield et al. (2004) found that for 2002-2003, overall graduation rates and the gaps between racial groups varied according to region of the United States. The highest overall graduation rates were found in the Midwest (75%), followed by the Northeast
(71%), the West (68%), and the South (62%). The biggest racial gaps were in the Northeast, where white students had a graduation rate of 79% compared to black student rate of 35%. Florida had a narrower racial gap, due to the low white graduation rate of 55.2%; however, the rate for blacks, 41%, was low. In cities such as Jacksonville (Duval County, Florida), the overall 2002-2003 rate was 46.3%, with subgroups of Native Americans (29.9%), Asians (76.3%), Hispanics (64.7%), blacks (35.7%), and whites (53.5%) varying in their graduation rates.

A national growth of attrition in the ninth grade coincides with standards based reform and high-stakes testing and may affect graduation rates (Haney et al., 2004). Students held back or asked to leave school altogether in the ninth grade never take the high-stakes test in the 10th grade, thus possibly improving the school grade. In 2000, Florida had 32% more students in the ninth grade than in the eighth grade, which was a higher percentage than any other state. In addition, Florida had the highest attrition rate from the ninth grade in the nation with 23.8% of 9th graders leaving school (Haney et al., 2004).

*Studies of High Stakes Testing and Graduation Rates*

Controversy exists about whether high stakes testing policies have a negative effect on high school graduation rates or the prospect of a student graduating from high school. One area of uncertainty is loss of students before high school. In Florida, there are approximately 30,000-40,000 fewer students enrolled in high schools than in elementary schools (Florida DOE, 2002b). As Florida is increasing in population, this probably reflects students leaving school prior to high school due to expulsion or dropping out.
A body of research has indicated high stakes testing policies, while positively impacting test scores, have had little effect on students dropping out of high school. An analysis of the effects of Texas Assessment of Academic Skills (TAAS) on educational outcomes, including achievement, retention, and graduation rates, concluded that the test had a large effect on raising average TAAS scores, but at best a small positive impact on retention rates and graduation rates (Carnoy, Loeb, & Smith, 2000). The rise in average TAAS scores did not coincide with a rise in SAT scores, which raises questions about transferability of learning to other contexts besides the TAAS or the lack of sufficient rigor on the TAAS needed to measure the same level of learning as the SAT.

Carnoy, Loeb, and Smith (2000) acknowledged that high retention and attrition rates existed in Texas, especially for minority groups, and the test has not proved to be an effective way to improve these key indicators of educational quality. Many students of color dropped out in the 9th grade; however for those students who made it through the 10th grade, the dropout rate was reduced in the years since TAAS was implemented. In another study, Carnoy and Loeb (2002) used an index to rate the strength of the state’s accountability systems for the purpose of studying the effect of state high stakes accountability on graduation rates and other measures. They found no relationship between the strength of a state accountability system and graduation rates. However, Carnoy and Loeb (2002) did not directly measure the effect of high stakes tests on graduation rates; instead they examined retention rates and the strength of the accountability system with implications for graduation rates.

Jacob (2001) studied the relationship between high stakes tests and student achievement including graduation using National Educational Longitudinal Study
(NELS) data. He found no differences in math and reading achievement to be related to the test, but he found dropout rates were about 6.5% higher for students in the bottom quintile of achievement tests for students in states with high stakes tests compared to students without the tests.

Reardon and Galindo (2002) also used 1988 and 1990 NELS eighth grade data to study the relationship between high stakes testing in the eighth grade and eventual failure to graduate. They found that the presence of the eighth grade promotion tests was strongly associated with an increased probability of dropping out before the tenth grade. The study controlled for socio-economic class, test scores, GPA, minority percentage, total enrollment, percentage of English as a Second Language and bilingual education students, student teacher ratio, percentage of teachers with advanced degrees, community type, and region.

The relationship between demographic variables of racial/ethnic identity and socio-economic class and success on high stakes tests leading to graduation is not random, according to a case study of 5,205 10th grade students in Duval County, Florida (Stranahan & Borg, 2005a). Using a probit regression model, the study showed that a key indicator for success on the FCAT and graduation success was the socioeconomic status of the students' households, with the positive coefficient for income over $70,000 having the largest magnitude of any variable in the model. Other significant indicators were racial/ethnic identity for African Americans and Hispanics compared to whites, withdrawals from school (mobility), attendance at schools with higher percentages of teachers who were newly hired, and attendance at a magnet program schools. In addition, the study analyzed the 9.5 percent of students who would have been affected by the
planned, higher cut scores for the FCAT due to begin the next year. Students negatively impacted by these higher cut scores were disproportionately African American and those with high mobility.

Clarke, Haney, and Madaus (2000) examined evidence of high stakes testing policies and high school completion and derived five suggestive lines of evidence (with the evidence mainly correlational), inferring that the high stakes testing policies are linked to decreased rates of high school completion.

- The first line of evidence concerns the era of minimum competency testing, such as took place in Florida in 1977. In states with the lowest dropout rates, no or weak minimum competency testing was in place. In states with the highest dropout rates, 90% had strong testing policies with mandatory promotion and graduation decisions.

- The second line of evidence concerned low socio-economic status schools that used the minimum competency test, with these schools having dropout rates from 4 to 6 percentage points higher than similar schools with no test requirement.

- A third line of evidence indicated a complex relationship between test scores and dropping out. To support this was a study by Griffin and Heidorn (1996), who found that in Florida, only students with moderately good grades and failure on the high stakes tests were associated with a significant increase in dropping out. No difference between white students and minority students in dropping out was associated with the test.
• A fourth line of evidence suggested that because of the high stakes test requirement in Texas, some 40,000 sophomores in 1993 dropped out of school. The dropout rates for black, Hispanic, and white students were about 25%, 23%, and 13% respectively. This is part of a 20 year pattern that shows increases in ninth grade dropouts for all three groups, but with rates about 50% higher for minorities than white students.

• A fifth line of evidence comes from research on grade retention and dropout rates. Being older than other students in grade because of retention predicts dropout better than low test scores.

The ninth grade year involves challenges for students, such as a change to a less nurturing environment and increased academic demands. Retention in the ninth grade is common in order to give students additional time to learn the required state standards and to prepare for the 10th grade high stakes tests. Haney (2001) conducted enrollment progression analysis to examine patterns of high school completion in Texas and possible effects of Texas Academic Assessment System (TAAS). Since TAAS was implemented, the racial gap between many students of color and white students as measured by progression from 9th grade to graduation has increased by 50%. Analysis also indicated that for black and Hispanic students, their grade progression was halted in the ninth grade rather than at graduation time. Ninth grade attrition rates for minorities in Texas first began to rise in the 1980s during the first wave of high stakes tests. For all students that made it to the 12th grade, larger proportions are graduating; however, for minority students, smaller proportions are progressing to the 12th grade due to increased retention and dropping out before grade 12. In the same study, Haney (2001) also studied 9th grade
retention and high school completion rate nationally and found that for every 10 students retained to repeat the 9th grade, 7 did not complete high school.

For many students, the prospect of graduating from high school has been diminished while high stakes testing has flourished. The relationship between raised graduation and promotion standards and graduation was studied by longitudinally tracking a cohort of students' progress in New York City (New York City Board of Education, 2001). The dropout rate increased a full 2 percentage points from 17.5 in 1999 to 19.5 in 2000. More than two-thirds of dropouts were older than their classmates due to being retained when entering the ninth grade. In addition, over 15,500 students in the cohort were projected to fail to graduate because of not passing the Mathematics Regents exam. Approximately 44% of the Class of 2001 was projected to graduate on time compared to 50% of the Class of 2000.

Unfortunately, many students never make it out of the ninth grade, and the problem seems to have worsened since high stakes testing was implemented. In 2000, Florida had the worst ninth grade attrition rate in the country, with 23.8% of ninth graders leaving school (Haney et al., 2004). A historical analysis of Florida (Haney et al., 2004) showed that the growth of attrition in the ninth grade during the 1970s coincided with the advent of minimum competency testing of the 1970s. The growth of attrition in the 1980s coincided with the academic standards movement (Haney et al., 2004). In addition, while the influence on graduation of being African American had disappeared by 1970s, racially correlated effects on graduation reemerged in the 1980s and 1990s (Dorn, 2003).
A broader insight into the graduation success phenomena can be gained from enrollment progression through the grades. Researchers calculated the percent increase or decrease in enrollment for each grade starting with kindergarten, for all 50 states, in cohort progression analysis (Haney et al., 2004). There were four major findings: (a) kindergarten attendance is becoming more universal, (b) there is an increasing attrition of ninth grade students, (c) there is an increasing bulge of ninth grade students, and (d) there are declines in high school graduation rates compared to rates in the 1970s. The rate at which students leave school before graduation has more than tripled over 30 years, from 4% in the early 1970s to nearly 12% in 2000.

Other researchers, such as Amrein and Berliner (2003), blame high stakes testing for lowered graduation rates, claiming that higher dropout rates are found in 88% of the states with high school graduation tests compared to states without graduation tests. Students increasingly are taking alternative routes to high school graduation, such as the special diploma for students with disabilities or the General Educational Development (GED) test (Amrein & Berliner, 2003). Special diplomas and the GED are inferior to a regular high school diploma in terms of social consequences or future earnings (Vernon, Baytops, McMahon, Padden, & Walter-Thomas, 2003).

Other anecdotal evidence coincides with the current high stakes testing period. In 2001, New York City schools asked 55,000 students to leave the school system or "pushed them out" in grades 9-12 (Gaylro et al., 2004). These were students who were struggling academically before being counseled by the school to leave school, transfer to private schools, or pursue a GED. Dropouts were disproportionately black and Hispanic, and although some went to private schools, many were never counted as dropouts. An
advocacy group, Advocates for Children, sued the New York City school system and
won, allowing students who had been pushed out to return to school. The group claimed
the “pushout” phenomenon was due to pressure to improve scores on the high stakes test
– the Regents exam.

Exit Exams and Graduation

Some researchers have focused specifically on high school exit exams to
determine effects on the graduation rate of this particular type of high stakes test.
Difficulties exist with isolating the effects of exit exams that are typically embedded in a
comprehensive program of high stakes testing and other educational initiatives.
Proponents of exit exams have argued that graduation exams make diplomas more
meaningful and motivate students to work harder, while opponents typically argue that
the tests lower graduation rates and narrow the curriculum.

Some studies found no overall relationship between exit exams and higher
dropout rates (Greene & Winters, 2004; Jacob, 2001). Greene and Winters (2004) used a
fixed effects regression model to measure graduation rates before and after
implementation of graduation tests. Results showed a statistically non-significant
negative effect, $p = 0.423$, of graduation tests on graduation rates. Jacobs examined the
effects of the bottom decile, quintile, and half of the distribution according to
standardized test scores (2001). Graduation testing appeared to have little effect on the
general student population dropout rate; however, students in the bottom quintile (lowest
25% in achievement according to standardized tests) of high stakes testing states had
about a 25% higher chance of dropping out than students from non-test states (Jacobs,
2001). Jacobs suggested that policymakers reexamine the goals, structure, and outcomes
of high stakes graduation tests, because graduation exams have the most effect on poor and minority students.

Some studies have found negative relationships between graduation exams and educational outcomes. Marchant and Paulson (2005) found, using regression analysis, that states that required graduation exams had lower graduation rates and lower Scholastic Aptitude Test (SAT) scores. Analysis by Amrein and Berliner (2002b) found that after Florida implemented the high school graduation exam, the graduation rate decreased compared to the nation, and there was weak evidence that the statewide dropout rate increased. There was strong evidence that the rate students enrolled in GED programs increased in the time period. More recently, media reports indicated graduation exams prevented 10% or 14,500 high school senior students from graduating in Florida in 2004, because they did not pass the FCAT graduation exam (Viahos, 2004).

The class of 2003 in Massachusetts was the first in that state to endure the results associated with the state's high-stake test graduation test, the Massachusetts Comprehensive Assessment System (MCAS). Although the state reported that 95% of the class passed the MCAS, a study of the enrollment patterns and high stakes testing found that this high rate of passing the test was caused by approximately 17,000 students who either previously dropped out or disappeared from the calculations (Haney, Madaus, & Wheelock, 2003; Vogler, 2004). Even if the 95% was accurate, this meant that 3,282 students were denied graduation on the basis of the one high-stakes test, even though they completed all other requirements (Haney et al., 2003). Similar problems were reported in Texas, where adjustments to the accounting system to reflect students
dropping out or otherwise being removed from accounting changed the rates significantly (Haney, 2000).

Overall, the evidence is inconclusive as to whether exit exams negatively affect lower graduation rates. Differences exist between states who employ graduation tests and those who do not, which may account for some of the differences in graduation rates. States with graduation tests most often had lower academic achievement as measured by standardized tests compared to states without graduation tests, according to tests at the beginning of high school and the end of high school (Jacobs, 2001). The states with graduation tests also had high dropout rates and higher proportions of African American, poor, or immigrant students compared to the entire student population than states without graduation tests. In addition, states with graduation tests tended to require more credits for graduation and to spend less per pupil each year.

Retention in Grade and Special Education Placement Research

As in many states, retention in grade is required in Florida for students failing the high stakes test, FCAT. In addition, both retention and special education placements change the composition of the pool of students taking the FCAT. Increases in both these student placements have an effect on graduation from high school.

In Florida, exceptional student enrollment has risen 34.1% since 1993, compared to 24.4% increase in Florida’s total enrollment (Florida DOE, 2003b). Although Florida has always retained a significant number of students, 2002-2003 was the first year that the Florida Comprehensive Achievement Test (FCAT) was mandatory for use in student retention and graduation decisions. In 2003, Florida retained in grade 98,772 boys (or 7.9%) and 63,715 girls (or 5.4%) (Florida DOE, 2004b). Florida law currently requires
retention in the 3rd grade for students failing the FCAT, with the provision that all students who fail the 3rd grade FCAT are required to participate in summer camps and then retake the test (Smith, 2004). Over 50,000 3rd graders failed the FCAT in 2003.

Retention in grade

A body of research (e.g. Shepard & Smith, 1986; Jimerson, 2001) fails to support the assumption that retaining students in grade will better prepare them for academic challenges and achievement. Furthermore, studies indicate that students retained in grade because of the requirements of high stakes testing policies may suffer destructive consequences to their future because of the risk of failure to graduate (Holmes & Saturday, 2000; Shepard, 1989).

Numerous studies have shown that retention in grade is the single biggest predictor of failure to graduate, and being retained more than once increases the odds of failure (Grissom & Shepard, 1989; Valenzuela, 2002). Rumsberger (1995) concluded that students retained at least once in grades 1-8 had four times the chance of dropping out between grades 8 and 10, when statistically controlling for socioeconomic status, eighth grade school performance, and other background and school factors. Research by Wehlage and Rutter (1986) found that being more than a year older than most other students in grade was a better predictor of dropping out than below average test scores. Once a student has been retained once, they are more than twice as likely to be retained again, and often become disengaged and drop out, especially under the pressures of high-stakes testing.

When student background, gender, and achievement were controlled for statistically, a statistically significant effect for grade retention on failing to graduate was
obtained in the Holmes study (1989). In this meta-analysis, 63 studies of retained children versus similar children who were not retained were analyzed. Results were that 54 studies found negative results of retention compared to 9 that had positive results, and the better controlled studies favored the promoted group (Holmes, 1989). Furthermore, in the Holmes meta-analysis (1989), there was a difference in mean achievement of -.31 standard deviations for retained children compared to promoted children.

Studies have shown that dropout rates were more severely impacted than academic achievement by retention (Hauser, 2002). Smith and Shepard (1989) concluded that retention in grade had the following likely results: (a) no benefits for either school achievement or social adjustment, (b) a strong relationship to later dropping out, (c) inadequate solution to school readiness in kindergarten, and (d) considerably conflict-laden and hurtful consequences for the student. After reviewing the research findings on retention, Darling-Hammond (1994) concluded that retention had negative impacts on children’s self-esteem and social adjustment as well as increasing the likelihood of dropping out.

Academic achievement or social adjustment may not be positively impacted by retention in grade. Studies conducted by Shepard and Smith (1986) concluded that academic achievement was not positively impacted substantially after grade retention. A meta-analysis of 20 studies explored the efficacy of grade retention in the 1990s; however, only 20% reached favorable conclusions about the student’s post retention achievement or social adjustment (Jimerson, 2001). Most positive outcomes for the dependent variable achievement were for the year immediately after the retained year with a mean effect size of .09 in favor of the retained students compared to un-retained
students. In subsequent years, the mean effect size for achievement was -.31 in favor of the un-retained students.

