1-1-2012

The impact of no child left behind (nclb) on school achievement and accountability

Glenn Maleyko
Wayne State University,

THE IMPACT OF NO CHILD LEFT BEHIND (NCLB) ON SCHOOL ACHIEVEMENT AND ACCOUNTABILITY

by

GLENN MALEYKO

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2011

MAJOR: EDUCATIONAL
LEADERSHIP AND
POLICY STUDIES

____
Advisor Date

____

____

____

DEDICATION

I dedicate this dissertation to my outstanding family whom I love and cherish. My supportive and understanding wife, JoAnne, stood with me through the duration of my classes, exams, and research. She is an outstanding mother to our children. There were several years where I was not home on many nights as I was taking my doctoral classes and working on the data analysis that I was required to conduct at WSU with this study. I also want to stress that my daughter, Ashley, and my son, Justin, are the joy and love of our lives and they were also supportive during my research.

ACKNOWLEDGEMENTS

There are many people that I want to thank for their support with my research. However, as mentioned in the dedication, it starts with my wife and children who made many sacrifices over the years so that I could complete my studies.

Dr. Marytza Gawlik was an outstanding advisor and she provided me with phenomenal support throughout my studies. She was my mentor through my classes, qualifying examinations, and subsequent research. She gave me a great deal of on-going guidance, encouragement, support, and motivation toward a standard of excellence. I have a tremendous amount of respect for Dr. Gawlik along with my committee members Dr. Michael Addonizio, Dr. Joslen Letscher and Dr. Craig Roney. They provided me with some excellent guidance and insights with my research and they continually encouraged me to move forward with my studies. Dr. Michael Addonizio and Dr. Craig Roney provided me with excellent support during my classes and with my research at Wayne State. Dr. Joslen Letscher, my outside committee member, has been involved with my research during the achievement of my Masters Degree and Specialist Degree at the University of Detroit Mercy, she was also my supervisor with the achievement of my teacher certification in 1997.

I would further like to thank my father Mike Maleyko and my mother Sharon Jacob along with my father-in law Joe Copot, mother-in-law Mary Copot, and step-father Dan Jacob. They have each provided me with support throughout my studies which included watching our children when necessary. A further thank you goes out to my siblings and their spouses, Joan and Ryan Gelinas, Jennifer and Chris Mills, Jason and Genevieve Maleyko, along with my in laws, Chris and Cathy Nepszy and Joe Copot. They have all been supportive of my studies. I have 10 nieces and nephews, Katie, Mateo, Xavier, Ellie, Isaac, Maija, Sophia, Alexa, Emily, and Brody. Our family gatherings provided me with a positive release from my studies when needed. My grandparents Tony and Anne Trombka have always been supportive of my academic and professional career. My late grandparents, Benjamin and Charlotte Maleyko were supportive of me during the earlier years of my life.

I would further like to thank all of the teachers, parents, and students who were very supportive of my studies at both Salina Intermediate School and DuVall Elementary School in Dearborn, Michigan. I would also like to thank my former superintendent Dr. John Artis, Dr. Cheryl Kreger my former Associate Superintendent, and my former Principal Irene Blackburn - each of whom provided me with support and guidance during my initial years as a school administrator along with an inspiration towards the achievement of my doctoral degree. I would further like to thank Mr. Brian Whiston along with Mrs. Norma Jean Sass. I learned a great deal from them along with the individuals mentioned above who were an inspiration to me. Mrs. Sass was extremely flexible with my schedule so that I could complete my studies. Mr. Whiston was also supportive and provided me with my first opportunity for a central office Director Position in School Administration. Early in my teaching career, I completed an internship with Mr. Karl Steuf and I also worked on a research project for Dr. Jeremy Hughes and I would like to thank both of them. Dr. Hughes helped with my initial interest in school accountability through my work on the "Leading Academic Indicators Project."

Although I cannot name them in the person due to confidentiality provisions, I would like to thank the principals of the four schools where I conducted my qualitative research along with the sixteen teachers that participated in the semi-structured interview protocol. I would further like to thank the representatives at NCES in Washington for their support with the restricted level data. Mr. Kevin Carroll, the Systems Security Officer for my NCES data also provided me with outstanding support.

A special thank you goes out to my friend and colleague, Bob Attee, who spent many hours assisting me with editing and revising my writing. My friend and colleague Shirley Leffew also supported me on numerous occasions when I had doctoral classes over the years.

TABLE OF CONTENTS

Dedication: ii

Acknowledgements: iii

List of Figures: xiii

List of Tables: xiv

CHAPTER I: INTRODUCTION.. 1

Statement of the Problem 3

Adequate Yearly Progress 5

Purpose of the Study 8

Research Questions 11

Independent and Dependent Variables 11

Significance of the Study 15

Limitations of the Study 16

Important Definitions 17

Organization of the Dissertation 19

CHAPTER II : LITERATURE REVIEW..21

Review Focus 22

Organization of the Literature Review 22

History of Accountability in the Four Sample States 23

California 23

Michigan 24

North Carolina 25

Texas 25

Measurement Concerns with the Use of AYP to Evaluate School Effectiveness **26**

Statistical Issues, Manipulations of the AYP data, and Lack of Consistency across the Country 26

Lack of Consistency with AYP Across the Country 37

Reliability Issues with a Single Measure Accountability System 39

High Flying Schools and the One Year Phenomena 40

AYP Measures Social Capital, Minority Status, and Racial Subgroups 42

Measuring English Language Learners (ELL) 44

Inevitability that a High Percentage of Schools will be Labeled as Failing 46

Summary of Measurement Concerns with the Use of AYP to Evaluate School Effectiveness 48

AYP Benefits and Unintended Consequences **49**

Philosophical intent 49

Consequences of Sanctions and Accountability 51

Summary of AYP Benefits and Unintended Consequences 62

The Use of the NAEP 62

External Accountability vs. Internal Accountability 64

Summary, Implications, and Discussion **65**

Effectiveness of AYP in Improving Student and School Achievement 65

Demographic Variables, School Improvement, and Classroom Instruction 66

Conclusion **68**

CHAPTER III: METHODOLOGY......69
Research Questions......70
Quantitative Research Design......71
Sample......71
Data Collection......74
Data Analysis......76
Statistical Hypothesis......77
Causal Hypothesis......80
Statistical Tests and Analysis......83
Research Question One......83
Research Question Two......90
Research Question Three......95
Statistical Tests: Alpha Level......97
Statistical Power (Sample Size)......97
Validity and Reliability......98
NAEP......99
Quantitative Limitations......100
Qualitative Research Design......101
General Design......102
Sample......102
Data Collection and Analysis......103
Analytic and Constant Comparative Method......106
Validity and Reliability......109

Qualitative Limitations 111

Research Design Significance 113

CHAPTER IV: RESULTS 114

Presentation of Results 114

AYP Formulas in the Four Sample States 114

California 115

Michigan 118

North Carolina 121

Texas 123

Quantitative Results and Analysis 126

Research Question One 127

Research Question Two 155

Research Question Three 223

Qualitative Results and Analysis 249

Description of Sample Schools 250

Semi-Structured Interview Findings 253

Overview of Participants Views on the NCLB Accountability reform 257

AYP Identification and Sanctions 259

Economically Disadvantaged Status and School Funding. 265

Positive Consequences 278

Negative Impact or Unintended Consequences 293

District Impact 301

Summary for Qualitative Findings 303

CHAPTER V: DISCUSSION AND CONCLUDING REMARKS **306**

The Use of Growth Data for AYP 311

A Call for Multiple Measures 315

National Curriculum and the Use of the NAEP.. 320

What We Still Need to Know **323**

Appendix A: Requirements for Schools Not Making AYP in Michigan **325**

Appendix B: NCLB Purpose Statement **328**

Appendix C: NCES Selection Procedures for Sample States..**330**

Appendix D: NAEP Subgroup Ranges and Median Values used in the study **331**

Appendix E: NAEP Item Response Theory...**332**

Appendix F: NAEP Plausible Values...**331**

Appendix G: NAEP Restricted Access Data Regulations **336**

Appendix H: AM Statistical Software **335**

Appendix I: NAEP Weighting Procedures **336**

Appendix J: NAEP Reading Achievement Levels by Grade...**337**

Appendix K: NAEP Mathematics Achievement Levels by Grade.................................**339**

Appendix: NAEP Non Cognitive Items and Questionnaires...**342**

Appendix M: NAEP Validity and Reliability **347**

Appendix N: California: Standardized Testing and Reporting (STAR) Validity and Reliability...**348**

Appendix O: Michigan: Michigan Educational Assessment Program (MEAP) **349**

Appendix P: North Carolina: End of Grade Test Validity and Reliability.................**353**

Appendix Q: Texas: Texas Assessment of Knowledge and Skills (TAKS)..................**359**

Appendix R: The Number of Test Items in Each Assessment 361

Appendix S: Documentation of Consent to use the Teachers' Voice Survey 362

Appendix T: NCLB Teacher Interview Questions 363

Appendix U: NCLB Administrator Interview Questions..365

Appendix V: Research Information Sheet..367

Appendix W: California Tables and Figures for Research Question One, Pearson Correlation...369

Appendix X: Michigan Tables and Figures for Research Question One, Pearson Correlation...375

Appendix Y: North Carolina Tables and Figures for Research Question One, Pearson Correlation...383

Appendix Z: Texas Tables and Figures for Research Question One, Pearson Correlation...391

Appendix AA: California Tables and Figures for Research Question Two, Pearson Correlation...399

Appendix BB: California Tables and Figures for Research Question Two, Logistic Regression. 415

Appendix CC: Michigan Tables and Figures for Research Question Two, Pearson Correlation...443

Appendix DD: Michigan Tables and Figures for Research Question Two, Logistic Regression. 459

Appendix EE: North Carolina Tables and Figures for Research Question Two, Pearson Correlation...484

Appendix FF: North Carolina Tables and Figures for Research Question Two, Logistic Regression. 500

Appendix GG: Texas Tables and Figures for Research Question Two, Pearson Correlation. 528

Appendix HH: Texas Tables and Figures for Research Question Two, Logistic Regression. 544

Appendix II: California Tables and Figures for Research Question Three, Logistic Regression. 572

Appendix JJ: Michigan Tables and Figures for Research Question Three, Logistic Regression. 588

Appendix KK: North Carolina Tables and Figures for Research Question Three, Logistic Regression. 604

Appendix LL: Texas Tables and Figures for Research Question Three, Logistic Regression. 620

References……...…………………………………………………………………… 636

Abstract ……………………………………………………………………………. 653

Autobiographical Statement ……………………………………………………….. 656

LIST OF FIGURES

Figure 1. Independent and dependent variables for research question No. 1........................ 12

Figure 2. Independent and dependent variables for research question No. 2......................... 13

Figure 3. Independent and dependent variables for research question No. 3......................... 14

Figure 4. Independent and dependent variables for research question No. 4........................ 14

Figure 5. Confidence interval example 28

Figure 6. AYP proficient schools in Michigan.. 30

Figure 7. Subgroup minimal size in a sample of states... 32

Figure 8. Straight line trajectory.. 34

Figure 9. Straight line trajectory with plateaus... 35

Figure 10. Back end loaded trajectory. 35

Figure 11. AYP Proficient schools in Kentucky with statistical manipulation...................... 38

Figure 12. Schools in Michigan that are projected to make AYP in 2014............................. 47

Figure 13. Racial/Ethnic Subgroups in the Sample States 72

Figure 14. Mathematics trajectory for California 116

Figure 15. ELA trajectory for California 117

Figure 16. Mathematics trajectory for Michigan.. 120

Figure 17. ELA trajectory for Michigan 121

Figure 18. Mathematics Trajectory for North Carolina 122

Figure 19. Reading trajectory for North Carolina 123

Figure 20. Mathematics trajectory for Texas 125

Figure 21. Reading trajectory for Texas 125

Figure 22. Logistic Regression Observed and Predicted Probabilities for California.......... 163

LIST OF TABLES

Table 1. Important definitions 17

Table 2. Agency website databases 77

Table 3. School level proficiency targets in sample states, AYP requirements..................... 85

Table 4. State assessments with validity and reliability information................................. 100

Table 5. Summary of research questions, instruments, and data analysis tools.................. 112

Table 6. The alternative AYP indicator: API index in California 118

Table 7. Reading ELA AYP percentage proficiency requirements in Michigan................ 119

Table 8. Mathematics AYP percentage proficiency requirements in Michigan 120

Table 9 Pearson correlation analysis r value defined.. 126

Table 10. Descriptive statistics for the 2005 California 4th grade dataset........................... 130

Table 11 Pearson correlation matrix for the 2005 California 4th grade dataset.................... 131

Table 12 Logistic regression: observed and predicted frequencies, constant model............ 158

Table 13 Logistic regression: observed and predicted frequencies, predicted model.......... 159

Table 14. Logistic regression: omnibus test of model coefficients 159

Table 15. Logistic regression: model summary assessing goodness of fit 160

Table 16. Logistic regresion: analysis table 162

Table 17. Componential analysis for qualitative dataset. 255

CHAPTER I

INTRODUCTION

The No Child Left Behind (NCLB) reform may be the most significant legislation affecting public education that has been enacted by the federal government during the past 35 years (Peterson & West, 2003). The implementation of public education reform has historically and constitutionally been the responsibility of individual states; however, the NCLB educational policy at the federal level establishes mandates for performance standards and consequences if they are unmet (Hess & Finn, 2004). Under NCLB, the federal government has the opportunity to reduce funds to states that fail to make progress and has the option of rewarding states that make progress with achievement. States have the option to opt out of NCLB, but if they make that choice, they could forfeit their federal Title I funding.

The NCLB bipartisan reform, signed into law in January 2002, aims to address the public and political pressure prompting school reform (Hess & Petrilli, 2006). One of the major campaign issues in the 2000 election was public educational reform and the need to improve the quality of education for all students. "Both George Bush and Al Gore embraced accountability in the 2000 election campaign" (Peterson & West, 2003, p. 7). The timing of the legislative debate was critical in receiving bipartisan support, as Congress wanted to show a united government working together in the aftermath of the terrorist attacks on the United States on September 11, 2001. Congress worked hard to get the legislation passed in an expeditious manner during the final months of 2001 because of increased political pressure for bipartisan support (Sunderman, Kim, & Orfield, 2005, p. 26).

The NCLB legislation is an overhaul of the Elementary and Secondary Education Acts (ESEA) of 1965. Public pressure toward school reform intensified in the 1980s with the report

"A Nation at Risk" and continued into the 1990s during the Clinton Administration (Irons & Harris, 2006). Many schools in the United States were struggling with providing quality educational programs for all students. This was especially prevalent in urban and low-income areas with diverse student populations. Some of those populations, otherwise known as categorical populations(or disadvantaged populations), as they are defined by NCLB include: schools with large numbers of English-language learners (ELL), special education students, students living below the poverty level (economically disadvantaged [ED]), and students of diverse racial and ethnic backgrounds, including Black/African American, Latino/Hispanic, American Indian/Alaskan Native, Asian/Pacific Islander, and Caucasian (United States Department of Education [USDOE], 2002, Section 1111, (2) C). The purpose of the NCLB reform is to provide quality education for all students, with the specific focus on disadvantaged student populations in an effort to ensure a uniform achievement level according to state standards (Borowski & Sneed, 2006; Haycock, 2006).

NCLB advocates believe that school accountability and the standards-based movement is the means to implement wide-scale public education reform that will transform the public school system into a more beneficial model for all students. This includes providing a spotlight on traditionally underperforming students. The philosophical intent of the reform is noble; it stands for an important principle that no child will be left behind and all children will receive a high-quality education despite their disadvantaged status (Borowski & Sneed, 2006; Chubb, Linn, Haycock & Wiener, 2005; Guilfoyle, 2006; Haycock, 2006; Hess, 2006; Hess & Petrilli, 2005; Kane et al. 2002; Lewis, 2006). The work by Kozol (1992) in "Savage Inequalities" brought attention to problems with schools in the United States, especially when it comes to the education of disadvantaged students. Public pressure placed educational reform as a major priority at the turn of the 21st century. As a result, NCLB was enacted and signed into law.

Statement of the Problem

The most important part of the NCLB reform is the set of accountability standards for schools, school districts, and states because it is the mechanism in which the framers of NCLB believe that school improvement will occur on a national level. This study examines a portion of the NCLB accountability system in order to measure the impact that the legislation is having on school reform efforts.[1] The study emphasizes the measurable accountability provisions that are found within the NCLB reform. The NCLB accountability system is based on the premise of providing rewards and sanctions. Under NCLB, schools are made accountable through measuring the achievement level of all students in the school. One of the greatest benefits of the reform is the philosophical premise that all children will count, as NCLB sets ambitious goals for what students must learn in order to close the achievement gap (Borowski & Sneed, 2006; Chubb et al. 2005; Guilfoyle, 2006; Haycock, 2006; Hess, 2006; Hess & Petrilli, 2006; Kane et al. 2002; Lewis, 2006). Guskey (2007) contends that accountability through a focus on student learning results is a positive aspect in NCLB that should be embraced by educational leaders as a means to improve the quality of education in schools.

The philosophical intent of the reform is being implemented through the Adequate Yearly Progress (AYP) measurement system. This study analyzes the effectiveness of the AYP accountability system by measuring the impact of the reform on school performance. The study also analyzes the level of consistency in which the reform is being implemented in states

[1] That impact is the real dependent variable even though other dependent variables will be measured in this study.

throughout the country. Finally, an examination of the positive and negative impact that the reform is having on school improvement and classroom instruction is addressed.

Under NCLB, every state is required to develop specific grade-level benchmarks, and each state must also administer assessments to evaluate the percentage of proficient students in specific schools and school districts as identified by their achievement of grade-level benchmarks (Shaul & Ganson, 2004). Achievement-level targets must be established by each state. Those achievement level targets increase at least every three years with the provision that by the year 2014, 100 percent of the students should meet proficiency standards in mathematics, reading, and science (USDOE, 2002, Section 1111 (2) B). The literature review in chapter two provides information that questions the statistical probability that the 100% achievement targets are realistic. Furthermore, achievement data must be disaggregated based upon specific subgroups or categorical populations. If schools or school districts fail to reach the specific proficiency targets as a whole or in a specific subgroup of their student population, they will be labeled as a struggling or failing school or school district (USDOE, 2002, Section 1111 (2) C).

This study measures the impact that the reform is having on the categorical populations at the school level and addresses the problem of whether or not the NCLB reform is successful as intended by using accountability measures in improving the quality of schooling for categorical populations. It further addresses the problem of consistency with the implementation of AYP in different states with a comparison of the impact that the reform is having on a sample of four states, 1) California, 2) Michigan, 3) North Carolina, and 4) Texas. These four states were selected due to their strong history of implementing school accountability which began prior to NCLB along with their different geographic locations and diverse student populations that are present within each state. Since a major portion of this study analyses disadvantaged and diverse student populations, it was critical to choose states with diversity.

Adequate Yearly Progress

Adequate Yearly Progress (AYP) is the statistical measure used to evaluate school effectiveness under NCLB. The AYP formula is used to calculate the achievement of schools and school districts in order to determine if schools are meeting the NCLB-designated achievement standards. The AYP measurement leads to a determination of school effectiveness and then labels the schools as successful or failing. Individual states are responsible for calculating AYP and imposing sanctions on schools (USDOE, 2002, NCLB ACT, Section 1111(2) B).

AYP is calculated by determining the percentage of students in a school who have met proficiency levels on state-administered tests in reading, mathematics, and science. Schools must test 95% of their students on an annual basis to achieve AYP. The analysis of AYP also involves disaggregated achievement data in eight subgroup populations otherwise defined as categorical populations in NCLB. The reform requires a 95% participation rate on the state accountability assessments as calculated by the school as a whole and within each of the subgroup populations. There is one additional measure that a state must implement to determine the AYP status of schools and districts. States have flexibility with choosing their additional measure and many states choose to use the graduation rate as one example. The additional measure must be approved by the USDOE. There are 37 analyses under the AYP formula, including subgroup populations, whole school, participation in mathematics and reading or language arts, as well as achievement in mathematics and reading or language arts (Lamitina, 2006).

When a school fails to make AYP, the school is labeled as “needing improvement” and will receive assistance in the early stages. The later stages require sanctions and a possible restructuring of the school. NCLB imposes several corrective actions for schools that fail to make AYP. Some of the early actions include availability of schools of choice and supplemental services from outside agencies (USDOE, 2002; Peterson & West, 2003). After failing to meet AYP for a fourth consecutive year, one of the possible sanctions is replacing the staff that is relevant to the failure. Restructuring or state takeover of the school is mandated in year five (Peterson & West, 2003, p 26). Appendix A provides information on the corrective actions and sanctions that are mandated if a school fails to achieve successful AYP status in the state of Michigan.

There are several questions that arise regarding the AYP formula in the NCLB reform. The most important question is whether or not the AYP formula is successful at mandating improvements in public education. The answer to this question is difficult to determine in a single study because there are multiple variables at play regarding the attainment of successful educational programs and the evaluation of school effectiveness. However, many researchers in the field posit that accountability formulas such as AYP that rely on sanctions and punishments are faulty because they fail to provide schools with the internal capacity to make change (Abdelmann, Elmore, Even, Kenyon & Marshall, 1999; Elmore, 2002, 2004; Hill & Crevola, 2006; Maleyko & Gawlik, 2011; Schoen & Fusarelli, 2008). In a study on the impact of high-stakes accountability systems in the United States, Abdelmann et al. (1999) found that accountability systems will not have a positive impact on curriculum and instruction unless it can be relevant to the development and implementation of the internal accountability system at the school level. An internal accountability system is developed when the stakeholders, including teachers, administrators, students and parents, develop a sense of responsibility toward the achievement of high standards for all students.

A problem addressed in this study is determining the level of school achievement and progress using the National Assessment of Educational Progress (NAEP) in comparison to the state accountability assessments from the sample states, 1) California, 2) Michigan, 3) North Carolina, and 4) Texas. The state accountability assessments that were used in this study include, 1) The California Standardized Testing and Reporting (STAR), 2) Michigan Educational Assessment Program (MEAP), 3) North Carolina End-of-Grade Test, and 4) Texas Assessment of Knowledge and Skills (TAKS).

Each state develops and implements its accountability measurement and design to meet NCLB's mandates. The state accountability systems must be approved by the USDOE to comply with the AYP requirements in NCLB. However, the only academic measurement in K-8 education that provides a standard achievement analysis that is implemented in all states is the NAEP. The NAEP is administered to students every two years in various subjects and is administered uniformly using the same test in states across the nation. NAEP results serve as a consistent measurement for all states. The assessment stays essentially the same from year to year, providing information on academic progress (National Center for Educational Statistics [NCES], 2009). The NAEP was used in this study to examine and compare the mathematics and reading assessments administered to students at the fourth-grade and eighth-grade level on a two-year cycle against the state accountability assessments in the four sample states. The students who take the NAEP are randomly selected by each state as a requirement of NCLB. Because AYP is the federal measurement tool that is administered by the state governments, it is feasible to measure the external validity and reliability of the AYP by conducting a study that compares school AYP results in a sample of states throughout the United States.

Meier et al. (2004) believe that NCLB is flawed because it focuses on measuring achievement levels but provides little support to improve schools. Another issue addressed in this study is what if any effect that NCLB had on improving the conditions of schools, especially schools with large numbers of categorical students. Finally, this study examines factors associated with building internal school capacity in order to determine the effective methods schools used to increase student achievement on state and national accountability measurements. A qualitative methodology was implemented in this study to answer the final research question regarding what if any impact that NCLB is having on the development of an internal school accountability system. The combination of the quantitative and qualitative methodology used in this study provides for richer, more valid and more reliable results than only using one research methodology (Gawlik, 2005). The information from this study is beneficial to educators and policy makers with an objective to improve the instructional environment at the school level for the benefit of public school students.

Purpose of the Study

The purpose of this study is to evaluate the effectiveness of AYP at establishing the conditions for school reform at the school level and classroom level. This is measured using qualitative and quantitative data analysis. In the qualitative portion of the study, the AYP status of schools is the independent variable while the impact on school reform and classroom instruction is the dependent variable. The evaluation the AYP formula includes an analysis of the external validity and the reliability of AYP as it relates to measuring school success across the country. Researchers maintain that there is a lack of consistency with AYP that results in reliability issues in using AYP to measure school effectiveness (Elmore, 2002; Guilfoyle, 2006; Kane and Douglas, 2002; Linn & Haug, 2002; Maleyko & Gawlik, 2011; Porter et. al, 2005; Scheon & Fusarelli, 2008; Wiley et al. 2005). Porter et al. (2005) found that there was little consistency with using AYP as a measure from state to state. Linn (2008) argues that inconsistencies among states in the categorization of a school as successful or failing may have more to do with the methodology that is employed by a state with the calculation of AYP rather than the relative effectiveness of schools.

Some researchers (Elmore, 2002; Harris, 2007; Maleyko & Gawlik, 2011; Mathis, 2004; Mathis, 2006) contend that AYP is measuring school demographics and the social capital that students bring to the school instead of school effectiveness. This study addresses that contention by measuring school demographics in relation to their AYP status. Data was collected and analyzed regarding the characteristics of schools that meet AYP standards and the schools that do not meet AYP standards in the sample states. The analysis of the data according to school demographic characteristics allowed for the prediction of the general profiles of schools that are successful under the AYP provisions.

A significant purpose of this study includes an evaluation of the effectiveness of AYP at establishing the conditions for school reform at the school level and classroom level, which is another dependent variable. This evaluation includes implementation of both qualitative and quantitative research methodologies in order to determine the level of impact that AYP is having on school reform and classroom instruction. As was previously stated, the philosophical intent of NCLB is to improve the educational conditions for disadvantaged students. That intent is identifiable according to the heading in Title I of the NCLB legislation: Improving the Academic

Achievement for the Disadvantaged (USDOE, 2002, Section 101). If AYP is successful, then the sanction provisions in NCLB should lead to school improvement and increased student achievement. However, after sanction provisions are enacted, if schools continue to be labeled as failing and no improvements are being made, then the interpretation should be that NCLB is not successful in achieving the goals set out in the reform.

It is important to measure the number of schools that have been positively affected by AYP under the NCLB provisions. It is probable that increases made by schools under AYP are a result of improvements in classroom instruction. The classroom instruction variable is a significant predictor of school success as the research in the field is clear that the quality of classroom instruction has a direct impact on student achievement (Brookover, 1985; Edmonds, 1979a, 1979b, 1982; Lezotte, 1992; Lezotte & Jacoby, 1991; Marzano, 2003; Nye, Konstopolous, & Hedges, 2004; Sanders, 2000; Sanders & Rivers 1996). Therefore it is critical to measure the impact that AYP is having on the implementation of instructional strategies. No literature has been found that provides data on the impact of AYP on classroom instruction. The role of leadership in the implementation of successful school improvement efforts is also proven to have a positive influence on student achievement, especially when administrators focus on student achievement and instructional strategies (Brown, 2001; Connell, 1996; Hallinger & Murphy, 1986; Hattie, 2009; Henchey, 2001; Marzano, Waters, & McNaulty, 2005; Teddlie & Springfield, 1993). Hence, this study investigates the impact that AYP is having on classroom instruction and school improvement strategies that are implemented by school leaders.