Heubert (2001) reported that in New York City, Chicago, and other cities, hundreds of thousands of predominately black, Latino, and English language learners failed promotion tests and were retained in grade. Strong relationships have been found between groups of students, such as the poor and non-white, and the use of tests for promotion and retention (Hauser, 2002). Hauser (2002) studied retained students and found that for black and Latino students fifteen to seventeen years of age, 40% to 50% were older than average for their grade, while only 35% of white students were older than average.

When students face retention in grade combined with the challenges faced by students entering high school, many do not graduate. Studies by Haney (2000, 2001) indicated a strong association between retention in the ninth grade and failure to graduate, where nationally, 7 out of 10 students who are retained in the ninth grade did not graduate. In Texas, only 20% of students graduate who were retained in the ninth grade. Livingston and Livingston (2002) examined retention policies in Georgia and reached the conclusion that the majority of poor rural blacks will fail to graduate due, in part, to mandatory retention policies based on high stakes testing.

Many educators have reached the conclusion that retention in grade contributes to failing to graduate. The increase in high school students who fail to graduate in North Carolina has been blamed by North Carolina's education leaders on the higher standards causing greater retention and dropping out (Jones, Jones, & Hargrove, 2003). North Carolina's graduation rate had dropped from about 66% to 56% from 1990 to 2000. A
review of literature indicated that virtually all studies supported the conclusion that retention in grade, including the lower grades, was related to an increase in students failing to graduate (Rumberger, 2001). Retention in the ninth grade can mean the end of students' high school career; however, passing to the 10th grade does not guarantee success. In Florida, about 30-40% of 10th grade students did not pass either the reading or math portion of the 2002 FCAT (Abrams, 2004).

Special education and remedial placements

Retention in grade and placement into remedial programs or special education usually has negative consequences for graduation (Gopaul-McNicol & Armour-Thomas, 2002; Jones, et al., 2003). Students with disabilities are failing some states' high stakes graduation tests at rates up to 70% to 90% (Heubert, 2002). According to a survey given to over 200 students at two large, urban high schools, students placed in special education perceive school as alienating at a significantly higher rate than non-special education students (Brown, Higgins, Pierce, Hong, & Thoma, 2003).

Low income and minority children are more likely to be placed into remedial, special education, or low-track classes than middle class, white children (Heubert, 2002; Jones, Jones, & Hargrove, 2003; Olivos, 1990; Parrish, 2002). Racial disparities in special education exist in 45 states (Parrish, 2002), with blacks and American Indians extensively overrepresented, Hispanics sometimes overrepresented, and Asians underrepresented. Overrepresentation of minority students in special education and remedial tracks is consistent with other significant disparities between racial groups that exist on nearly every key indicator, including dropout rates, disciplinary rates, graduation rates, and college entrance rates (Gordon, Piana, & Keleher, 2000). Data collected from a
dozen school districts (Gordon et al., 2000) indicated that the cumulative effect of racial profiling by the education system removes African American, Latino/Latina, and Native American students from challenging coursework, retains them in grade, suspends, expels, and finally pushes them out of school without a diploma in far greater numbers than white students.

"Soft disability" (termed soft due to the subjective criteria for placing children compared to regular medical diagnosis) categories of mental retardation, emotional disturbance, and specific learning disability have the greatest racial disparities, even when controlling statistically for poverty, poor health, and inferior schools. In Florida, the risk of a black child being labeled mentally retarded is 3.9 times that of a white child, and only 23% of special education students graduate (Orfield et al., 2004).

Removing special education and low performing students from the regular high stakes testing can have the effect of raising school grades. A study of 12 elementary schools located in 12 school districts in New York State examined the increased pattern of retention, remedial education, and special education placements that co-occurred with the increase in high-stakes testing during three school years: 1978-1979, 1984-1985, and 1988-1989 (Allington & McGill-Franzen, 1992). The incidence of increased special education placement was statistically significant for all three time periods, $F(2, 22) = 8.03, p < .01$, which mirrored the state and national trends. A significant Time x Grade interaction effect for the incidence of increased grade and special education placements, focusing on grades K-2, was $F(2,22) = 9.19, p < .01$. While it is difficult to determine the causes of concurrent reported increases in student achievement as measured by increased high-stakes test scores in New York, removing significantly low performing
students from the accountability measures creates a spurious result that appears to be increased academic achievement. Thus, according to Allington and McGill-Franzen (1992), pressures to improve school's chances of receiving higher school grades may increase inappropriate referrals.

Factors in Dropout Behavior

One key to understanding dropout behavior or failure to graduate is to recognize that graduation is a process, not an isolated event (Beatty, Neisser, Trent, & Heubert, 2001; Jimerson, Egeland, Strouse, & Carlson, 2000; Rumberger, 2001). Schools are not an isolated entity or playground, but a part of larger society, including family, friends, the community, the school, and society. Demographic characteristics of gender, racial/ethnic identity, immigration status, and language background are correlated to school dropout (Rumberger, 2001). Multiple factors interact over time to make isolating a single cause for failing to graduate nearly impossible.

School factors contribute in many cases to students failing to graduate. Large high schools have a higher dropout rate. For example, schools enrolling between 1,500 and 2,500 students had after the 10th grade a dropout rate of 12% compared to 7% for medium or small high schools (Lee & Burkam, 2001). One study found that students often drop out in the 9th grade because of reasons encountered in the 9th grade, rather than simply as a result of preparation from previous grades (Neild, Stoner-Eby, & Furstenberg, 2000). Structural factors of tracking and grade retention in the school have been found to contribute to dropping out (Velez & Saenz, 2001). Schools themselves may initiate a significant proportion of the dropouts, sometimes by strict enforcement of attendance laws (Fine, 1991). The relative size of the racial/ethnic minority is also important, with
the larger proportion of student body being non-white constituting a risk factor for dropping out and lower pass rates on high stakes tests (Borman, Eitle, Michael, Lee, Johnson, Cobb-Roberts, Dom, & Shircliffe, 2004). Other school factors, such as high concentration of low-achieving students and less-qualified teachers, are associated with failure to graduate (Beatty, Neisser, Trent, & Heubert, 2001).

Multiple factors from individual student histories, characteristics, and dispositions are related to whether a student will graduate from high school. Individual characteristics, such as residential and school mobility, home environment, and economic contexts influence student dropout decisions (Rumberger, 2001). School related reasons for dropping out, such as low academic achievement, poor student engagement, and absenteeism, are common reasons. Low academic achievement, retention in grade, and disengagement as early as the first grade are also predictive of eventual dropping out (Campbell, 2004).

Some students express themselves and their cultural identities through resistance to schooling. School alienation and poor student-teacher relationships have been found to contribute to students leaving school (MacDonald, 1992; Wayman, 2002). Students may perceive that teachers treat students differently or are biased according to racial or ethnic background, and these perceptions are one component of student disengagement from school. Wayman (2002) employed logistic regression analysis to study the relationship of ethnicity, social status, and gender to teacher bias. Results were that based on perceptions of students, ethnic bias was the greatest facet of teacher bias, followed by status bias.

Resistance to schooling can be a formidable obstacle for students, especially for disenfranchised groups. A qualitative study of black students' disengagement from
schools concluded that racial/ethnic identity was a central construct in students' lives, and that racism, conformity, and perceptions of dominance were pervasive factors in student resistance to schooling (Sefa Dei, Mazzuca, McIsaac, & Zine, 1997). Also cited as reasons for black students’ departure from school were the school culture of meritocracy, differential treatment by teachers, and the negative culture of difference realized from white male power, white privilege, and the institution of otherness for people of color.

Relationships between students and teachers can be important in students' decisions to stay in school. A study of 3,840 students in 190 urban and suburban high schools found positive or negative teacher-student relationships to be an important predictor for student dropout (Lee & Burkam, 2001). Sometimes these relationships are negative according to studies that found that oppositional behaviors and adversarial subcultures engaged some students and contributed to failure to graduate, especially for minorities such as Latino/Latina or African American (Valenzuela, 2000; Velez & Saenz, 2001).

Research on Perspectives and Efficacy of Educators

Those who created high-stakes testing policies have not usually been educators; instead, they have been politicians and business people. Teachers and school administrators are groups intimately involved in the implementation of high-stakes testing policies. These professional educators are in a position to see and experience the results and be personally affected by high stakes policies.

Qualitative studies of African American and Hispanic principals and superintendents in Texas found that some educators believed the tests provided increases
in equity and evidence that all children can succeed (Scheurich, Skrla, & Johnson, 2004; Skrla & Scheurich, 2004). However, another qualitative study using critical dialogue practices and narrative inquiry focused on perceptions of two African American teachers in Texas concerning the impact of standardized curriculum on African Americans (Stephens, et al., 2002). The teachers thought that the curriculum perpetuated tracking and reproduced social and educational inequities for African Americans. Emphasis on test scores they believed led to a negative learning climate for African Americans.

Other studies showed that high stakes testing had a major effect on teaching. A study of the influence of high stakes testing and teacher professional development indicated that teachers with high stakes accountability systems focused more on academic standards because of high stakes testing and that professional development for these teachers had improved (Berry, Turchi, Johnson, Hare, Owens & Clements, 2003).

A survey study of teacher efficacy, or the belief in their own ability to help students learn and pass the high stakes tests, indicated that although most teachers had a negative attitude towards the tests, this attitude was not associated with lower student achievement (Mulvenon, McKenzie, Connors, & Williams, 2003). Teachers who taught mathematics and reading in the grades that were tested had a significantly higher efficacy in reading, $F(1,141) = 8.97, p = .0032$, and mathematics, $F(7,129), p = .0596$, than teachers who were not required to test with for reading and for math efficacies.

A national survey of teachers indicated that the severity of the consequences of high stakes testing policies influences the instruction students receive (Pedula, Abrams, Madaus, Russell, Ramos, & Miao, 2003). Teachers reported nationwide that the state testing program led them to teach in ways contrary to their own ideas of sound
educational practices (Abrams, 2004; Flores & Clark, 2003; Kozol, 2005; Rex & Nelson, 2004). Greater numbers of Florida teachers than the national average (79% v. 64%) reported pressure to raise test scores, especially from external sources (Abrams, 2004).

A survey by Fielding (2004) of 1,049 Texas educational diagnostician’s perceptions in 2001-2002 indicated that over 78% of respondents answered that more than 50% of all special education referrals were driven primarily by low performance on the state high-stakes test, TAAS. Over 86% reported some level of pressure from administrators to recommend that students who failed the TAAS qualify for special education services. Findings of a study by Allington and McGill-Franzen (1992) indicated that special education referrals were increased when high stakes tests were implemented.

Conclusions

After examining the theoretical and empirical research base, the answer to whether high stakes testing policies are related to graduation rates is unclear. The complexity of the environment, the multifaceted character of high stakes testing policies, the many stakeholders, and the broad and contradictory research base imply to researchers that analysis should be carefully designed, implemented, and interpreted. The importance of research on the high stakes testing phenomena and graduation cannot be understated, because the quality of our schools and the education of our children are at stake.

High stakes testing policies are broad and far reaching in scope; therefore, a macro study focusing on broad societal impacts and consequences is appropriate. The long history of the different policy periods for 28 years in Florida makes Florida the ideal
place to study the policy relationships to graduation rates. A longitudinal examination of
Florida high stakes testing policies and graduation rates can enable identification of
differences in graduation rates attributable to different high stakes testing policy periods.

In addition, a micro study of individual student characteristics and achievement history can help stakeholders, such as students, parents, educators, and concerned members of society, discern which factors are most essential for graduation success for individual students. The individual level is most important for educational researchers, because often this level is where the efforts of all stakeholders can be most effective. In contrast to the macro level that is largely controlled by people unconnected to the everyday life of most stakeholders, the micro level is characterized by the accessibility to concerned students, parents, educators, and members of society.
CHAPTER 3
METHODS AND PROCEDURES

The study consisted of two phases. A study from the macro perspective examined broad relationships between seven different high stakes testing policy periods and graduation rates for 28 years. The micro perspective focused on the individual student's graduation success in Florida during the 1999 to 2003 time frame. Those achievement and demographic factors that were most predictive of individual student's graduation success were examined in the micro study.

Research Design

The macro analysis utilized a quantitative trend analysis procedure to evaluate the relationship between high-stakes testing policies and graduation rates in Florida from 1978 to 2005. In the macro phase, the study utilized a longitudinal, exploratory design (McMillan & Schumacher, 2001). The study determined graduation rates with the Cumulative Promotion Index (CPI) using State of Florida DOE aggregated enrollment counts and graduation counts. The macro phase examined the relationship of graduation rates to seven different high stakes testing periods from 1978 to 2005. Broad sociological explanations involving the use of high stakes testing policies and the phenomena of graduation rates and changes in graduation rates over 28 years were sought from the macro perspective.
The micro analysis focused on the likelihood of an individual student's graduation related to high stakes testing scores, achievement data, and demographic factors. Explanations of individual student graduation success were sought through the description of the relationship of specific student variables to graduation success. This phase utilized a quantitative methodology in a longitudinal, exploratory cohort design (Huerta, 2003). Data from the school records of a random sample of 3,000 high school students in Florida was utilized. In addition, the sample was divided, by a quasi random process, into 3 samples of 1000 students each. Cross validation analysis was performed on the three random samples to determine if confidence in the generalizability of the equation was warranted (Osborne, 2000).

**Research Questions**

Three research questions were posited for investigation

1. To what extent are high-stakes testing policies related to graduation rates from high school in Florida?
2. Can changes in high-stakes testing policies be useful in predicting a change in the number of students graduating from high school?
3. What relationships may be found between graduation from high school and students’ high-stakes test scores, history of retention in grade, attendance, grades, membership in racial groups, gender, socio-economic class, limited English proficiency, and exceptional student education classification?

In order to answer the research questions, four hypotheses were posited for analysis. Hypothesis 1 was designed to test if a significant amount of variance in the dependent variable could be attributable to the demographic variables: percentages of
white identity, percentages of exceptional student education, and percentages of limited English proficiency. After removing the variation attributable to the demographic variables, Hypothesis 2 was tested to find out if a significant amount of variance in the dependent variable graduation rate could be attributable to the periods of high stakes testing. Research questions 1 and 2 concerning the extent that high stakes testing policies were related to graduation rates and whether changes in the high stakes testing policies could be useful in predicting changes in graduation rates were answered by the testing of Hypothesis 2, which was clarified by the testing of Hypothesis 1.

Hypothesis 3 was designed to test if the achievement variables of grades, retention history, attendance, and the demographic variables of gender, racial/ethnic identity, free lunch status, limited English proficiency, and exceptional student education could be used to create a model that would predict a significant amount of the variation in the dependent variable, graduation success. Hypothesis 4 was designed to test if the addition of high stakes testing scores would improve or weaken the model developed in Hypothesis 3. The third research question was answered by both Hypotheses 3 and 4. The use of 2 hypotheses clarified and isolated the variation and relationships of high stakes test scores to graduation success.

Hypothesis 1

H1: A noteworthy portion of the variance in the dependent variable, high school graduation rate, will be accounted for by the set of the independent variables {racial/ethnic identity, exceptional student education, limited English proficiency}. 
H0₁: No variance in the dependent variable, high school graduation rate will be accounted for by the set of the independent variables {racial/ethnic identity, exceptional student education, and limited English proficiency}.

**Hypothesis 2**

H2: The residualized dependent variable scores (adjusted for variance explained by the set of the independent variables) will be related to the time periods when different high stakes tests were used. The test will utilize 7 time periods when different high stakes testing programs were used: (a) 1978, during the basic skills test, SSAT-1, but before any graduation test; (b) 1979, the functional literacy test, SSAT-11 was required for graduation; (c) 1980 - 1982, during the suspension of withholding high school diplomas due to the Debra P. v. Turlington (1979) court case; (d) 1983 - 1990, implementation of a revised Secondary School Admission Test, SSAT-11 with increased standards based rigor, mandatory retention, and graduation consequences; (e) 1991 - 1997, during the High School Competency Test (HSCT) of the 1990s; (f) 1998 - 2002, during the FCAT testing from 1998 to 2002; and (g) 2003-2005 during the FCAT testing and the inclusion of AYP for school grades.