Research Questions

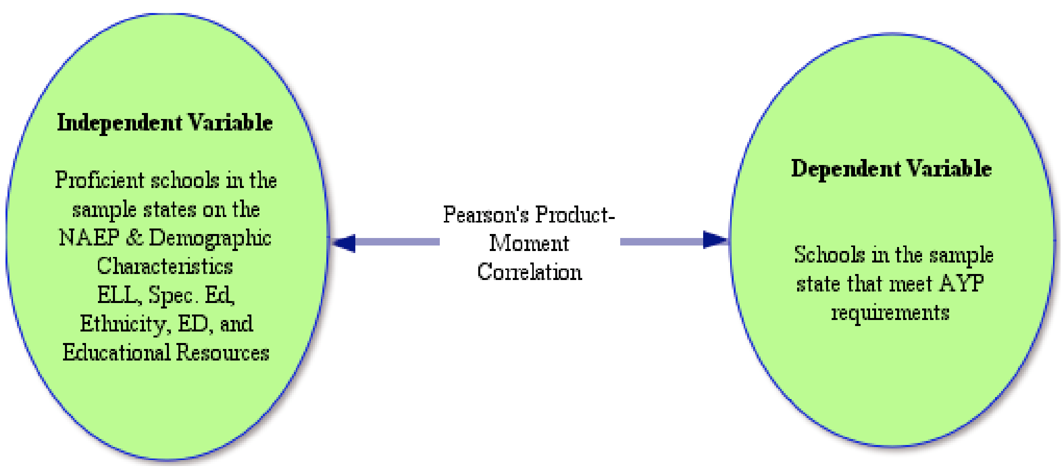
1. Is there a significant correlation between the schools that meet proficiency standards on the NAEP and the schools that make AYP in the sample states? Are there significant differences between subgroups?
2. Is there a significant correlation between the schools that meet proficiency standards on the NAEP and schools that meet proficiency standards on the state accountability assessments (STAR, MEAP, End-of-Grade Test, and TAKS) in the sample states? Is there a combination of factors that best predicts proficiency status on the NAEP and state accountability assessments?
3. Are the demographic (categorical) characteristics and educational resources significant predictors of the schools that make AYP and fail to make AYP in the sample states?
4. What impact is AYP having on school improvement initiatives and classroom instruction?

Independent and Dependent Variables

The variables that are examined in this study include an evaluation of the impact of AYP at the school level. The first three research questions were addressed by evaluating a sample of four states across the country, 1) California, 2) Michigan, 3) North Carolina, and 4) Texas. As it was previously stated, the selection of the four sample states is tied to the historical nature of the accountability systems implemented in those states prior to the NCLB legislation. The selected sample states also provide representation from different geographic regions in the country. This study examines the state accountability system that complies with the AYP requirements in NCLB for each of the sample states.

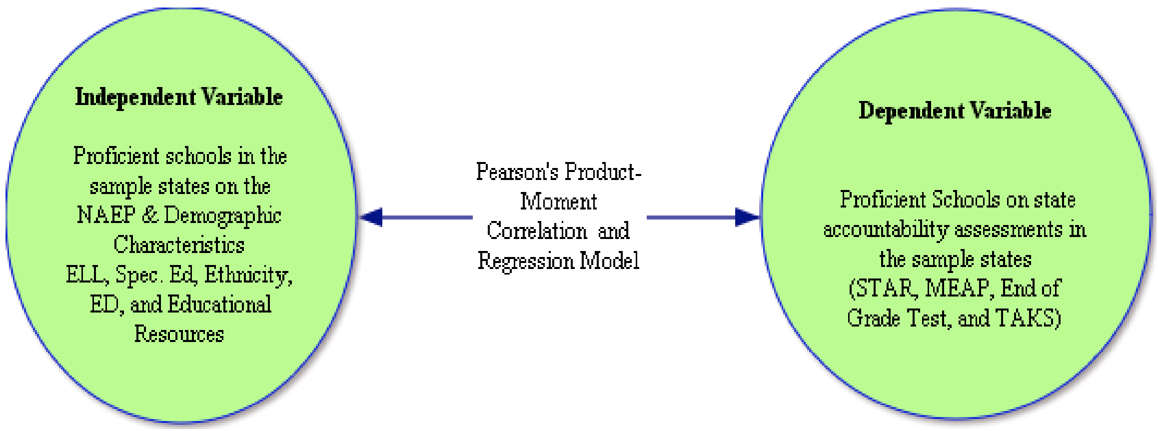
The independent variables in the first research question are; whether or not schools achieve proficiency on the NAEP reading and mathematics assessments in the sample states at the fourth- and eighth-grade level for the years 2005 and 2007, the educational resources that a school receives, and the demographic characteristics of schools, including socioeconomic status, ethnicity, special needs, and ELL status. The years 2005 and 2007 were selected because the NAEP assessment is administered every two years and those datasets were the most current NAEP assessments available during the time that this study was conducted. The dependent variable is whether or not the schools meet AYP proficiency at the fourth- and eighth-grade level in the sample states for the years 2005 and 2007. As such it is a categorical variable.

Figure 1. Independent and dependent variables for research question No. 1



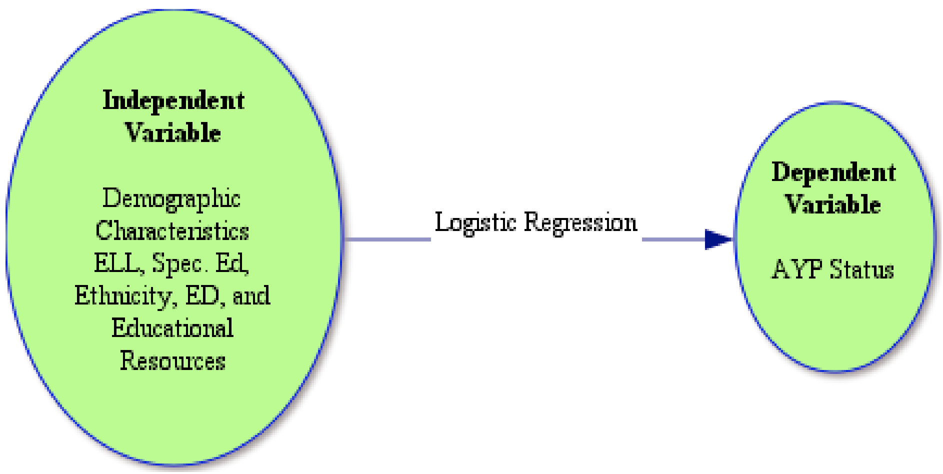
The independent variables in the second research question are: the schools in the sample states that achieve proficiency on the NAEP reading and mathematics assessments at the fourth- and eighth-grade level for the years 2005 and 2007, the educational resources that a school receives, and the demographic characteristics of schools, including socioeconomic status, ethnicity, special needs, and ELL status. The dependent variable is where or not the schools in the sample states meet proficiency status on the state accountability assessments (STAR, MEAP, End of Grade Test, and TAKS) in reading and mathematics at the fourth- and eighth-grade level for the years 2005 and 2007.

Figure 2. Independent and dependent variables for research question No. 2



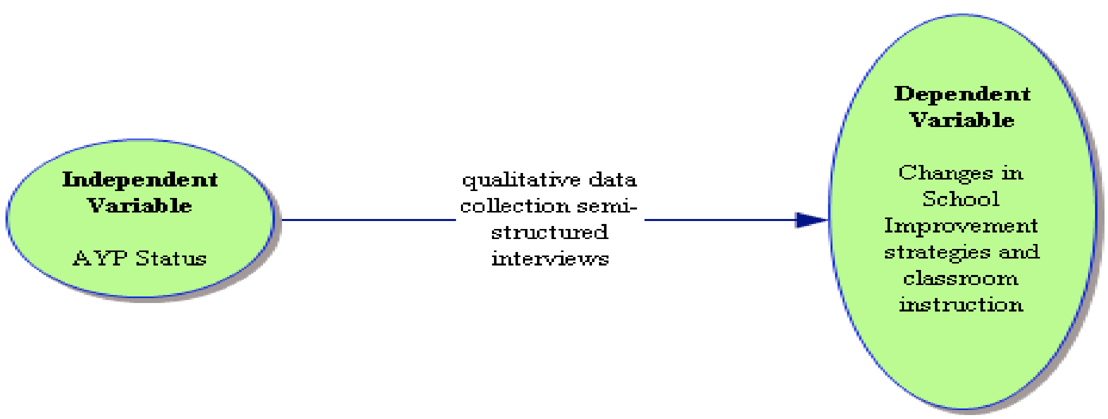
The variables in the third research question include an examination of the demographic characteristics of schools that meet AYP standards and fail to meet AYP standards during the 2005 and 2007 years in the four sample states. The independent variables are the educational resources that a school receives along with the demographic characteristics of schools, including socioeconomic status, ethnicity, special needs, and ELL status. The dependent variable is the AYP status of schools.

Figure 3. Independent and dependent variables for research question No. 3



The fourth research question involves an analysis of the impact that AYP is having on school improvement and classroom instruction. The AYP score in the sample schools is the independent variable. The dependent variables are potential changes in school-improvement efforts and classroom instruction that are a result of AYP. If AYP is working, then it is logical to assume that schools have implemented strategies that led to improvement.

Figure 4. Independent and dependent variables for research question No. 4



Significance of the Study

The NCLB reform and the AYP accountability provisions have an impact on all public schools throughout the United States. The reform provides high-stakes incentives for schools that make AYP and consequences for schools that fail to make AYP. This study is significant because it measures the characteristics of schools that make AYP and do not make AYP. Researchers question the validity (Mathis, 2006; Ravitch, 2010) and the reliability (Elmore, 2002; Harris, 2007; Hess & Petrilli, 2006; Kane, Douglas, & Geppert, 2002; Linn & Haug, 2002; Sunderman et al. 2005) of using AYP as a measure of school effectiveness. This study examines the issues associated with AYP through a comparison of AYP assessment data and NAEP data. An analysis of the relationship between NAEP and AYP in a sample of schools within states provides information regarding potential reliability issues that are associated with the AYP measurement. Moreover, the qualitative data enriches this study by producing a means to analyze how schools are reacting to the high stakes accountability measures that are mandated by NCLB. Furthermore, it provides information on the positive and negative impacts that NCLB is having on schools in the sample. The qualitative data gathered through interviews with principals and teachers in a sample of schools reveals the effect that AYP is having on the implementation of school improvement strategies and classroom instruction.

In addition, this study provides scientific data regarding the reliability of using the current AYP formula along with information on the characteristics of schools meeting AYP. This data should be valuable to legislators and school officials as the AYP provisions in NCLB are debated and reauthorized in Congress. The significance of the NCLB legislation is far reaching and the specific provisions in the reform are expected to be around for several years to come. However, because the research in the field (Borowski & Sneed, 2006; Elmore, 2002; Harris, 2007; Kane et al. 2002; Lewis, 2006; Sunderman et al. 2005) questions the use of the AYP formula, the information from this study should provide legislatures with important information about potential changes that are needed with NCLB and the AYP provisions. This information is useful in determining if the NCLB reform is effective with the achievement of the goals that were set out by the legislature and implemented by government officials. A critical issue that needs to be addressed in NCLB is the alignment of the AYP measurement with state or national standards. This study addresses the impact of school accountability assessments by estimating the correlation between school achievement on the NAEP and school achievement on the state proficiency assessments in the sample states.

Limitations of the Study

NCLB is a large and complicated piece of legislation (Sunderman et al. 2005). This study focuses on a small, but important portion of the reform. The examination of the AYP provisions in this study is somewhat limited in scope. One limitation in this study is the use of four sample states. The choice of the four sample states is due to the magnitude in managing a study involving additional states.

There is a limitation with four schools that were sampled for the fourth research question during the examination of the impact that AYP is having on classroom instruction and school improvement initiatives. The data may not be generalized to the larger population of schools across the state or the country due to the small sample size. However, the qualitative interviews did yield rich and important information about school improvement initiatives in the school building and classroom level.

Important Definitions

The information in Table 1 identifies important terms addressed in this dissertation proposal. These terms will be referenced throughout the study.

Table 1

Important Definitions

Important Terms	**Definition**
Adequate Yearly Progress (AYP)	The statistical formula that is used in NCLB to evaluate schools, school districts and state-level effectiveness. The measurements include assessment in English/language arts, mathematics, science, and one other additional measure. All students must meet proficiency standards as a whole and in the disaggregated subgroup populations in order for a school or school district to meet AYP requirements.
Confidence Interval	A range of values that we are confident but not certain that it contains the population parameters (Hinkle, Wiersma & Jurs, 2003). The confidence interval is constructed at the confidence level which is determined by the researcher and is often 95% in the social sciences. This means that if sampled numerous times the data estimate would reflect the true populations 95% of the time. Thus, there is a 5% margin or error.
Demographic characteristics; otherwise known as categorical populations or subgroups;	NCLB defines specific populations of disadvantaged students that must be disaggregated within a school or school district in order to determine their AYP proficiency rate and AYP participation rate. If a school meets standards as a whole school but fails to meet standards in one subgroup, the entire school is labeled as failing to meet AYP standards. The specific populations that are analyzed in AYP include: special education, ELL, economically disadvantaged (ED), and students from major racial or ethnic subgroups.
Educational Resources (school finance)	The resources that are available to a school or district which is defined as the amount of money that the school receives for each student in order to pay for the cost of education in the school or district.
External validity	Generalizing studies or measures to other samples of the same population (Rudestam & Newton, 2007, p. 114).

Major racial or ethnic subgroups as they are defined by NCLB.	NCLB uses the national definition to include the following populations, Black/African American, Latino/Hispanic, American Indian/Alaskan Native, Asian/Pacific Islander, and Caucasian.
Margin or error	Expresses the amount of random sampling error. A larger margin or error provides less confidence that the results reflect the true population.
National Assessment of Educational Progress (NAEP)	This assessment is administered in all states by selecting schools in a random manor to be assessed in specific content areas. The grades assessed are generally 4th, 8th, and 12th. (National Center for Education Statistics [NCES], 2009). For the purpose of this study, the focus of the NAEP data will be in reading and mathematics for grades 4 and 8.
NCLB sanctions for failure to meet AYP	There is a series of six improvement steps that schools or school districts must take in order to improve their AYP status. Please refer to Appendix A.
No Child Left Behind (NCLB)	The Federal legislation that was a reauthorization of the ESEA (1965). The legislation mandates high standards of achievement for all students through testing and accountability measures.
Reliability	The ability of a measure to produce consistent results (Rudestam & Newton, 2007, p. 96).
Safe harbor	Schools that do not meet the target can still be considered to have achieved Adequate Yearly Progress if they meet the Safe Harbor criteria, which is generally a 10% decrease of the percent of students not proficient in comparison to the previous year. For example, if a school had 50% proficient in 2009 and then 60% proficient in 2010, they would meet AYP standards even though the 60% proficient was below the overall cut score of for example 72%. This is because the result was a 10% reduction in the number of students not proficient from the previous year. Each state implements a defined safe harbor formula. However, the above definition provides a hypothetical example with the general parameters for the Safe Harbor provision.
School accountability	Teachers, administrators and schools are responsible for student performance as it is measured on achievement tests in academic areas. The school performance is evaluated when it is compared to a standard that defines target levels of achievement set by state or local governments. The system will often involve rewards, sanctions, and intervention strategies that motivate schools to improve performance. (Abdelmann et al. 1999; Ladd, 1996).

School improvement strategies and initiatives	Changes in the implementation of school processes that include curriculum, assessment, and instruction on a school or district-wide level with the purpose to increase academic achievement.
School reform	"Planned efforts to change schools in order to correct perceived social and educational problems. Sometimes broad social crises triggered school reforms and sometime reforms were internal improvements initiatives by professionals" Kuban & Tyack, 2005, p. 4).
Social capital	Parents' education level, the values the family places on education, the socioeconomic status, the peer group influence, and similar assets and liabilities (Elmore, 2002; Mathis, 2004).
Stratification sampling techniques as they are defined by the National Center for Education Statistics (NCES)	The division of a population into parts, or strata, each of which is more homogeneous than the population as a whole. If sample sizes for these strata are set proportional to the stratum share of the population, then the resultant sample will be more efficient than a simple random sample of the population disregarding the strata, as the simple random sample will have resultant sample sizes for each stratum that are randomly smaller or larger than the stratum share (too much in one stratum, too little in others, by chance).
Value added (growth model)	A measurement of the amount of progress that students make from one year to the next. The measurement usually occurs on an annual basis. Schools receive credit for student progress that occurs on state standardized assessments while taking account of the level at which students enter the school. Thus, students may not meet the achievement standard according to the status (cut score) measurement but they would be proficient if they make progress on trajectory toward status level proficiency.

Organization of the Dissertation

Chapter Two of this dissertation provides an examination of relevant research studies in the field that analyze the use of AYP as a school accountability formula. The research that is referenced includes empirical studies and theoretical references from well-respected researchers who address the implementation of AYP by the federal government. The purpose was to examine the research in order to synthesize a new perspective that supports the design of this study (Boote & Beile, 2005; Galvan, 2006).

Chapter three provides a detailed description of the research methodology. This methodology includes a mixed-methods approach with the collection and analysis of both quantitative and qualitative data. Chapter three is divided into two subsections, one addressing the quantitative methods and the second addressing the qualitative methods. The quantitative data analysis provides answers to the first three research questions while the qualitative data analysis provides answers to the fourth research question. Rudestam and Newton (2007) argue that a mixed model of quantitative and qualitative methodology combines the rigor and precision of the quantitative data with the depth of understanding from the point of view of those who live through experiences with the qualitative data.

Chapter four is a presentation of the results. Similar to chapter three, it is divided into two subsections, one addressing the quantitative results and the second addressing the qualitative results. Chapter four also provides a detailed description of the AYP formulas that is implemented by each of the four sample states. Chapter five provides concluding remarks on the overall study, recommendations for improvement with AYP, and areas that are recommended for analysis with future studies.

CHAPTER II

LITERATURE REVIEW

The NCLB reform aims to hold educational agencies and states accountable for improving the quality of education for all students. It further aims to identify and transform low-performing schools that have failed to provide a high quality education to their students, into successful schools. The accountability provisions in NCLB intend to close the achievement gap between high and low achieving students and especially the achievement gaps between minority and non-minority students along with the advantaged and disadvantaged students. The reform expects to accomplish this goal using state assessment systems that are designed to ensure that students are meeting state academic and grade level content expectations (NCLB, 2002, section 101).

The implementation of NCLB involves the use of high standards that are measurable for all students. There is no doubt that NCLB has provided for an increased focus on student populations that have traditionally performed at low levels (Borowski & Sneed, 2006; Guilfoyle, 2006; Haycock, 2006; Hess, 2006; Hess & Petrilli, 2006; Kane et al. 2002; Lewis, 2006). The measurement tool used to accomplish this is AYP. However, there are some faults with the AYP measure and there is a question as to whether or not AYP will be able to provide an accurate assessment of the goals that are stated in the Title One Purpose Statement of the NCLB legislation. (Appendix B provides a copy of the purpose statement that is found in NCLB). Some of those faults include states being allowed to develop their own standards, test score proficiency levels, and statistical measurement formulas under AYP (Harris, 2007; Olson & Jacobson, 2006; Popham, 2005a; Porter et al. 2005; Wiley et al. 2005). Cronin, Dahlin, Adkins, and Kingsbury (2007) found that fifty different educational measurement standards are implemented across the United States. Students in one state could meet proficiency standards on the state AYP assessment, while if their achievement was at the same level in another state, they would fail to meet the state AYP standards.

Review Focus

NCLB is a large and complex piece of legislation that focuses on public education. Chapter two specifically focuses on AYP and the accountability provisions that are found in Title One of the NCLB legislation. The review addresses the complexities involved with establishing school accountability in order to examine effective and ineffective provisions of the NCLB school accountability reform. The references include empirically based literature and theoretical literature that is available in the field.

Organization of the Literature Review

The first section of the review provides information on the state accountability history from the four sample states in this study, 1) California, 2) Michigan, 3) North Carolina, and 4) Texas. The second section identifies the measurement concerns with the implementation of AYP. The literature is clear regarding inconsistencies with the implementation of AYP across the country (Harris, 2007; Olson & Jacobson, 2006; Popham, 2005a; Porter et al. 2005; Ravitch, 2010; Wiley et al. 2005). States have the ability to statistically manipulate their AYP implementation which could give a false impression to the public that AYP is a consistent measure of school effectiveness across the country. One of the most prevalent concerns identified in the literature centers on the lack of reliability with using a single high stakes test to measure school effectiveness.

The third section of the review deals with the benefits of AYP and the unintended consequences. The biggest advantage of the NCLB reform has to do with the philosophical intent that all children count including students in specific categorical subgroups (Borowski & Sneed, 2006; Guilfoyle, 2006; Haycock, 2006; Hess, 2006; Hess & Petrilli, 2006; Kane et al. 2002; Lewis, 2006). The last section summarizes the arguments that are presented in this review and it provides justification to conduct this study. Some of the areas that require empirical research include the impact of AYP on school improvement and classroom instruction as well as a measurement of success using AYP to improve school programs for the benefit of students.

History of Accountability in the Four Sample States

The four states that are included in this study -- California, Michigan, North Carolina, and Texas -- have a strong history of school accountability, which began prior to the implementation of the NCLB reform. Each of the sample states represents different regions of the country and two of the states (Michigan and North Carolina) have been implementing a growth model with their accountability systems for at least two years. North Carolina was one of the first two states to receive approval for the growth model in 2006 and Michigan received approval in 2008. Texas has since been added to the list of states that are approved for the implementation of a growth model starting with the year 2009. The remainder of this section provides information on the history of the accountability systems in the four sample states.

California

The origin of the school accountability system in California began in 1988 with the passage of Proposition 98 (California Department of Education [CDE], 2009). The state legislature mandated that a School Accountability Report Card (SARC) would be initiated in order to provide the public with information about public school performance. The mandated information that was included in the SARC included an assessment of student achievement in the areas of reading, writing, arithmetic, and other academic goals. In 1993, school of choice options for parents were included as part of the school accountability system. The California accountability system continued to evolve in the 1990s and in 1998 there was the addition of the achievement results by grade level from the assessment tool used by the school district and the statewide assessment (CDE). In the year 2000, the results of the state report cards were published on the internet and it was required that all school districts ensure that parents receive the SARC report.

Michigan

The history of the Michigan School Accountability System started in 1990 with the approval of Public Act 25, which initiated an accountability system (Education Policy Center [EPS], 2000). The accountability system included a mandate on school improvement initiatives by schools, the creation of a core curriculum and learning outcomes for all students, school accreditation, and an annual report that was required for completion by all schools. In 1995, the law was amended to include pupil performance on the Michigan Educational Assessment Program (MEAP), which was used as part of the school accreditation process in Michigan.

The accountability system in Michigan continued to evolve and in the year 2000, a task force was created to evaluate the accountability system and make recommendations for reforming the accreditation process in Michigan. The recommendations by the task force resulted in legislation that created the Michigan Education YES Report Card System that started in the year 2002. The evaluation tool used in the report card system placed a heavy reliance on the MEAP results, which was weighted as 2/3s of the grade. The MEAP was used to measure achievement status, achievement change, and achievement growth. The other 1/3 of the grade was based on eleven performance standards in the form of a report that schools are required to complete (Michigan State Board of Education, 2003). Michigan was one of the first states to implement the AYP formula prior to NCLB in order to meet the goals of the Improving America's Schools Act of 1994 (Center on Education Policy[CEP], 2005).

North Carolina

North Carolina has a long history of school accountability. There was an emphasis on the improvement of teacher quality along with high stakes testing accountability (Febey & Louis, 2008). North Carolina was known to have one of the most comprehensive accountability systems prior to NCLB. The framework for the accountability system in North Carolina was established in 1977 as students were required to take standardized tests. In 1984, there was a mandate for a basic education program with a core curriculum and assessment. In 1990, the school report card system was established along with End of Grade Tests for grades three through eight which began in 1992 (Febey & Louis, 2008). In 1995, there was the adoption of the Accountability, Basics, and Control (ABCs) Plan that started with the measurement of both student status levels along with student growth. It also provided incentives for schools to perform well on standardized tests and it provided support for struggling schools (Public Schools of North Carolina, 2006).

Texas

The foundation for the Texas Public School Accountability System that emphasizes student achievement began in 1990 with the adoption of the Texas Assessment of Academic Skills (TAAS) test. In 1993, schools were held accountable through a state grade, which was based on the measurement of student achievement on the TAAS along with graduation rates (Peterson & West, 2003). In 1999, the accountability system in Texas was revised to increase the emphasis on student measurements with standardized tests as it provided for a disaggregated analysis of student test data in specific subgroups. Campuses and districts received an accountability rating each year which was based on the percentage of all students that met proficiency standards along with the four student groups (White, Hispanic, African American and ED) that meet the proficiency standards on the state assessment in grades three through eleven. The rating also considered the overall student dropout rate and each individual student group (Texas Education Agency [TEA], 2006). Texas, like North Carolina, was generally regarded to have a strong accountability system (Peterson & West, 2003). The Texas accountability system that evolved in the 1990's is believed to be the foundation for NCLB and regarded as the birthplace of the accountability reform (Nelson, McGhee, Meno & Slater, 2007).

Measurement Concerns with the Use of AYP to Evaluate School Effectiveness

Statistical Issues, Manipulations of the AYP data, and Lack of Consistency across the Country

There are a number of problems associated with the usage of the current AYP formula in order to measure and evaluate school effectiveness. The first problem area concerns the manner in which statistical manipulations are occurring with the implementation of AYP across the country. Borowski and Sneed (2006) conclude that the manner in which AYP is determined is arbitrary in nature. States have the ability to lower standards and manipulate statistical measures of AYP that could result in the lack of improvement in instruction and student achievement. They conclude that the only benefit would be the public relations appearance that improvements have been met according to the final analysis of AYP. Borowski and Sneed emphasize that the sanction provisions were created at an ideological or political level instead of a scientific level that would assist in improving schools or school districts.

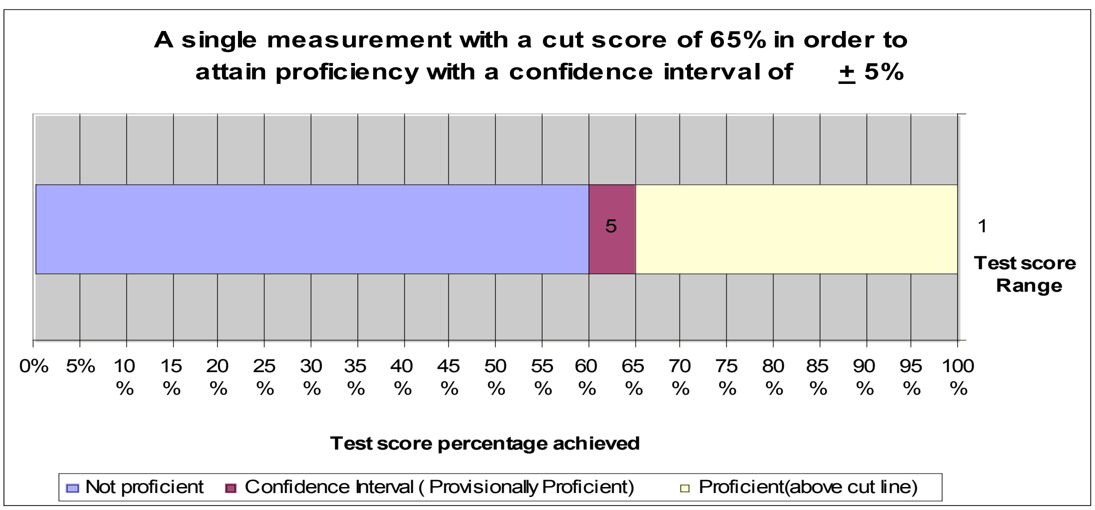
Porter et al. (2005) found that state flexibility in setting their own achievement tests has an impact on whether or not schools or school districts make AYP. The focus of their study was on three specific areas of measurement that have a major impact on AYP. Those areas include: the line of trajectory that states establish en route to a 100% proficiency rate by the year 2014, the minimum number of students that are necessary in order for there to be a subgroup population that will count for AYP purposes, and whether or not the state uses a confidence interval along with how wide the confidence interval is in determining if schools or districts reach the proficiency targets that are required under AYP. The implementation of NCLB by the USDOE provides states with flexibility in determining how they measure AYP in the three above-mentioned areas of measurement.