H0₂: The residualized dependent variable scores, adjusted for variance explained by the set of the independent variables, will not vary according to the time periods when different high stakes tests were used in Florida. The test will utilize 7 time periods when different high stakes testing programs were used: (a) 1978, during the basic skills test, SSAT-1, but before any graduation test; (b) 1979, the functional literacy test, SSAT-11
was required for graduation; (c) 1980 - 1982, during the suspension of withholding high school diplomas due to the Debra P. v. Turlington (1979) court case; (d) 1983 - 1990, implementation of a revised Secondary School Admission Test, SSAT-11 with increased standards based rigor, mandatory retention, and graduation consequences; (e) 1991 - 1997, during the High School Competency Test (HSCT) of the 1990s; (f) 1998 - 2002, during the FCAT testing from 1998 to 2002; and (g) 2003-2004 during the FCAT testing and the inclusion of AYP for school grades.

**Hypothesis 3**

H3: The independent variables high stakes test scores, history of retention, attendance, grades, gender, racial group membership, socio-economic status, limited English proficiency, and exceptional student education status can predict the dependent variable high school graduation success.

H03: The independent variables high stakes test scores, history of retention, attendance, grades, gender, racial group membership, socio-economic status, limited English proficiency, and exceptional student education status cannot predict the dependent variable high school graduation success.

**Hypothesis 4**

H4: A noteworthy portion of the variance in the dependent variable, high school graduation success will be accounted for by the set of the independent variables, {history
of retention, attendance, grades, gender, low socio-economic status, limited English proficiency, exceptional student education status, and racial group membership}.

$H_0$: No variance in the dependent variable, high school graduation success will be accounted for by the set of the independent variables, {history of retention, attendance, grades, gender, low socio-economic status, limited English proficiency, exceptional student education status, and racial group membership}.

**Context and Access**

For the first part of the study, data in the form of archival enrollment counts, graduation counts, and demographic counts were obtained from the published records of the Florida Department of Education, the *Florida Statistical Abstracts* (Floyd, 1996, 1997; Shermyen, 1991, 1992; Shoemeyer, 1985, 1986; Terhune, 1984; University of Florida, 1998) and the National Center of Education Statistics (U. S. Department of Education, 2002). As the basis for that data, the staff at local schools and districts completed annual surveys of basic demographic and educational information at the state, district, and school level and the data were tabulated and published by the Florida Department of Education.

For the second part of the study, the data were provided by the Florida Department of Education K-20 Data Warehouse in the form of data from archival school records. Access was secured through application to the Florida K-20 Data Warehouse and involved a lengthy process of numerous email communications and submissions of written documentation. Personal interviews involving high level personnel from the Florida Department of Education were required. The Dean of the College of Education
and Human Services at the University of North Florida was required to attest to the need for the study.

A random sample taken from a population of all ninth graders enrolled in public high schools in Florida in Fall 1999 was performed by the K-20 Data Warehouse. The school records of this randomly derived cohort of high school students provided data in the form of FCAT test scores, retention history, attendance, grades, and demographic information. Data were provided for the students longitudinally for 4 years until the graduation event in 2003.

**Ethical Consideration**

The well being and respect for the participants from which the data were drawn must come first in the research process. Anonymity was maintained because no names were attached to the data to protect the privacy of the individual’s student records. No experimental procedures were used. The K-20 Data Warehouse used randomly generated unique identification numbers for individual student records and person identified data, such as birth dates were not loaded into the database. Random selection of records further ensured a lack of sample bias. Applicants for data from the K-20 Data Warehouse must go through a high level approval process before release of data. Only the study applied for in that process may be undertaken with the data released in that manner.

**Data Collection**

For the macro study, data were collected from aggregated enrollment counts and graduation counts for public schools in Florida for grades 9, 10, 11, and 12 for each academic year from 1978 to 2005, including data for racial groups, gender, exceptional student education (ESE), and limited English proficiency (LEP). The data are available
from the published reports, briefs, and statistical abstracts (Florida DOE, 1977-2004; Floyd, 1996, 1997; Shermyen, 1991, 1992; Shoemeyer, 1985, 1986; Terhune, 1984; University of Florida, 1998). As the basis for that data, the staff at local schools and districts completed annual surveys of basic demographic and educational information at the state, district, and school level and the data were tabulated and published by the Florida Department of Education.

No consistent data for the demographics of poverty were available for the 28 year time period; therefore a poverty variable, although relevant, was not included in this phase of the study. For the variable, percentage of students in limited English proficient (LEP), data were missing from the years 1982, 1983, 1984, 1985, and 1987; therefore, data from the National Center of Educational Statistics (U. S. Department of Education, 2002) was used for the those years for the LEP variable. Data for racial/ethnic subgroups were not available for all years; therefore, graduation rates for these subgroups were not calculated for all years.

For the micro study, data were provided from the K-20 Data Warehouse at the Florida Department of Education in the form of school records of a random sample of 3,000 students in the 9th grade public schools in 1999-2000. The data followed the students in the sample for four years until the graduation or failure to graduate event. Data on graduation success, retention history, school attendance, grades, FCAT reading and mathematics scores in the 8th and 10th grades, limited English proficiency, exceptional student education, gender, and racial group membership were collected.
Data Analysis

The first two research questions, corresponding to the macro analysis, were as follows: a) To what extent are high-stakes testing policies related to graduation rates from high school in Florida?, and b) Can changes in high-stakes testing policies be useful in predicting a change in the number of students graduating from high school? For this phase of the study, Florida consolidated enrollment counts for high school for grades 9, 10, 11, and 12, graduation counts for years 1978 through 2005, and state education demographics were utilized for descriptive and analytic statistics in order to gain understanding of who graduates from high school, who does not graduate from high school, how many graduate and whether high stakes testing is related to graduation.

Statistical procedures included descriptive statistics such as calculation of the Cumulative Promotion Index (CPI) for schools in the state of Florida during the 28 year time period from 1978 to 2005 (see Appendix 2). CPI was calculated for six categories, an overall group, and five different ethnic groups (white, black, Hispanic, Asian/Pacific, and American Indian/Native Alaskan).

The relationship between periods of the high stakes testing policy and graduation were analyzed through multiple regression analysis and univariate analysis of variance. Each year, from 1978 to 2005, was considered a unit of analysis. Cohort data were used over multiple years from 1978 to 2005. Seven different time periods corresponding to the five different high stakes tests that were used in Florida constituted one independent variable. The seven time periods when different high stakes testing programs were used were: (a) 1978, during the basic skills testing SSAT-1, but before any graduation test; (b) 1979, the functional literacy test, SSAT-11 was required for graduation; (c) 1980 - 1982,
during the suspension of withholding high school diplomas due to the Debra P. v. Turlington (1979) court case; (d) 1983 – 1990, implementation of a revised Secondary School Admission Test, SSAT-11 with increased standards based rigor, mandatory retention, and graduation consequences; (e) 1991 -1997, during the High School Competency Test (HSCT) of the 1990s; (f) 1998 – 2002, during the FCAT testing from 1998 to 2002; and (g) 2003-2005 during the FCAT testing and the inclusion of AYP for school grades. Other independent variables were percentages of white identity, exceptional student education, and limited English proficiency The dependent variable for Hypothesis 1, graduation rate, was adjusted for the variation due to percentages of white identity, exceptional student education, and limited English proficiency to produce a non-standardized residual, which became the dependent variable for Hypothesis 2. Correlation, regression analysis, and univariate analysis of variance were used to explore and predict possible relationships between high stakes testing and graduation rates.

The research question corresponding to the micro study was as follows: What relationships may be found between graduation from high school and students’ high-stakes test scores, history of retention in grade, attendance, grades, membership in racial groups, gender, socio-economic class, limited English proficiency, and exceptional student education? In this phase, a longitudinal cohort design used a random sample of 3,000 records of 9th grade students from Florida public schools in the year 1999/2000 and then tracked the same students every year until the graduation event in 2002/2003. In addition, the sample of 3,000 students was divided into three groups, using a quasi random procedure, in order to cross validate the study. These three random samples were
used to perform cross validation analysis to determine if confidence in the
generalizability of the equation is warranted (Osborne, 2000).

The data were analyzed with SPSS software and statistical procedures included
descriptive statistics, such as calculation of the Basic Completion Rate (BCR) for the
sample of 3,000 students. The BCR is number of graduates/9th grade enrollment and was
calculated for eight categories, an overall group, five different racial/ethnic groups
(white, black, Hispanic, Asian/pacific, and American Indian/Native Alaskan), special
education students, limited English proficiency students, and free or reduced lunch
students. The BCR was adjusted to include only first time 9th graders in 1999 and for
estimated out migration, which is 7.38% from Florida for the four year period from 1999
to 2002 (Perry, 2003).

Correlations to indicate the degree and direction of relationships between
independent variables and the dependent variable were computed using SPSS. Then,
discriminant analysis (Klecka, 1980) was used to find out how well attainment of a high
school diploma is predicted by using the independent variables history of retention,
attendance, grades, racial/ethnic group, socio-economic status, limited English
proficiency, and exceptional student education status membership. Further analysis
determined if including the high stakes test scores in the discriminant analysis made a
difference to the percent of variance predicted by the model. The multivariate effect size
was measured via Wilks’ Lambda. Which variables were more important to the
prediction was determined, by using a discriminant structure coefficient to correlate
variables to the prediction. A test of multivariate significance (F test) was also
performed.
For the macro analysis, the dependent variables were high school graduation rate and the non-standardized residual measured by the Cumulative Promotion Index (CPI). The non-standardized residual is the amount of variation in graduation rate left over after removing the variations from percentages of limited English proficiency, exceptional student education, and white identity. Limited English proficiency (LEP) refers to the percentage of students who qualify for English language learning programs, such as English for Speakers of Other Languages (ESOL). Exceptional student education (ESE) refers to the percentage of students who have a written Individualized Education Plan (IEP). Gender refers to number of male or female high school students in the state for the year. The linear regression procedures will use categories of percentage white as an independent variable, because categorization with fewer terms is useful for producing greater variation in the regression equation.

For the micro analysis, the dependent variable was individual student graduation success in four years with a standard diploma. Independent variables included Florida high stakes test (FCAT) scores in reading and mathematics from the 8th and 10th grade (these are the only years from which the cohort of students had taken the FCAT). History of retention was the total times the student had been retained according to the records provided by K-20 Education Warehouse. Attendance was the days absent in the year 2000 according to the records provided by the K-20 Education Warehouse. Grades were the high school grade point average (GPA). The five racial and ethnic categories were white, black (non-Hispanic), Hispanic, American Indian/Alaskan Native, and Asian/Pacific Islander. Dichotomous categories of racial/ethnic identity, white and non-white, were used for the discriminant analysis with the following groups: (a) groups of
color included those classified as black (non-Hispanic), Hispanic, American Indian/Alaskan Native, or Asian/Pacific Islander and (b) the white group included those students classified as white. Socio-economic status was determined by the free/reduced lunch classification. Limited English proficiency (LEP) referred to students who qualify for English language learning programs, such as English for Speakers of Other Languages (ESOL). Exceptional student education (ESE) referred to the students who have a written Individualized Education Plan (IEP). Gender referred to either male or female status.

Chapter Summary

The study was designed to provide information and further understanding of how high stakes testing policies may be related to graduation rates in Florida. The macro analysis used aggregated data in a longitudinal, exploratory design to study broad graduation trends from a 28 year period. The independent variables were percentages of the white identity racial group, percentages of limited English proficient students, percentages of exceptional student education, and the period of high stakes testing. The dependent variables were the graduation rates, calculated using the Cumulative Promotion Index (CPI), and the non-standardized residual.

The micro analysis used individual student data in a longitudinal, exploratory, cohort design. The data were taken from the four year period when the student was in high school. The micro analysis examined how achievement and demographic variables were predictive of individual student graduation. The independent variables were individual retention history, school attendance, grades, FCAT reading and mathematics scores in the 8th and 10th grades, membership in limited English proficiency education,
membership in exceptional student education, gender, and racial group membership were collected. The dependent variable was individual graduation success.

Chapter 4 presents the findings of the descriptive and inferential statistical analysis for both the macro analysis and the micro analysis. Linear regression was used in the macro study to test Hypothesis 1 and Hypothesis 2. Discriminant analysis was used in the micro study to test Hypothesis 3 and Hypothesis 4.

Chapter 5 contains a summary of the findings, discussion, recommendations for practice, further study, as well as conclusions drawn from the study.
CHAPTER 4

RESULTS

High-stakes testing policies have broad and sometimes unintended consequences on the graduation prospects of students. The purposes of the present study were to analyze information with descriptive and inferential statistics focusing on high-stakes testing policies over a period of 28 years and examine the relationship to students' prospects for graduation from high schools. Examination of demographic and achievement determinants of graduation success clarified what, if any, relationships existed between graduation from high school and the high stakes testing variables, achievement variables, and demographic variables.

Three research questions were posited for investigation: (a) To what extent were high-stakes testing policies related to graduation rates from high school in Florida?, (b) Could changes in high-stakes testing policies be useful in predicting a change in the number of students graduating from high school?, and (c) What relationships were found between graduation from high school and students' high-stakes test scores, history of retention in grade, attendance, grades, racial/ethnic identity, gender, socio-economic class, limited English proficiency, and exceptional student education?

The macro analysis examined relationships between periods of high stakes testing policies and graduation rates, longitudinally over a period of 28 years. This part of the study used published records of the Florida Department of Education (Florida DOE,

Descriptive statistics, linear regression, and univariate analysis of variance were used to analyze the relationship between high stakes testing policies and graduation rates.

The micro analysis examined individual student records over a four year span during which a randomly selected cohort of 3,000 high school students in Florida attended high school. Demographic and achievement variables, including four sets of high stakes test scores and the dependent variable graduation success, were analyzed with descriptive statistics and discriminant analysis to obtain a clearer picture of which variables were predictive of high school graduation success.

**Findings**

*The Macro Analysis: High Stakes Testing Policy Periods and Graduation Rates*

The average CPI from Florida public high schools for the 28 year period was 58.69, which means that there was an average 58.69% chance over the 28 year period for students entering the ninth grade to graduate four years later with a standard diploma (see Table 1). The highest overall CPI was in 1982 at 68.84%, occurring at the end of a period characterized by the Debra versus Turlington court case, when the graduation test, SSAT-11, was not used to deny students graduation. The lowest overall graduation rate, at 47.66%, was in 1999, a period at the beginning of the increasingly difficult high stakes
test, FCAT. The range between the highest and lowest was 21.18. The difference between the 1978 and 2005 was 4.51, with the chance of students graduating in 2005 somewhat lower than in 1978.