Confidence Intervals

Porter et al. (2005) provide empirical data on the number of states that are not using confidence interval vs. those states that are using confidence intervals and the size of the confidence intervals that they are using (Table 1 defines the confidence interval). As they examined the confidence interval decisions by states, they identified 2 states that use a confidence interval lower than 95%, 14 that use a 95% with a one-tailed approach, 1 uses a 98% trajectory with a one tailed approach, 16 use a 99% two-tailed confidence interval approach, 6 are unspecified, and 11 do not use a confidence interval.

Popham (2005a) explains that it is not statistically sound to implement the use of a confidence interval for AYP calculations. The manipulation of the confidence interval has a dramatic impact on the number of students and schools that pass proficiency tests and it influences the AYP status of schools. Figure 5 provides an illustration of the confidence interval which is hypothetical and not based on an actual state example. In Figure 5, the hypothetical cut score on the assessment is 65% with a confidence interval of $\pm$ 5% and the students who scored 65% or above are deemed to be proficient. The students who scored from 60% to 65% did not meet the proficiency cut score, but they fell within the confidence interval and they are deemed to be proficient due to the possibility of measurement error. When a state implements the use of the confidence interval students who have score within the confidence interval are calculated as proficient under the AYP formula.

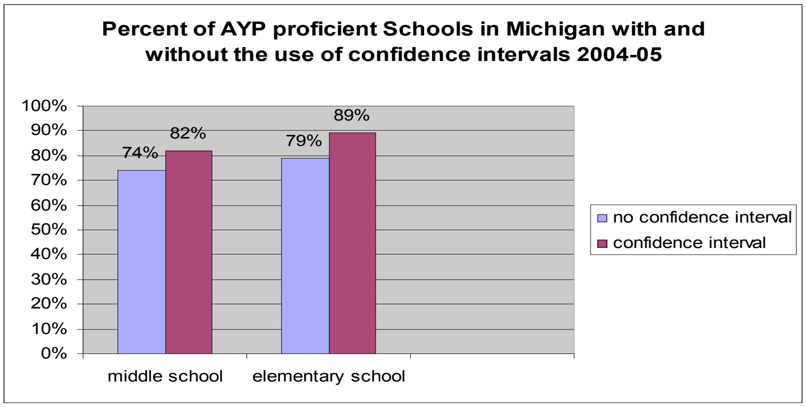
***Figure 5*. Confidence interval example**



Note. This is a hypothetical example of implementation of the confidence interval by states. The percentage proficient is not adjusted per the actual scale that states use as cut scores.

The end result is the perception that schools are successful according to AYP when in fact it might have been the implementation of the confidence interval that allowed them to meet AYP standards. For example, starting in the 2004-05 school year, Michigan began to use a confidence interval for determining proficient test scores on the MEAP (Center on Education Policy [CEP], 2007; Michigan Department of Education [MDE], 2008a). Thus, students who fall within the confidence interval qualify as proficient even if they do not meet the specific cut score on the state assessment. The Michigan Consolidated State Application for the implementation of AYP (2005) that was approved by the USDOE indicates that the use of the confidence interval resulted in more schools making AYP in Michigan (CEP, 2007). For example, in the year 2004-05 without the usage of a confidence interval in Michigan, 79% of elementary schools made AYP and 74% of middle schools made AYP (CEP, 2007). With the use of the confidence interval, 89% of elementary schools in Michigan made AYP and 82% of middle schools made AYP. Figure 6 illustrates that the use of a confidence interval can have an impact on the number of schools in a state that make AYP.

***Figure 6*. AYP proficient schools in Michigan.**



Note. The data was taken from the Center for Education Policy Studies (2007) report along with AYP data from the Michigan Department of Education.

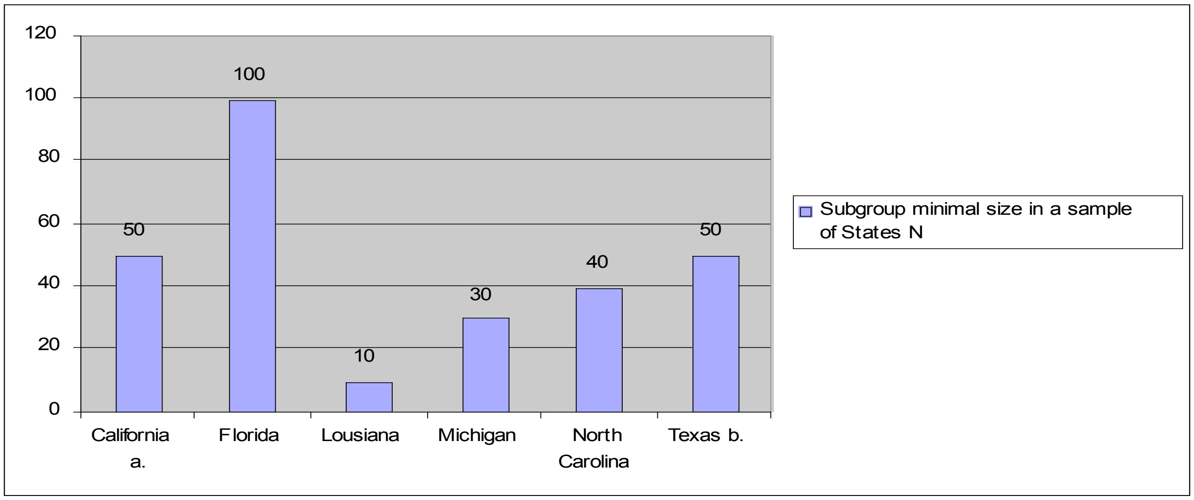
Subgroup Size (N)

Some studies (Meier, Kohn, Darling-Hammond, Sizer, & Wood, 2004; Olson & Jacobson, 2006; Porter et al. 2005) found that there are significant inconsistencies from state to state with the size (N) of subgroups that count for AYP purposes. For example, Olson and Jacobson found that the minimum size (N) for a subgroup in Louisiana is 10 in contrast to California where it is 50 and it must be an equivalent of 15% of the school population until the number reaches 100. The study includes data that was taken from the Associated Press for the 2003-04 school year enrollment figures. The study analyzed students in grades 3 through 8 and grade 10 since those are the grades that are tested each year under NCLB. The study examined five major subgroups that count under the NCLB AYP provision which includes: White, Black, Asian, Native American, and Hispanic. They did not look at the other subgroups such as ELL students, ED students, or special education students because there would have been the possibility that those students could have been counted more than once as they might fit into more than one category.

Based on the AYP data, Olson and Jacobson (2006) found that over 1.9 million students across the United States are not being counted for AYP purposes due to the minimal subgroup size requirement. This includes over 400,000 minority students in California that are not being counted. The number of students that were not counted in the five subgroups that were examined includes fewer than 2% of White students, 10% of Black and Hispanic, 33% of Asian students, and 50% of American Indians. They also found that over 93% of the students are included in both an overall school population and in a racial or ethnic category.

The findings by Olson and Jacobson (2006) with the statistical manipulation of the subgroup size are consistent with the findings by Porter et al. (2005) who concluded that different (N) size minimal subgroup calculations create inconsistencies from state to state in regard to the number of schools that make AYP. Porter et al. found that the subgroup sizes that states are using for AYP calculations range from 0 to 100 across the United States. There are nine states that use a minimal subgroup size that is greater than 42, fourteen states use a minimal subgroup size of 40, and twenty-seven states define a minimal subgroup of 34 or less. Figure 7 provides a sample of state minimal subgroup levels and it illustrates the disparities with the minimal subgroup size implementation among the states.

***Figure 7*. Subgroup minimal size in a sample of states.**



[a.] In California, the subgroup must equal at least 15% of the population otherwise the minimal could rise to 100.

[b.] In Florida, the subgroup can be as low as 50 students as long as they represent 10% of the school population.

Note. This data is available in the article by Olson and Jacobson (2006) and on the state department of education websites via their consolidated school AYP applications that were submitted to the United States Department of Education.

These studies (Olson & Jacobson, 2006; Porter et al. 2005) demonstrate that the federal government allows for an inconsistent measurement standard across the county, which produces misleading information to the public when it comes to AYP and the comparison of achievement levels among schools in different states. This misleading information includes identifying successful and failing schools in different states under AYP. These studies provide professionals and the public with important information and it gives the perception that statistical manipulation is being used so that students are not counted in specific subgroups. The referenced studies uncover one of the problems with the AYP calculation and the inconsistency from state to state.

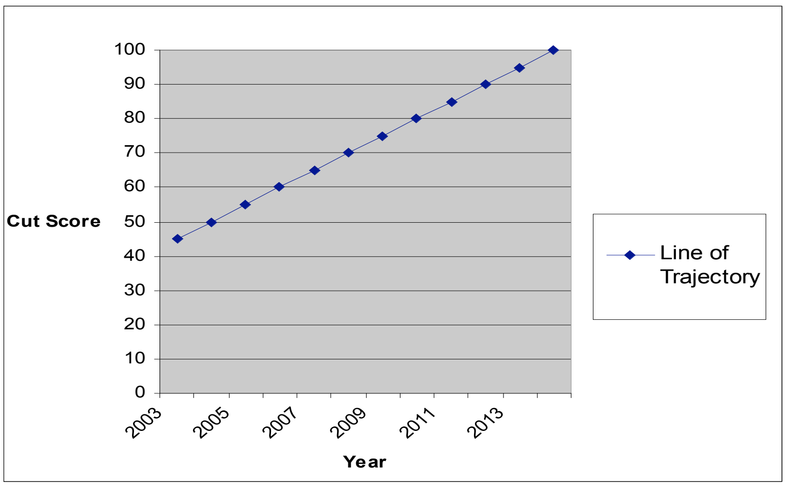
This dissertation study investigates a specific aspect of AYP. The study supports and refutes the referenced studies as to whether or not the AYP measurement provides a false or misleading impression with the labeling of failing and successful schools.

Lines of Trajectory

Past research (Popham, 2005a; Porter et al. 2005; Wiley et al. 2005) show that the manipulation of the AYP line of trajectory has a major impact on the number of schools that make AYP. Porter et al. provide an analysis of the number of states that are adhering to different measurement designs. They identify three specific forms or target lines of trajectory that states use with the evaluation of school effectiveness according to AYP. Porter et al. identified 24 states that use the back loaded trajectory, 19 that use the straight trajectory with plateaus, 4 that use the straight trajectory, no states use the front loaded trajectory, and three states had an unidentified trajectory. This further exemplifies the inconsistent measures that are being implemented across the country when it comes to determining the AYP status of schools. The following figures provide examples of the line of trajectory that states might choose to use en route to a 100% proficiency level by the year 2014. Figure 8 provides an illustration of the straight line trajectory, Figure 9 illustrates the straight line trajectory with plateaus, and Figure 10 illustrates the back end loaded trajectory.

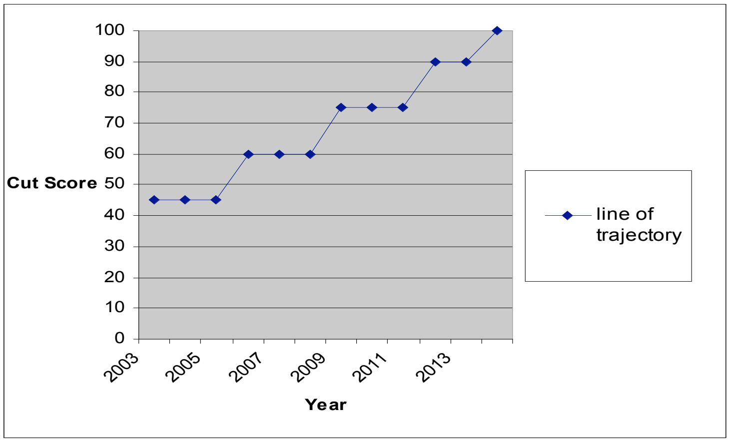
Mathis (2004) calls it a fallacy to believe that schools have the ability to make steady progress on achievement tests. According to Mathis, the implementation of the line of trajectory that is required by NCLB is not founded in any research. He argues that the implementation of the balloon effect (or back loaded trajectory as it is defined by Porter et al. 2005) where there is a moderate amount of progress made initially and then a high level of progress in the later years, is only delaying the inevitable that all schools will eventually fail (Figure 10 provides and illustration of the backend loaded trajectory). Mathis believes that changes through the reauthorization of NCLB might be the only solution to the problem that eventually all schools will fail by 2014.

***Figure 8*. Straight line trajectory.**



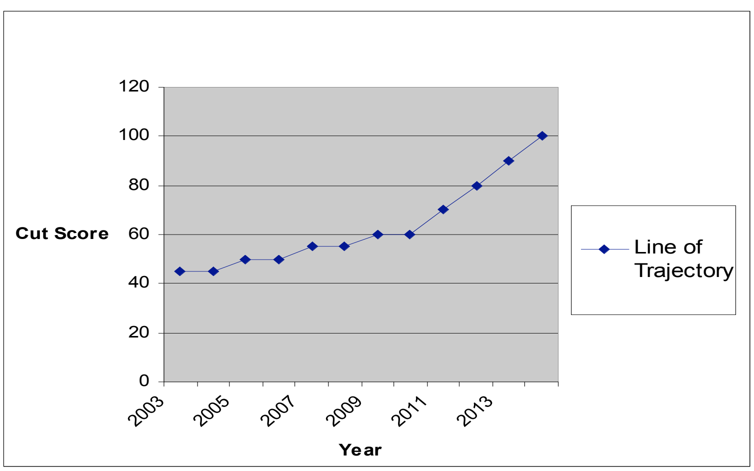
Note. This figure is based on the description presented by Porter et al. (2005) and Wiley et al. (2005) which is aligned with the example presented by the Minnesota Department of Education in their Consolidated State Application.

***Figure 9.* Straight line trajectory with plateaus.**



Note. This figure is based on the description presented by Porter et al. (2005) and Wiley et al. (2005) which is aligned with the Illinois example for their initial and final years of AYP implementation in their Consolidated State Application

***Figure 10.* Back end loaded trajectory.**



Note. This figure is based on the description presented by Porter et al. (2005) and Wiley et al. (2005) which is aligned with data presented by the Michigan Department of Education (2008) in their Consolidated State Application.

Cut Score Manipulation

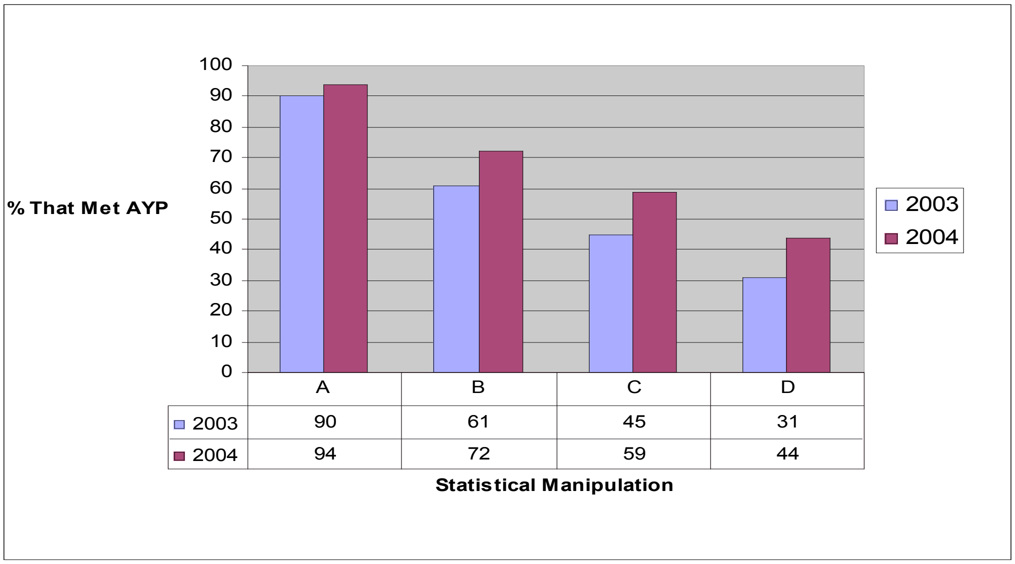
Studies and researchers in the field (Darling-Hammond, 2007a; Darling-Hammond, 2007b; Guilfoyle, 2006; Harris, 2007; Ravitch, 2010; Sunderman et al. 2005) call into question the use of a single measurement cut score with the analysis of school effectiveness. Harris (2007) examined the use of cut scores in the NCLB accountability measures using data sets from Michigan and Florida to illustrate how the different levels of cut scores can affect the number of students that are deemed high performing according to their poverty levels. Harris shows that when states use different cut scores like those in Michigan and Florida, the level that is used can directly impact the number of students that will be deemed proficient. Harris found that if a state uses a cut score that is halfway between the mean achievement of high minority students and low minority student populations, the result can be a dramatic shift in the numbers of students who do not make the proficiency level in the low minority schools. Harris discovered that states can manipulate the number of schools and students that are proficient by changing the cut scores for the achievement of standards on their state assessments. The result is a dramatic impact on the proficiency averages of schools that have high percentages of poverty and minority students. The motivation for states to do this would be the perception that school accountability is working in their state as more students and schools are deemed proficient under AYP.

The cut score manipulation that states might implement can provide the appearance that gains have occurred without any improvement in the conditions of school programs or classroom instruction. The findings by Harris identify another method in which states can use to modify the percentage of schools that meet AYP standards that is not a result of improvements in school quality. The manipulation of the statistical data can have a dramatic impact on AYP scores.

Lack of Consistency with AYP Across the Country

The implementation of the AYP accountability provisions in NCLB provide for a low level of consistency among the 50 states due to the wide variety of statistical methods that states use to calculate AYP. Porter et al. (2005) analyzed how the state of Kentucky was implementing their AYP data according to subgroup size, confidence intervals, and the line of trajectory. They found that the state manipulation of those statistical measures in combination can have a dramatic impact on the number of schools that made AYP in Kentucky. For example, Kentucky holds schools accountable for subgroup achievement when they have a minimum of 60 in the subgroup, they use a two-tailed 99% confidence interval, and they have a back loaded trajectory. With the implementation of those measurement procedures, in 2003 the state of Kentucky had 90% of their schools meet the AYP requirements. In 2004, 94% of the schools met the AYP requirements under the same measurement procedures. When the researchers eliminated the use of the confidence interval, they determined that only 61% of the schools in Kentucky would have made AYP in 2003 and 72% would have made it in 2004. When they further created a straight line trajectory and eliminated the confidence interval, the researchers determined that only 45% of schools would have made AYP in 2003 and 59% would have made AYP in 2004. When they modified the minimal number of students for subgroups to be disaggregated from 60 to 30, used the straight line of trajectory, and eliminated the confidence interval, they determined that 31% of schools would have made AYP in 2003 and 44% in 2004. Figure 11 illustrates the data findings by Porter et al. with significant changes that occur as the state modifies the statistical methodology for the achievement of AYP in the state of Kentucky for the 2003 and 2004 school years.

***Figure 11.* AYP Proficient schools in Kentucky with statistical manipulation**



[a] Includes the use of the confidence interval, minimal subgroup size of 60, and a back end loaded trajectory.

[b] Includes the use of the minimal subgroup of 60 and the back end loaded trajectory with no confidence interval.

[c] Includes the use of the minimal subgroup size of 60 with a straight line trajectory and no confidence interval.

[d] Includes a minimal subgroup size of 30 with a straight line trajectory and no confidence interval.

Note: This figure is based on data that was presented in the study by Porter et al. (2005) along with information from the Kentucky Department of Education.

A number of implications are apparent from the results of the study by Porter et al. (2005). The first implication is that there are different standards that are being implemented by different states in order to measure the AYP accountability provisions in NCLB. The result is a low level of consistency between the states when it comes to setting AYP requirements. Another implication is the different measurement designs that states use with the evaluation of AYP has a dramatic impact on the number of schools that meet AYP requirements. This includes making changes in the confidence interval measurement, the number of students in a specific subgroup, and the type of trajectory that a state uses. It is difficult to perform a fair comparison of the AYP data from state to state without examining the measurement decisions that a state has taken while evaluating their schools according to AYP.

Since failure to meet AYP comes with a cost through the implementation of sanctions that are required by NCLB for Title I schools, it is possible that states would continue to choose less stringent designs when measuring schools according to AYP. The Porter et al. (2005) findings had an impact on the design of this dissertation proposal that deals with measuring AYP and examining the impact that AYP has on school improvement and classroom instruction. This dissertation study supports many of the studies in the field that point to the inconsistencies with the use of AYP as a reliable nationwide accountability measure.

Reliability Issues with a Single Measure Accountability System

There is a consensus among researchers in regard to the reliability issues that are associated with the use of a single measurement to evaluate school effectiveness. Several studies and theoretical articles (Darling-Hammond, 2007a: Darling-Hammond, 2007b; Elmore, 2002; Harris, 2007; Hess & Petrilli, 2006; Kane, Douglas, & Geppert, 2002; Linn & Haug, 2002; Ravitch, 2010; Sunderman et al. 2005) call into question the reliability of a single test method for the evaluation of school effectiveness. Elmore argues that there are many technical difficulties with the reliance on a single test to evaluate school effectiveness and the achievement of state curriculum standards. Elmore claims that the creation of an accountability system at the state level was essentially a political act. He believes that the push for state accountability at the local level is a risky business for state politicians. Since most politicians do not have expertise when it comes to high stakes assessments, they are unaware of the reliability issues that are involved in using a single measurement to assess an entire school or district.

Harris (2007) agrees with Elmore (2002) when he concludes that the current accountability system is a measure of the minority and poverty level of students in a school vs. the academic achievement that the school is providing. Harris criticizes reports that use one year of data to conclude that high poverty schools and/or high minority schools are high achieving. Harris found reliability issues with using one year of measurement to classify schools as successful or failing. Kane et al. (2002) also point out a weakness with using a single year test to measure school effectives because different cohorts of students are measured against each other in order to determine the school AYP status. The study by Kane et al. emphasizes that there are different variables at play within the cohort groups which ultimately have an impact on standardized test achievement. The result is unreliable data when it comes to measuring the effectiveness of schools.

High Flying Schools and the One Year Phenomena

Past research (Harris, 2007; Kane et al. 2002) found that there can be fluctuations in annual test scores by schools in a single year that have nothing to do with gains in school performance. The study by Kane et al. found that school enrollment has an impact on the statistical probability that a school meets accountability status. Larger schools are more stable when comparing annual test scores. Smaller schools tend to experience greater fluctuations in year to year scores. Thus, measuring school progress according to annual gains based on a cut score design is statistically unreliable as the size of the school has an impact. “With test scores being so volatile, school personnel are at a substantial risk of being punished or rewarded for results that are beyond their control" (Kane et al. p. 60). Linn and Haug (2002) found that random fluctuations often resulted in schools being recognized as outstanding and others that were labeled as needing improvement which had nothing to do with changes in the quality of education and performance at the school level.

Harris (2007) provides a rebuttal to research reports by the Education Trust and Heritage Foundation that have attempted to identify high-flying schools. High-flying schools are defined as those schools that help students reach high levels of achievement despite high poverty levels or high levels of minority students. Harris evaluated the data that was used by the Education Trust study to define schools that are labeled as high-flying. He analyzed the high flying definition that was used by the Education Trust at 1-1-1. This means that the school needs to achieve at a high level in either year of the study, in any of two subjects, and in any of two grades that were measured. He then analyzed this data in a graphic form by making more stringent requirements with the use of the classifications 2-1-1, 2-2-1, and the 2-2-2. For example, the last definition 2-2-2 means that schools needed to perform at a high level for two years, in two subjects, and in two grades.

Harris (2007) criticizes the information that was provided in the Education Trust study, which defines schools as high-flying while only taking into account a single year data set. When multiple year datasets are used that take into account multiple tests at different grade levels, Harris found that schools with low levels of poverty are 22 times more likely to become high performing when compared with the high poverty schools. He also found that low poverty and low minority schools are 89 times more likely to become high performing vs. a high poverty and high minority school. The findings by Harris point out another problem with AYP as there is the possibility of the statistical chance where some schools will do well on a single assessment for an individual school year. However, the probability of achieving that same result for multiple years and over multiple tests is less likely to occur which reduces the possibility of statistical error. This is one of the reasons why two years of data (2005 and 2007) were selected for this study.

AYP Measures Social Capital, Minority Status, and Racial Subgroups

Literature in the field (Elmore, 2002; Harris, 2007; Wiley et al. 2005; Maleyko & Gawlik, 2011; Meier et al. 2004; Mathis, 2004a; Reeves, 2006) has shown that poverty and minority status have a major impact on the achievement levels of students in schools. Elmore (2002) points to the implementation of AYP as measuring the social capital of students in schools. Elmore argues that standardized test scores show the social capital (Table 1 defines social capital) of the students in schools. Most high-performing schools elicit higher performance by relying on the social capital of their students and families rather than on the internal capacity of the schools themselves. Most low-performing schools cannot rely on the social capital of students and families and instead must rely on their organization capacity. Hence, with limited investment in social capacity, low-performing schools get worse relative to high-performing schools.

The primary focus of the NCLB reform is the belief that school accountability measurement and sanctions will motivate students and schools. The argument that Elmore (2002) frames leads to the conclusion that while the intent of NCLB is to improve schools for underperforming students; it will actually work in reverse as it broadens the gap between the schools with a high social capacity and a low social capacity. Elmore indicates that the NCLB philosophy is based upon a faulty premise. The accountability sanctions do not provide organizational capital to the schools that require this capital in order to achieve a level of success with underperforming students.

The identification of what Elmore (2002) refers to as social capital is an interesting perspective because he argues that in order to improve schools, there is a need to improve the social and organization capital of the underperforming schools. He maintains that the social capital of the students and their families is what is being measured by the standardized, single measurement accountability system. The result is a need for extensive studies in order to analyze his hypothesis along with an analysis of the possible positive and/or negative impact that the NCLB accountability provisions are having on public education in the United States. This dissertation study examines some of the concerns that Elmore poses with the collection and analysis of empirical data.