The available data from archival records of the Florida Department of Education that were needed to calculate CPI for racial/ethnic subgroups were incomplete. However, from 1977 to 1981 and 1992 to 2002, the CPI for subgroups could be calculated (see Table 1). Using available data, white students had variable chances for graduation over the 28 year period, with a CPI of 66.06 in 1978, a high in 2002 of 72.21, and a low in 1999 at 52.74. Asian students had consistently high CPI, ranging from 70.65 in 1978, 98.5 in 1989, and 95.87 in 2003. Black students’ CPI generally fell over the 28 year period, with 52.50 in 1978, 56.61 in 1980, 34.24 in 1999, and 44.17 in 2003. Hispanic students had CPI scores of 56.44 in 1978, 82.17 in 1980, 48.10 in 1997, and 53.69 in 2003. American Indian students’ graduation rates were an exception, with a highest CPI in 1997 at 56.96 and the lowest rate of 12.78 in 1977. The small sample size of only a few hundred students with about 60 graduates for this group indicates that further research using a larger sample is necessary before drawing conclusions. In addition, in the year 1978, the CPI for American Indians was over 128, indicating that students enrolled and left the school in an uneven pattern. Possibly, more American Indians were recorded as graduating than were recorded in the ninth grade.
Table 1
Cumulative Promotion Index (CPI) in Florida from 1978 to 2005

<table>
<thead>
<tr>
<th>School year</th>
<th>Overall CPI</th>
<th>Period average CPI</th>
<th>CPI white</th>
<th>CPI black</th>
<th>CPI Hispanic</th>
<th>CPI Asian</th>
<th>CPI American Indian</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978</td>
<td>62.37</td>
<td>(1) 62.37</td>
<td>66.06</td>
<td>52.60</td>
<td>56.44</td>
<td>70.65</td>
<td>12.78</td>
</tr>
<tr>
<td>1979</td>
<td>60.02</td>
<td>(2) 60.02</td>
<td>62.23</td>
<td>49.46</td>
<td>58.35</td>
<td>*177.64</td>
<td>*128.78</td>
</tr>
<tr>
<td>1980</td>
<td>65.60</td>
<td>(3) 65.47</td>
<td>66.89</td>
<td>56.10</td>
<td>82.17</td>
<td>96.21</td>
<td>24.64</td>
</tr>
<tr>
<td>1981</td>
<td>63.78</td>
<td>-</td>
<td>66.15</td>
<td>56.61</td>
<td>60.32</td>
<td>91.13</td>
<td>48.11</td>
</tr>
<tr>
<td>1982</td>
<td>66.84</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1983</td>
<td>68.82</td>
<td>(4) 61.59</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1984</td>
<td>62.36</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1985</td>
<td>63.74</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1986</td>
<td>61.01</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1987</td>
<td>63.40</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1988</td>
<td>55.41</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1989</td>
<td>60.42</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1990</td>
<td>62.02</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1991</td>
<td>64.07</td>
<td>(5) 57.87</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1992</td>
<td>64.21</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1993</td>
<td>59.43</td>
<td>-</td>
<td>60.63</td>
<td>51.62</td>
<td>61.81</td>
<td>98.51</td>
<td>56.09</td>
</tr>
<tr>
<td>1994</td>
<td>56.53</td>
<td>-</td>
<td>58.51</td>
<td>47.37</td>
<td>59.40</td>
<td>93.27</td>
<td>55.15</td>
</tr>
<tr>
<td>1995</td>
<td>56.20</td>
<td>-</td>
<td>58.23</td>
<td>47.21</td>
<td>59.07</td>
<td>94.77</td>
<td>-</td>
</tr>
<tr>
<td>1996</td>
<td>55.30</td>
<td>-</td>
<td>58.36</td>
<td>48.61</td>
<td>54.16</td>
<td>84.63</td>
<td>-</td>
</tr>
<tr>
<td>1997</td>
<td>53.13</td>
<td>-</td>
<td>57.93</td>
<td>42.32</td>
<td>48.10</td>
<td>84.14</td>
<td>-</td>
</tr>
<tr>
<td>1998</td>
<td>50.26</td>
<td>-</td>
<td>54.61</td>
<td>39.46</td>
<td>48.40</td>
<td>82.53</td>
<td>56.96</td>
</tr>
<tr>
<td>1999</td>
<td>50.30</td>
<td>(6) 49.39</td>
<td>53.63</td>
<td>38.93</td>
<td>53.04</td>
<td>79.82</td>
<td>50.82</td>
</tr>
<tr>
<td>2000</td>
<td>47.66</td>
<td>-</td>
<td>52.74</td>
<td>34.24</td>
<td>48.60</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2001</td>
<td>50.12</td>
<td>-</td>
<td>56.31</td>
<td>37.75</td>
<td>50.85</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2002</td>
<td>50.12</td>
<td>-</td>
<td>54.74</td>
<td>37.65</td>
<td>51.46</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2003</td>
<td>61.61</td>
<td>(7) 58.99</td>
<td>72.21</td>
<td>44.17</td>
<td>53.69</td>
<td>95.87</td>
<td>44.55</td>
</tr>
<tr>
<td>2004</td>
<td>57.58</td>
<td>-</td>
<td>63.44</td>
<td>43.09</td>
<td>57.47</td>
<td>88.69</td>
<td>17.23</td>
</tr>
<tr>
<td>2005</td>
<td>57.86</td>
<td>-</td>
<td>63.13</td>
<td>43.39</td>
<td>57.36</td>
<td>91.10</td>
<td>88.45</td>
</tr>
</tbody>
</table>

Average 58.69

*Note: These values over 100% indicate uneven enrollment in this years for the subgroups of Asian and American Indian.
Using available data from 1978 to 1981 and from 1993 to 2005, the achievement gap between white students and black students changed, from the CPI being about 10 points higher for white students than black students in the earlier periods until 1995, when the black student graduation rate started to fall. In 2003, the achievement gap was 28 points, almost a 200% increase. The white student and Hispanic student achievement gap also increased from 1993 to 2003 by 19.51 points in favor of white students. In 2005 the achievement gap narrowed to about 20 points between black and white students and to about 6 points between Hispanic and white students.

The 28 year span from 1978 to 2005 was divided into 7 periods of high stakes testing policies (FL DOE, 2003a) (see Table 2). The first period, in 1978, featured the SSAT-1 basic skills assessment test, but this test was not used for graduation decisions. The CPI for the first period was 62.37. The second period was in 1979, when the new functional literacy high stakes test, the SSAT-11, was given, with an average CPI of 60.02. The third period, from 1980-1982, was characterized by a suspension of graduation consequences of the SSAT-11, with an average CPI of 65.47. During the fourth period, from 1983 to 1990, the state required SSAT-11 with increased rigor and consequences. The average CPI of the fourth period was 62.17. The increased standardized and rigorous assessment trend continued with the High School Competency Test (HSCT) from 1991 to 1997, with an average CPI of about 57.39. During the sixth period, from 1998 to 2002, the state used the Florida Comprehensive Assessment Test (FCAT), with an average CPI of 49.74. Finally, during the seventh period, 2003 to 2004, the state required the FCAT and the addition of the No Child Left Behind Accountability
measure known as the Adequate Yearly Progress (AYP), which included graduation rates as a measure of school performance. The seventh period had an average CPI of 59.7.

Table 2
Periods of Florida High Stakes Testing

<table>
<thead>
<tr>
<th>Period</th>
<th>Test Description</th>
<th>High Stakes</th>
<th>Required for Graduation</th>
<th>Average CPI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1978</td>
<td>The 1975-77 statewide assessment test (SSAT-1) basic skills test was not a requirement for graduation or a high stakes test.</td>
<td>no</td>
<td>no</td>
<td>62</td>
</tr>
<tr>
<td>1979</td>
<td>The functional literacy graduation test requirement, the State Student Assessment Test, Part II (SSAT-II)] was required for graduation.</td>
<td>no</td>
<td>yes</td>
<td>60</td>
</tr>
<tr>
<td>1979-1982</td>
<td>The Debra P. v. Turlington case court case led to suspension of the withholding of diplomas on the basis of a student's failure to pass the SSAT-II.</td>
<td>no</td>
<td>suspended</td>
<td>65</td>
</tr>
<tr>
<td>1983-1990</td>
<td>A new scale score was used for a revised SSAT-II test, with mandatory retention and graduation consequences.</td>
<td>no</td>
<td>yes</td>
<td>62</td>
</tr>
<tr>
<td>1991-1997</td>
<td>The High School Competency Test (HSCT) was given with increased standards based difficulty.</td>
<td>no</td>
<td>yes</td>
<td>56</td>
</tr>
<tr>
<td>1998-2002</td>
<td>Florida Comprehensive Achievement Test (FCAT), based, on the Sunshine State Standards was implemented.</td>
<td>yes</td>
<td>yes</td>
<td>50</td>
</tr>
<tr>
<td>2002-2005</td>
<td>FCAT and AYP are used for accountability for graduation rates.</td>
<td>yes</td>
<td>yes</td>
<td>60</td>
</tr>
</tbody>
</table>
Demographics for the Macro Analysis

Changes in aggregate demographics from 1978 to 2005 followed linear trends (see Figure 2). The percentage of white students fell from 68.89% in 1977 to 48.80% in 2005. During the same period, the percentage of students enrolled in exceptional student education increased from 8.95% in 1977 to 19.58% in 2005. The percentage of students in limited English proficiency programs also increased from 2.1% in 1977 to 9.5% in 2005. Overall, a higher percentage of graduates were girls (52.78%) than boys (47.22%).
Figure 2. Changes from 1978 to 2005 of white identity, exceptional student education, and limited English proficiency.


The expected pattern was for historically low academically performing groups, such as students with a special education classification, students with limited English proficiency, and students who were not white, to have lower graduation rates. As shown in Figure 3, graduation rates were erratic over the 28 years. Graduation rates fell overall
4.51% from 62.37% in 1977 to 57.86% in 2005. A similar erratic pattern did not occur in the demographic trends.

Figure 3. Raw Overall CPI Graduation Rates in Florida 1977-2005

When analyzed by year, the CPI shows that a decline in graduation success corresponded to stricter high stakes testing policies, such as in 1979, 1984, 1988, and 1993 through 2003. When periods of the testing were not used for graduation decisions, the graduation rates spiked (see Figure 3). For example, in 1979, when the SSAT-11
counted for graduation, the CPI was 60.02; whereas, in 1980, the SSAT-11 did not count for graduation, and the graduation rate spiked to 65.80. The 2nd spike in the graduation rate chart occurred at the end of the decade of the 1980s, when the state of Florida informally suspended the stricter consequences of retention for several years due to lack of positive results in test scores (Morris, 2001). The lowest period of graduation success was during the FCAT period from 1998 to 2002, when high stakes testing and strict consequences for students, teachers, schools, and communities was mandated and enforced. The CPI of this period ranged from 48 to 50. There was a substantial upturn in CPI from 2003 of more than 11 points to 61.61. In 2004 and 2005, the CPI leveled off at about 58.

Correlations for the Macro Analysis

All of the demographic variables, percentage of exceptional student education membership (ESE), percentage of white students (racial/ethnic identity), and percentage of limited English proficiency membership (LEP) were highly correlated with each other according to the Pearson correlations of |.928| or more (see Table 3). The variable graduation rate (grad rate) was slightly less correlated with the demographic variables but was still highly correlated with a Pearson correlation of |.672| or more, p < .001. Correlations show that lower values of graduation rate coincided with the increase in exceptional student education, the reduction in the percentage of white students, and the increase in limited English proficiency. Graduation rate was moderately correlated with time period at -.512. The unstandardized residual, which was the variation left over in graduation rate after removing the variations attributable to ESE, percentage white identity, and LEP, was correlated with graduation rate at .684.
Table 3
Correlations for the Macro Analysis.

<table>
<thead>
<tr>
<th></th>
<th>Period</th>
<th>ESE</th>
<th>Identity</th>
<th>LEPE</th>
<th>Grad Rate</th>
<th>Unstandardized Rate</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period</td>
<td>1</td>
<td>.941**</td>
<td>-.927**</td>
<td>.881**</td>
<td>-.512**</td>
<td>.194</td>
<td></td>
</tr>
<tr>
<td>ESE</td>
<td>-</td>
<td>1</td>
<td>-.970**</td>
<td>.928**</td>
<td>-.691**</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Identity</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-.962**</td>
<td>.672**</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>LEP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>-.708**</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Grad Rate</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>.684**</td>
<td></td>
</tr>
<tr>
<td>Unstandardized</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**. Correlations were statistically significant, p < .005 level (n = 28)

For major subgroups, correlations between black student graduation rate, time period, ESE, white identity, LEP, white student graduation rate, and Hispanic student graduation rate were large and statistically significant (> .691, p < .01) (see Table 4). Hispanic student graduation rates were moderately and negatively correlated with ESE (- .581, p = .011), white identity (.546, p = .023), LEP (- .588, p = .013), white student graduation rates (.529), and black student graduation rates (.748). White student graduation rates were not significantly correlated with the demographic variables. Thus, black student graduation rates were most affected of the three major subgroups. Black and Hispanic subgroups showed that corresponding to increasing percentages of ESE and
LEP, their graduation rates were lower. Where high percentages of white students existed, higher graduation rates for blacks and Hispanics occurred.

Table 4
Correlations for Subgroups in Macro Analysis

<table>
<thead>
<tr>
<th>Period</th>
<th>%ESE</th>
<th>%White</th>
<th>%LEP</th>
<th>White GR</th>
<th>Black GR</th>
<th>Hispanic GR</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Period</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 28</td>
<td>.941*</td>
<td>-.927*</td>
<td>.881*</td>
<td>-.271</td>
<td>-.701**</td>
<td>-.393</td>
</tr>
<tr>
<td>ESE</td>
<td>-.970*</td>
<td>.928*</td>
<td>-.428</td>
<td>-.801*</td>
<td>-.581***</td>
<td></td>
</tr>
<tr>
<td>% White</td>
<td>-.962*</td>
<td>.325</td>
<td>.804*</td>
<td>.548***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LEP</td>
<td>-.407</td>
<td>-.831*</td>
<td>-.588***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White GR</td>
<td>-.407</td>
<td>-.831*</td>
<td>-.588***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black GR</td>
<td>-.407</td>
<td>-.831*</td>
<td>-.588***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic GR</td>
<td>-.407</td>
<td>-.831*</td>
<td>-.588***</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: * p < .001, ** p < .01, *** p < .05

Testing of Hypothesis 1

H1 states a noteworthy portion of the variance in the dependent variable, high school graduation rate, will be accounted for by the set of the independent variables {racial/ethnic identity, special education, and limited English proficiency}. Linear regression was performed to test the null hypothesis, HO1 (see Table 5).

Table 5
Linear Regression Effect Size for HO1

<table>
<thead>
<tr>
<th>Model</th>
<th>R</th>
<th>R Square</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.730*</td>
<td>.532</td>
</tr>
</tbody>
</table>

a. Independent Variable: ESE, LEP, racial/ethnic identity
b. Dependent Variable: grad rate
Results showed that the effect size or $R^2$ was .532, thus 53.2% of the variation in graduation rate was explained by the demographic variables in the regression equation for Hypothesis 1 (see Table 5). This means that 46.8% of the variation can be explained by something other than the demographic variables of Hypothesis 1. The regression $R$ was statistically significant ($F(3, 27) = 9.110, p < .001$) (see Table 6).

Table 6
Sum of Square Breakdown for HO:

<table>
<thead>
<tr>
<th>Model</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Regression</td>
<td>.046</td>
<td>3</td>
<td>.015</td>
<td>9.110</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Residual</td>
<td>.041</td>
<td>24</td>
<td>.022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>.087</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a: Predictors: (Constant), LEP, ESE, white identity
b. Dependent Variable: grad rate

The regression equation indicates that the predicted CPI was lower when percentages of exceptional student education and limited English proficient students were higher.

Hypothesis 1 Regression Equation (see Table 7)

Graduation Rate = 1.288 - .011(ESE) - .007 (Racial/ethnic identity) - .019 (LEP)
Table 7

**Coefficients for Linear Regression HO₁**

<table>
<thead>
<tr>
<th>Model</th>
<th>B</th>
<th>t</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Constant)</td>
<td>1.288</td>
<td>2.389</td>
<td>.025</td>
</tr>
<tr>
<td>ESE</td>
<td>-.011</td>
<td>-1.257</td>
<td>.221</td>
</tr>
<tr>
<td>Racial/ethnic identity</td>
<td>-.007</td>
<td>-1.094</td>
<td>.285</td>
</tr>
<tr>
<td>LEP</td>
<td>-.019</td>
<td>-1.678</td>
<td>.106</td>
</tr>
</tbody>
</table>

Note: Dependent variable is grad rate

In the regression equation, the predicted CPI also was lower for increasing percentages of white students; however, there was a positive correlation (.672, p < .001) between racial/ethnic identity (percentage of white students) and graduation rate. In addition, the graduation rate achievement gap radically widened after 1997, with white students benefiting. Lower percentages of white students coincided with a rising white graduation rate and widening racial achievement gap favoring white students.

Interpretation of these findings, as with all correlation studies, should be done with caution. One possible interpretation is that white students performed better academically in later years than in earlier years, because white students benefited from the increased usage of high stakes testing policies.

The null Hypothesis 1, that no variance in the dependent variable, high school graduation rate, would be accounted for by the set of the independent variables {racial/ethnic identity, special education, and limited English proficiency}, was rejected because a significant amount, 53.2% of the variation was explained by the set of independent variables.
Testing of Hypothesis 2

H2 stated that the residualized dependent variable scores, adjusted for variance explained by the set of the independent variables, would vary according to the time periods when different high stakes tests were used in Florida. Seven time periods (see Table 2) when different high stakes testing programs were used: (a) 1978, during the basic skills test, SSAT-1, but before any graduation test; (b) 1979, the functional literacy test, SSAT-11 was required for graduation; (c) 1980 - 1982, during the suspension of withholding high school diplomas from the SSAT-11 due to the Debra P. v. Turlington court case; (d) 1983 - 1990, implementation of a revised Secondary School Admission Test, SSAT-11 with increased standards based rigor, mandatory retention, and graduation consequences; (e) 1991 -1997, during the High School Competency Test (HSCT) of the 1990s; (f) 1998 - 2002, during the FCAT testing from 1998 to 2002; and (g) 2003-2005 during the FCAT testing and the inclusion of AYP for school grades.