Researchers (Mathis, 2004b; Meier et al. 2004; Wiley et al. 2005) argue that schools cannot do everything to fix societal problems. Mathis (2004a) argues that the schools cannot create other conditions in society where the students attend school with high levels of readiness, motivation, a high level of academic ability, and support from the home. He states that the use of test scores to close the achievement gap is an ideological belief rather than a belief that is founded in research. Mathis further states that not all students start at the same level. He argues that it is unrealistic to believe that all students will be able to achieve the same proficiency levels. Mathis refers to personal capital which is the readiness and ability to achieve on tests. In conjunction with Elmore (2002), Mathis argues that factors outside of the school known as social capital have an impact on student achievement. He defines social capital as a combination of the following: parents' education level, the values the family places on education, socioeconomic status, the peer group influence, and similar assets and liabilities. Mathis argues that there is a great gap in the educational equity in schools in the United States as he cites a study by the Organization for the Economic Cooperation and Development which ranked the United States twenty-second out of twenty-five industrialized countries when it comes to educational equality. As schools are labeled as failing, the NCLB reform will help to distinguish the disparities among the high poverty students and low poverty students (Mathis, 2004a).

Harris (2007) found that African-American kindergarten students are achieving at a rate of 34 percentile points below the levels of white kindergarten students. He believes that the current political climate has evolved in a manner in which school systems are blamed for poor levels of student achievement. He states that the politicians are neglecting to look at other factors that impact student achievement. Harris argues that a fault in NCLB is the failure to account for learning growth by students which is in contrast to the sole measure of cut score proficiency achievement. He asserts that the AYP measurement system cannot be solely attributed to learning in schools. Harris (2007) illustrates his position with the following quote:

> Consider the foot-race analogy made by President Lyndon Johnson when he argued for affirmative action and compensatory education. Johnson said that undernourished students would lose the vast majority of the running races, not because other students or track coach failed to try hard enough, but because the students were undernourished. (p. 373)

This analogy illustrates that other variables are at play when students come to school with inadequate social capital. The problem with AYP is that it does not take into account the amount of social capital or the starting point at which the students enter school.

Measuring English Language Learners (ELL)

Researchers (Abedi, 2004; Darling-Hammond, 2007b; Guilfoyle, 2006; Meier et al. 2004) propose the need to make revisions in NCLB when it comes to measuring the achievement levels of ELL students. The USDOE needs to consider changes with the ELL subgroup because the current system is not fair to schools with large ELL populations as it provides an artificial ceiling on ELL achievement, making it statistically impossible to achieve 100% proficiency by 2014. (Darling-Hammond, 2007b; Guilfoyle, 2006; Meier et al. 2004). It would be prudent to allow students to remain in the ELL subgroup for their entire educational careers which would allow for the tracking of those students over time and give credit to schools that have success with ELL students. Since the purpose of ELL programs is to provide students with English language proficiency, allowing them to count in the ELL subgroup throughout their public education years would allow for a long term evaluation of ELL programs. ELL students are currently only allowed to count in that subcategory for two years following the attainment of proficiency (Darling-Hammond, 2007b; Guilfoyle, 2006; Meier et al. 2004).

There is a need to change the NCLB provision which requires that students are tested after one year in the country. The research on English language acquisition demonstrates that it takes much longer than one year in order to acquire academic language. Research (Hakuta, Butler & Witt, 2000; Pray & McSwan, 2002) has shown that ELL students require anywhere from three to seven years to achieve grade-level academic performance in English. Abedi, Courtney, and Leon (2003) found that there are serious reliability issues with using standardized assessments with ELL students. When the determination was made to include ELL students in the AYP formula, the politicians did not follow scientifically based research which has shown that it takes much longer than one year for ELL student to achieve proficiency in English. The result is that ELL students are expected to perform at levels that are equal to those students whose primary language is in English.

Inevitability that a High Percentage of Schools will be Labeled as Failing

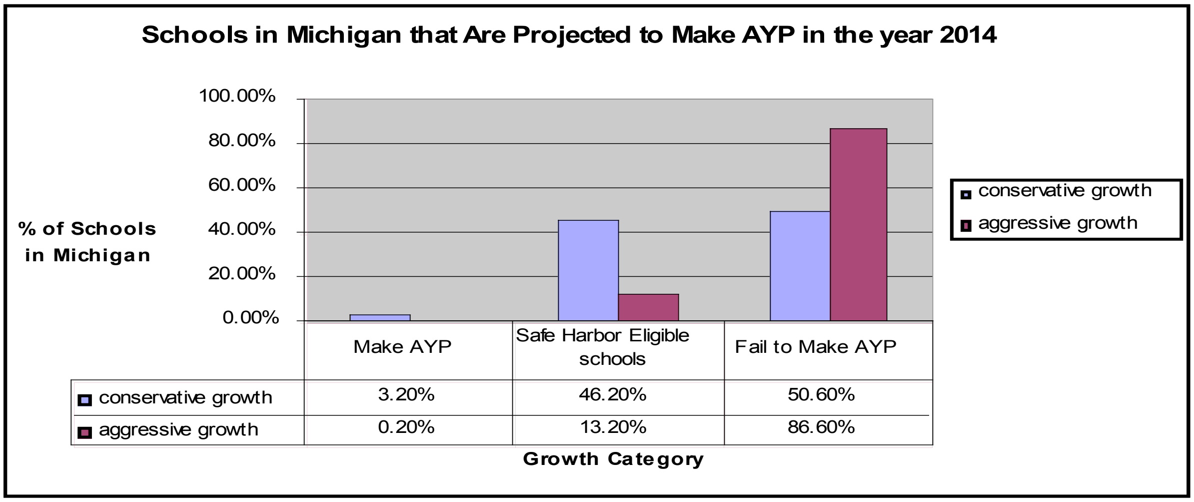
Wiley et al. (2005) found that a large number of schools will be labeled as failing by the year 2014. Their research is based on an examination of six states in the Great Lakes Region; 1) Illinois, 2) Indiana, 3) Michigan, 4) Minnesota, 5) Ohio, and 6) Wisconsin. The authors found that due to the different requirements and parameters that are mandated by AYP in each of the six states, each state exhibits a different but increasing pattern of having more schools not make AYP by the year 2014.[2]

Even with the most conservative estimate, Wiley et al. (2005) found that the number of failing schools is projected to total well over 50%. For example, in Michigan they found that in 2004, 25% of the schools failed to make AYP. Michigan uses the stair step trajectory method until 2010 when it became a back-loaded trajectory. The minimum size for a subgroup is 30 and there was no confidence interval that was used by the state in 2005. According to the more conservative growth estimates, Michigan is predicted to have 3.2% of schools make AYP in 2014, while 46.2% will be safe harbor eligible (Table 1 provides a safe harbor definition), and 50.6% will not make AYP and will not be safe harbor eligible. With a more aggressive or low growth prediction, the averages in the same order change to 0.2%, 13.2%, and 86.6%. The sharp decline in schools that will not make AYP was predicted to occur in 2010 for Michigan due to the line of trajectory design. Figure 12 illustrates the percentage of schools that are projected to make AYP and fail to make AYP in the year 2014.

---

[2] The Wiley et al. study did not take into consideration the implications of the subgroup issues or the integration of the science test into the AYP equation.

***Figure 12*. Schools in Michigan that are projected to make AYP in 2014.**



Note: This data is based on the study by Wiley et al. (2005).

The findings by Wiley et al. (2005) have major implications for the current state of the NCLB accountability system. As the proficiency bar rises, a large majority of schools are predicted to be labeled as failing by 2014 under the NCLB reform. When looking at the AYP trajectory that leads to 2014, a reasonable assumption is that the number of schools that fail to make AYP will increase similar to the predictions in this study. This conclusion is made by looking at the number of schools that have currently failed to make AYP and aligning it with the line of trajectory increase by each state until the year 2014. The findings in this study show that the only thing that could change that projection is revisions in Congress through the reauthorization process.

Researchers in the field (CEP, 2005; Darling-Hammond, 2007b; Hess & Petrilli, 2006; Lewis, 2006; Linn, 2003; Mathis, 2004a; Ravitch, 2010) argue that the target of having 100% of all students achieve high standards is statistically impossible and schools are set up for ultimate failure under NCLB. Mathis (2004a) references the fact that in 2003, 87% of schools in Florida and 57% in Delaware were labeled as failing schools. He concludes that an increasing number of schools will be punished as failing. The politicians in Congress need to look at these studies and similar studies during the reauthorization process in order to avoid a catastrophic phenomenon with the majority of schools in the United States getting labeled as failing due to the statistically impossible 100% proficient AYP provisions in NCLB.

Summary of Measurement Concerns with the Use of AYP to Evaluate School Effectiveness

An analysis of the empirical research and theoretical literature in the field has provided a consensus that there are numerous problems with the NCLB accountability provisions. Every state has the autonomy to set up their own AYP proficiency standards and state administered tests in reading, mathematics, and science. The reason for this autonomy is that it would have been politically difficult to get legislation passed in NCLB that provided a mandated consistent assessment for all students in all states. The constitutional power over education lies with the state governments. The federal government can not force participation in NCLB, but those states that fail to participate could forfeit their federal funding. Therefore it is difficult to use AYP to accurately compare the quality of schools in different states across the United States. The statistical manipulations of the AYP measurement systems at the state level provide the public with a false impression of school effectiveness. The research in the field leads to a conclusion that modifications to the accountability provisions in NCLB will be necessary in order for the legislation to effectively improve the conditions of school programs and classroom instruction on a consistent basis.

AYP Benefits and Unintended Consequences

Philosophical intent

There is a consensus with the literature in the field (Borowski & Sneed, 2006; Darling-Hammond, 2007b; Guilfoyle, 2006; Haycock, 2006; Hess, 2006; Hess & Petrilli, 2006; Kane et al. 2002; Lewis, 2006), that the most beneficial part of the NCLB reform is the philosophical premise that all children will count and no child will be left behind. Borowski and Sneed state the following:

> In our view, NCLB's most important potential benefits lie in its recognition of the fundamental premise that all children can learn and have a right to be taught, and the corollary principal that educators, parents and students should receive periodic assessments of how students are progressing toward the attainment of high academic standards. (p. 504)

Borowski and Sneed's (2006) arguments have been molded through the litigation of desegregation plans for schools in the United States. They often found that districts could not provide disaggregated data by race, students with disabilities, or the number of students in honors programs on state assessments. They assert that accountability systems have the potential to have a positive impact on public education as parents are informed about how schools are achieving on assessments and within particular subgroups. Borowski and Sneed (2006) state the following:

> NCLB's principal strength is its focus on improving students' academic performance, and doing so for all children of all backgrounds. The act requires that all states establish standards in reading, mathematics, and science and develop assessments to determine whether students are proficient in meeting the standards. NCLB requires that all students meet these standards by the 2013-2014 school year. (p.506)

The philosophy found in NCLB is a focus on high academic standards and success for all students in all areas and subgroups that have a great potential to provide educational equity

(Borowski & Sneed, 2006). The potential is powerful because it mandates improvement not just in the aggregate, but also for all subgroups through disaggregated data. Borowski and Sneed argue that the intent of the legislation must go beyond having the data disaggregated in order to inform classroom instructional practices with the goal to make improvements for all students in schools.

Haycock (2006) was one of the few researchers found in the literature which took an exceptionally positive stance in regard to the impact of NCLB on public education. Haycock states that while educators are often critical about the NCLB provisions, she has consistently found that based on her interactions with educators, they believe the legislation is having a positive impact on students. The biggest benefit is what she terms to be “no more invisible kids”. She argues that NCLB has provided a spotlight on the performance of poor and minority children including ELL and students with disabilities. She claims that the result is that schools are focusing more attention on the education of these students. Achievement for all students becomes critical under NCLB (Haycock, 2006).

Haycock’s positive views on NCLB are somewhat supported by Dee and Jacob (2009) who found that NCLB is having a positive impact on raising the achievement of mathematics scores across the county as it is measured by the NAEP. However, benefits in the area of reading could not be found. In their study, they used a control group and experimental group of states. The states that had a strong history of implementing accountability prior to NCLB were considered to be the control group, while the states that did not have a strong history of accountability prior to NCLB were the experimental group. The states from the experimental group had greater gains than the control group. However, some researchers criticize the methodology used in this study with labeling states as control and experimental (Zehr, 2009).

Consequences of Sanctions and Accountability

Impact on Subgroup Populations

Haycock (2006) found that at Centennial Place Elementary School in Atlanta, administrators, and teachers are using data and information to assess the needs of those students. Haycock claims that the school disaggregated the data due to NCLB and they were able to pinpoint a weakness with special education students. She posits that the result was growth in mathematics and reading that produced a 30 point gain. She also references Woodville Elementary in Richmond, Virginia where there was a 20 point increase in students meeting the standards on the state assessment. In order to support her conclusions from the schools that she references, Haycock refers to the Hall and Kennedy (2006) study which found that in elementary reading, 27 of 31 states showed improvement and 22 of 29 narrowed the gap between African American and White students, while 24 of 29 narrowed gaps between Latino and White students. In elementary math, 29 of 32 states showed improvement, 26 of 30 narrowed the gap between African American students and White students, while 22 of 30 narrowed the gap between Latino and white students.

Since Haycock (2006) takes a positive view towards NCLB, it was important to reference her work in order to provide an accurate picture of the wide variety of research and critical reviews of NCLB that are prevalent in the field. While there is an agreement in the literature with the author regarding the focus that NCLB has placed on subgroups populations in NCLB and the philosophical intent of the legislation, the data that is presented in this article is not as powerful as other empirical studies. Haycock (2006) claims that a 30 point growth was made by a school in Atlanta as a result of the positive impact that NCLB is having on the school.

However, there are numerous variables which could have an impact on the achievement of that school. Some of the previous literature has illustrated that the manner in which the states manipulate their variables like subgroup (N) totals, confidence intervals, and the proficient cut scores on assessments, could also have an impact on the annual gains that are referenced.

Poverty and racial diversity

Research in the field (Mathis, 2004a; Meier et al. 2004; Sunderman et al. 2005) show that the poor and ethnically diverse schools will be punished first through the sanction provisions. Rose (2004) found that smaller schools with White populations below 250 students have a much better chance of meeting AYP vs. larger schools with multiple subgroups. Schools with more subgroups and higher percentages of racial subgroups of students will have a greater chance of failing to meet AYP standards which will result in the diversion of funding from schools to pay for the mandated sanctions.

Porter et al. (2005) found that schools with a 25% or more Black student population in Kentucky were much more likely to not meet AYP proficiency at a proportion of 25 to 30%. They also discovered that schools with higher percentages of Economically Disadvantaged (ED) populations were more likely to fail to meet AYP proficiency. When they developed a demanding AYP design by eliminating the confidence interval and using a straight line trajectory, there were a greater proportion of schools (73%) with a high ED population that failed to make AYP. This was in contrast to only 30% of schools failing to make AYP with lower ED populations. A study by Springer (2008) found that schools with larger Hispanic populations were 8% more likely to fail AYP. Schools with larger ED populations were more likely to fail AYP by 9%. However, it is difficult to reliably conclude that these results are transferable to other states since there were lower numbers of racial subgroups and lower numbers of ED students in the state that Springer studied when compared to other states across the country.

Educational Triage

Educational triage involves focusing on the students who are just below the proficiency level in order to accelerate their achievement so that they meet the state standards. Those students are often labeled as the “bubble kids”. Guilfoyle (2006) argues that the NCLB accountability provisions provide an incentive for schools to focus on the students who are achieving just below the cut line in order to improve their high stakes test scores. Springer (2008) suggests that if educational triage is occurring in schools, then the students in the low categories have a less chance of passing and those students in the high categories might receive less attention and resources. The “bubble kids” benefit by receiving resources that might have otherwise been used for the highest or lowest achieving students.

The findings by Springer (2008) indicate that educational triage is not occurring in schools. Springer analyzed the test scores from over 300,000 students from a western state in the United States. He used the academic achievement data from the fall and spring state test scores in the area of mathematics. The state that he used for his analysis has two assessments that are given to students in one academic year as opposed to the one assessment that is mandated by NCLB. Springer could not find any evidence that schools were involved in educational triage with the reallocation of their resources for the bubble kids. Some of his findings include a conclusion that the lowest students gained more than those near the proficiency line for the schools that failed to make AYP during the previous school year. There was a 0.2 standard deviation gain vs. the students who were near the proficiency line. The students in the highest categories did not have a negative consequence as their achievement growth was also relative to the “bubble kids”. With the schools that made AYP the previous year, Springer found that the students in the lowest category from the previous year, showed the highest gains. He also found that proficient students in failing schools gained more than proficient students in non-failing schools. These results suggest that the higher achieving students and lower achieving students are not losing out on educational resources or interventions for the benefit of the “bubble kids”. However, since there are very small numbers or racial subgroups in the state that was studied, it would be difficult to conclude that these results were prevalent in other states with larger populations of racial or ethnic diversity. One weakness with the study is that Springer only used mathematics test data and not language arts or reading test scores which is an equally important part of the AYP accountability formula. As a result, the findings can only be attributed to mathematics instruction in the sample state.

Narrowing the Curriculum

Scholars posit that NCLB is narrowing the curriculum as it discourages the usage of instructionally useful forms of assessment that involve extensive writing and analysis (Darling-Hammond, 2007a; Ravitch, 2010). Some states were forced to eliminate the usage of sophisticated performance-based assessment systems that resemble those used in other nations around the world that score well on international assessments. Guilfoyle (2006) argues that testing and measurement is a major emphasis in NCLB and since mathematics and reading are being tested, there is an unprecedented focus on those two areas. She believes that the areas of the curriculum that are not being measured by NCLB are suffering as schools start to preclude or reduce their focus in those areas. This includes history, art, civics, music, physical education, health, and other cultural areas (Ravitch, 2010). Guilfoyle argues that although these areas are not being measured, they do have a tremendous impact on the education of students and they are extremely important. Since those areas are not being tested, students and teachers are not placing an emphasis on those content areas.

Guilfoyle also indicates that standardized assessments provide us with a limited amount of information. The testing provisions in NCLB are unable to provide information about prevalent instructional improvement. A weakness in this point of view is that empirical data was not presented in order to determine if there is a narrowing of the curriculum. However, since NCLB measures growth in mathematics and reading, it is possible that schools would place more emphasis on mathematics and reading curriculum while reducing the other areas of the curriculum in order to account for the emphasis on reading and mathematics.

Borowski and Sneed (2006) state the following: “we have repeatedly observed that what gets measured gets done, and conversely, what is hidden is usually bad news” (p.505). This quote suggests that if things are not measured there is the potential for harm and since schools are only tested on mathematics, language arts/reading and science (starting in 2007), there is a lack of accountability with other curricular areas and the potential for neglect. Lewis (2006) references the Center on Education Policy Reports in making a claim that classroom instruction is being narrowed in focus. The result is NCLB created an educational system that failed to produce improved achievement results for schools (Lewis).

Sunderman et al. (2005) collected qualitative survey data from teachers in two school districts, Fresno, California, and Richmond County, Virginia. They found that the teachers acknowledged a narrowing of the curriculum as they indicated that more focus was given to the test areas in contrast to those areas that were not tested. The process of providing more focus on language arts and mathematics while reducing or eliminating other areas of the curriculum could be labeled as a form of curricular triage as other areas of the curriculum get neglected in favor of the areas where the school is being evaluated by the AYP accountability provisions.

Focus on Basic Skills vs. Higher Level Thinking and Innovation

A strategy that educators could use to improve tests scores on the NCLB accountability measurements is to teach to the test where the sole focus is on basic skills and test taking methods (Guilfoyle, 2006; Meier et al. 2004). This strategy can result in increases on standardized assessments. However, teaching to the test will not improve the school's overall educational program in order to prepare students for society. The United States public education system was created with the intent to support democracy and societal goals. A heavy emphasis on test taking methods will deny students the opportunity to engage in authentic learning opportunities or higher level critical thinking methods of instruction which will be more beneficial to them once they leave the K-12 educational system.

Schoen and Fusarelli (2008) along with Guilfoyle (2006) argue that there is a need to create a system that rewards educators for innovations in schools instead of the current system that places the utmost importance on mastery of core content and basic skills. Schoen and Fusarelli believe that the constructivist approach which supports the acquisition of the skills that are needed for students to become successful in the 21st century are in competition with the accountability provisions in NCLB that emphasize content based learning and the need to do well on a single measure. Schoen and Fusarelli along with Darling-Hammond (2007a) claim that there is a need to change the current assessment system that is supported by NCLB. They believe that new assessments must be designed through the reauthorization of NCLB. Those assessments should measure higher order critical thinking skills, problem solving, and the ability to process information in a relevant manner.

Schoen and Fusarelli (2008) reference work that has been done in Finland with the success that the country has achieved with the creation of an educational system that has produced some of the highest scores on the Paris-based Organization for Economic Cooperation and Development (OECD) international comparative study in 2000 and 2003. Some of the skills that are advocated under the 21st century schools movement include: a partnership with educators, businesses, and policy makers, an emphasis on critical thinking and problem solving skills, using creativity and innovation skills, emphasizing collaboration skills, using information and media literacy skills, and highlighting contextual learning skills. The NCLB legislation has narrowed the curriculum and essentially made it difficult for schools to promote innovation and move towards the ideals of the 21st century schools movement (Darling-Hammond, 2007a; Darling-Hammond, 2007b; Meier et al. 2004; Ravitch, 2010; Schoen & Fusarelli, 2008).

Admonishment for Individual Students

Schoen and Fusarelli (2008) point out that twenty-two states use a high stakes test in conjunction with student promotion. They reference data from Cleveland in 2007 where 43% of high school seniors failed to pass the Ohio Graduation Test and did not graduate. Researchers (Darling-Hammond, 2007a; Elmore, 2002; Schoen & Fusarelli, 2008) argue that the current high stakes testing system found in NCLB does not provide an accurate measurement of the type of knowledge and skills that are needed for success beyond K-12 education. Elmore states that the single most damaging element of the legislation is the threat of sanctions which includes sanctions against students. He argues that it is extremely damaging to use a single test to determine the graduation status or the advancement of grade levels for students because multiple measures are required for valid results.

Teacher Turnover

Sunderman et al. (2005) found that teachers from schools that are labeled as needing improvement indicated that they were more likely to transfer within a five year period vs. those teachers from schools that met the AYP requirements. The researchers conclude that teacher turnover might be an unintended consequence of NCLB as high quality teachers leave schools with high needs due to the sanction provisions in NCLB. This finding is consistent with the discovery by Mintrop (2004) who surveyed teachers in Maryland and Kentucky in order to determine their reasons for remaining in a school or leaving a school. Approximately 73% of the teachers in Kentucky listed the additional pressure of the probation status of the school as their number one reason for leaving, while 59% of the teachers in Maryland put the probationary pressure as their fourth ranked reason for leaving. Mintrop indicates that based on the results from his quantitative and qualitative study, highly performing and skillful teachers were just as likely to leave as less engaged and less skillful colleagues. Mintrop found that being on probation motivated some of the highly motivated teachers and it also upset many of them.

The findings by Sunderman et al. (2005) and Mintrop (2004) suggest that the accountability provisions in NCLB might further entrench the low performing schools into a continuous pattern of low achievement as teacher turnover increases because highly skilled teachers prefer not to deal with the pressure of probation and sanctions. This unintended consequence is damaging to schools in high need areas that are labeled as failing by NCLB. The research in the field shows that classroom instruction and teacher quality have a dramatic impact on student achievement (Darling-Hammond, 2007b; Hattie, 2009; Marzano, 2003; Nye et al. 2004; Sanders, 2000; Sanders & Rivers, 1996). The retention of highly skilled teachers is a critical component of a successful school especially in schools with high levels of ED students (Nye et al. 2004).

Financing AYP

State governments provide the majority of funding for public education. However, since states are struggling to raise tax revenue in these tough economic times, they have been forced to reduce funding for public education. Mathis (2005) found that the cost of administration of NCLB which only includes the test administration alone requires between a 2% to 2.5% increase for a total of $11.3 billion at the federal level. Archer (2005) shows that massive funding is required to implement the mandated AYP assessments but that the federal government is not providing for the amount of funding needed to implement the testing provisions. As a result, states need to absorb the additional cost in their educational budget or opt for the implementation of cheaper forms of assessment that include more multiple-choice questions and less open-ended questions. For example, the state of Michigan has reduced the use of the writing assessment which will now only be administered in grade four and seven (MDE, 2009). They also drastically reduced the number of open-ended questions in their reading, science, and social studies assessments. Ohio has totally eliminated the implementation of writing and social studies assessments since they are not mandated by NCLB and they will result in a cost savings for the state (McMillan, 2010). Thus, an unintended consequence of the legislation is the reduction of state educational resources due to testing requirements in NCLB along with the implementation of state accountability assessments that de-emphasize higher order thinking requirements and open-ended writing assessments.

Providing a "standards-based" NCLB education for all children will require massive new investments in education spending (Mathis, 2003). Mathis references ten different studies and found that substantial increases in financing is needed in the range of 20% to 35 % in order to achieve the goals set out in NCLB (Mathis 2003, p. 682).

Mathis (2005) came to the following conclusion:

> Based on 40 separate adequacy studies, additional new costs to give all students standard-based opportunities are conservatively estimated at 27.5% or $137.8 billion in new money. Thus, implementation of the administrative and learning opportunities aspects of the law would require a new sum of $144.5 billion or an increase of 29% in educational spending. (p.90)

Murnane (2007) recommends an increase in federal funding through target grants in the range of 2.75 billion dollars along with a modification to the AYP formula in order for NCLB to meet the needs of disadvantaged students. The referenced studies call into question the ability of NCLB to meet the goals set out in the reform with the current educational resources that are provided to schools which has been exacerbated in these difficult economic times.

The Impact of School Resources

Odden and Busch (1998) found that school resources and school based financing is a key to improved school performance and the resources are not being distributed fairly and spent effectively among schools. Hattie (2009) in a meta-analysis concluded that money does not matter with regard to achievement. Hanusek (1989) found that there is no consistent statistical relationship with educational expenditure and measures of student performance. However, other research has found that there were systematic, positive relations between educational resources and student outcomes (Addonizio, 2009; Greenwald, Laine & Hedges, 1996; Hedges, Laine & Greenwald, 1994). Addonizio (2009) along with Phelps and Addonizio (2006) found that incentives, motivation, culture, and other organizational characteristics have far greater implications for a school's performance vs. the allocation of physical resources.

Green, Huerta, and Richards (2007) emphasize the need to balance resources with educational outcomes. Their research shows that investment in human capital and professional learning is critical. They found that it is important to have teachers work together with different structures including team teaching. They also conclude that it is essential to establish teacher evaluation systems that are linked to student performance outcomes. Compensation for teachers should be linked to multiple measures that involve student outcomes. The quantity of additional funding resources will only lead to improvements if effective implementation is implemented (Green et al. 2007).

Grubb, Huerta, and Goe (2006) argue that increased funding and resources alone, will not lead towards improvements. Their review of literature found that resources encompass a lot more than just funding or capital improvements. They believe it is hard to measure some of the effective practices in relationship to school funding which includes high expectations and strong school leadership. An adequate level of funding is necessary for school success but it is not a determinant factor in itself. An outcome based approach that measures school effectiveness is a critical component for measuring the impact of school resources (Grubb et al. 2006). Miles and Darling-Hammond (1998) concluded that the particular use of resources including time was critical. Certain instructional conditions need to be met where teachers work in collaboration with one another that include common planning time. Teachers in effective schools know their students well and implement complex resources in the form of effective instructional conditions.

Summary of AYP Benefits and Unintended Consequences

The referenced literature is in agreement that the philosophical intent of NCLB, which focuses on providing all students with a high quality education (including those in the subgroup populations), is an important mandate in the NCLB reform. A powerful aspect of NCLB is the requirement that data is disaggregated by subgroup population. If there was a sole reliance on the aggregate data, then it is probable that students in many subgroups would not receive the attention that they need in order to make improvements. However, the literature provides a consensus that there are numerous problems with the AYP accountability provisions in NCLB. The result is the potential to harm the public school system as many schools and their programs will be weakened through the implementation of the accountability sanctions and the labeling of schools as failing. Some of the unintended consequences of AYP accountability include: a narrowing of the curriculum, a focus on basic skills, a reduction of higher level thinking opportunities and curricular innovation, danger to subgroup populations and individual students, and the potential for teacher turnover in schools with high numbers of disadvantaged students.

The Use of the NAEP

This literature review provides empirical data that leads to the synthesized argument that the AYP measure in NCLB is inconsistently implemented among the states across the country (Ravitch, 2010). Since it would have been politically difficult to pass legislation that mandated a single consistent measurement for all students in all states, the legislators came up with a compromise to use the NAEP assessment with a sample of schools in all states on a biannual basis. The purpose for including NAEP in the NCLB legislation was to evaluate state assessment standards through a comparison with state level NAEP data (Hombro, 2003). States that choose not to participate in the NAEP assessment on a biannual basis would have to forfeit their Title I funding.

Sunderman et al. (2005) found that the AYP provisions were not consistent from state to state as different states had different cut scores and their trajectories were often much different. They found that even when there was the appearance that one state might have higher standards when compared to another state based on the proficiency levels that the state requires, the proficiency targets could be deceiving because the quality of the assessments used might be much different in each state. The researchers suggest that an analysis of the standards need to be compared against the NAEP scores as the only consistent measure because standards from one state to another state were not necessarily relative to one another. Bandeira de Mello, Blankenship and McLaughlin (2009) also found that states vary widely in where they set their proficiency standards in 4^{th} and 8^{th} grade reading and mathematics. Fuller, Wright, Gesicki, and Kang (2007) concluded that states varied in their proficiency standard but that the basic level NAEP scale was more closely aligned with proficiency on state accountability assessments. Thus most state accountability assessments do not have a profciency level standard that is aligned with the NAEP scale proficiency level standard. This inconsistency provides a justification for using the NAEP in this dissertation study to provide a statistical analysis in comparsion to the state accountability systems in order to support or refute the referenced literature. A comparison of the NAEP with state accountability assessments in this study provided an analysis of the rigor that each of the four sample states employed with their state accountability assessment and AYP standard.

External Accountability vs. Internal Accountability

One of the goals in NCLB is to improve the conditions of schools for disadvantaged students. It is critical that empirically based research guides the methods that are used to shape the programs and policies that lead to school improvement. A common concept that has been referenced in the literature (Abdelmann et al. 1999; Elmore, 2002; Fullan, Hill & Crevola, 2006; Schoen & Fusarelli, 2008) has to do with the capacity of schools to make improvement in the quality of instruction. The authors emphasize that an internal accountability system will have a greater impact on classroom instruction and school improvement instead of the external accountability measure that is mandated through NCLB. Abdelmann et al. (1999) found that external accountability systems will have little impact on the improvement of schools unless there is a changed concept of individual responsibility and collective expectations within a school.

Elmore (2002) asserts that the NCLB accountability theory is faulty because an external accountability system will do little to change the capacity of the school to make improvements in the classroom. The internal accountability process and investment in teacher capacity is, according to Elmore, where the research points when it comes to making meaningful improvements for students. He explains that the current accountability system in NCLB will have little or no impact on meaningful school improvement. There is the need for further research with the internal accountability system concept which is related to the success rates with the school improvement process. This dissertation research aims to address this issue through the use of qualitative methodologies. Conducting interviews with educators at the school level provided detailed data about school improvement initiatives within schools that are a result of the NCLB reform.

Summary, Implications, and Discussion

The literature in the field emphasizes that multiple problems are associated with the implementation of the accountability provisions in NCLB. This review establishes the importance of the NCLB legislation as one of the most sweeping educational reforms at the federal level. The review also provides an analysis of the specific statistical problems, inconsistencies, and measurement issues that are involved with the implementation of AYP. The third section focuses on the philosophical intent of NCLB and the unintended consequences of the reform. The researchers cited in this review are in agreement that the intent of the legislation is noble and the inclusion of achievement levels for all students is a positive outcome. However, they are equally critical of the implementation of the accountability provisions as it currently exists. Some of the unintended consequences include; a narrowing of the curriculum, a reduction or elimination of higher level thinking skills and innovation within the curriculum, the danger to individual students who fail high stakes tests which impacts graduation rates, and the potential for higher levels of teacher turnover as a result of the pressure of probation and the labeling of a school as failing.

Effectiveness of AYP in Improving Student and School Achievement

The literature provided limited studies on the impact that the AYP provisions had on improving the conditions of schools for students. Since the purpose of NCLB is to improve the conditions of schools for disadvantaged students, it is critical to measure the impact that AYP is having in that area. For example, an analysis of the number of schools that are making AYP versus those that are not making AYP according to the subgroup demographic data is an important area for further research. If students in the subgroup populations are not making improvements, then an assertion could be that the NCLB reform is not improving school conditions for the disadvantaged students.

Another area for potential research is an analysis of school success stories and the number of schools that were once on the failing list but have since met the AYP standards. The accountability sanctions were created with the purpose to improve schools for disadvantaged students and the expectation is that schools should make improvements. If they are not, then changes might be needed with AYP. There is a lack of data in those aforementioned areas.

Demographic Variables, School Improvement, and Classroom Instruction

Demographic variables have an influence on the success level of schools with standardized testing. Coleman (1966) found that student backgrounds and not the school was the biggest contributor to student success. This includes family poverty levels, minority status, and the parents' level of education. Reeves (2004) argues that the research indicates that poverty, linguistic differences, and culture are important variables which have an effect on student achievement. However, they are not the only important factor as variables in teaching, curriculum, and leadership have a greater influence over student achievement. Poverty levels of students do have an impact on overall achievement but they do not impact the gains that students make (Reeves). Other research in the field shows that quality classroom instruction has a positive impact on student achievement. Marzano (2003) found that "schools which are highly effective produce results that almost entirely overcome the effects of student backgrounds" (p.7). Marzano further found that schools control 20% of the variance in student achievement.

Marzano's research is in alignment with the effective schools research that was prevalent by several prominent researchers in the field during the 1980's, which provided a counterargument to the findings by Coleman. Researchers (Brookover, 1985; Edmonds, 1979a, 1979b, 1982; Levine & Lezotte, 1995; Lezotte, 1992; Lezotte & Jacoby, 1991) conclude that when the correlates of effective schools are in place, the school has a positive impact on student achievement despite the minority or poverty levels of students. In their research, they highlight examples of what they describe as effective schools.

Since the research in the field shows that the quality of classroom instruction has a direct impact on student achievement (Brookover, 1985; Darling-Hammond, 2007b; Edmonds, 1979a; 1979b; Edmonds, 1982; Hattie, 2009; Lezotte, 1992; Lezotte & Jacoby, 1991; Marzano, 2003; Nye et al. 2004, Sanders, 2000; Sanders & Rivers, 1996**),** it is critical to measure the impact that AYP is having on classroom instruction. Literature could not be found that provides empirical data regarding the impact that AYP is having on classroom instruction.

The role of leadership and the importance of the principal with the implementation of successful school improvement efforts are also proven to have an impact on student achievement (Brown, 2001; Connell, 1996; Hallinger & Murphy, 1986; Hattie, 2009; Henchey, 2001; Marzano et al. 2005; Teddlie & Springfield, 1993). This is an area where there is a need to conduct empirical research. There is a need to collect qualitative data in the form of interviews with teachers and administrators in order to effectively evaluate the impact that AYP is having on classroom instruction. Research could not be found in that area and the absence of that research provides a direction, motivation, and need for this dissertation study.

Conclusion

The current accountability provisions in NCLB are not sufficient in evaluating school performance. It is evident that there are numerous problems with the implementation of AYP in order to motivate effective school improvement efforts. Reforms to the legislation are needed in order to improve the reliability of measuring school effectiveness. This includes the need to standardize the AYP measure across the country while implementing multiple forms of measurement which include growth data and a measurement of school improvement efforts through organizations like NCA or Advanced Education. Since the key to school improvement is through the enhancement of the internal capacity of the school while increasing the internal accountability mechanisms, it is critical that research is conducted in this area. Research is needed regarding the impact that AYP is having on student learning, classroom instruction, and school improvement. The research should not solely rely on the standardized test scores as a basis for evaluating the effectiveness of the NCLB accountability provisions. It should also examine the impact that AYP is having through other measures which include qualitative data analysis that examines how teachers and administrators are responding to AYP. This study aims to fill that gap with the current literature in the field and it provides rich results that measure the effectiveness with AYP in order to answer the research questions in this study.

CHAPTER III

METHODOLOGY

As stated in chapter one, the purpose of this study is to evaluate the effectiveness of AYP at establishing the conditions for school reform at the school level and classroom level. This includes an analysis of the external validity and the reliability of AYP as it relates to measuring school success across the country. Researchers maintain that there is a lack of consistency with AYP that results in reliability issues in using AYP to measure school effectiveness (Elmore, 2002; Guilfoyle, 2006; Kane and Douglas, 2002; Linn & Haug, 2002; Porter et. al, 2005; Scheon & Fusarelli, 2008; Wiley et al. 2005). Since a purpose of the NCLB reform is to improve the educational conditions for underperforming students, then the sanction provisions in NCLB should lead to school improvement and increased student achievement. If this is found, then a logical conclusion is that NCLB is successful. However, if schools are labeled as failing and no improvements are being made, then an interpretation should be that NCLB is not successful in achieving the goals set out in the reform.

The intent of the NCLB reform is to improve the conditions for underperforming students with a specific emphasis on subgroup or categorical populations. NCLB mandates that achievement data must be disaggregated by categorical populations (USDOE, 2002, Section 1111 (2) C). This study investigates potential correlations between achievement on the NCLB mandated state accountability assessments, NAEP, and school demographics. An analysis of the data according to school demographic characteristics allows for the prediction of the general profiles of schools that are successful under the AYP provisions. Researchers (Elmore, 2002; Harris, 2007; Mathis, 2004; Mathis, 2006) contend that AYP is measuring school demographics and the social capital that students bring to the school instead of school effectiveness. If that contention is true, then the external validity of the AYP measurement formula could be flawed. This study quantifies that contention by measuring school demographics in relationship to AYP status. Data was collected on the characteristics of schools that make AYP and the characteristics of schools that do not make AYP in the sample states.

This study employs a triangulated methodology of mixed methods as it purposefully combines the collection of quantitative and qualitative data to provide an accurate picture of the impact that NCLB is having on school achievement and school accountability (Rudestam & Newton, 2007). The first three research questions require the collection and analysis of quantitative data while the fourth research question requires the collection and analysis of qualitative data. "Mixed method designs can yield richer, more valid, and more reliable findings than research based in either qualitative or quantitative method alone" (Gawlik, 2005, p.62). The methodological approaches are divided according to the research questions and the data collection and analysis techniques are described in this chapter.

Research Questions

The research questions as stated in chapter one are the following:

1. Is there a significant correlation between the schools that meet proficiency standards on the NAEP and the schools that make AYP in the sample states? Are there significant differences between subgroups?
2. Is there a significant correlation between the schools that meet proficiency standards on the NAEP and schools that meet proficiency standards on the state accountability assessments (STAR, MEAP, End of Grade Test, and TAKS) in the sample states? Is there

a combination of factors that best predicts proficiency status on the NAEP and state accountability assessments?

3. Are the demographic (categorical) characteristics and educational resources significant predictors of the schools that make AYP and fail to make AYP in the sample states?

4. What impact is AYP having on school improvement initiatives and classroom instruction?

Quantitative Research Design

The first part of the quantitative study investigates the AYP formula that each of the four sample states use to determine school AYP scores. The subsequent portions of the quantitative study use descriptive statistics to implement statistical models that include correlation relationships, and logistic regression, which answer the first three research questions in this study.

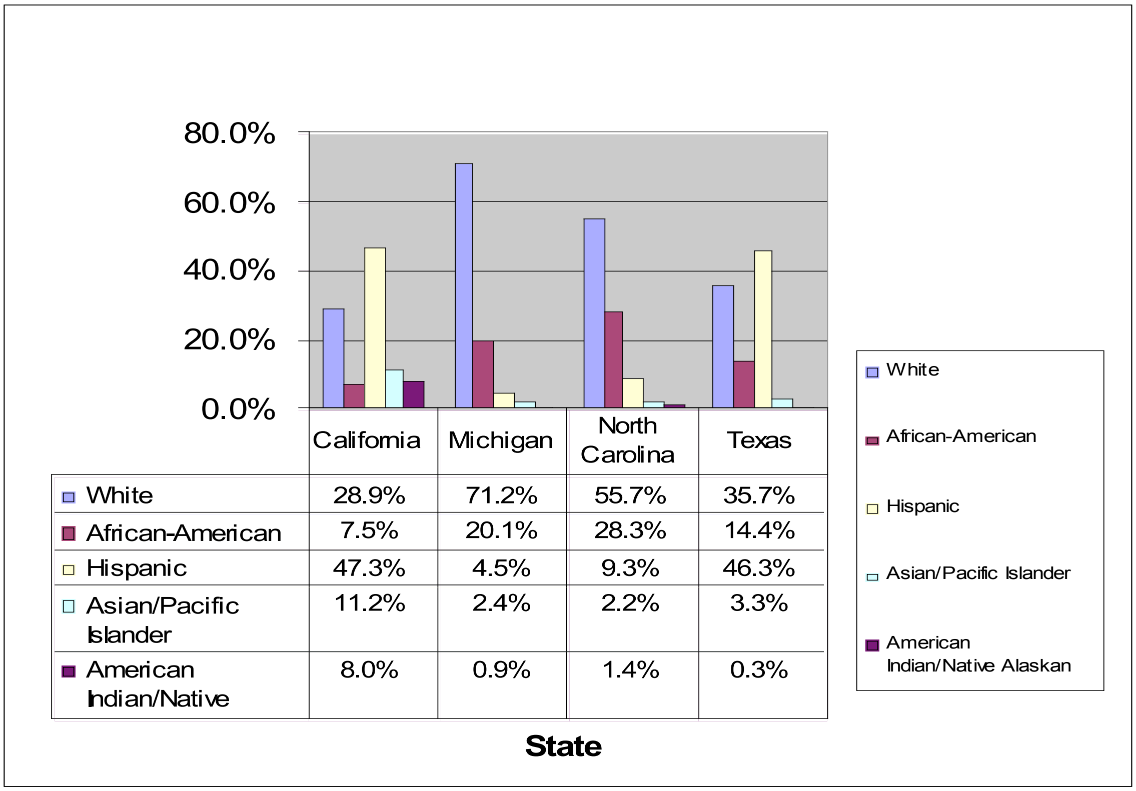
Sample

The sample addressing the first three research questions in this study includes 3430 elementary and middle schools that completed the NAEP assessment in reading and mathematics during the school years 2005 and 2007 from four sample states, 1) Michigan, 2) California, 3) Texas, and 4) North Carolina. The four sample states were selected based on several factors which include their different geographic locations, their diverse racial and ethnic populations, and their strong history of school accountability.

The sample states provide differences in student populations which include varying percentages of the subgroup populations that are part of the AYP measurement tool such as racial/ethnic subgroup categories. North Carolina and Michigan have large White/Caucasian and

African-American populations while California and Texas have large Hispanic populations. Figure 13 provides a breakdown of the state populations in schools by racial/ethnic subgroup.

Figure 13. **Racial/Ethnic Subgroups in the Sample States**



	California	Michigan	North Carolina	Texas
White	28.9%	71.2%	55.7%	35.7%
African-American	7.5%	20.1%	28.3%	14.4%
Hispanic	47.3%	4.5%	9.3%	46.3%
Asian/Pacific Islander	11.2%	2.4%	2.2%	3.3%
American Indian/Native	8.0%	0.9%	1.4%	0.3%

Another important factor is that each of the four states has a solid history of school accountability prior to NCLB. The selection of the states with a strong history of school accountability allows for a comparison of the states that are deemed as the nation's best with the requirement of school accountability. Future research may choose to compare states with strong accountability systems and weak accountability systems. However, the selection of these four states was purposeful as the research interest involves analyzing the differences in states that have strong school accountability systems in order to provide information that could allow for replication of the nations' most effective state accountability systems.

Two of the sample states, Texas and North Carolina, are recognized for implementing strong accountability models that assisted policy makers with the development of the NCLB accountability provisions (Nelson et al. 2007; Peterson & West, 2003). North Carolina is a critical state in this study because it was one of the first states to implement a value-added growth model with the calculation of AYP. North Carolina began collecting and analyzing growth data in 1995. Recent pressure from educational professionals and researchers (Elmore, 2002; Hess & Petrilli, 2006; Hoff, 2007; Koretz, 2008; Peterson & West, 2006; Popham, 2005b; Schoen & Fusarelli, 2008; Sunderman et al. 2005; Wiley et al. 2005) is mobilizing to the point where the growth model could become part of the AYP formula in a greater number of states throughout the nation as the process of reauthorizing NCLB occurs. Michigan and Texas have subsequently been approved for the growth model and there are fifteen states that received approval for the growth model throughout the United States (USDOE, 2009).

The sample from the first three research questions relies on the methodologies that NCES uses to select schools that participate in the NAEP assessment through stratified random sampling techniques. (Table 1 in the first chapter provides the definition that NCES uses for random stratified sampling). This involves sampling between 100 to 450 schools from each of the four sample states, 1) California, 2) Michigan, 3) North Carolina, and 4) Texas, for the years 2005 and 2007 in each of the subject areas tested (reading and mathematics) and at each grade (fourth and eighth). NCES randomly selects schools from each state in order to administer the NAEP assessment in reading and mathematics. The schools that were selected for each assessment for the year 2005 are different than the schools that were selected for the year 2007. NCES uses a three step process with the selection of sample public schools to take the NAEP assessment. This three step process involves the following: 1) Selecting public schools within designated areas, 2) selecting students in the relevant grades within the designated schools, and 3) allocating selected students to assessment subjects (NCES, 2009). Appendix C provides a breakdown of the specific process that NCES uses to select students that participate in the NAEP assessment.

Data Collection

The collection of data for the quantitative portion of the study addressing the first three research questions involves accessing data from the state educational agency websites, the NCES website, the National Longitudinal School-Level State Assessment Score Database (NLSLSASD), and restricted access secondary use NAEP data from NCES. The state agency websites were used to download data files that provide information on the demographic characteristics of schools, AYP status scores from schools, school percentage proficiency scores on the state mathematics and reading assessments at the fourth-grade and eighth-grade level, and the percentage of proficient students in the sample schools from the sample states. All of the data was collected from the 2005 and 2007 school years.

The NLSLASD website was used to assist in the collection of school level proficiency data on state accountability assessments. The state website databases and NLSLASD database were used in combination to provide accurate data on the state accountability assessments. For example, California institutes an algebra exam in mathematics for 8th grade and that data was not available on the NLSLASD website. However, the NLSASD site provided data that was imported into SPSS. Some of the state agency websites did not provide data in the format that allowed for SPSS import (ex. Michigan only provided data for the 2005 year in an html format).

The NCES website was used for the collection of quantitative data including demographic characteristics and school district financial data. A large portion of the data that was collected through NCES is restricted access secondary use data. The collection of this data involved submitting an application to NCES that was reviewed and approved by the Wayne State Legal Department in the Special Programs Administration Division prior to submission. All individuals who had access to the restricted access NAEP data were required to sign affidavits of non-disclosure. A security plan was outlined in order to ensure the security of the physical and electronic data that was received from NCES. The application was filed on the NCES website and all of the required affidavits, legal application, and security plan were sent to NCES via certified mail. State agency websites only provided district level data on special education populations and ELL populations for the 2005 and 2007 school years.

Since school level demographic data on special education and ELL students was not available via the state agency websites, the NCES restricted access data was used to identify the calculations in those subgroups for each of the sample schools. However, state agency websites and the NCES public access database provided accurate calculations on school level subgroup populations. The NCES restricted access data only provided a range or estimate of the population percentage for each subgroup. A median range was used to identify ELL and special education population percentages of schools. Refer to Appendix D for information on the NAEP subgroup range estimates.

The NCES restricted access data provided information regarding individual student achievement on the NAEP according to the NCES Item Response Theory[3] and Maximum Marginal Likelihood (ML)[4] estimation that results in the calculation of five plausible values[5] for each individual student. Each individual student's five plausible value NAEP assessment results in reading and mathematics during the 2005 and 2007 school years at the fourth-grade and eighth-grade level were provided on the restricted access database from NCES. Since NCES uses a school identification number, the students were then sorted and grouped by school identification number which allowed for a determination of school level achievement. Listed in Table 2 is the information on where to locate the website databases that were used to access data for this study.

Data Analysis

The procedures outlined in this section involved the analysis of NAEP, AYP, and state accountability assessments (STAR, MEAP, End of Grade Test, and TAKS) in reading and mathematics at the fourth-grade and eighth-grade level for the years 2005 and 2007 from four states, 1) California, 2) Michigan, 3) North Carolina, and 4) Texas.

Analysis of the AYP accountability provisions of each state occurred through an examination of the data that is available on the state education agency websites. This analysis includes an investigation regarding the methods that the four sample states use to comply with the accountability

---

[3] Item Response Theory (IRT) is a set of statistical models that describe the relationship between assessment items and proficiency in a subject or skill area. (USDOE, NCES, 2001). Refer to Appendix E for a detailed explanation.

[4]Maximum Likelihood (ML) estimation is applied to situations when the variables of interest are only partially observed. MML estimation provides estimates of marginal (i.e., aggregate) parameters that are the most likely to have generated the observed sample data. MML estimation proves useful when one is interested in estimating the mean and variance of a scale from survey data that provides only imperfect measurement of the target construct (USDOE, NCES, 2001).

[5] Plausible values are proficiency estimates for an individual NAEP respondent, drawn at random from a conditional distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses (USDOE, NCES, 2001). Refer to Appendix F for a detailed explanation.

provisions in NCLB at the elementary and middle school level. The introductory portion of the analysis involved a description of the accountability procedures that each state uses to evaluate schools at the elementary and middle school level.

Table 2

Agency Website Databases

Education Agency	Location	Website URL
California Department of Education (CDE)	Fresno, California	http://www.cde.ca.gov/
Michigan Department of Education (MDE)	Lansing, Michigan	http://www.michigan.gov/mde
Public Schools of North Carolina (PSNC)	Raleigh, North Carolin:	http://www.dpi.state.nc.us/
National Longitudinal School-Level State Assessment Score Database (NLSLSASD)	American Institute for Research, Washington, D.C.	http://www.schooldata.org/
Texas Education Agency (TEA)	Austin, Texas	http://www.tea.state.tx.us/
U.S. Department of Education, National Center for Education Statistics (NCES)	Washington, D.C.	http://nces.ed.gov/

Statistical Hypothesis

A statistical hypothesis is an assumption about a population parameter that may or may not be true. A sample of the population is often used in inferential statistics in order to determine if the hypothesis is correct. Inferential statistics are used to draw conclusions about parameters because it is often not feasible or desirable to examine an entire population. Thus random samples are used in order to draw conclusions about the population based on the results from the sample (Hinkle et al. 2003). When conducting research, a hypothesis provides information regarding the study and delineates the problem along with the variables that were examined in the study. The sample data that is collected is then analyzed and if it is consistent with the hypothesis, the hypothesis is accepted and if not, it is rejected. The result is two hypotheses, a null hypothesis and an alternative hypothesis. If the null hypothesis is accepted, then the data results are due to chance. However, if the null hypothesis is rejected, then the alternative provides an explanation that is not the result of random cause, but rather the influence of the variables under study (Hinkle et al. 2003).

In order to enhance the quality of this study, a null hypothesis and alternative hypothesis are provided for each of the three quantitative research questions. The usage of three hypotheses provides a focus and guidance for the implementation of the methodology for this study. Each of the following subsections is organized according to the research questions. The null hypothesis and alternative hypothesis are listed along with a justification for the stated predictions.

Hypothesis for Research Question One

Due to the different AYP accountability systems that are being implemented in the four sample states, it was predicted that there would not be a significant correlation between the schools that meet proficiency standards on the NAEP and those that meet AYP. The literature review established the variability in standards among the states. However, it was predicted that there would be significant differences among schools with high percentages of categorical populations. The subsequent section in this dissertation titled causal hypothesis provides a justification for the prediction. Listed below are the null (H_0) and alternative (H_1) hypotheses for research question one:

H_0 : AYP status of schools = proficiency status of schools on the NAEP.

H_1: AYP status of schools ≠ proficiency status of schools on the NAEP.

$H_{0.2}$: AYP proficiency status of schools with high categorical populations and limited educational resources = achievement of proficiency on the NAEP.

$H_{1.2}$: AYP proficiency status of schools with high categorical populations and limited educational resources ≠ achievement of proficiency on the NAEP.

Hypothesis for Research Question Two

The AYP status of schools in each state is based on an AYP formula that relies on the usage of the state accountability assessments in mathematics and reading. Thus, the hypothesis prediction in research question one is aligned with the hypothesis prediction in research question two because the AYP formula is based on the usage of the mathematics and reading assessments. Listed below are the null and alternative hypotheses for research question two:

H_0 : proficiency status of schools on the state accountability assessments = proficiency status of schools on the NAEP.

H_1: proficiency status of schools on the state accountability assessments ≠ proficiency status of schools on the NAEP.

$H_{0.2}$: Schools with high categorical populations and limited educational resources = achievement of proficiency on the NAEP and state accountability assessments.

$H_{1.2}$: Schools with high categorical populations and limited educational resources ≠ achievement of proficiency on the NAEP and state accountability assessments.

Hypothesis for Research Question Three

Research question three identifies the relationship between the demographic characteristics of schools and their AYP status. This research question relies on the demographic characteristics of schools and it examines the potential relationship that schools with high numbers of categorical populations have with the AYP status of the schools. It was predicted that schools with high percentages of categorical populations and limited school resources had less of a chance to achieve a successful AYP status. This prediction was based on the review of literature in the field as explained in the subsequent section titled causal hypothesis.

Listed below is the hypothesis for research question three:

H_0 : Schools with high categorical populations and limited educational resources = achievement of proficient AYP status.

H_1: Schools with high categorical populations and limited educational resources ≠ achievement of proficient AYP status.

Causal Hypothesis

When examining the relationships between two variables, the goal is to establish that changes in the explanatory variable causes changes in the dependent variable. Despite the fact that there might be a strong association present, the conclusion that this is due to a causal link is often hard to determine as correlation does not equal causation (Gawlik, 2005). Nonetheless, causal hypotheses[6] may be used to give some determination of the relationship that is observed. The first set of relationships in this study are between the AYP status of school (dependent variable), school resources or school funding (independent variable), and the influence of the demographic characteristics of school populations (independent variable), which include the influence of ED status, race/ethnicity, ELL, and special needs. Schools with limited school resources that have large numbers of students in the above mentioned demographic areas are

[6] Is a hypotheses that predicts a cause-and-effect relationship among the variables to be studied.

predicted to have a negative impact on the AYP status of schools and the achievement of schools on both the NAEP and state accountability assessments.