The correlation between time periods (year) and the graduation rate was -.512, \( p = .004 \), indicating that the graduation rate as measured by CPI fell moderately over the time periods chronologically (see Table 3). In addition, the correlations between time periods and the demographic variables were high. See Figure 2 for a graphic display of the direction of change in the demographics over the years. See Table 8 for descriptive statistics on the means for the seven time periods and the standard deviations that show the variation for graduation rates with the time periods to be mostly normally distributed. The means of the graduation rates by period ranged from 65.47% (Period 3) to 49.74% (Period 6).
Table 8
Descriptive Statistics for Means and Standard Deviations across Time Periods

<table>
<thead>
<tr>
<th>Time Period</th>
<th>N</th>
<th>Mean Graduation Rate</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>62.37</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>56.63</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>65.47</td>
<td>1.51</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>62.16</td>
<td>3.75</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>57.39</td>
<td>4.96</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>49.74</td>
<td>1.42</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>59.70</td>
<td>1.95</td>
</tr>
</tbody>
</table>

Table 9 shows that Period 1 (no high stakes test given in 1978), Period 2 (when the SSAT-11 was first given in 1979), and Period 6 (when the FCAT was given in 1998-2002) were periods of falling non-standardized residual means. Periods 4, 5, and 7 had rising non-standardized residual means.

Table 9
Descriptive Statistics of Mean and Standard Deviation for Time Period and Non-Standardized Residual for H02

<table>
<thead>
<tr>
<th>Year</th>
<th>Mean</th>
<th>St. Deviation</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-.031764</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>-.63729</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>.023000</td>
<td>.0190520</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>.001691</td>
<td>.0338541</td>
<td>8</td>
</tr>
<tr>
<td>5</td>
<td>.003510</td>
<td>.0375985</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>-.039600</td>
<td>.0106479</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>.047761</td>
<td>.0301948</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>.000000</td>
<td>.0387813</td>
<td>28</td>
</tr>
</tbody>
</table>

Results of the univariate analysis of variance for Hypothesis 2 may be seen in Table 10. The independent variable was the time period for the high stakes testing. The
dependent variable was the non-standardized residualized variation, which was the
variation in graduation rate that was left over after removing the variation from
percentages of exceptional student education (ESE), white identity, and limited English
proficient (LEP).

Table 10
Test of Between-Subjects Effects for $H_0^2$

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F</th>
<th>Sig.</th>
<th>Partial Eta Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year</td>
<td>.020</td>
<td>6</td>
<td>.003</td>
<td>3.362</td>
<td>.018</td>
<td>.490</td>
</tr>
<tr>
<td>Error</td>
<td>.021</td>
<td>21</td>
<td>.001</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>.041</td>
<td>28</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: $R^2 = .490$

In testing the null hypothesis, $H_0^2$, the univariate analysis of variance was
statistically significant ($R^2 = .490, p = .018$), which means 49% of the variation in the
residual variation of the graduation rate after the variation from ESE, percent white
identity, and LEP can be explained by the independent variable time period. About 51%
of the residual variation can be explained by factors other than the time period.

The null hypothesis $H_0^2$ was rejected because the residualized dependent variable
score, adjusted for variance explained by the set of independent variables of limited
English proficiency, exceptional student education, and racial/ethnic category, varied
significantly with 49% of the variation attributed to the time period variable.

The Micro Analysis: Individual Student Variables and Graduation Success

The micro study focused on the relationship between individual student variables,
such as achievement and demographic variables, and the individual’s chance for
graduating from high school. Data from a random sample of 3,000 ninth grade students in
1999 from the State of Florida during the four years of high school until the graduation event was analyzed with descriptive statistics and discriminant analysis.

Descriptive Statistics for the Micro Study

The random sample (N=3,000) taken by the Florida Department of Education from the population of all 9th grade Florida public school students in 1999-2000 was composed of slightly more males than females (see Table 11). Slightly less than half (49.5%) were white, black students were 26.7%, Hispanic students were 19.9%, Asian students were 1.9%, American Indian/Islander were .4%, and 1.4% were other racial/ethnic groups. About 12.7% were enrolled in limited English proficient programs and 39.5% were enrolled in the free/reduced lunch program. 21.8% were enrolled in exceptional student education. The population of all K-12 public school students in Florida in 2000 (Florida DOE, 2005b) consisted of 52.6% white, black students were 24.8%, Hispanic students were 19.1%, Asian students were 1.9%, American Indian students were 3%, and other racial/ethnic identity groups were 1.4%. LEP enrollment of the total population in 1999 was 7.1%. ESE enrollment of the total population in 1999 was 19.1%. Free/reduced lunch enrollment in 1999 was 44%.

Only 42.6% of the 3,000 students graduated in four years with a standard diploma. The following racial/ethnic percentages graduated: white students had 48.4%, black students had 32.2%, Hispanic students had 41.5%, Asian students had 51.7%, and American Indian/Pacific Islander had 25%. When the 404 students who had been retained in the ninth grade prior to 1999-2000 were removed from the calculations, the basic completion rate (BCR) for students entering the ninth grade for the first time in 1999-2000 was 47%.
Limits in the data supplied by the Florida DOE made it impossible to determine how many students transferred out of the Florida public school system to another state’s school system. Data on which students withdrew from school was provided, but there was insufficient detail to discern transfers out of state from transfers within the state because the classification system for transfers used by the Florida Department of Education did not discriminate between transfers in state or out of state. However, according to data obtained from the U.S. Census (Perry, 2003), the out migration for 15-19 year old children in Florida in the time period 1999-2002 was about 7.38%. When adjusted for estimated out migration from Florida and to include only first time 9th graders, the basic completion rate was 51% in 2003. This rate is consistent with the 53% graduation rate reported by Orfield et al. (2004) for the same year of 2003.
Table 11
Demographics of Random Sample of 3,000 Florida 9th Grade Students 1999-2000

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>1,606</td>
<td>53.5</td>
</tr>
<tr>
<td>Female</td>
<td>1,394</td>
<td>46.4</td>
</tr>
<tr>
<td>Racial/ethnic identity-white</td>
<td>1,488</td>
<td>49.5</td>
</tr>
<tr>
<td>Racial/ethnic identity-black</td>
<td>802</td>
<td>26.7</td>
</tr>
<tr>
<td>Racial/ethnic identity-Hispanic</td>
<td>598</td>
<td>19.9</td>
</tr>
<tr>
<td>Racial/ethnic identity-Asian</td>
<td>58</td>
<td>1.9</td>
</tr>
<tr>
<td>Racial/ethnic identity-American Indian/Islander</td>
<td>12</td>
<td>0.4</td>
</tr>
<tr>
<td>Racial/ethnic identity-other</td>
<td>42</td>
<td>1.4</td>
</tr>
<tr>
<td>Limited English proficient</td>
<td>381</td>
<td>12.7</td>
</tr>
<tr>
<td>Free/reduced lunch</td>
<td>1,184</td>
<td>39.5</td>
</tr>
<tr>
<td>Exceptional student education</td>
<td>654</td>
<td>21.8</td>
</tr>
<tr>
<td>Earned diploma-overall</td>
<td>1,279</td>
<td>42.6</td>
</tr>
<tr>
<td>Earned diploma-w/o retained 9th graders in 1999</td>
<td>1,220</td>
<td>47.0</td>
</tr>
<tr>
<td>Earned diploma- retained in 1998 9th graders</td>
<td>59</td>
<td>14.0</td>
</tr>
<tr>
<td>Earned diploma-adjusted BCR</td>
<td>1,220</td>
<td>51.0</td>
</tr>
<tr>
<td>Earned diploma-white students unadjusted</td>
<td>720</td>
<td>48.4</td>
</tr>
<tr>
<td>Earned diploma-black students unadjusted</td>
<td>258</td>
<td>32.2</td>
</tr>
<tr>
<td>Earned diploma-Hispanic students unadjusted</td>
<td>248</td>
<td>41.5</td>
</tr>
<tr>
<td>Earned diploma-Asian students unadjusted</td>
<td>30</td>
<td>51.7</td>
</tr>
<tr>
<td>Earned diploma-Am. Indian/Pac. Islander unadjusted</td>
<td>3</td>
<td>25.0</td>
</tr>
</tbody>
</table>

The BCR of the 404 students who had been retained in the ninth grade before the 1999-2000 school year was much lower at 14%. This group was not included in the regression analysis for Hypothesis 3, because they did not have scores for the 9th grade FCAT (the 8th grade FCAT was given for the first time in 1998). Of the 62 students from
this group who took the 10th grade mathematics FCAT, 45.5% received a failing grade of 1 on the Mathematics FCAT. Of the 55 students who took the 10th grade reading FCAT, 74.19% received a failing grade of 1 on the Reading FCAT. Table 12 shows the racial/ethnic composition of the retained 9th graders to be disproportionately black and Hispanic, with these groups graduating at a lower rate.

Table 12
Graduation and Racial/Ethnic Identity Subgroups of 1998 Retained 9th Graders

<table>
<thead>
<tr>
<th></th>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
<th>Am. Indian</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did not Graduate</td>
<td>128</td>
<td>128</td>
<td>78</td>
<td>5</td>
<td>2</td>
<td>4</td>
<td>345</td>
</tr>
<tr>
<td>Graduated</td>
<td>27</td>
<td>19</td>
<td>12</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>59</td>
</tr>
<tr>
<td>Total</td>
<td>155</td>
<td>147</td>
<td>90</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>404</td>
</tr>
</tbody>
</table>

N = 404

In addition, many other students did not have a complete set of FCAT scores. For example, Table 13 shows that many students never completed the 10th grade FCAT in mathematics (all four FCAT variables had similar missing data percentages). The discriminant analysis that included the high stakes test scores, FCAT, consisted of a much smaller sample (N=1,156), due to many students being eliminated because they did not have a complete set of four FCAT scores (8th grade mathematics, 8th grade reading, 10th grade mathematics, and 10th grade reading). In addition, the group of students without complete FCAT scores was disproportionately black, male, ESE, and poor (see Table 13).
Reasons for not having complete FCAT records might include poor school or state record keeping or students not taking the test. However, the percentage of students with no FCAT records was disproportionately black and American Indian/Islander. For the 10th grade FCAT, 59% of black students had missing records, and 75% of American Indian/Islander had missing records. White students had much higher test taking percentages with only 39% having missing records, as well as Asian students with only 33% having missing records, and 42% of Hispanic students having missing records.

More students with missing records were 61% male versus 39% female. Disproportionate numbers of students with missing records were poor (70.7%) and exceptional student education (30.8%). Low percentages of available data on the FCAT are a limitation of this study. Although low percentages of available FCAT scores represent a limitation of the study, this may be caused by lack of participation in the testing. Such low participation may be an important finding, especially for subgroups.
For the 8th grade FCAT reading, 16% of those received a failing score of 1 and 42.2% did not have a score. For the 8th FCAT mathematics, 17.8% failed with a score of 1 and 42.4% had no score. These students were all promoted to the 9th grade, but the mechanism of their promotion is not known. The 10th grade FCAT reading was required for graduation in 2002-2003, and 16.8% failed with a score of 1 and 44.9% had no score. For mathematics in the 10th grade, 10.4% received a 1 and 44.9% had no score. Thus, the majority of students from the original random sample of students either did not take the FCAT or failed it for each of the four tests (See Table 14).

Table 14
FCAT Score Frequencies and Percentages of 3,000 Florida 9th Grade Students

<table>
<thead>
<tr>
<th>FCAT Level</th>
<th>FCAT Reading 8th Grade</th>
<th>Math 8th Grade</th>
<th>FCAT Reading 10th Grade</th>
<th>Math 10th Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>480 (16%)</td>
<td>536 (17.8%)</td>
<td>501 (6.8%)</td>
<td>312 (10.4%)</td>
</tr>
<tr>
<td>2</td>
<td>483 (16.1%)</td>
<td>378 (12.6%)</td>
<td>548 (18.2%)</td>
<td>347 (22%)</td>
</tr>
<tr>
<td>3</td>
<td>540 (18%)</td>
<td>503 (16.7%)</td>
<td>328 (10.9%)</td>
<td>328 (10.9%)</td>
</tr>
<tr>
<td>4</td>
<td>213 (7.1%)</td>
<td>191 (6.4%)</td>
<td>122 (4.1%)</td>
<td>405 (14.5%)</td>
</tr>
<tr>
<td>5</td>
<td>22 (.7%)</td>
<td>126 (4.2%)</td>
<td>156 (5.2%)</td>
<td>158 (5.3%)</td>
</tr>
<tr>
<td>Missing</td>
<td>1,269 (42.2%)</td>
<td>1,273 (42.4%)</td>
<td>1,350 (45%)</td>
<td>1,349 (45%)</td>
</tr>
</tbody>
</table>

Measures of central tendency for students' FCAT scores were between 2 and 3, which are passing scores near the center of a positively skewed curve, except for 10th
grade mathematics, which was slightly negatively skewed (see Table 15). However, this
does not account for the 42.2% to 45% students who did not take the FCAT. Thus, the
validity of the measures of central tendency produced from analyzing FCAT scores in
this sample is questionable.

Table 15
Measures of Central Tendency for FCAT Scores

<table>
<thead>
<tr>
<th></th>
<th>Reading</th>
<th>Math 8th</th>
<th>Reading 10</th>
<th>Math 10th</th>
</tr>
</thead>
<tbody>
<tr>
<td>8th</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N Valid</td>
<td>1,738</td>
<td>1,734</td>
<td>1,657</td>
<td>1,658</td>
</tr>
<tr>
<td>N Missing</td>
<td>1,269</td>
<td>1,273</td>
<td>1,350</td>
<td>1,349</td>
</tr>
<tr>
<td>Mean</td>
<td>2.32</td>
<td>2.42</td>
<td>2.31</td>
<td>2.86</td>
</tr>
<tr>
<td>Median</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Skewness</td>
<td>.252</td>
<td>.451</td>
<td>.796</td>
<td>-.023</td>
</tr>
<tr>
<td></td>
<td>.059</td>
<td>.059</td>
<td>.060</td>
<td>.060</td>
</tr>
</tbody>
</table>

N=3,000

Testing of Hypothesis 3

In order to test H03, discriminant analysis was conducted using the random sample of 3,000 high school students in the 9th grade in Florida public schools in 1999/2000, of which 1,152 had no missing independent variables (see Table 16).

Table 16
Discriminant Analysis Case Processing Summary for H03

<table>
<thead>
<tr>
<th>Unweighted Cases</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>1152</td>
<td>38.4</td>
</tr>
<tr>
<td>Excluded – Missing at least one variable</td>
<td>1848</td>
<td>61.6</td>
</tr>
<tr>
<td>Total</td>
<td>3000</td>
<td>100.0</td>
</tr>
</tbody>
</table>
The discriminant analysis for HO₃ yielded a Wilks' Lambda coefficient of .777, \( p < .001 \); therefore, 22.3% of the variation in graduation success was attributed to the set of independent variables (see Table 17). The result was statistically significant (\( p < .001 \)). Independent variables were racial/ethnic identity, gender, exceptional student education status, limited English proficient status, free or reduced lunch status, and four separate FCAT scores for reading and math in the 8th and 10th grades. The dependent variable was graduation success in four years from high school with a standard diploma (0 = did not graduate; 1 = graduated).

Table 17
Wilks' Lambda for HO₃

<table>
<thead>
<tr>
<th>Function(s)</th>
<th>Wilks' Lambda</th>
<th>Chi-Square</th>
<th>Df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.777</td>
<td>288.548</td>
<td>12</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

The discriminant analysis indicated that GPA was the greatest predictor in the discriminant function with a moderately strong structure coefficient of -.756 (see Table 18). Total retentions were the second most influential predictive variable with a moderately strong structure coefficient of .679. The other variables that were moderately strong predictors were FCAT math in the 10th grade, absences in the year 2000, FCAT reading in the 10th grade, FCAT reading in the 8th grade, and FCAT math in the 8th grade. All of the independent variables that were moderately significant with a structure coefficient of more than \(|.300| \) were academic achievement variables, except for the absences from the year 2000. Interestingly, the demographic variables of free/reduced
lunch, exceptional student education, racial/ethnic identity, limited English proficiency, and gender were not strong predictors of graduation success in the micro study. This indicates that a primary concern from the micro perspective is achievement variables, particularly grades and promotion. Other essential components of student success from the micro perspective are passing the FCAT and school attendance. These findings support the position that school achievement variables are the determinants of high school graduation for individuals, rather than demographic variables.