The causal hypothesis in this study is aligned with theoretical and empirical publications (Mathis, 2004; Meier et al. 2004; Porter et al. 2005; Springer, 2008; Sunderman et al. 2005) with the prediction that a high level of racial or ethnic diversity would increase the probability that schools will fail to meet the AYP requirements in NCLB. There is a vast amount of literature in the field that establishes a strong correlation between poverty and student achievement (Brooks-Gunn & Duncan, 1997; Camara & Schmidt, 1999; Chamberlein, 2007; Duncan & Brooks-Gunn, 2000; Guo & Harris, 2000; McGee, 1997; Phelps & Addonizio, 1996; Rothstein, 2001, 2004, 2008; The Connecticut Commission on Children, 2004). Researchers (Camara & Schmidt, 1999; Jencks & Phillips, 1998; Harris & Herrington, 2004; Lee, 2002, 2004; Lin & Harris, 2008; Taylor, 2005) further show that race and ethnic diversity also create a student achievement gap. Educating children with minority status is often difficult due to high poverty levels (Gawlik, 2005).

ELL students often take anywhere from three years to seven years to attain English language proficiency at an academic level (Hakuta et al. 2000; Pray & McSwan, 2002). NCLB mandates that ELL students are tested after one year in the country for reading and upon entry to the school in the areas of mathematics and science. Zehr (2008) argues that AYP may not be the most effective accountability tool for measuring the progress of ELL students. Research (Abedi & Dietal, 2005; Abedi, Hoftstetter & Lord, 2004; Menken, 2008) shows that schools with large ELL populations will have a greater chance of failing to meet AYP due to the achievement gap with ELL students.

Studies that address the achievement gap between special education and regular education students were not as common. This might be due to the fact that until NCLB was enacted, the special education students did not take the same assessments as regular education students and there is no baseline for comparison. However, a report by the USDE (2002) provides data which indicates the special education student dropout rate is double that of regular education students. The report further indicates that 40% of the students are labeled as special education because they cannot read. The special education subgroup was analyzed in this study to determine if it is a predictor of the AYP status of schools. It is critical to study the special education subgroup because policy makers determined that there are problems with the services provided to these students and it is a component in the NCLB reform (USDOE, 2002, 111 (2) C).

As mentioned in the literature review, Odden and Busch (1998) found that school resources and school based financing is a key to improved school performance and the resources are not being distributed fairly and spent effectively among schools. Hanusek (1989) found that there is no consistent statistical relationship with educational expenditure and measures of student performance. However, the results of other research indicate that there were systematic, positive relations between educational resources and student outcomes (Addonizio, 2009; Greenwald, Laine & Hedges, 1996; Hedges, Laine & Greenwald, 1994). Alternatively, Addonizio (2009) and Phelps and Addonizio (2006) found that incentives, motivation, culture, and other organizational characteristics have far greater implications for a school's performance than the allocation of physical resources.

Even though there is conflicting research on the impact that school finance is having on student achievement, the reality is that resources are not being disseminated equitably across the

United States (Odden & Busch, 1998; Odden & Clune, 1998). Thus it was critical to examine the educational resources variable when it comes to the AYP status of schools to determine if it has an impact on school achievement. It is difficult to imagine that money does not make a difference in school achievement (Hattie, 2009). As a result, this study measures the impact of educational resources by evaluating the funding levels that schools receive in relationship to their achievement on the state accountability assessments, AYP status, and the NAEP. The school resource variable was also addressed in the qualitative portion of this study.

This study hypothesizes that schools with high percentages of categorical populations and limited school resources, would result in a negative influence on their AYP status, achievement on the NAEP, and state accountability assessments. This negative influence was predicted to be statistically significant with the use of the Pearson correlation and logistic regression model in this study.

Statistical Tests and Analysis

Research Question One

1. Is there a significant correlation between the schools that meet proficiency standards on the NAEP and the schools that make AYP in the sample states? Are there significant differences between subgroups?

This part of the study involved the collection of school level restricted access data from NCES[7] and public access data from the state department of education websites. Data was collected from the four sample states, 1) California, 2) Michigan, 3) North Carolina, and 4) Texas, on the NAEP assessments in reading and mathematics at the fourth-grade and eighth-

[7] Appendix G provides an outline of the regulations that NCES adheres to with the usage of restricted access data.

grade level for the 2005 and 2007 school years. Data was also collected from the state agency websites regarding the schools that met and failed to meet AYP during the years 2005 and 2007 at the elementary and middle school level.

The data that was collected from the NCES restricted access database, the NCES public access website, the NLSLSASD website database, and the state agency websites were entered into a new database using the Statistical Package for the Social Sciences (SPSS). This included entering the demographic characteristic data, educational resources data[8], AYP status of the schools, school level NAEP scores in mathematics and reading, and the state accountability assessment data in reading and mathematics (STAR, MEAP, End of Grade Test, and TAKS) from each of the schools. All of the school-level cases were weighted in SPSS based on school enrollment statistics. A separate database was created for each state in the sample (California, Michigan, North Carolina, and Texas), grade level assessed (4th and 8th grade), and year of assessment (2005 and 2007). The result was the creation of 4 separate databases for each state for a total of 16 databases in this study. The mathematics and reading results were entered into the same data base for the state and year of assessment.

The state formula for AYP and state level AYP reports were used to determine if a school met AYP proficiency standards. The AYP formula involves mathematics and reading plus one other measure. Table 3 provides information on the percentage of students that are required to be proficient in a school on each content area in order for a school to meet AYP requirements.

---

[8] This was calculated as the per pupil average funding based on school district statistics provided by NCES.

Table 3

School Level Proficiency Targets in the Sample States: AYP requirements [a.]

Assessment Year 2005	California AYP target (%)	Michigan AYP target (%)	North Carolina AYP target (%)	Texas AYP target (%)
4th Grade Reading	24.4	48	76.7	53
8th Grade Reading	24.4	43	76.7	53
4th Grade Mathematics	26.5	56	81	42
8th Grade Mathematics	26.5	43	81	42
Assessment Year 2007	California AYP target (%)	Michigan AYP target (%)	North Carolina AYP target (%) b.	Texas AYP target (%)
4th Grade Reading	24.4	59	43.2	60
8th Grade Reading	24.4	54	43.2	60
4th Grade Mathematics	26.5	65	77.2	50
8th Grade Mathematics	26.5	54	77.2	50

a. Each state determines their own assessment and cut score for proficiency on each assessment plus the trajectory towards 100% proficiency by 2014

b. North Carolina made revisions which changed their trajectory target scores.

NCES representatives indicated that the purpose of the NAEP is to assess individual students in order to analyze state level data that is disaggregated according to demographics and subgroup populations. Since the interest in this study is the performance of individual schools, it was necessary to develop a statistical formula (cut score) for determining school level proficiency on the NAEP. The administration of the NAEP is statistically complex. An individual student only takes a portion of the entire NAEP and an individual determination of student proficiency is never made on the NAEP assessment. This is much different than the state level accountability assessments where there is a determination of individual student proficiency which allows for the calculation of school level average proficiencies in the school. As it was previously mentioned in this chapter, each student who takes the NAEP assessment is assigned a series of five plausible values which are not actual scores. (Refer to Appendix F for information on plausible values). The five plausible values assigned to a student are estimates of student achievement on a posterior distribution which is based on the student demographics (USDOE, NCES, 2010). Therefore, it is impossible to come up with a school level percentage of proficiency on the NAEP. It is not statistically recommended to average the five plausible values that are assigned to a student in order to come up with an individual score (Davier, Gonzalez & Mislevy, 2009).

Since the administration of the NAEP involves a complex statistical formula, the determination of school proficiency was calculated based on the school level average or mean scaled score[9] on the NAEP for each subject and year administered among the students in a school who took the NAEP assessment. When NCES reports the results for states or populations of students, they often use the scaled score average. The use of plausible values and the NAEP sample weighting procedures increases statistical reliability while limiting problematic influences from skewed results.

The average NAEP scaled score for a school was calculated by taking the datasets from the restricted access database in SPSS and then importing the data into AM statistical software which is available to NAEP secondary researchers as a result of a federal grant (Refer to

[9] Scaling is the process of assigning numbers to reflect students' performance on an assessment. Chapter 12 of The NAEP 1998 Technical Report (2001) describes the scaling models employed by NAEP: Three distinct scaling models, depending on item type and scoring procedure, are used in the analysis of NAEP data. Each of the models is based on item response theory (IRT; e.g., Lord, 1980). Each is a “latent variable” model, defined separately for each of the scales, which expresses respondents’ tendencies to achieve certain scores (such as correct/incorrect) on the items contributing to a scale as a function of a parameter that is not directly observed, called score (θ) on the scale (p. 229), (USDOE, NCES, 2001).

Appendix H for more information on AM software). The calculation of the average school level scaled scored occurred by using the five student plausible values that are assigned to an individual student on the NAEP assessment, school level identification numbers, and student weights[10] that were provided from the NCES restricted access database.

The individual student data was grouped by schools using the NCES school identification numbers. The AM statistical software was used to calculate the mean for each of the sample schools. Schools were considered to have met sufficient standards on the NAEP if the average scale score for the school met or exceeded the proficient achievement level as it was established by NCES for each grade and subject tested. (Refer to Appendix J [Reading] and Appendix K [Mathematics] which provides information on the proficiency scales on the NAEP that have been established by NCES).

Since research in the field (Bandeira de Mello, et al. 2009; Fuller et al. 2007; Gewertz, 2010) shows that achievement on the NAEP has a closer correlation with the basic level standard on the NAEP scale, the analysis of school level proficiency was done using both the proficient level scale and the basic level scale in order to calculate school level proficiency on the NAEP. Therefore, the statistical analysis occurred twice, first using the proficient level NAEP scaled score as a school level proficiency cut score, and second using the basic level NAEP scaled score as a school level proficiency cut score to determine school level proficiency on the NAEP.

A Pearson's Product-Moment Correlation test was conducted to determine if there was a statistically significant correlation between the proportion of schools in a state that meet NAEP proficiency standards on the mathematics and reading assessments at the fourth-grade and eighth-grade levels in relation to the proportion of schools in the state that meet AYP during the

[10] The use of the NAEP student weights helped to increase the validity with the calculation of average school level scaled scores on the NAEP. Refer to Appendix I for information on NAEP weighting procedures.

years 2005 and 2007. The Pearson's Product-Moment Correlation was also conducted to estimate correlations of school populations (defined in terms of subgroup populations) in relationship to proficiency status.

Pearson's Product-Moment Correlation

$$r = \frac{\sum zxzy}{n-1}$$

r = the direction and degree of closeness between the two linear variables

zx = the independent variable in z score form.

zy = the dependent variable in z score form.

n = the number of cases.

Variables:

This list includes the variable codes that were used in SPSS and are present in the variable tables and figures that are presented in chapter four. Since proficiency status is a binary (dichotomous variable), if a school was deemed proficient on the NAEP and state AYP, they were given a code of "1" in SPSS and on the tables presented in chapter four. Schools that were not proficient were given a "0" code.

Independent Variables:

x_1 = NAEPCOMBPROFprof = the proportion of schools that meet proficiency standards on the NAEP mathematics and reading assessments at the fourth and eighth grade level during the years 2005 and 2007. A school is determined to be proficient if they meet proficiency standards on both the mathematics and reading NAEP. Proficiency was determined using the NAEP proficient level.

x_2 = NAEPCOMBPROFbasic = the proportion of schools that meet proficiency standards on the NAEP mathematics and reading assessments at the fourth and eighth grade level during the years 2005 and 2007. A school is determined to be proficient if they meet proficiency standards on both the mathematics and reading NAEP. Proficiency was determined using the NAEP basic level.

x_3 = SPECIALEDPER = % of special education students (based on a median range provided by NCES)[11].

x_4 = EDPER = % of ED students.

x_5 = BLACKPER = % of Black/African American students.

x_6 = HISPANICPER = % Latino/Hispanic students.

x_7 = NATAMPER = % American Indian/Alaskan Native students.

x_8 = ASIANPER = % Asian/Pacific Islander students.

x_9 = WHITEPER = % White/Caucasian students.

x_{10} = TOTAL REVENUE PER STUDENT = Educational resources (average per pupil funding based on school district data).

x_{11} = ELLPER = % of ELL students (based on the same median range that was used for the special education subgroup population which is outlined in Appendix D).

[11] Data on special education populations and ELL populations was not available via the state agency websites. Thus the NCES data on those subgroup populations were used for this study. Appendix D provides information on the ranges which are used by NCES. NCES collects Non-Cognitive Data through a school level questionnaire. Please refer to Appendix L for information on the questionnaire.

Dependent Variable:

y = STATEAYP = the proportion of schools that meet AYP standards in the sample states.

Research Question Two

2. Is there a significant correlation between the schools that meet proficiency standards on the NAEP and schools that meet proficiency standards on the state accountability assessments (STAR, MEAP, End of Grade Test, and TAKS) in the sample states? Is there a combination of factors that best predicts proficiency status on the NAEP and state accountability assessments?

This part of the study involved an examination of the relationship between the NAEP assessments in reading and mathematics at the fourth-grade and eighth-grade level in specific schools and the state proficiency assessments (STAR, MEAP, End of Grade Test, and TAKS) in mathematics and reading at the fourth-grade and eighth-grade levels in specific schools. The NAEP restricted access data from the sample schools was entered into the SPSS databases that were divided according to the state, year of assessment, and grade level. All of the datasets from schools were weighted in SPSS based on school level enrollment statistics. A Pearson's Product-Moment Correlation was used to analyze if there is a relationship between the state proficiency scores on the fourth-grade and eighth-grade reading and mathematics assessments and the NAEP proficiency scores in reading and mathematics among the sample schools from each of the four sample states. The analysis of the proficiency scores was done by analyzing state reading (in North Carolina and Texas) or ELA (in California and Michigan) with NAEP reading and state mathematics with NAEP mathematics assessments (the California 8th grade state assessment is an algebra assessment). The proficiency status of schools was used as the measurement for the calculation and analysis of research question two. School proficiency status was determined using the same proficiency status cut scores on the state accountability assessments and the NAEP[12] that were used for research question number one (Refer to Table 3). A logistic regression[13] analysis was conducted in order to determine if there were statistically significant relationships among the categorical populations, educational resources, and NAEP proficiency status in relation to the state accountability assessment proficiency status. The logistic regression equation is listed below.

Logistic Regression:

$$P = \frac{e^{a+bX}}{1 + e^{a+bX}}$$

P = the probability of a proficient school score on the state accountability assessment.

e = the base of the natural logarithm (about 2.718).

a and b = the parameters of the model.

x = the independent variables and is defined in the subsequent section titled variables.

* The value of a yields P when X is zero, and b measures how quickly the probability changes with changing X a single unit.

[12] In alignment with research question one, NAEP proficiency was tested twice. Once using the Proficient level Scale and the second time using the Basic Level Scale.
[13] Binary logistic regression was used for this analysis as it measures a dichotomous variable with 2 categories. Multinomial logistic regression is more appropriate for variables with more than 2 categories.

Variables:

Similar to research question one, this list includes the variable codes that were used in SPSS and are present in the variable tables and figures that are presented in chapter four. Since proficiency status is a binary (dichotomous variable), if a school was deemed proficient on the NAEP and state accountability assessments in reading and then mathematics, they were given a code of “1” in SPSS and on the tables presented in chapter four. Schools that were not proficient were given a “0” code.

Independent Variables:

x_1 = NAEPMATHPROFprof = NAEP proficiency status of schools in mathematics at the fourth-grade and eighth-grade levels during the years 2005 and 2007. (The NAEP proficient scale was used to determine proficiency).

x_2 = NAEPMATHPROFbasic = NAEP proficiency status of schools in mathematics at the fourth-grade and eighth-grade levels during the years 2005 and 2007. (The NAEP basic scale was used to determine proficiency).

x_3 = NAEPREADPROFprof = NAEP proficiency status of school in reading at the fourth-grade and eighth-grade levels during the years 2005 and 2007. (The NAEP proficient scale was used to determine proficiency).

x_4 = NAEPREADPROFbasic = NAEP proficiency status of school in reading at the fourth-grade and eighth-grade levels during the years 2005 and 2007. (The NAEP basic scale was used to determine proficiency).

x_5 = SPECIALEDPER = % of special education students (based on a median range provided by NCES, refer to Appendix D).

x_6 = EDPER = % of ED students.

x_7 = BLACKPER = % of Black/African American students.

x_8 = HISPANICPER = % Latino/Hispanic students.

x_9 = NATAMPER = % American Indian/Alaskan Native students.

x_{10} = ASIANPER = % Asian/Pacific Islander students.

x_{11} = WHITEPER = % White/Caucasian students. Note: For the North Carolina 2005 dataset and the Texas 2005 and 2007 datasets, it was necessary to create a DUMMYWHITE variable in place of the WHITEPER variable. The coding range for this variable is reported in Appendix D in alignment with the range that was used for the SPECIALEDPER variable and ELLPER variable. This range is consistent with the range that NCES uses for coding the percentage of schools in a specific subgroup on the surveys that are administered to schools that take the NAEP assessment. The coding of the DUMMYWHITE variable was necessary because there was a conflict with the interaction between the WHITEPER variable and the HISPANICPER variable. The logistic test created a redundancy and the WHITEPER variable was not reported in the output table. However, the creation of the DUMMYWHITE variable allowed for inclusion of this variable in the equation. It was only necessary to use the DUMMYWHITE variable with the 2005 North Carolina dataset and all of the Texas datasets.

x_{12} = TOTAL REVENUE PER STUDENT = Educational resources (average per pupil funding based on school district data). Note: For the logistic regression, this variable was changed to DUMMYREV which is a range of revenue allocated per student. This was necessary in order to allow for an appropriate calculation with logistic regression because it is a monetary value. The categories for the DUMMYREV variable are the following: 1= $6,000 to $ 7,999, 2 = $8, 000 to $9, 999, 3 = $10,000 to $11, 999, and 4 = $12, 000 and above. The decision to use this range was made after analyzing the number of schools in the sample that were at different levels. It was important to ensure that there were enough schools in each range in order to accurately group the higher and lower funded schools. The average difference among per pupil funding among the four states was another factor. The selected range provided a consistent and effective analysis of the school funding variable in all four states.

x_{13} = ELLPER = % of ELL students (based on a median range provided by NCES, refer to Appendix D).

Dependent Variables:

y_1 = STATEMATHprofstatus state assessment proficiency status of the schools in mathematics at the fourth-grade and eighth-grade levels during the years 2005 until 2007.

y_2 = STATEREADprofstatus (North Carolina and Texas) and STATEELAprofstatus (California and Michigan) state assessment proficiency status of the schools in reading at the fourth-grade and eighth-grade levels during the years 2005 until 2007. (Michigan and California use an English Language Arts) state proficiency assessment while North Carolina and Texas use a reading assessment for determining proficiency status which is the reason for the different coding of this variable).

Research Question Three

3. Are the demographic (categorical) characteristics and educational resources significant predictors of the schools that make AYP and fail to make AYP in the sample states?

As it was stated in the data collection section of this chapter, the data used for this portion of the study was collected from the NCES restricted access database, the NCES public access website, the NLSLSASD website database, and state agency websites. The schools that were selected for analysis include the same schools that were used for research questions one and two in accordance with the NCES sampling procedures via the NCES restricted access NAEP database. The analysis of data during this portion of the study included the demographic characteristic data, educational resources data, and the AYP status data from each of the schools. The demographic characteristics data that was analyzed individually and separately for associations included: ELL students, special education students, ED students, and students of diverse racial and ethnic backgrounds, which include Black/African American, Latino/Hispanic students, American Indian/Alaskan Native, Asian/Pacific Islander, and White/Caucasian. A logistic regression was administered to analyze the impact that the independent variables had on the dependent variable state AYP status. All of the datasets from schools were weighted in SPSS based on school level enrollment statistics. Listed below is the identification of the independent and dependent variables in the equation.

Independent Variables.

The demographic characteristics of the schools are the independent variables in the third research question. They include the following:

x_1 = SPECIALEDPER = % of special education students (based on a median range provided by NCES, refer to Appendix D).

x_2 = EDPER = % of ED students.

x_3 = BLACKPER = % of Black/African American students.

x_4 = HISPANICPER = % Latino/Hispanic students.

x_5 = NATAMPER = % American Indian/Alaskan Native students.

x_6= ASIANPER = % Asian/Pacific Islander students.

x_7 = WHITEPER = % White/Caucasian students. Note: For the North Carolina 2005 dataset and the Texas 2005 and 2007 datasets, it was necessary to create a DUMMYWHITE variable in place of the WHITEPER variable identical to what was done for research question two.

x_8 = TOTAL REVENUE PER STUDENT = Note: Identical to research question two, this variable was changed to DUMMYREV following the same range that was identified and explained in research question two with the implementation of the logistic regression.

x_9 = ELLPER = % of ELL students (based on a median range provided by NCES, refer to Appendix D).

Dependent Variable

Y = STATEAYP = The proportion of schools that meet AYP standards in the sample states.

Statistical Tests: Alpha Level

The p value is the probability that the null hypothesis is true. We will infer the sample value to the population when statistically significant data is found (Hinkle et al. 2003). The alpha level, which is predetermined by the researcher, provides a confidence level in the statistical data that is analyzed. An alpha level of .05 is the common acceptable practice in the social sciences. Thus an alpha level of .05 was used to determine if empirical findings are statistically significant with the logistic regression.

Statistical Power (Sample Size)

Power is the probability of rejecting the null hypothesis when it is false and it is defined as the formula 1-ß (Hinkle et al. 2003). This study relies on data that is used from NCES. There are between 100 to 450 sample schools in each state dataset and at each grade level according to each sample assessment. Since this study uses NCES school level data that has been tested for reliability by the federal government, the usage of this sample size provides for a sufficient level of statistical power. NCES used stratified random sampling techniques with the selection of the schools that were sampled in each state.

Validity and Reliability

Validity refers to the degree that a measure is quantifying what it claims to measure (Rudestam & Newton, 2007, p. 96). The schools that were selected from each of the sample states are the same ones that were selected by NCES through stratified random sampling techniques for the NAEP administration in mathematics and reading during the 2005 and 2007 school years. The selection of these schools for the sample ensures a high level of validity due to the sampling techniques that were used by NCES.

"Reliability refers to the ability of a measure to produce consistent results" (Rudestam & Newton, 2007, p. 96). Reliability was increased in this study due to the following factors: 1) four different states were used for the sample from different regions across the country, 2) two years of data sets 2005 and 2007 were used, 3) two different content areas were assessed mathematics and reading, and 4) two different grade levels fourth-grade and eighth-grade were used during the data analysis. It is important to use several different datasets in this analysis since past research (Harris, 2007; Kane et al. 2002, Linn & Haug, 2002) shows improvements in scores are often a result of random fluctuations. Harris (2007) found that it is more reliable to use multiple years of measurements, in multiple content areas, and in multiple grade levels vs. one set of data.

The high level of validity and reliability in this study is supported by the usage of the assessment instruments that have been tested for validity and reliability through the NCES, the CDE, the MDE, the PSNC, and the TEA. The following subsection on the NAEP and the Appendices referenced in Table 4 include descriptions that are intended to support the validity and reliability of the assessments used in this study.

NAEP

NCES protects the validity of the NAEP assessment by performing studies to ensure that the assessment adheres to specific standards with the testing of NAEP. NCES has created a validity study panel in order to evaluate the validity of the NAEP assessment. Several studies are available that monitor and lead towards an increase in test validity of the NAEP. The random sampling that NCES adheres to increases the validity of the results (NCES, 2008). The central finding of the validity study is that the NAEP assessment is robust to support the main conclusions about U.S. and state progress in mathematics since 1990 (Daro, Stancavage, Ortega, DeStefano, & Linn, 2007). Appendix M provides information on how NCES ensures validity and reliability with the implementation of the NAEP assessment.

Each of the sample states institutes procedures that increase the validity and reliability of the state assessments in reading and mathematics. A review board was instituted in each state to evaluate the validity of each of the state assessment tools that are used in this study. Listed in Table 4 is information that outlines the state assessments that are used in this study along with a summary of where to find validity and reliability measures taken by each state which is located in the Appendices of this study.

Table 4

State Assessments with Validity and Reliability Information

State	Assessment Tool	Validity and Reliability Appendices Location
California	Standardized Testing and Reporting (STAR)	Appendix N
Michigan	Michigan Educational Assessment Program (MEAP)	Appendix O
North Carolina	End of Grade Tests	Appendix P
Texas	Texas Assessment of Knowledge and Skills (TAKS)	Appendix Q

Quantitative Limitations

Some of the limitations for this study were identified in chapter one. There are additional limitations with the usage of the of the quantitative dataset in this study. The analysis of only the 2005 and 2007 achievement data is a limitation. The decision to use those datasets is due to the administration of the NAEP on a two-year basis. The 2009 NAEP school-level restricted data was not available at the time the application was filed. Since several states now use a growth model (Table 1 provides information that defines the growth model) for AYP, the achievement of AYP might have been different for those states had the growth model been approved for the 2005 and 2007 school years. For example, Michigan was approved under that growth model for AYP in 2008 and the growth formula is not in the statistical measurement of AYP for 2005 and 2007.

Another limitation has to do with the usage of the NAEP school level data. NAEP uses an extremely complicated sampling model in order to determine the percentage of proficiency within a population. However, as mentioned previously, the entire NAEP is never administered to an individual student. Instead NAEP uses sampling procedures where students within similar demographics are matched together in order to come up with the proficiency percentages within a state or a population. Thus the calculation of school level proficiency on the NAEP was done using the portions of the NAEP that were administered to an individual school which usually range in test items anywhere from 20 to 38 test questions depending on the grade and subject tested math or reading. The entire NAEP assessment test items range from 100 to 180 items depending on the grade and subject area tested. Thus the NAEP school level proficiency data is being calculated with the potential that different test items are used in different schools. However, due to the manner in which the NAEP is administered and calculated, this is the only method that could be used to calculate school level proficiency on the NAEP. The state accountability assessment test items range between 31 to 75 test items depending on the grade and subject area tested. Refer to APPENDIX R for information on the number of test items administered for each assessment that is used in this study.

Qualitative Research Design

This part of the study addresses research question four which involves qualitative methodologies. Four schools were selected from schools in the United States in order to conduct interviews with administrators and teachers. The implementation of the semi-structured interview protocol was used to get a better understanding of the impact that AYP is having on classroom instruction and school improvement efforts in the sample schools. The selection of the schools was purposeful as the schools were identified through an analysis of state level AYP data. It was advantageous to identify outlier schools that have been successful under AYP even though they have disadvantaged student populations. It was further advantageous to select a variety of schools with different subgroup populations and experiences with the AYP provisions.

General Design

This study contains a qualitative component because of the need to discover and understand a phenomenon. A phenomenon is the process, perspectives, and worldwide views of people involved. (Merriam, 1998). The studied phenomenon process includes the extent to which and how schools are responding to the accountability components in NCLB. The case study design was employed to acquire an in-depth understanding of the type of strategies that are implemented in schools and the meaning for those involved. The constant comparative method is appropriate because it is inductive and provides for concept-building qualitative research. This research methodology involves taking a single piece of data (one interview statement) and comparing it with all others that may be similar or different which assists with the development of conceptualization of the relations between different forms of data. (Gawlik, 2005).

Sample

The selection of four schools in the United States was purposeful in order to answer research question 4. The schools that were selected to participate in the semi-structured interview protocol were public schools (non-charter). The four schools that were selected among the sample were chosen based on (a) their AYP status, and (b) based on the number and type of specific subgroup populations within the NCLB definition. Some of the schools and interview participants have experience with the restructuring process when schools fail to meet AYP for several consecutive years.

The principal and four teachers from each of the four schools were selected to participate in the semi-structured interview protocol. The teachers who were interviewed were selected based on their involvement with the school improvement plan (SIP) and knowledge that they have in implementing both classroom and building level strategies in response to AYP in NCLB. It was important to select teachers that had some knowledge on the SIP process in the school in order to collect informative and rich data through the semi-structured interview process. Site selection was purposeful and the schools were selected based on selective sampling that included representation of different percentages of subgroup populations within each of the four sample schools. Selective sampling of schools involves identifying and seeking out the schools that represent the widest possible range of interest including student population and district location (Gawlik, 2005).

Data Collection and Analysis

Research Question 4: What impact is AYP having on school improvement initiatives and classroom instruction?

The data collection in this section involved interviewing teachers and principals with a specific set of questions that were selected in order to determine if AYP is having an impact on classroom instruction and school improvement initiatives. All of the questions were created based on the literature review with the purpose to answer the fourth research question in this study. Several of the questions were based on "The "Teachers' Voice Survey" which was part of the Civil Rights Project at Harvard University (Sunderman et al.2004). Appendix S provides the documented permission that was obtained by Dr. Sunderman to use the survey for this study.

Appendix T is a copy of the questions for the semi-structured interviews with the teachers and Appendix U is the administrator interview questions.

A visit to each of the schools occurred in order to conduct the interviews. The participants were not asked to give their names and the information they provided was recorded and coded through the usage of pseudonyms to protect the identification of participants. Consent to participate in this study was acquired through the usage of the Research Information Sheet Template that is provided by the Wayne State University Human Investigations Committee (HIC) Internal Review Board (IRB). Appendix V provides a copy of the information sheet that was used to gain informed consent of the participants in this study.

Participation in this study was voluntary. The interviews lasted for approximately 18 to 35 minutes per participant. The administrator interviews were closer to 30 minutes in total while the teacher interviews averaged about 20 minutes. The interviews were recorded using an ipod and saved on a wave file format, they were then transcribed following the interview.

An analysis of the transcription data occurred via the strategies described in the rest of this section.

The potential influence of the NCLB school accountability on the implementation of school strategies and school improvement process at the school level is a critical question that is answered in this study. This also leads to answering the question regarding the impact that NCLB is having on school achievement. Thus a case study methodology is appropriate as it allowed for the collection and analysis of data in order to triangulate the findings through multiple sources of data that include the interviews, the quantitative data, and the building visitations that included a tour of the school.

The data that was collected and analyzed from the four schools in this sample allowed for a comparative analysis of the differences among the schools in the sample. (Merriam, 1998). Since the schools have different demographic characteristics and have experienced varying levels of success and hardships under AYP, the comparative analysis provides for an increase in reliability through the triangulation of the data. Triangulation of the data distinguishes and compares the rich findings that were discovered in each of the four sample schools.

The comparisons among the schools allowed for the development of tentative categories or domains that were coded, compared, and analyzed against each other and with the quantitative data analysis in this study. This analytic tool was then used to evaluate the type of school improvement initiatives and strategies that were implemented in relation to their AYP status. The findings show that there are some similar categories and differences, which might be directly attributable to the type of AYP label and the sanctions that schools were required to implement under the NCLB legislation. The interview method with the objective to triangulate the data provided for some rich findings. It allowed for the acquisition of in-depth knowledge and experience as it provided insight into the thoughts and ideas of experienced educators working in the schools. In fact, some of the educators have worked in multiple states and schools over the years and have been exposed to different levels of school accountability. The interviews allowed for an acquisition of the knowledge and experience from the participants

Field notes and records of strategies were kept along with notes in relationship to the observations that were discovered during the school tours. This information was valuable during the coding of the information through the domain analysis. Documents for the schools regarding accountability were found on the state websites and school level websites. This included an analysis of the school report cards that are required by NCLB along with the analysis of multiple years of data and in multiple content areas. Artifacts were also collected from the school, which included the school improvement plan and other relevant documents that provided information into the type of strategies that schools were implementing which often supported the participant comments during the interviews. Many of the documents were requested based on answers provided by the participants during the semi-structured interview protocol.

Analytic and Constant Comparative Method

As mentioned, the data analysis involved finding emerging categories or domains that were discovered during the interviews in relation to school achievement and the AYP status of schools. Interviews and documents were the two sources that provided analysis of multiple perspectives. They were then compared to other incidents in the data, and comparisons were made that led to tentative categories. As the categories got refined and subcategories were created, a framework for patterns and relationships among the coded categories began to develop (Gawlik, 2007). NVIVO 9.0 was used to support the coding of the domains in the qualitative analysis.

The emerging categories or domains were analyzed in comparison with the documents (artifacts), field notes (which included tours of the schools and multiple classrooms), and the quantitative dataset analysis. The domain categories and findings were also analyzed in comparison to the schools success level under NCLB and the AYP accountability provisions. As the domains were refined through the qualitative analysis, subcategories or (semantic domains) were created and patterns or relationships among the categories were found and coded. This allowed for the triangulation of data to support the overall findings in this study. The inclusion of excerpts from the interview and documents provided further support for the analysis.

Generally, interview questions asked focused on the influence of accountability on the school improvement plan, classroom instruction, and school level responses. Moreover, questions were asked that sought the meaning of accountability and its impact on school performance, professional development, and everyday life at the school building. The wording and sequence of questions were pre-determined and followed the same order for all teachers and principals. This reduced interviewer bias; however, standardized questions tend to constrain the responses of the interviewee. Therefore, probes were used when appropriate to expound on existing responses or to clarify terms that were not understood by the interviewer. By relying on standardized, semi-structured interview questions, the research was able to compare responses across schools, teachers, and administrators. (Gawlik, forthcoming; Patton 1990).

The general inductive approach was used where findings emerged from the dominant or significant themes inherent in the raw data, without the restraints imposed by more structured methodologies. This type of qualitative data analysis is intended to aid an understanding of the meaning in complex data through the development of summary themes or categories from the raw data (Miles & Huberman, 1994; Pope et al., 2000). Hence, data analysis for this study was determined by both the research objectives and multiple readings and interpretations of the raw data. The coding began with the close readings of the text followed by consideration of the multiple meanings inherent in the responses. The upper level or more general codes were derived from the research aims while the lower level or specific categories were derived from multiple meanings of the raw data (N Vivo coding). Coding was assisted using specialist software known as NVIVO 9.0 that highlighted relevant text for each code and stored the text. Coding was further implemented in a very visual manner using cue cards and a large board that allowed for a large visual analysis of the comparative data. After coding was complete, the search for subtopics, including contradictory points of view and new insights that conveyed a theme or category were created. As the categories became refined and subtopics created, a framework for patterns and relationships among the coded categories began to develop. The outcome of the process created themes that captured aspects of the raw data and assessed to be the most important themes in light of the research objectives (Gawlik, forthcoming).

Spradley (1980) believes that domain analysis is a systematic examination of something to determine its parts, the relationship among parts, and their relationship to the whole (p.85). A cultural domain analysis was completed from the interviews with teachers and administrators in the four schools. "A cultural domain is a category of cultural meaning that includes other smaller categories" (Spradley, 1980, p. 88). When cultural domains are created, the focus of the analysis is on the people or actors, the places, and the events or social activities. The organization of the domain included a cover term which was organized according to semantic relationships with included terms. The description of cultural domains involved the use of language. Cover terms, included terms, and semantic relationships are all words and phrases that will define and give meaning to the objects, events, and activities observed (Spradley, 1980, p. 89). The collection of artifacts in the form of quantitative data assisted with the organization of the domain analysis. The domain analysis from the interviews allowed for the development of organizing patterns of information. As mentioned, artifacts that were collected include the school AYP report card that is provided through the databases along with a copy of the school improvement plan, school annual reports, and information that is provided on the school website. The data from the quantitative portion of this study was also used to triangulate the data during the analysis of the interviews.

The second type of analysis that was used is a taxonomic domain analysis. A taxonomy is a set of categories organized on the basis of a single semantic relationship. The taxonomic analysis allowed for a deeper analysis than the creation of the cultural domain analysis. The focus of the taxonomic analysis involved dividing the included terms and analyzing relationships within the cultural domains (Spradley, 1980, p. 112). The taxonomic analysis involved looking at several factors with regard to how individuals react to the AYP status of schools.

The third analysis strategy that was used in this research study is a componential analysis. When the interviewer discovers contrasts among members of a domain, the contrasts are best described as attributes of meaning (Spradley, 1980, p. 131). Wolcott (1990) believes that sorting the data involves both analysis and interpretation. Both of these techniques were used during the collection and analysis of data for this research study. "Analysis, used in this narrower sense, follows, standard procedures for observing, measuring, and communicating with others about the nature of what is 'there', the reality of the everyday work as we experience it" (Wolcott, 1990, p. 33).

Validity and Reliability

Lincoln and Guba (1985) suggest that there are three major activities that increase the probability that credible findings will be produced. They include prolonged engagement, persistent observation, and triangulation (p. 301). The credibility from this study is established through a triangulation of the data. The data was triangulated through three methods of data collection and subsequent analysis. The first method is the semi-structured interview protocol and the second method was the tour of the school and visitations to a sample of the classrooms in the school. This also involved interactions with faculty and students during the tour. The collection of artifacts in the form of the school improvement plan and the school report card that provides the school AYP grade is a third piece of data that was collected to support triangulation. Since the qualitative portion of this study is not as extensive as the quantitative sections, the analysis of the quantitative data which includes the state accountability assessments, and the demographic characteristic data further supports the triangulation of data. In order to further support valid findings, connections were also made with the literature review throughout the analysis of this portion of the qualitative study.

Even though the qualitative section was not the larger portion of the study, the data collected from this analysis provides important findings that would otherwise not be identified through the quantitative data analysis. For example, the quantitative data does not provide any information with regard to the strategies that schools are implementing which result in either successful or unsuccessful state accountability test scores. The quantitative data does not provide information regarding how administrators and teachers implement school improvement strategies or specific instructional strategies that are a response to AYP and the NCLB reform. Since educators in the school are responsible for making the improvements, attaining their viewpoints was important to determine the impact that NCLB is having on classroom instruction and school improvement initiatives.

The interview data from this portion of the study was triangulated with the quantitative data in order to support and/or refute whether or not AYP is having a positive impact on classroom instruction and school improvement initiatives. As it was indicated in the literature review, improvements in test scores could be a result of numerous factors that have nothing to do with the improvement in classroom instruction. Thus the interviews assisted with determining the level of success with implementing strategies at the school level and classroom level in response to the AYP provisions in NCLB.

The validity for this part of the study is further supported through the usage of the interview questions that were based on the questionnaire that was published by Sunderman et al. (2004) in the Harvard Civil Rights Project. The validity of this study was increased by piloting the semi-structured interview protocol with 10 teachers and 2 administrators in several schools.

Qualitative Limitations

There are limitations to the reliability of the study since only four schools are in the qualitative sample. Also, five educators were interviewed from each school for a total of 20 participants which includes 16 teachers and 4 principals. It is possible that the purposeful selection of specific teachers that are involved with the school improvement plan could provide results that might not otherwise be found with the overall faculty. However, selection of teachers involved with the school improvement plan was critical in order to acquire an adequate understanding of the impact that AYP is having on the school improvement process. Also, the participants in the interviews appeared to provide honest and informative answers when they described the impact that AYP was having on other teachers in the school.

The findings from the study can not be used to generalize the impact that AYP is having on classroom instruction and school improvement efforts in other schools in the United States. However, through a triangulation of the data between the four schools there are some rich findings that provide insight into how schools are reacting to the AYP formula.

Table 5 outlines the research questions for this study, instruments, and data analysis techniques.

Table 5

Summary of Research questions, instruments, and data analysis tools

Research Questions	Instruments	Data Analysis Tools
Is there a significant correlation between the schools that meet proficiency standards on the NAEP and the schools that make AYP in the sample states? Are there significant differences among the subgroups?	State AYP data in the sample states along with NAEP data from grades 4 and 8 in the sample states.	Pearson's Product-Moment Correlation
Is there a significant correlation between the schools that meet proficiency standards on the NAEP and schools that meet proficiency standards on the state accountability assessments (STAR, MEAP, End of Grade Test, and TAKS) in the sample states? Is there a combination of factors that best predicts proficiency status on the NAEP and state accountability assessments?	NAEP restricted school level data in mathematics and reading at the 4^{th} and 8^{th} grade level. State proficiency assessment data from the sample of schools where NAEP data was collected in the sample states for the years 2005 and 2007 available through state databases and websites.	Pearson's Product-Moment Correlation Logistic Regression
Are the demographic (categorical) characteristics and educational resources significant predictors of the schools that make AYP and fail to make AYP in the sample states?	State AYP school data and demographic data that is available through state database websites and the National Center for Education Statistics (NCES).	Logistic Regression
What impact is AYP having on school improvement initiatives and classroom instruction?	Qualitative research techniques through interviewing principals and teachers in four schools in one of the sample states.	Semi-structured interview protocol and the triangulation of interview data.

Research Design Significance

This study involves the collection and analysis of empirical data regarding the external validity and reliability of using AYP as a school effectiveness and accountability measurement tool. The study provides data and findings that support, call into question, and justify the need for revisions to the current AYP formula. As the literature review suggests, there is a problem with the effectiveness of implementing the AYP accountability provisions in NCLB. If educators at the school level are not taking steps to enhance the implementation of classroom instruction and school improvement efforts, then the intent of the reform is not successful with the achievement of the goals in NCLB. The findings from this research study provide information to inform the public, politicians, and educators regarding the strengths and weaknesses of the AYP formula, the need for potential changes and enhancements, including the areas that should be sustained with the implementation of the AYP formula.

CHAPTER IV

RESULTS

Presentation of Results

The presentation of the results in this chapter includes formal text that reports the objective findings. The first section of this chapter involves the identification and analysis of the AYP accountability formula in each state. The subsequent portions of this chapter follow a sequential model according to each of the four research questions. The first three research questions involve quantitative data and include the presentation of tables and graphs in order to provide an illustration of the findings in this study. Numerous output tables from SPSS are referenced in this chapter and displayed in the Appendices for reader analysis. Since a large portion of the data is restricted access data, some of the statistical tables that were created during the data analysis portion of this study were only included in the presentation of the results after receiving a formal approval from NCES to maintain the confidentiality agreement that was signed with NCES.

The qualitative data includes the presentation of formal text that illustrates the findings. Referenced portions from the semi-structured interviews were used to provide an illustration of the data and findings. The analysis of the qualitative data includes a display table to illustrate the findings that were discovered through a triangulation of the qualitative data.

AYP Formulas in the Four Sample States

This section provides information on the AYP formula in each of the four sample states. The information includes the state assessments in mathematics and reading that are used for the calculation of AYP at the elementary and middle school grades. All four states require a 95% participation rate among the students and subgroups in order to meet AYP standards. All four states use alternative assessments for ELL students[14] and special education students[15] which count for participation rates under AYP requirements. Information is presented in this section establishing the state formula for AYP that includes; the minimal student population size that is required to qualify for a subgroup in a school, a description of the confidence interval that is used, the line of trajectory that the state implements regarding the number of students that need to be proficient in a school, information on growth model formulas in the state, the alternative measure used as required by NCLB, and any other pertinent information for the calculation of AYP in the state.

California

The California Standardized Tests are part of the Standardized Testing and Reporting System (STAR) which provide the calculation of AYP in the state of California. The tests used for AYP are administered to students annually in the areas of mathematics and reading in grades two through eight at the elementary and middle school level (CDE, 2009). In order for a subgroup within a school to be calculated under AYP, a minimal group size of 50 is required and those students must represent at least 15 percent of the student body. Otherwise a subgroup will be counted when that subgroup size reaches 100 or 15% which ever one is lower. Thus California institutes a formula subgroup size that ranges between 50 to100 students. It is worth

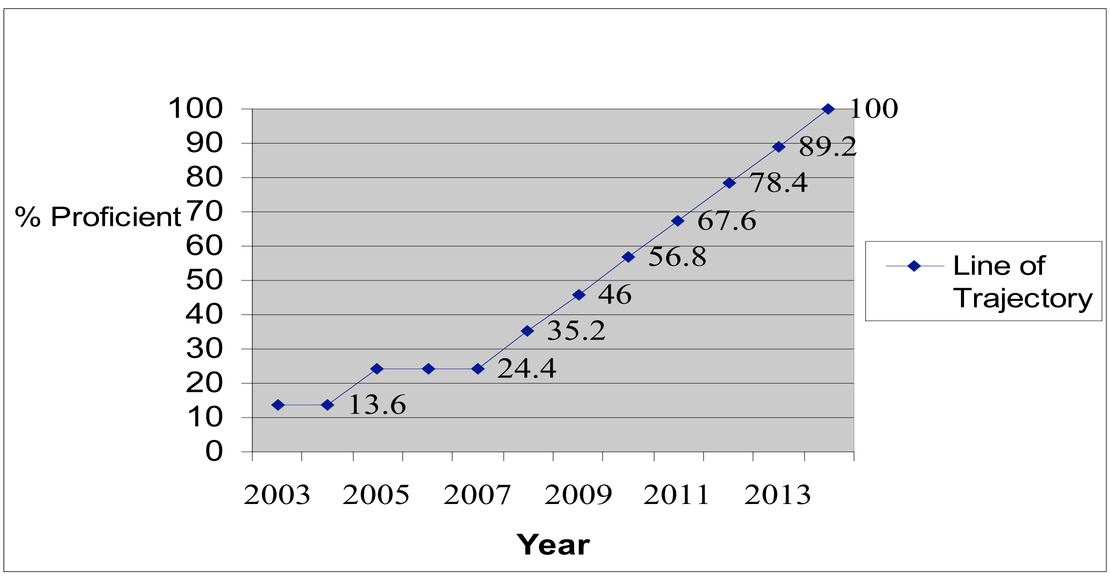
---

[14] Only students in the country for less than one year can use the ELL alternative assessments which count for participation rates only and not proficiency rates.

[15] There is a 2 to 3 percent cap regarding the number of students that can take the special education alternative assessment and count as proficient for AYP determination. All students who take the assessment above the cap are deemed non-proficient but they do count under the participation rate of 95%.

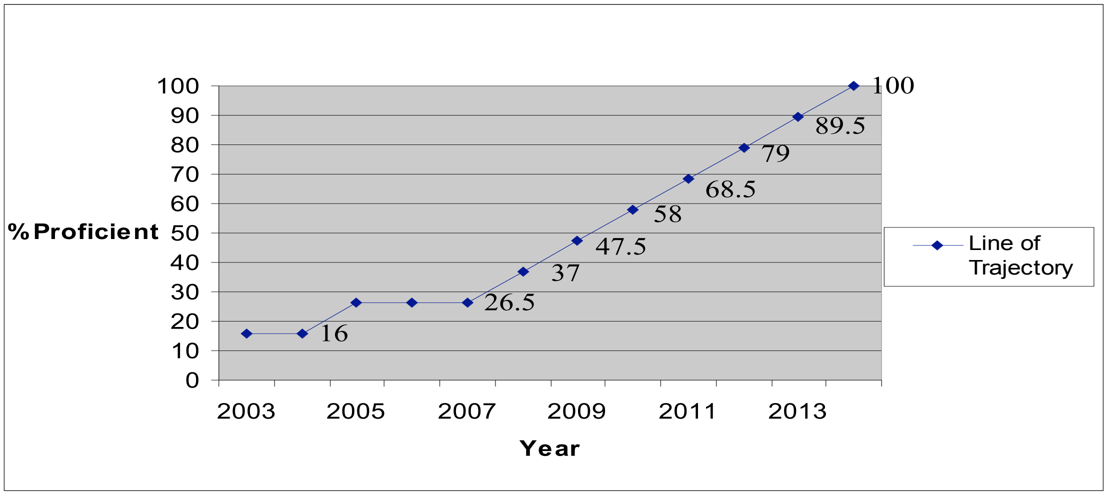
noting that students included in the ELL subgroup will count in that subgroup until they meet proficiency standards on the Alternative English assessment and regular state accountability assessment for multiple years. California applies a 99 percent confidence interval for schools and districts with less than 100 scores when determining AYP calculations. California institutes a stair step trajectory that is backend loaded until 2014 with a 100% proficiency requirement. Figure 14 illustrates the trajectory for mathematics and Figure 15 illustrates the trajectory for English language arts (CDE, 2009).

Figure 14. Mathematics trajectory for California



Note: This figure is based on information from the California Consolidated State application (2008).

***Figure 15*. ELA trajectory for California**



Note: This figure is based on information from the California Consolidated State application (2008).

The state of California has not implemented a growth model formula for the calculation of AYP in the state. However, the alternative indicator is based on the state's academic accountability report know as the Academic Performance Index (API). The API measures the performance and growth of schools due to statewide tests at grades two through eight at the elementary and middle school level. The API numeric scale ranges from a low of 200 to a high of 1000 that reflects a school's level of performance. The API target is 800 for all schools although schools can also meet proficiency standards by making annual gains. The tests that are used for the calculation of API (Alternative Indicator in California) are weighted. Those assessments include science and social studies in the measure even though they are not required by AYP (CDE, 2009). The list of tests and their weights for the calculation of the API index are listed in Table 6.

Table 6

The alternative AYP indicator: API index in California

Statewide Test Name	Grades	Weight
CST/CAPA in English Language Arts	2-8	0.48
CST/CAPA in Mathematics	2-8	0.32
CST in Science	5	.20
CST in HSS	8	.20
NRT Reading	3, 7	0.06
NRT Language	3, 7	0.03
NRT Spelling	3, 7	0.03
NRT Mathematics	3, 7	0.08
CST in Science	8	.20
CST in Mathematics	8	.10

Note: This table is based on information from the California Consolidated State application (2008).

Michigan

The Michigan Educational Assessment Program (MEAP) is used for the calculation of AYP which is administered to students annually in the areas of mathematics and English language arts/reading in grades three through eight at the elementary and middle school level. A minimal subgroup size of 30 is required in order for a subgroup to be calculated under the AYP formula. Michigan applies a 95% confidence interval for the calculation of schools that meet AYP proficiency standards. Michigan institutes a stair step trajectory that is backend loaded until 2014 with a 100% proficiency requirement. However, the state sets their trajectory at different levels for each grade three through eight. Table 7 and Table 8 provide information on the specific grade level requirements in mathematics and English language arts/reading. Figure 16 illustrates the trajectory for mathematics and Figure 17 illustrates the trajectory for English language arts/reading. Both figures are based on the 4th grade trajectory requirement since Michigan modified their trajectory requirements in 2005-06. Under the safe harbor provision, schools in Michigan can meet AYP if they score below the proficiency trajectory as long as they reduce the number of students who failed to meet proficiency standards from the previous year by 10%.

The state of Michigan was approved for the implementation of a growth model formula for the calculation of AYP starting in the 2008 school year. However, since the data used in this study is from the academic years 2005 and 2007 the growth model formula is not used for the calculation of AYP in this study. The alternative indicator that the state uses is an attendance rate of 90% at the elementary and middle school level.

Table 7

Reading ELA AYP percentage proficiency requirements in Michigan

School Year	Reading/ELA						
	Grade						
	3	4	5	6	7	8	11
2001-02		38%			31%		42%
2002-03		38%			31%		42%
2003-04		38%			31%		42%
2004-05		48%			43%		52%
2005-06	50%	48%	46%	45%	43%	41%	52%
2006-07	50%	48%	46%	45%	43%	41%	52%
2007-08	60%	59%	57%	56%	54%	53%	61%
2008-09	60%	59%	57%	56%	54%	53%	61%
2009-10	70%	69%	68%	67%	66%	65%	71%
2010-11	78%	77%	76%	75%	74%	73%	79%
2011-12	86%	85%	84%	83%	82%	82%	86%
2012-13	93%	92%	92%	91%	91%	91%	93%
2013-14	100%	100%	100%	100%	100%	100%	100%

Note: This table is based on information from the Michigan Consolidated State application (2009).

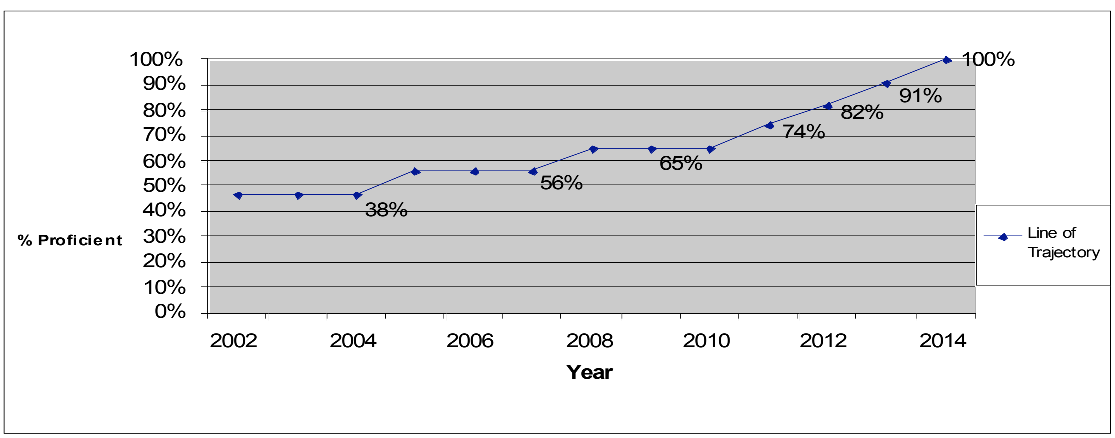
Table 8

Mathematics AYP percentage proficiency requirements in Michigan

School Year	Mathematics Grade 3	4	5	6	7	8	11
2001-02		47%				31%	33%
2002-03		47%				31%	33%
2003-04		47%				31%	33%
2004-05		56%				43%	44%
2005-06	59%	56%	53%	50%	46%	43%	44%
2006-07	59%	56%	53%	50%	46%	43%	44%
2007-08	67%	65%	62%	60%	57%	54%	55%
2008-09	67%	65%	62%	60%	57%	54%	55%
2009-10	67%	65%	62%	60%	57%	54%	55%
2010-11	75%	74%	71%	70%	67%	66%	67%
2011-12	83%	82%	81%	80%	78%	77%	78%
2012-13	91%	91%	90%	90%	89%	89%	89%
2013-14	100%	100%	100%	100%	100%	100%	100%

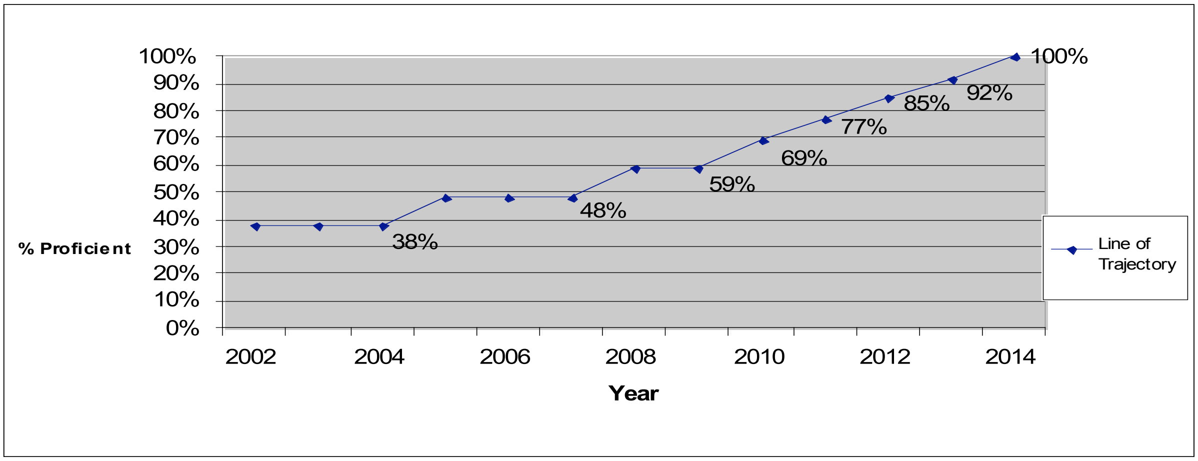
Note: This table is based on information from the Michigan Consolidated State application (2009).