Table 18
Coefficients and Structure Matrix \( HO_3 \)

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Standardized Canonical Function Coefficients</th>
<th>Structure Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>-.566</td>
<td>-.756</td>
</tr>
<tr>
<td>Total retentions</td>
<td>.486</td>
<td>.679</td>
</tr>
<tr>
<td>FCAT math 10(^{th}) grade</td>
<td>-.386</td>
<td>-.479</td>
</tr>
<tr>
<td>Absences 2000</td>
<td>.273</td>
<td>.449</td>
</tr>
<tr>
<td>FCAT reading 10(^{th}) grade</td>
<td>-.225</td>
<td>-.436</td>
</tr>
<tr>
<td>FCAT reading 8(^{th}) grade</td>
<td>-.103</td>
<td>-.372</td>
</tr>
<tr>
<td>FCAT math 8(^{th}) grade</td>
<td>.652</td>
<td>-.300</td>
</tr>
<tr>
<td>Free reduced lunch</td>
<td>.058</td>
<td>-.196</td>
</tr>
<tr>
<td>ESE</td>
<td>-.019</td>
<td>-.123</td>
</tr>
<tr>
<td>Racial/white</td>
<td>.041</td>
<td>-.141</td>
</tr>
<tr>
<td>Limited English</td>
<td>-.004</td>
<td>-.057</td>
</tr>
<tr>
<td>Proficient</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.089</td>
<td>.037</td>
</tr>
</tbody>
</table>

Note: Pooled within-group correlations between discriminating variables and standardized canonical discriminant functions. Variables ordered by absolute size of correlation with function.

Cross Validation of \( HO_3 \)

Cross validation (Osborne, 2000) was performed to test the confidence that should be placed in the results of the discriminant analysis by dividing the original sample of 3,000 students into three random samples of 1,000 students each (see Table 19). Results showed that the three sub-samples had Wilks' Lambda within 5% or less of
the original Wilks' Lambda (see Table 19). The structure coefficients of the three sub-
samples indicated the achievement variables, especially GPA, total retentions,
attendance, FCAT reading in the 10th grade, and FCAT math in the 10th grade, to be
essential to graduation success.

Table 19
Cross validation of H03.*

<table>
<thead>
<tr>
<th></th>
<th>Original N = 1152</th>
<th>Random Sample 1 N = 381</th>
<th>Random Sample 2 N = 386</th>
<th>Random Sample 3 N = 385</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks' Lambda</td>
<td>.777</td>
<td>.758</td>
<td>.789</td>
<td>.724</td>
</tr>
<tr>
<td>Structure Coefficient</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td>-.756</td>
<td>.746</td>
<td>-.605</td>
<td>-.801</td>
</tr>
<tr>
<td>Total retention</td>
<td>.679</td>
<td>-.687</td>
<td>.643</td>
<td>.603</td>
</tr>
<tr>
<td>FCAT Math 10th gr.</td>
<td>-.479</td>
<td>.558</td>
<td>-.389</td>
<td>-.415</td>
</tr>
<tr>
<td>Absences 2000</td>
<td>.449</td>
<td>-.387</td>
<td>.492</td>
<td>.407</td>
</tr>
<tr>
<td>FCAT Read 10th gr.</td>
<td>-.436</td>
<td>.527</td>
<td>-.448</td>
<td>-.273</td>
</tr>
<tr>
<td>FCAT Math 8th gr.</td>
<td>.300</td>
<td>.423</td>
<td>-.224</td>
<td>-.213</td>
</tr>
<tr>
<td>FCAT Read 8th gr.</td>
<td>-.372</td>
<td>.527</td>
<td>-.345</td>
<td>-.200</td>
</tr>
<tr>
<td>Free/reduced lunch</td>
<td>-.196</td>
<td>.394</td>
<td>.019</td>
<td>-.177</td>
</tr>
<tr>
<td>ESE</td>
<td>-.123</td>
<td>.093</td>
<td>-.178</td>
<td>-.089</td>
</tr>
<tr>
<td>Identity/white</td>
<td>-.141</td>
<td>.177</td>
<td>-.099</td>
<td>-.121</td>
</tr>
<tr>
<td>LEP</td>
<td>-.057</td>
<td>.061</td>
<td>-.158</td>
<td>.034</td>
</tr>
<tr>
<td>Gender</td>
<td>-.037</td>
<td>.045</td>
<td>-.015</td>
<td>.154</td>
</tr>
</tbody>
</table>

* Note: Structure coefficients for Random Sample 1 were "reflected" (i.e., had opposite
signs from those in the original, second, and third random samples.

H3 stated that the independent variables high stakes test scores, history of
retention, attendance, grades, racial group membership, socio-economic status, limited
English proficiency, gender, and special education status can predict the dependent
variable high school graduation success. The null hypothesis, HO3, was rejected because
a significant portion, 22.3%, p < .001, of the variance in the dependent variable,
graduation rate was predicted by the set of independent variables.
Testing of Hypothesis 4

Discriminant analysis was conducted in order to obtain a clearer picture of the relative importance of the independent demographic and achievement variables of racial/ethnic identity/ethnic background, gender, exceptional student education status, limited English proficient status, free or reduced lunch status, grade point average (GPA), total retentions, and absences in 2000 on the dependent variable, graduation success in four years from high school with a standard diploma (see Table 20). No FCAT scores were included with the goal of discerning differences between the discriminant model that included the FCAT and the model without the FCAT.

Out of 3,000 students in the original sample, 2,356 had no missing data and thus were included in the analysis of H04 (see Table 20).

Table 20
Discriminant Analysis Case Processing Summary for H04

<table>
<thead>
<tr>
<th>Unweighted Cases</th>
<th>N</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>2356</td>
<td>78.5</td>
</tr>
<tr>
<td>Excluded – Missing at least one</td>
<td>644</td>
<td>21.5</td>
</tr>
<tr>
<td>variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3000</td>
<td>100.0</td>
</tr>
</tbody>
</table>

The Wilks’ Lambda coefficient was .649, p < .001; therefore 35.1% of the variation in graduation success was explained by the set of independent variables (see Table 21).

Table 21
Wilks Lambda for H04

<table>
<thead>
<tr>
<th>Test of Function(s)</th>
<th>Wilks’ Lambda</th>
<th>Chi-Square</th>
<th>df</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.649</td>
<td>1015.119</td>
<td>8</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
The lower amount of explained variation found in the analysis that included FCAT scores (22.3%) versus the analysis without FCAT scores (35%) may be explained by the much smaller sample of students in $H_0^3$ (1,156 v. 2,356) that was no longer completely random. That the smaller sample of $H_0^3$ was no longer random is evidenced by the disproportionate numbers of male, black students, ESE, and poor excluded from the analysis due to missing FCAT scores.

The structure matrix indicated that the most important predictor was GPA, which was strongly related to graduation success, with a structure coefficient of .897 (see Table 22). The achievement variable, total retentions, was also a strong predictor with a structure matrix coefficient of -.616. Absences in 2000 were a moderate predictor with a structure coefficient of -.365. The demographic variable, free/reduced lunch, was a moderate predictor with a structure coefficient of -.337.

Table 22  
**Standardized Canonical Function Coefficients and Structure Matrix for $H_0^4$**

<table>
<thead>
<tr>
<th>Independent variable</th>
<th>Standardized Canonical Function Coefficients</th>
<th>Structure Matrix</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>.739</td>
<td>.897</td>
</tr>
<tr>
<td>Total retentions</td>
<td>-.367</td>
<td>-.616</td>
</tr>
<tr>
<td>Absences 2000</td>
<td>-.100</td>
<td>-.365</td>
</tr>
<tr>
<td>Free reduced lunch</td>
<td>.096</td>
<td>.337</td>
</tr>
<tr>
<td>ESE</td>
<td>.187</td>
<td>.258</td>
</tr>
<tr>
<td>Racial/white</td>
<td>-.049</td>
<td>.189</td>
</tr>
<tr>
<td>Gender</td>
<td>.046</td>
<td>-.138</td>
</tr>
<tr>
<td>LEP</td>
<td>.084</td>
<td>.107</td>
</tr>
</tbody>
</table>

Note: Pooled within-group correlations between discriminating variables and standardized canonical discriminant functions. Variables ordered by absolute size of correlation with function.

$H_0^4$ Cross Validation

Three quasi-random samples of 1,000 students each was taken from the original sample of 3,000 students in order to cross validate the original analysis (Osborne, 2000).
The group of independent variables without FCAT scores and the dependent variable graduation success underwent discriminant analysis. Results showed that all three quasi-random samples had less than 2% difference in Wilks’ Lambda from the original sample (see Table 23). In addition, the strongest predictors in the four samples were GPA and total retentions, which are both achievement variables. Another noteworthy predictor was absences in the year 2000 (10th grade). In the original sample and one quasi sample, free/reduced lunch was a noteworthy predictor. Similarly to the original sample, demographic variables of exceptional student education, racial/ethnic category, gender, and limited English proficiency were not noteworthy predictors indicating that in the mechanics of high school graduation success, demographics of racial/ethnic identity, special education, and limited English proficiency were not very important.

Table 23
Cross validation of H04

<table>
<thead>
<tr>
<th></th>
<th>Original N=2356</th>
<th>Random Sample 1 N=786</th>
<th>Random Sample 2 N=803</th>
<th>Random Sample 3 N=766</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilks’ Lambda Structure Coefficients</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPA</td>
<td>.897</td>
<td>.870</td>
<td>.875</td>
<td>.923</td>
</tr>
<tr>
<td>Total retention</td>
<td>-.616</td>
<td>-.607</td>
<td>-.627</td>
<td>-.596</td>
</tr>
<tr>
<td>Absences 2000</td>
<td>-.366</td>
<td>-.410</td>
<td>-.346</td>
<td>-.348</td>
</tr>
<tr>
<td>Free/reduced Lunch</td>
<td>.337</td>
<td>.469</td>
<td>.238</td>
<td>.293</td>
</tr>
<tr>
<td>ESE</td>
<td>.258</td>
<td>.258</td>
<td>.290</td>
<td>.218</td>
</tr>
<tr>
<td>Racial/white</td>
<td>.189</td>
<td>.194</td>
<td>.134</td>
<td>.230</td>
</tr>
<tr>
<td>LEP</td>
<td>.107</td>
<td>.115</td>
<td>.132</td>
<td>.076</td>
</tr>
<tr>
<td>Gender</td>
<td>-.138</td>
<td>-.118</td>
<td>-.193</td>
<td>-.100</td>
</tr>
</tbody>
</table>

The null hypothesis, H04, stated that no variance in the dependent variable, high school graduation success will be accounted for by the set of the independent variables,
The null hypothesis, $H_0$, was rejected because a noteworthy part of the variance in high school graduation success, 35.1%, was accounted for by the set of independent variables (history of retention, attendance, grades, low socio-economic status, limited English proficiency, special education status, gender, and racial group membership).

**Chapter Summary**

All four of the hypotheses tested resulted in rejecting the null hypotheses. Thus, results of the macro analysis of graduation rates over 28 years indicated that graduation rates were significantly related to the high stakes testing policy periods from 1977 to 2005. Over the 28 years, the graduation rate was negatively correlated to the year indicating a reduction of graduation success over the time span. Graduation rates fell 4.51 percentage points over 28 years. In addition, specific periods of testing with more stringent policies were characterized by lower graduation rates and other periods of testing with loosened policies or increased accountability for schools were characterized by higher rates.

In addition, black and Hispanic students suffered a more profound and negative loss of graduation rate with the advance of 28 years of seven different periods of high stakes testing. The CPI for black students fell from 56.61 in 1981 to 41.17 in 2004 (range of -15.44). The CPI for Hispanic students fell from 60.32 in 1981 to 53.69 in 2004 (range of -6.63). In contrast, the white student CPI rose from 66.15 in 1981 to 72.21 in 2004 (range of 6.06).
In terms of the achievement gap for graduation success, the gap in 1981 was 9.54 percentage points between black and white students and 11.89 percentage points between Hispanic and white students. The achievement gap worsened to 31.04 percentage point gap between black and white students in 2004 and 18.52 percentage point gap in 2003 between Hispanic students and white students. Although the achievement gap narrowed in 2004-2005 to be about 20 percentage points between black students and white students and 6 percentage points between Hispanic students and white students, a worsening trend in the achievement gap over the 28 year time period, especially in the most recent 12 years is of concern to educational leaders.

Changes in the demographic variables of percentage white identity, ESE, and LEP were predictive of changes in graduation rates in the macro analysis. However, on an individual student level, all of the important predictors, such as GPA, retention, FCAT scores, and attendance, focused on student achievement. The free/reduced lunch variable was an exception, being a moderate predictor for graduation success when the FCAT scores were not included in the analysis. However, this does not take into account the over 50% test mortality for the micro analysis. When analyzed separately, the group of student not included in the analysis because they did not have FCAT scores was disproportionately poor, black, and exceptional student education. This is a hidden bias against poor, black, and exceptional student education students that did not show up in the discriminant analysis.
CHAPTER 5
SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

SUMMARY

Of concern in the present study was whether high stakes testing policies in Florida from 1978-2005 were related to graduation rates and whether changes in the policies could predict changes in the graduation rate. Whether graduation rates were rising (or falling) and the status and patterns of graduation rates for different racial/ethnic groups was of interest. Results from this study, using the cumulative promotion index (CPI), were that graduation rates declined 4.51 percentage points from 62.37 in 1978 to 57.86 in 2005.

The progression of graduation rates was not chronologically linear over the years (unlike the demographics for the percentage of white students, limited English proficiency, and exceptional student education). Analysis of the changes in graduation rates over 28 years supports the hypothesis that the high stakes testing policies, demographic, and graduation rates were related. High stakes testing policies were predictive of a significant part of the variation in graduation rates.

Interesting relationships were found between students' graduation success and students' FCAT scores, history of retention in grade, attendance, grades, membership in racial/ethnic groups, gender, socio-economic class, limited English proficiency, and exceptional student education. The discriminant analysis in the micro analysis clearly
demonstrated that the achievement variables of GPA, retention history, FCAT scores, and attendance were powerful predictors of graduation success for individual students. Thus, on the individual student level, getting good grades and achieving in school were the most important considerations for students in planning for graduation success. A major finding of this study is that students who do not graduate may be described as failing students on all achievement indicators (GPA, history of retention, FCAT scores, attendance).

Demographics were not good predictors for graduation success in the micro analysis using discriminant analysis. The exception was the low socio-economic variable, which was a moderately good predictor in some tests. However, conclusions drawn from this result may be questionable, due to over 50% test mortality from the analysis. The mortality was primarily due to missing FCAT scores for students who may have avoided the test because of fear of failure or whose scores were missing because of poor record keeping. Of this group, descriptive statistics indicated that the poor, black, and exceptional student education students were disproportionally represented in the mortality group.

Discussion

*Florida Graduation Rates*

Overall graduation rates were lowest during the first FCAT period from 1998 to 2002, with an average CPI of 49.39. A large rise in the graduation rate in 2003 to a CPI of 58.99 might be attributed to several factors: (a) additional accountability spurred efforts by the school systems to maximize student graduation to comply with the NCLB federal law requiring graduation rates to be included in school accountability in 2003; (b)
public attention to Florida’s low graduation rates by researchers, such as Greene and Winters (2002), caused schools and the Florida DOE to help students graduate and count every graduate to a greater extent than in the previous years; (c) students improved their academic performance resulting in higher graduation rates; and (d) students became accustomed to the FCAT resulting in more students being able to perform well and graduate. Research by Linn, Graue, and Sanders (1991) found that test scores often improve after an initial drop, a pattern that occurred in this case. Results of this study did not evidence improved learning and higher academic achievement or student preparation for testing as a basis for improving graduation rates. In the micro study, over half the students either did not take the FCAT or did not pass.

Inconsistencies in graduation rates for 2003 were found in the following four studies: the macro analysis, the micro analysis, the state by state study of graduation rates by Orfield et al. (2004), and the state of Florida official graduation rate (Florida DOE, 2004b). The graduation rate found in the macro analysis was 61.61%, the micro analysis was 51%, the Orfield et al. (2004) study was 53%, and the state of Florida official rate was 69%.

Explanation for the large differences between the macro and micro analyses includes different data sources for the two studies, with the macro study using state of Florida official published enrollment and graduation numbers and the micro study using unpublished, raw data obtained from the Florida Department of Education in the form of a random sample of 3,000 Florida high school 9th graders. Different formulas were used for the macro (Cumulative Promotion Index - CPI) and micro (adjusted Basic Completion Rate - BCR) studies.
Interestingly, both the micro analysis of the present investigation using the random sample data and the Orfield et al. (2004) study had almost the same overall graduation rates of 51% and 53% respectively, and similar graduation rates for subgroups. The Orfield et al. (2004) study resulted in a black student rate for 2003 of 41% and a white student rate of 55%. The micro analysis resulted in a black student rate of 42% and a white student rate of 58%.

Florida reported a standard diploma graduation rate for 2003 of 69%, which was greater than the 51% for the micro study or 61.61% for the macro study. One source of discrepancy might be the elimination of students with transfer classifications from the graduation rate calculations by the state of Florida formula for graduation rate (Florida DOE, 2002b). Transfers out of school are treated as students who migrated out of state, and these students are not included in the state of Florida graduation rate calculations. The analysis of the present study did not eliminate students with transfer classifications.

Although the state of Florida longitudinally tracks students after transferring to ascertain their status (Florida DOE, 2002b), the longitudinal system supporting the classifications was not included in the data used in this study. In some cases, students from the random sample with the transfer classification had more than one transfer on record and in other cases, students with transfers on record graduated from a Florida high school. Thus, elimination of all students with the transfer classification(s) would have been inappropriate, because it was impossible to tell which students actually left the state and which left the school they were attending but stayed in Florida.

In this study, instead of eliminating transfers, in order to approximate the effect of out migration from Florida, a rate of 7.38% out migration for the years 1999 to 2003
from Florida for ages 14-18 was derived from the U. S. Census (Perry, 2003). The original BCR of 42.6 was adjusted to include only first time ninth graders and for the out migration of 7.38%, resulting in an adjusted BCR of 51%.

For the macro study, no adjustments were used for migration. However, the CPI was used, partially because of the two year time window that minimized the effect of migration (see Appendix B). The 61.61% CPI obtained in the macro study was lower than the State of Florida formula rate of 69%. Two sources for the discrepancies could be the different formulas used or the different databases used, with the macro study using Florida Department of Education published data and the Florida Department of Education using unpublished raw data.

Demographics, High Stakes Testing, and Graduation Success

The demographic profile demonstrated over the 28 year period consistently increasing percentages of limited English proficiency and exceptional student education and consistently decreasing percentages of white students; however, graduation rates over the periods of high stakes testing were not consistently rising or falling chronologically. The 28 year demographic analysis demonstrated the following moderately strong correlations: (a) the percentages of students who were white and graduation rates (.672), (b) the percentages of students with exceptional student education and graduation rates (-.691), (c) the percentages of students with limited English proficiency and graduation rates (-.762). High stakes testing policies were moderately correlated with graduation rates.

Peaks and low points were apparent upon examining the graph of 28 years of graduation rates (see Figure 3). The periods of increased graduation rates corresponded to
relaxation of the stringent requirements, increases in accountability for graduation rates, or students growing accustomed to a particular test. This occurred in 1980 during the court case of Debra P. vs. Turlington (CPI of 65.8), in 1991 (CPI of 64.07) when the retention policies were relaxed (Morris, 2001), and in 2003 (CPI of 61.61) when added accountability of the NCLB was used at the same time the FCAT had been used for several years. Decreased graduation rates corresponded to the following periods of stringent policies and new tests: during the SSAT-11 in 1979 (CPI of 60.02), the revised SSAT-11 in 1988 (CPI of 55.41), the HSCT in 1993 (CPI of 59.43), and the FCAT in 2000 (CPI of 47.66). Apparently, changes in graduation rates can be predictably influenced by changes in high stakes testing policies.

Discussion of Findings for Hypothesis 1

Hypothesis 1 stated that a noteworthy portion of the variance would be accounted for by the set of the independent variables {racial/ethnic identity, exceptional student education, and limited English proficiency}. When these demographics in the form of percentages of the total school population were analyzed in the linear regression, racial/ethnic identity (white), exceptional student education, and limited English proficiency accounted for 53.2% of the variation in graduation rates over the 28 year periods in the macro study. This result was statistically significant ($p < .001$) and indicated that changes in demographics could be used to predict changes in graduation rates.

As expected, increased percentages of exceptional student education and limited English proficient students corresponded to lower graduation rates. Research has shown that exceptional student education placements usually have negative consequences for
graduation (Gopaul-McNicol & Armour-Thomas, 2002; Heubert, 2002; Jones et al.,
2003). Limited English proficient students face challenges in an English orientated test
and curricula environment that can reduce graduation rates (Valenzuela, 2002). As the
percentage of white students was lowered over the 28 years, the graduation rate was
lowered. Interestingly, higher percentages of white students corresponded to higher
graduation rates for black students. Higher graduation rates for black students attending
schools with large percentages of white students were found than when black students
attended segregated schools, according to studies by Borman et al., (2004) and Orfield
(2001). In particular, Borman et al. (2004) found that increased racial segregation was
related to lower achievement for black students in Florida.

Discussion of Findings for Hypothesis 2

Hypothesis 2 stated that the non-standardized residualized dependent variable
scores, obtained by adjusting the variance by the set of the independent demographic
variables, would be related to the time periods when different high stakes testing policies
were used. Seven different time periods were used: (a) 1978, during the basic skills test,
SSAT-1, but before any graduation test; (b) 1979, the functional literacy test, SSAT-11
was required for graduation; (c) 1980 - 1982, during the suspension of withholding high
school diplomas due to the Debra P. v. Turlington (1979) court case; (d) 1983 – 1990,
implementation of a revised Secondary School Admission Test, SSAT-11 with increased
standards based rigor, mandatory retention, and graduation consequences; (e) 1991 -
1997, during the High School Competency Test (HSCT) of the 1990s; (f) 1998 – 2002,
during the FCAT testing from 1998 to 2002; and (g) 2003-2004 during the FCAT testing
and the inclusion of AYP for school grades.
The null hypothesis $H_0$ was rejected; thus, this study found it unlikely that the time periods for different high stakes testing policies in Florida were not responsible for a significant portion of the variance in graduation rates. The amount of variance was 49% of the residual that could be attributed to different test periods, which is about 25% of the total variance in graduation rates over 28 years. These results clearly indicate a strong relationship between graduation rates and high stakes testing policies.

**Discussion of Findings for Hypothesis 3**

Hypothesis 3 stated that the independent variables high stakes test scores, history of retention, attendance, grades, racial group membership (white), socio-economic status, limited English proficiency, gender, and exceptional student education status could predict the dependent variable high school graduation success. Confirming Hypothesis 3, a statistically significant ($p < .001$) amount of the variance (22.3%) was accounted for by the discriminant model.

On an individual student level, achievement variables of GPA (grades), history of retention, FCAT scores, and attendance, rather than classification of individual students into demographic groups of racial/ethnic, gender, exceptional student education, or limited English proficiency were the significant predictors of graduation success. The poverty variable, free/reduced lunch, was a moderate predictor in one test. GPA and total retentions were the most important predictors. FCAT scores, although important, were not as strong as GPA or retentions. These results were expected with respect to the achievement variables being good predictors. However, the demographics were expected to be good predictors and were not. The study focused on the individual student; therefore, measurable influences of demographics may have been difficult to capture.
statistically. Another explanation is that lack of participation or school record keeping of groups most negatively affected by demographics eliminated members disproportionately from the analysis.

The cross validation of the three random samples indicated the data used to test HO3 was consistent and of acceptable quality. This indicates that more confidence in the generalizability in the results of the statistical testing is warranted. Even though the study had the limitation of lacking FCAT records for many students in the sample, the results are of acceptable quality and the study should be replicated in other years that may have increased percentages of FCAT records in the sample.

Discussion of Findings for Hypothesis 4

Hypothesis 4 stated that a noteworthy portion of the variance in the independent variable, high school graduation success would be accounted for by the set of independent variables, \{history of retention, attendance, grades, low socio-economic status, limited English proficiency, exceptional student education status, gender and racial group membership\}. The null hypothesis was rejected, because 35.1% of the variance was attributed to these variables. The result was statistically significant \(p < .001\).

The results showed that more variance was attributed to the discriminant model in HO4 than HO3, perhaps due to the more intact sample used in HO4. About half of the random sample of 3,000 was lost due to missing FCAT scores in HO3, which weakened the strength of the discriminant analysis. Similarly to HO2, the achievement variables of GPA and history of retention were the strongest predictors of graduation success, with attendance a significant predictor also. The low socio-economic indicator, free/reduced
lunch was a moderate predictor. The other demographic factors of white identity, limited English proficiency, exceptional student education, and gender were not good predictors, which was not the expected result. The cultural deficit perspective for low achieving students, which states that cultural, personal, or social lacking of diverse populations are to blame for low achievement, was not supported by this results. The exception was the low socio-economic indicator; however, this was the weakest of the significant predictors in the analysis.

The cross validation of the three random samples indicated the data used to test \( H_0 \) was consistent and of acceptable quality. This allows more confidence in the generalizability to be placed in the results of the statistical testing.

**Discussion of Demographic Interrelationships**

Although the white identity indicator was not a significant predictor of graduation success in the testing of micro analysis, a strong relationship between racial/ethnic identity and graduation success (similarly to the macro analysis) was verified in the descriptive statistics for the micro analysis. The following wide disparities were noted in the unadjusted racial/ethnic percentages that graduated: white students had 48.4%, black students had 32.2%, Hispanic students had 41.5%, Asian students had 51.7%, and American Indian/Pacific Islander had 25%.

Demographic factors were studied for influence on graduation rates in the macro study. The demographics of racial/ethnic identity (percent white), exceptional student education, and limited English proficiency accounted for a significant 53.2% of the variation in graduation rates, \( F(3, 27) = 9.110, p < .001 \). A variable for poverty or economic class, such as the free/reduced lunch program, could not be included in the
analysis because of the inconsistency of the variable measures over the 28 year period. However, research by Stranahan and Borg (2005a) indicated that poverty, if measured validly and reliably, may have accounted for an additional significant part of the variation in graduation rates. In addition, the analysis of demographics in the micro study indicated that poverty was a significant predictor of lack of graduation success.

The racial/ethnic identity (white), exceptional student education, and limited English proficiency influences on graduation; however, were not significant in the micro analysis testing. The free/reduced lunch variable was moderately predictive. However, the short time period of four years may not have resulted in measurable effects of the many years of disadvantages in schooling for poor and racial/ethnic or linguistic minority students. On an individual student basis, the best predictors for graduation success were the success variables of grade point average, promotion in grade (lack of retention), FCAT scores, and good school attendance. Historically, students without poverty or linguistic minority characteristics in schools experience these success factors in schools in the United States. Those without such disadvantaging characteristics also participate more in standardized testing.

The micro analysis results must be questioned because of the more than 50% test mortality that excluded disproportionate numbers of poor, black, and exceptional student education from the analysis. The mortality was primarily because of missing FCAT scores due to poor record keeping by schools or students who may have avoided the test or were withheld from it by their families. For the 10th grade FCAT, 59% of black students had missing records, and 75% of American Indian/Islander had missing records. This pattern indicates a lack of connectedness between the schools and the needs of the
students. White students had much higher test taking percentages with only 39% having missing records, as well as Asian students with only 33% having missing records, and 42% of Hispanic students having missing records. More students with missing records were male (61%) versus female (39%). Disproportionate numbers of students with missing records were poor (70.7%) and exceptional student educated (30.8%).

Problems were found in the racial disparity in the key achievement indicator of graduation rate between black students, Hispanic students, and white students in the macro study. White students experienced success through graduation at a disproportional higher rate than the average student, especially poor and/or black students. Graduation rates from 1978 to 1981, from 1993 to 1994, and from 1998 to 2004 were examined for these groups. Early year disparities between black students and white students averaged 11.87 percentage points. Disparities between Hispanic students and white students averaged 1.1 percentage points. In 1993, disparities between black students and white students were 9 percentage points and Hispanic students had a higher graduation rate than white students by over 1 percentage point.

During the two FCAT periods from 1998 to 2002 and 2003 to 2005, the racial/ethnic disparities between black students and white students increased steadily. An increased disparity of 200% from the 1993 disparity of 9 percentage points to a disparity of 28 percentage points was reached in 2003. The white student graduation rate increased to 72.21% with the black student rate at 44.17%. Hispanic students lost almost 20 percentage points to white students in this period. The disparities were reduced to 20 points (black/white) and 6 points (Hispanic/white) in 2004 and 2005.
One possible explanation for the increasing disparities in racial/ethnic graduation rate during the FCAT period includes greater increases in poverty in black and Hispanic students than poverty in white students. However, according to the United States Census Bureau (DeNavas-Walt, Proctor, & Lee, 2005), no significant change in the national racial/ethnic disparities in poverty occurred from 2003 to 2004.

Another possible explanation is that the reported trend towards re-segregation in Florida schools (Borman et al., 2004) reversed some progress achieved from the court ordered desegregation eras of the 1970's and 1980's (Orfield, 2002). Borman et al. and Orfield found that more positive academic outcomes, including high stakes test scores and additional opportunities for black and Hispanic students occurred at desegregated schools than at schools with a minority populations over 50%.

A third explanation for the large rise in the racial disparity could be that a cultural bias within the FCAT administration or test items itself favored white students (Gopaul-McNicol & Armour-Thomas, 2002; Tellez, 2003). Additionally, cultural interactions within the schools (faculty, students, or staff), including instruction, method, and curriculum content that favored white students were capable of influencing schooling success in Florida (Abedi, 2004; Venenzuela, 2000). These possible explanations are neither mutually exclusive nor exhaustive of all possibilities.

Relationships of High Stakes Testing and Graduation Rates

The literature is replete with conflicting empirical research results concerning whether high stakes testing impacts high school graduation rates. The present study indicated that the set of high stakes testing policies did impact graduation rates. Results of the macro analysis indicated that after removing the variation from the demographics
of racial/ethnic identity, limited English proficiency, and exceptional student education, the residual variations were related to changes in graduation rates, with 49% of the residual variations explained by changes in high stakes testing policies. The univariate analysis of variance indicated a significant predictive ability of high stakes testing policies on graduation rates ($F (4, 22), = 13.962, p < .001$). Thus, the high stakes testing policies were shown to be powerful programs with far reaching, profound positive or negative results for students and society in terms of graduation rates.

Apparently graduation rates can be predictably changed by high stakes policies, with tighter, more stringent policies or periods with new tests corresponding to lower rates and relaxed policies or policies that entail accountability for graduation rates corresponding to higher rates. In 2003, graduation rates jumped 11.5 points in one year when the No Child Left Behind accountability (that includes graduation rates) began in 2003. The positive impact of the federal NCLB has been noted by some scholars (Achieve, 2004; Carnoy & Loeb, 2002; Westchester Institute for Human Services Research, 2003). The NCLB requirements for state reporting by subgroups and added school funding for disadvantaged populations has other unquestionable benefits. For example, in the decade of the 1980s, few records exist on enrollment and graduation counts by subgroups of racial/ethnic identity. Record keeping for subgroups of racial/ethnic identity groups had been suspended during the Reagan Administration, but is required under NCLB.

However, the concept that unintended low graduation rates, corresponding to stringent and coercive high stakes testing policies, can be overcome with even more coercive centralized policies may have unintended consequences as well. Effects of using
coercion assessment to obtain results from students, teachers, families, and educators may result in school avoidance responses (Sefa Dei, Mazzuca, McIsaac, & Zine, 1997; Sirotnik, 2002) and corresponding undesirable consequences for educational engagement and development by students and faculty.

Dropping out is a complex phenomena with multiple causes (Beatty et al., 2001; Jimerson et al., 2000; Rumberger, 2001). This study showed that many students drop out and/or refuse to participate in the FCAT. The factor most predictive of dropping out was poor academic achievement. Studies have shown that high stakes testing policies may result in teachers being forced to teach in ways contrary to their own ideas of sound educational practice (Abrams, 2004; Flores & Clark, 2003; Rex & Nelson, 2003). Scholars have noted that curricula may be narrowed, critical ideas may be ignored, multicultural and student-created knowledge may not used in many standardized curricula under the influence of high stakes testing policies (Apple, 2001; Grant, 2004; Kozol, 2005; Ladson-Billings, 1994; McNeil, 2000; Orphal, 2000; Sleeter, 2003). Special education referrals may increase (Allington-McGill, 1992; Fielding, 2004). Teachers, administrators, and school of education university faculty may develop more concern for avoiding negative effects of high stakes testing policies than working for other educational attainments.

Re-examining Some Assumptions of High Stakes Testing

The findings of this study indicated that high stakes testing did not produce high academic achievement by many students. About 20% to 30% of students who did take the FCAT failed the test. When combined with the students who did not have scores for the FCAT, many either failed or did not have scores for the test. Positive academic
benefits from increases in the emphasis and consequences of academic achievement for students in the high stakes testing policy periods were not assessed by this study. Students may have benefited from extra attention to academics from teachers, parents, and the community who desired high student performance on standardized tests. However, a positive skew to the FCAT scores in the micro study indicated that many students did not score well on the FCAT in 2003. Therefore, the assumption that the high stakes testing policies produce high academic achievement is questionable. In view of this evidence, the assumption of improved student learning was not supported. To be fair, this study examined a period using a new high stakes test, the FCAT. It follows that higher test scores might be expected in a period that did not have a new test.

The assumption that high stakes testing policies produced increased motivation to learn and perform well may be questioned in terms of possible lack of participation from over half of the students in the micro analysis. Possible lack of participation was especially prevalent for traditionally disadvantaged groups of black, poor, and exceptional student education students. In the case of the 1,342 students who are assumed to have not taken the 10th grade mathematics FCAT, the students were 61% male, 44% white (compared to 49.5% in the original sample), 37% black (compared to 26% black in the original sample), 31% exceptional student education (compared to 21.8% in the original sample), 71% poor (compared to 39.5% in the original sample), and with a 11% graduation rate (compared to 42.6% in the original sample). It is evident in the results of this study that the high stakes testing did not produce an equitable result in terms of participation.
Questions about the assumption that high stakes testing policies are based on reliable, research based, and valid education policies can be found by examining the history of retention and graduation success relationship. The students with a history of retention(s) were not shown by this study to be effectively assisted in obtaining graduation success. A precursor for failure, number of retention(s) was the second best predictor for failing to graduate. The group of students, disproportionately black and poor, who had been retained in the 9th grade before 1999, had only a 14% graduation rate. This result is consistent with results from other studies (Grissom & Shepard, 1989; Rumsberger, 1995; Valenzuela, 2002).

Also questionable were the validity of statistical measures of central tendency derived from the FCAT scores. From the random sample of 3,000 students, from 42.2% to 45% of scores were missing; therefore these students were excluded from statistical analysis. In the micro analysis, the mean and median were around the passing mark of 2 to 3 for all four FCAT tests, implying a successful average on the test that did not represent the entire study body. Measures of central tendency can be misleading to consumers of information who may be evaluating the success of the high stakes testing policies.

While addressing the needs of failing students may seem to be in line with the assumption of increased equity of high stakes testing policies, this study provides evidence that many failing students who need help the most are not being reached effectively by the high stakes testing policies. Disproportionately lower graduation rate for poor, black and exceptional student education students were especially prevalent.
The assumption that the testing policies produce increased economic productivity must be questioned because 49% of a randomly selected sample of 3,000 9th graders in Florida in 1999 were left behind not to graduate in 2003. Gradation is an important indicator of the future earning potential of citizens, an advantage that 49% of the students in the micro analysis did not receive. Contrary to the stated purpose of abolishing inequity in the schools of No Child Left Behind, school reform via high stakes testing leaves many students behind. These students are at risk for poverty and experiencing other socially destructive conditions. Thus, the assumption that high stakes testing policies produce increased equity for all students must be questioned further.
Conclusions and Recommendations

Failure to graduate is clearly a phenomenon that is a threat to individual, social, and national well being. Educational leaders must find a new perspective and consider the entire population of high school students when calculating graduation rates, rather than a trimmed down population that excludes many for the analysis. Ideally, student learning and promotion success would be documented from early elementary grades and the entire population of students that enter school would be studied longitudinally. Then, assessments of programs, such as high stakes testing, remediation, and retention, would yield increased information for educational leaders. Possibly, many students are lost from the education system before the 9th grade and such studies would provide that information.

Classroom teachers, students, and parents must understand that to allow a student to fail in the classroom may very well tragically impact the child's future chance for graduation from high school. Of course, retentions are related to passing grades; however, teachers, parents, and administrators should be aware of the substantial research base (Darling-Hammond, 1994; Grissom & Shepard, 1989; Haney, 2000, 2001; Holmes, 1989; Holmes & Saturday, 2000; Jones, Jones, & Hargrove, 2003; Livingston & Livingston, 2002; Rumsberger, 1995; Shepard, 1989; Smith & Shepard, 1989; Valenzuela, 2002; Wehlage & Rutter, 1986) indicating that retained students stand a lower chance of graduating. Success for school completion on mandatory tests is essential also (including taking the tests), as is regular attendance in school.
Poor students are at increased risk for failing to graduate; therefore, progressive economic and education programs are needed to help these students. Blaming the students for their own academic and promotion failure in schools (which can start as early as kindergarten) or blaming them for not taking the high stakes test or for not coming to school is a way to shift responsibility from adult decision makers, professionals, and parents to children who are not only too young but are also powerless socially and economically. Whose purposes are served is a relevant question. Positive impact of high stakes testing in terms of student academic achievement has been found by scholars and the required improvements in record keeping, especially for subgroups is an unquestionable benefit (Achieve, 2004; Carnoy & Loeb, 2002; Westchester Institute for Human Services Research, 2003). However, some people who benefit from high stakes testing policies are not the students.

People benefiting from the current educational policy agenda in the form of tax cuts, increased political power, cultural dominance, and a labor force specifically prepared to work in stratified positions are the wealthiest of the population. Educational theorists, such as Cole (2005), have stated that globalization drives capitalism that sorts and cultivates students for their labor. Political and economic programs supporting business interests have resulted in insufficient funding for school infrastructure, reduction of innovative curriculums not centered on test results, reduction of social programs for the poor children, reduction or elimination of school programs to end racism, and the reliance on mono-cultural practices, culture, and knowledge in the schools.

Yet, the high stakes testing policies are popular with the voting public. Media reports have extensively covered school violence and low test scores. Politicians and
civic leaders have blamed frightening, dismal economic forecasts on the public system of education. High stakes test scores are often reported in the media as increasing every year while the equity gap is reported as steadily disappearing, when the present study and other studies (Dorn, 2003; Orfield et al., 2004; Stranahan & Borg, 2005a; Swanson, 2004) have shown this to be untrue. The high stakes testing policies have been rationalized to the public as the panacea for educational ills.

Scholars, such as Sherman Dorn (1998), have criticized high stakes testing policies for their effect on power and politics. Media attention, national policies, and school policies have resulted in an increase in narrow public attitudes towards public education. These public attitudes have been reflected by Congressional action. Social issues of poverty, hunger, lack of medical care, domestic violence, unequal funding in education, racism, segregation, and mono-cultural dominance in the schools have taken a back seat in the public policy arena as funding for programs to address these issues is decreased in almost every area. Corporations in the business of education have gleaned massive amounts of profit. These issues are relevant to educational leaders who have goals of quality education and equity for every child, because without public and the associated political support, quality education becomes difficult to obtain.

Contrary to the No Child Left Behind stated purpose of abolishing inequity in the schools, school reform via high stakes testing leaves many students behind. In the present micro analysis, 48% of the first time 9th graders in Florida in 1999 were left behind not to graduate in 2003. These students are at risk for a lifetime of poverty and other personally destructive conditions.
Socio-economic trends in Florida show that the poverty rate has grown every year for four years (National Poverty Center, 2004). Florida also has a large population, of which 41.5% are the working poor, defined as within 200% of the poverty rate (Research Institute on Social and Economic Policy, 2004). Not only are physical riches moving more to the most advantaged classes (Powell, 2005), but cultural capital in the form of education and graduation from high school are moving more to the advantaged classes. The present study shows that students who were white and not poor had a clear advantage in their chances for graduation from Florida high school and eligibility for the opportunities of college or good paying jobs.

Parallel trends exist in the global arena, in spite of the stated goals of the World Bank and International Monetary Fund. Globalization policies have not benefited the poor as much as the rich. For example, in New Zealand, globalization policies increased the wealth of the top 10% of people, while reducing the working and lower classes wealth and creating a poverty underclass where none had existed before (Elliott, 2002). To be fair, some reports have stated that individual poverty has decreased because of globalization, although within some nations, poverty has increased (Brooks, 2004).

The high stakes testing movement in education involves a more powerful impact than merely working to raise standardized test scores. School reform functioning in coercive ways does not support full functioning of critical thinking by teachers and students. This is unfortunate, because teaching functions to form the knowledge, attitudes, and beliefs of students and ultimately, society (Hill, 2000). Within the high stakes testing paradigm, the knowledge that is taught under government mandate has been channeled by predominately a conservative, financially oriented plurality from
outside the educational realm with the most profound benefits occurring for the wealthy and private sectors.

If some or all of the eight assumptions described in the review of literature in this dissertation are invalid, then the soundness of the high stakes testing policies should be questioned. Will the citizens who do not receive a high school diploma ever vote or be able to participate in the American dream of prosperity? What effects will this have on democracy and politics?

One possible effect is a population of under educated, poor, disproportionately black and Hispanic people, who are disenfranchised from the vote and have neither the education nor the critical thinking skills needed for democracy and prosperity. As social programs are increasingly denied through government budget cuts to the people who are most likely to never graduate from high school, tax breaks and political power have been reaped by the wealthiest in society. As citizens read in newspapers about the success of high stakes testing policies in increasing equity and achieving high standards in the schools, the graduation equity gap grows larger and, according to the results of this study, many citizens are receiving an inadequate education. A weakened democracy and a difficult economic and social future for individuals and society could well be the result.

Many people who do not graduate from high school experience poverty, prison, and a greater incidence of disenfranchisement from the political system. One indicator of social malaise is the rising prison population among youths, mostly children of color (Giroux, 2005). Over one third of black males spend time in prison. A symptom of the malaise that seems to characterize the United States and its youth is being one of the few
countries in the world to have increased the number and percentage of youth sentenced to life in prison without parole since 1990 (Human Rights Watch, 2005).

High stakes testing policies have far reaching consequences for the life chances of the entire population of the United States. Consideration of the benefits and risks for the entire population should be made and valued; rather than consideration of the benefits and risks for primarily business concerns. More stringent high stakes testing policies tend to mean fewer people obtain high school diplomas. A great number of people are being left behind in the high stakes education paradigm, as evidenced by the almost 50% examined in the micro analysis who did not even have records for taking the test.

Nevertheless, high stakes testing policies have added meaning to the high school diploma in terms of achievement. A milestone in academic achievement, satisfactory performance on high stakes testing represents an important and valuable cultural value. Without evidence of academic achievement, the high school diploma may be meaningless. Due to the negative and unfair consequences of high stakes testing; however, alternatives to the testing should be sought by educational leaders. Such alternatives may include a variety of tests in different subjects, the option of taking the tests in languages other than English, and the input on content and design of the tests by professional educators. No educational decision as important as graduation or promotion in grade should be made on the basis of one test. Other tests and alternative assessments must be considered in making important life altering decisions such as who graduates and who doesn’t. The question of what should be the meaning of a high school diploma may be discussed, developed, formulated, instrumented and evaluated in designing future high school curriculums and assessments.
Recommendations for students, parents, and educators must be to make certain that every child is protected from failure in school. Protection from academic failure, perhaps with culturally relevant curriculums, caring teachers, and changes in high stakes testing policies to be more supportive of student success, could help insulate children from persuasive negative effects of inequitable social capital or social class reproduction (Bourdieu, 1986) in schools. A fair and equitable educational accountability system must be developed to ensure students do not fail. Consequences of school failure are dire, including loss of graduation and future opportunities in life.

Further recommendation include that the high stakes testing paradigm be examined carefully to determine that assumptions supporting the policies are valid and that support for the policies is justified. The impact, consequences, and the value systems used to justify and support high stakes testing should be examined. The functioning of high stakes testing policies at this time is inequitable to many groups of people. The argument that governments and economically privileged tax payers have a legitimate interest in the results of education should be re-examined with a more equitable focus on all groups of people, instead of the focus on the needs of a wealthy group of people, such as those who authored the report *A Nation at Risk* (1983). Myopic visions of high stakes testing should be eliminated in favor of new education policies innovative and inclusive educational visions.

Recommendations for education research are to continue to examine the impact and all consequences of high stakes testing policies. More research into the validity of the assumptions supporting the policies should be conducted. Research for developing more understanding of the consequences of high stakes testing policies could be conducted
with qualitative studies of students after they leave high school. An economic impact of high stakes testing policies on those who graduate and those who do not graduate would be of interest. Another research direction would be longitudinal quantitative studies over longer periods of time using individual students focusing on relationships between graduation, tests, school achievement, retention, curriculum, and demographics. Other studies, similar to this study, should be conducted on the quantitative relationship of high stakes testing policies and graduation rates using different states as well as Florida. Low percentages of available data on the FCAT were a limitation of the study. Further studies similar to this micro analysis should be done of other years that may have higher FCAT participation and/or record keeping.

Importance should be given to understanding the broad set of policies, rather than just a piece of the policies, such as effects of the graduation test on graduation rates over a relatively short period of time. The history of high stakes testing policies and how these policies became so powerful in a changing, challenging, competitive, and multicultural world should be researched to provide valuable insight into current times. Discovering ideas, options, and possibilities for creating and implementing a high quality education system that serves the needs of all the people of the United States should be a foremost goal of educational leaders.
APPENDICES
Appendix A

The Basic Completion Rate (BCR) (Swanson, 2003)

\[
BCR = \frac{\text{Number of graduates in a given year with a regular high school diploma}}{\text{Number of 9th graders 3 years earlier}}
\]
Appendix B

The Cumulative Promotion Index (CPI) (Swanson, 2003).

The CPI approximates the probability that an entering 9th grade student will graduate in four years with a regular diploma. High school graduation is represented as a step-wise process composed of three grade promotions and the graduation event. The example uses a hypothetical class of 2001 to illustrate.

\[
\text{CPI} = \frac{E_{\text{grade 10, 2002}}}{E_{\text{grade 9, 2001}}} \times \frac{E_{\text{grade 11, 2002}}}{E_{\text{grade 10, 2001}}} \times \frac{E_{\text{grade 12, 2002}}}{E_{\text{grade 11, 2001}}} \times \frac{G_{\text{2001}}}{E_{\text{grade 12, 2001}}}
\]

Where

\( G_{\text{2001}} \) is the number of students who graduated with a regular high school diploma in 2001.

\( E_{\text{grade 10, 2002}} \) is the number of students enrolled in grade 10 at the beginning of the 2001-2002 year.

\( E_{\text{grade 12, 2002}} \) is the number of students enrolled in the 12th grade at the beginning of the 2001-2002 year.

\( E_{\text{grade 11, 2001}} \) is the number of students enrolled in the 11th grade at the beginning of the 2000-2001 year.

\( E_{\text{grade 10, 2001}} \) is the number of students enrolled in the 10th grade at the beginning of the 2000-2001 year.

\( E_{\text{grade 11, 2002}} \) is the number of students enrolled in the 11th grade at the beginning of the 2001-2002 year.

\( E_{\text{grade 9, 2001}} \) is the number of students enrolled in the 9th grade at the beginning of the 2000-2001 year.

\( E_{\text{grade 10, 2002}} \) is the number of students enrolled in the 10th grade at the beginning of the 2001-2002 year.
References


Florida Department of Education (DOE), Division of Public Schools. (1979, April). Students by racial/ethnic category: Staff by sex and racial/ethnic category 1978-1979 (Series 79-21). Tallahassee, FL: Author.


New York: Teachers College Press.


Laurel Stanley Vitae

Laurel Stanley was born , in Madison, Wisconsin

Education

Doctor of Education in Educational Leadership
University of North Florida
Degree awarded April, 28, 2006

Dissertation: Florida High Stakes Testing and Graduation Success

Master of Secondary Education
Eastern New Mexico University
Portales, NM August, 1996

Thesis: A Study of Self-Concept in an Educational Clinic

Bachelor of Business Administration
Major in Accounting
Eastern New Mexico University
Portales, NM December, 1990

Honors: Summa Cum Laude

Experience

University of North Florida
Working on Gifted Issues (WOGI)
Florida Department of Education grant

Research Assistant, 2005 –

Publications

Journal Articles


Other Publications


Conference Presentations