Figure 16. Mathematics trajectory for Michigan



Note: This figure is based on information from the Michigan Consolidated State application (2009).

Figure 17. ELA trajectory for Michigan



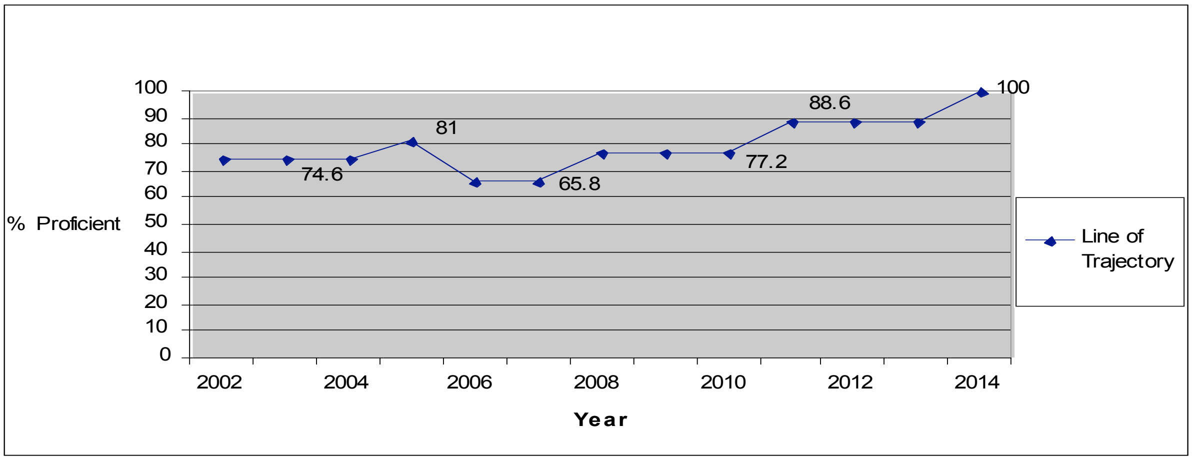
Note: This figure is based on information from the Michigan Consolidated State application (2009).

North Carolina

The End of Grade Tests is used in North Carolina for the calculation of AYP and the assessments are administered to students annually in the areas of mathematics and reading in grades three through eight at the elementary and middle school level in the state of North Carolina (State of North Carolina, 2008). If a school or district does not meet the participation rate of 95% in an individual year, the rate will be averaged over a two or three year period. In order for a subgroup within a school to be calculated under AYP, a minimal group size of 40 is required. North Carolina applies a 95% confidence interval for the calculation of schools that meet the AYP standards. North Carolina institutes a stair step trajectory that is somewhat complicated due to the fact that the assessments changed in the year 2007. Thus the stair step reduces in the middle years and then rises to 100% by 2014. Figure 18 illustrates the trajectory for mathematics and Figure 19 illustrates the trajectory for English language arts (State of North

Carolina, 2008). Under the safe harbor provision schools in North Carolina can meet the AYP requirements if they score below the proficiency trajectory as long as they reduce the number of students who failed to meet proficiency standards from the previous year by 10%.

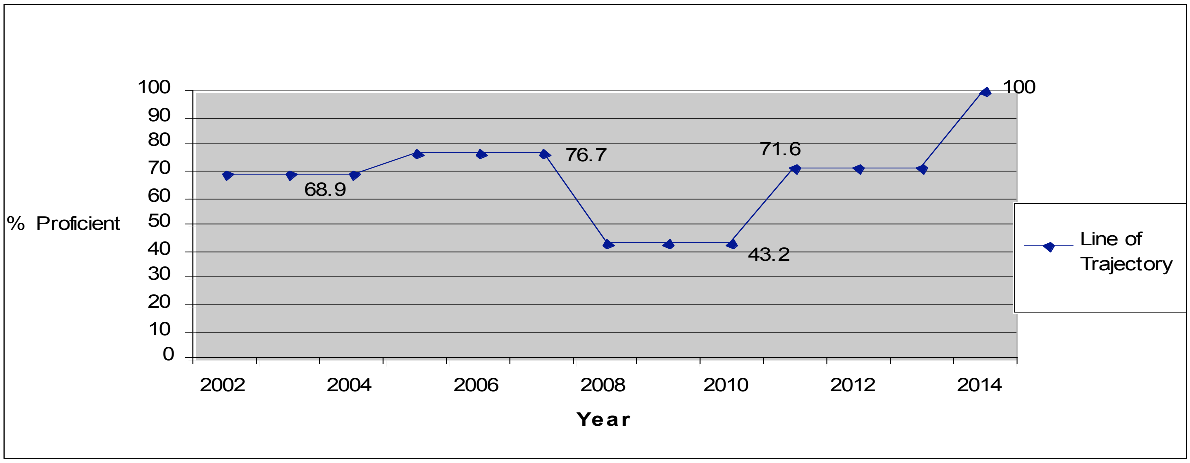
Figure 18. **Mathematics Trajectory for North Carolina**



Note: This table is based on information from the North Carolina Consolidated State Accountability Workbook (2008).

Along with Tennessee, the state of North Carolina was one of the first two states in the country that were approved for the implementation of a growth model formula for the calculation of AYP starting in the 2006 school year. Since the data used in this study is from the academic years 2005 and 2007 the growth model formula is only used for the calculation of AYP for the 2007 dataset. The alternative indicator that the state uses is an attendance rate requirement of 90% at the elementary and middle school level.

***Figure 19.* Reading trajectory for North Carolina**



Note: This table is based on information from the North Carolina consolidated state accountability workbook (2008).

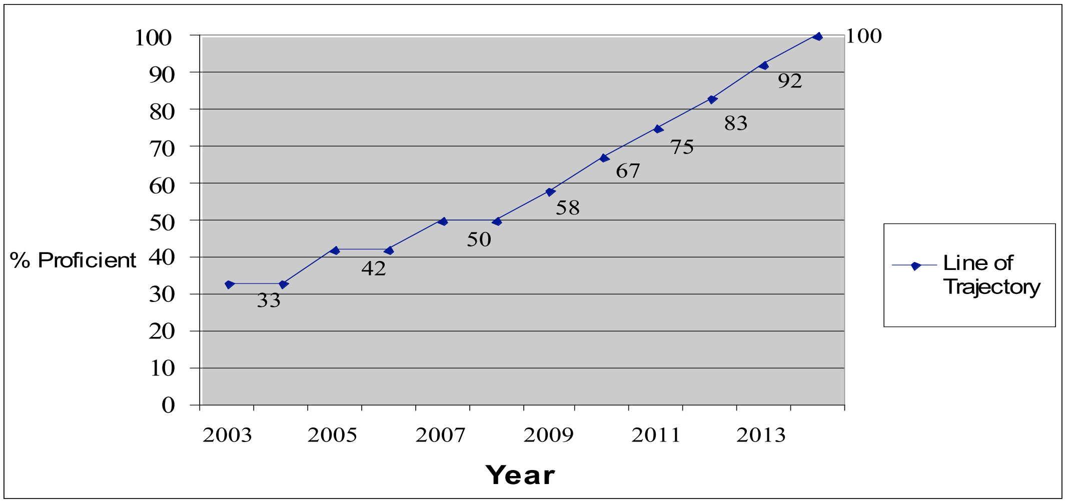
Texas

The Texas Assessment of Knowledge and Skills (TAKS) is used for the calculation of AYP which is administered to students annually in the areas of mathematics and reading in grades three through eight at the elementary and middle school level. The state requires a 95% participation rate for a school or campus to meet AYP proficiency standards, however, that ratio can be averaged over a two year period in order to meet the requirement. In order for a student subgroup to be included in the AYP performance calculation, a district or campus must have test results for 50 or more students in a subgroup at the elementary and middle school level. However, in order to use a minimum subgroup rate of 50 students, the subgroup must comprise at least 10% of all test takers in the subject. If the student subgroup reaches a total of 200 students, it will count as a subgroup regardless of whether or not it comprises a total of 10% of the school population. The result is that Texas uses a formula ranging from 50 to 200 students that qualifies as a minimal subgroup size. In the area of major racial or ethnic subgroups, there are only three ethnic subgroups that count for the calculation of AYP including, African American, Hispanic, and White. It is interesting that in the Texas Consolidated State Application (TEA, 2009) that was submitted for the USDOE, Texas uses six pages to justify the use of their minimum subgroup formula. In all other sections of the application they tend to use one or two pages to justify their AYP process. This suggests that the TEA knows it was not a sound recommendation with the minimal subgroup size. It also aligned with the literature review which showed that subgroup manipulation can have a major impact on school AYP results in specific states (Olson & Jacobson, 2006).

Texas applies a confidence interval for the calculation of AYP if the school population ranges from 10 to 50 students in total. Texas institutes a straight line trajectory, although there are two smaller stair steps in the earlier years. Figure 20 illustrates the trajectory for mathematics and Figure 21 illustrates the trajectory for reading. Under the safe harbor provision, schools in Texas can meet the AYP requirements if they score below the proficiency trajectory as long as they reduce the number of students who failed to meet proficiency standards from the previous year by 10%.

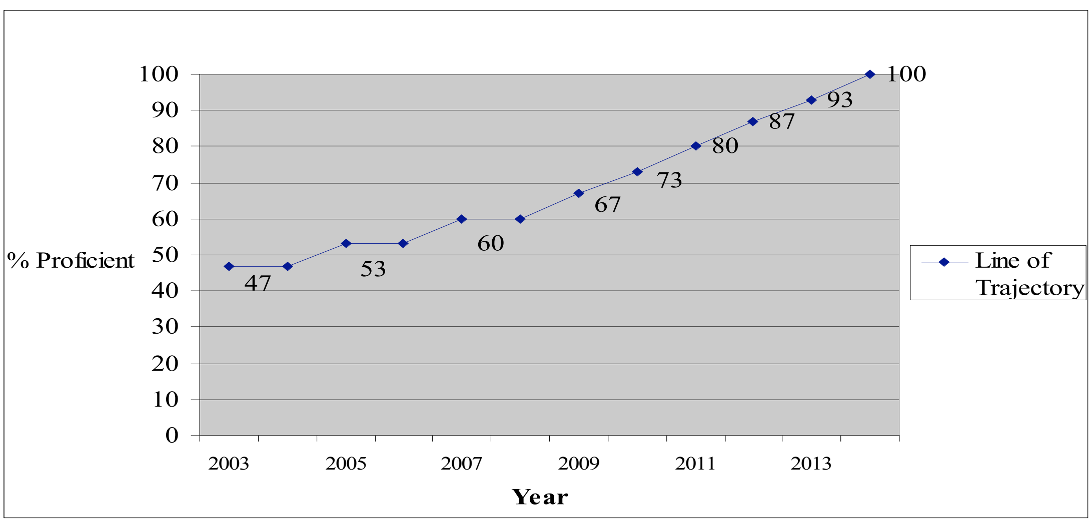
The state of Texas was approved for the implementation of a growth model formula for the calculation of AYP starting in the 2009 school year. However, since the data used in this study is from the academic years 2005 and 2007, the growth model formula is not used for the calculation of AYP in this study. The alternative indicator that the state uses is an attendance rate requirement of 90% at the elementary and middle school level.

Figure 20. Mathematics trajectory for Texas



* This figure is based on the TEA Consolidated State Accountability Workbook.

Figure 21. Reading trajectory for Texas



* This figure is based on the TEA Consolidated State Accountability Workbook.

Quantitative Results and Analysis

The purpose of the remainder of this section in chapter four is to report the results of the statistical analysis that is used to answer the first three quantitative research questions. The Analyses include Pearson's Product Moment Correlation and Logistic Regression. This section is organized first by research question and then according to the results in each state which are in-turn divided by year (2005 and 2007), grade level (4th and 8th), and content area of each assessment (mathematics and reading/ELA). Data tables and figures are displayed in the first section of the analysis to give a sample of the type of SPSS outputs that were analyzed. The remainder of the tables and figures were placed in the Appendices. Terminology is used to give meaning and explain the strength of the relationships with the data. Table 9 provides a definition for the terminology used to explain the r value with the Pearson Correlation Analysis.

Table 9

Pearson Correlation Analysis r value defined

Terminology/Definition	Approximate Pearson Correlation Range for the r value
Minor	(+ or -) .000 to .099
Moderate	(+ or -) .100 to .399
Large, Strong, or Sizable	(+ or -) .400 to .549
Extremely Large or Strong	(+ or -) .550 to 1.0

Research Question One

1. Is there a significant correlation between the schools that meet proficiency standards on the NAEP and the schools that make AYP in the sample states? Are there significant differences between subgroups?

The hypothesis for research question number one (as stated in chapter three) is that there would not be a strong relationship between the NAEP assessment school level proficiency and state AYP school level proficiency. For the second part of research question one, it was predicted that there would be a statistically significant relationship between state AYP results and the demographic characteristics (categorical variables) along with school funding. There was also a prediction that there would be a relationship between NAEP proficiency status and the demographic characteristics including school funding. In order to answer research question one and test the predicted hypothesis, a Pearson's Product-Moment Correlation was used to analyze each dataset.

California 2005 4th Grade

For the 2005 dataset at the 4th grade level, there was a moderate positive correlation between the two variables for the proficient level NAEP scale and state AYP results ($r = 0.192$, respectively, significant at the .01 level). There was a much stronger positive correlation for the basic level NAEP scale and state AYP results ($r = 0.563$, $p < .01$). These correlations were not surprising as the literature review established that there was a stronger relationship between the basic level NAEP scale and the state accountability assessments vs. the proficient level NAEP scale and the state accountability assessments.

The correlation results with state AYP and the subgroups included a large positive correlation with WHITEPER ($r = .571$, $p < .01$) and a moderate positive correlation with NATAMPER ($r = .365$, $p < .01$) and SPECIALEDPER ($r = .136$, $p < .01$). There was a minor positive correlation with ASIANPER ($r = .076$, $p < .01$). An extremely large negative correlation with EDPER ($r = -.623$, $p < .01$), HISPANICPER ($r = -.632$, $p < .01$), and ELLPER ($r = -.577$, $p < .01$) was found. There was a moderate negative correlation with TOTALREVENUE ($r = -.187$, $p < .01$) and a minor negative correlation with BLACKPER ($r = -.044$, $p < .01$).

It is not surprising that the WHITEPER variable had a strong positive association and the ASIANPER had a weak positive association. Nor was it surprising that the EDPER had a strong negative correlation. The results from subsequent datasets indicate that the BLACKPER, HISPANICPER, SPECIALEDPER, and ELLPER variables do not appear to be consistent with their associations although the predicted hypothesis and the causal relationship analysis (as discussed in chapter three) predict that there would be a negative association. However, it was somewhat surprising that the TOTALREVENUE variable had a negative association.

While the Pearson correlation test provides information on *whether there is a relationship between existing variables by offering a correlation coefficient, it does not indicate the direction of the correlation.* However, the analysis indicates whether the relationship is positive or negative in association to one another. If the results are negative, then the relationship is inverse. For example, as the EDPER variable increases to a higher level, the state AYP result would produce a lower score or vice versa. Since the state AYP score is a dichotomous variable with a score of 1 or 0 as indicated in chapter three, the California 2005 4th grade results can be interpreted to suggest that a larger EDPER value is more closely associated to schools that fail to meet AYP. If the Pearson Correlation value is positive, then the results can be interpreted to mean that as the value increases, there is a greater relationship with schools that met AYP proficiency as in the case of the WHITPER variable within this dataset. However, it must be noted that the Pearson Product-Moment Correlation is not as robust of a measure as the Logistic Regression analysis that is used to answer research question two and three.

As table 10 illustrates, the correlation with state AYP and the subgroup variables and the correlation between the NAEPPROFbasic level variable and subgroup variables are closer in value vs. the correlation with the NAEPPROFprof level variable and the subgroup variables. For example the correlation between state AYP and the WHITEPER variable produced an r value of .571 while the NAEPPROFbasic level variable and WHITEPER had an r value of .687. In contrast, the correlation between NAEPPROFprof level variable and WHITEPER value had an r value of .330. As Table 10 further indicates, there is a closer association with the state AYP results and the NAEPPROFbasic variable vs. the state AYP results and the NAEPPROFprof variable which is a consistent finding among the subgroups.

The results in this dataset support research in the field which shows that the basic level NAEP scale is more closely aligned with state AYP proficiency vs. the proficient level NAEP scale (Bandeira de Mello, Blankenship & McLaughlin, 2009; Gewertz, 2010). Table 9 provides descriptive statistics and Table 10 provides the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.[16]

---

[16] All of the r values signify the relationship to the dependent variable state AYP

Table 10

Descriptive statistics for the 2005 California 4th grade dataset

	Mean	Std. Deviation	N *
STATEAYP	.53	.499	280142
NAEPCOMBPROFbasic	.3600	.48001	284558
TOTAL REVENUE PER STUDENT	9224.61	1269.325	280172
EDPER	.6234	.30004	284558
ASIANPER	.0954	.12434	284558
NATAMPER	.0062	.01450	284558
BLACKPER	.0834	.10624	284558
HISPANICPER	.5701	.30628	284558
WHITEPER	.2314	.25221	284558
ELLPER	.4250	.29376	279937
SPECIALEDPER	.0933	.08661	269135
NAEPCOMBPROFprof	.0401	.19629	284558

* Different weighted sample sizes among the subgroups are present in Table 9 because some of the schools only took one of the NAEP assessments (math or reading) resulting in the inability to calculate a school level NAEP proficiency. Also, some of the schools did not have a recorded AYP proficiency score or state accountability assessment data for different reasons. A choice was made not to eliminate the schools with incomplete data because they provided information on the relationships between the subgroup populations and the proficiency assessment results (either NAEP or state accountability assessments) that were recorded in the database. In the places where missing datasets are present, SPSS does not use those sample schools in the calculation of the relevant outputs. This is consistent among all of the descriptive statistics tables that are reported in this study.

Table 11

Pearson correlation matrix for the 2005 4th grade listed variables (demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.563**	.192**	-.187**	-.623**	.365**
NAEP BASIC level	.563**	1	.273**	-.145**	-.733**	.367**
NAEP PROF level	.192**	.273**	1	.061**	-.391**	.229**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.076**	-.044**	-.632**	.571**	-.577**	.136**
NAEP BASIC level	.046**	-.126**	-.704**	.687**	-.644**	.079**
NAEP PROF level	-.025**	-.115**	-.334**	.330**	-.246**	-.059**

**. Correlation is significant at the 0.01 level (2-tailed)

California 2005 8th Grade

For the 2005 dataset at the 8th grade level, it was not possible to conduct a Pearson Product-Moment correlation with state AYP and the NAEPPROFprof level variable because none of the California 8th grade schools met NAEP scale proficiency at the proficient level in both reading and mathematics. However, there was a moderate positive correlation between the NAEPPROFbasic level variable and state AYP results (r = .336, p < .01). This finding further supports the earlier conclusion in the literature review which establishes a closer correlation between the basic level NAEP scale and state proficiency scores vs. the proficient level NAEP scale and state AYP.

The correlation results with state AYP and the subgroups included a large positive correlation with WHITEPER (r = .582, p < .01), and a moderate positive correlation with

NATAMPER (r = .149, p < .01), and ASIANPER (r = .360, p < .01). There was an extremely large negative correlation with EDPER (r = -.658, p < .01) and HISPANICPER (r = -.609, p < .01). There was also a large negative correlation with ELLPER (r = -.481, p < .01). There was a moderate negative correlation with TOTALREVENUE (r = -.293, p < .01), BLACKPER (r = -.163, p < .01), and SPECIALEDPER (r = -.186, p < .01). Consistent with the 2005 4th grade California dataset, the WHITEPER variable had a strong positive association and there was a moderate positive association with ASIANPER.

It was anticipated that the EDPER variable would have a strong negative correlation. The BLACKPER, HISPANICPER, SPECIALEDPER, and ELLPER variables were predicted to be negatively associated with the school AYP status based on the information that was reported in chapter three under the heading causal hypothesis. However, it was surprising that the TOTALREVENUE variable had a negative association as research in the field suggests that school funding may be an indicator of school success. Please refer to the Tables in Appendix W for descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and the financial variable.

California 2007 4th Grade

For the 2007 dataset at the 4th grade level in California, there was a moderate positive correlation between the two variables for the proficient level NAEP scale and state AYP results (r = 0.178, respectively, significant at the .01 level). There was a stronger positive correlation for the basic level NAEP scale and state AYP results (r = 0.393, p < .01). The correlation results with state AYP and the subgroups included a moderate positive correlation with WHITEPER (r = .300, p < .01) and ASIANPER (r = .316, p < .01). There was a minor positive correlation with NATAMPER ($r = .010, p < .01$) and BLACKPER ($r = .024, p < .01$) . There was a sizable negative correlation with HISPANICPER ($r = -.404, p < .01$) and ELLPER ($r = -.406, p < .01$). There was a moderate negative correlation with EDPER ($r = -.347, p < .01$) and a minor negative correlation with TOTALREVENUE ($r = -.088, p < .01$) and SPECIALEDPER ($r = -.060, p < .01$).

The results from this dataset are moderate in comparison to the results in the California 2005 4th and 8th grade datasets; however, the results are somewhat consistent with the positive and negative associations that were found in the other California datasets. Also, the correlation results with the state AYP and the subgroups and the correlation results with the NAEPPROFbasic and the subgroups are not as close in value as the findings with the 2005 California datasets. However, the correlation results with the NAEPPROFprof variable and the subgroups are closer in value to the correlation results with the state AYP and the subgroups which are not consistent with the findings in the 2005 California dataset.

For example the correlation between state AYP and the WHITEPER variable produced an r value of .300 while the NAEPPROFbasic level variable and WHITEPER had an r value .696 and the correlation between NAEPPROFprof level variable and WHITEPER value had an r value of .430. Also, the correlation between state AYP and the EDPER variable produced an r value of -.347 while the NAEPPROFbasic variable and EDPER had an r value -.779 and the correlation between NAEPPROFprof variable and EDPER value had an r value of -.494. This suggests that the subgroup variables had a closer association with the NAEP results vs. the state AYP results.

The EDPER and WHITEPER variables are the benchmark variables used throughout the analysis of the Pearson results in this section to report the subgroup associations between state

AYP, NAEPPROFprof, and NAEPPROFbasic. While the other correlation results are analyzed when making conclusions and they are reported in the correlation matrix, these two variables are used to report the consistency among the aforementioned variables. These two variables where chosen because the EDPER variable had a consistent negative association throughout almost all datasets and the WHITEPER variable has a consistent positive association. The output tables for the descriptive statistics and the correlation matrix can be found in Appendix W.

California 2007 8th Grade

For the 2007 dataset at the 8th grade level in California, there was a moderate positive correlation between the two variables for the proficient level NAEP scale and state AYP results ($r = 0.216$, respectively, significant at the .01 level). There was a stronger positive correlation for the basic level NAEP scale and state AYP results ($r = 0.480$, $p < .01$). All of the Pearson correlation results in California support the predicted hypothesis that there would not be a strong relationship between the proficient level NAEP scale results and the state accountability assessment results.

The correlation results with state AYP and the subgroups included a large positive correlation with WHITEPER ($r = .565$, $p < .01$) and a moderate positive correlation with NATAMPER ($r = .125$, $p < .01$) and ASIANPER ($r = .289$, $p < .01$). There was an extremely large negative correlation with EDPER ($r = -.583$, $p < .01$) and HISPANICPER ($r = -.542$, $p < .01$). There was a moderate negative correlation with TOTALREVENUE ($r = -.365$, $p < .01$), ELLPER ($r = -.370$, $p < .01$), BLACKPER ($r = -.181$, $p < .01$) and SPECIALEDPER ($r = -.193$, $p < .01$). It is not surprising that the WHITEPER variable had a strong positive association and the ASIANPER had a moderate positive association. Also, the EDPER and HISPANICPER had strong negative correlations which have been consistent in the other California datasets.

The correlation results with the BLACKPER, HISPANICPER, SPECIALEDPER, and ELLPER variables were fairly inconsistent in the other California datasets and in subsequent datasets throughout this study. It has been surprising that the TOTALREVENUE variable had a negative association in all of the California datasets. However, as previously stated, it is important to remember that the Pearson Correlation test provides information on *whether there is a relationship between existing variables by offering a correlation coefficient. However, it does not indicate the direction of the correlation.* The results from the logistic regression analysis under research question two and three provide richer and more robust results and analysis of the TOTALREVENUE variable.

As Table W6 in Appendix W indicates, and consistent with the 2005 California datasets, the correlation with state AYP and the subgroup variables and the correlation between the NAEPPROFbasic level variable and subgroup variables are closer in association vs. the correlation with the NAEPPROFprof level variable and the subgroup variables with the California 2007 8th grade dataset. For example the correlation between state AYP and the WHITEPER variable produced an r value of .565 while the NAEPPROFbasic level variable and WHITEPER had an r value .616. In contrast, the correlation between NAEPPROFprof level variable and WHITEPER value had an r value of .167. As Table W6 indicates, the closer association with the state AYP results and NAEPPROFbasic vs. the state AYP results and the NAEPPROFprof are consistent among the analysis of the subgroup variables.

Other than the 2005 4th grade California dataset, the results support the predicted hypothesis and referenced literature that there is a closer relationship with state AYP and the basic level NAEP scale vs. State AYP and the proficient level NAEP scale. Table W5 provides descriptive statistics and Table W6 provides the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable. The tables can be found in Appendix W.

Michigan 2005 4th Grade

It was not possible to compute the correlation between the state AYP data and NAEP proficient level data with the analysis of the 2005 4th grade Michigan dataset because none of the schools in the sample met NAEP scale proficiency at the proficient level. When the correlation coefficient was computed between the state AYP results and the basic level NAEP scale, there was a moderate positive statistically significant relationship ($r = 0.112$, $p < .01$). The 4th grade results indicate that there is no correlation between the state AYP results and the proficient level NAEP scale results. In fact, only one school failed to meet AYP standards at the 4th grade level while no schools met proficiency on the NAEP scale according to the proficient level measurement. The relationship with the basic level NAEP scale was not as strong as the relationships that were found in California.

The correlation results with state AYP and the subgroups included a moderate positive correlation with WHITEPER ($r = .165$, $p < .01$) and a minor positive correlation with NATAMPER ($r = .015$, $p < .01$), HISPANICPER ($r = .048$, $p < .01$), ASIANPER ($r = .049$, $p < .01$), ELLPER ($r = .011$, $p < .01$) and SPECIALEDPER ($r = .044$, $p < .01$). There was a moderate negative correlation with EDPER ($r = -.157$, $p < .01$) and BLACKPER ($r = -.182$, $p < .01$). There was a very moderate negative correlation with TOTALREVENUE ($r = -.031$, $p < .01$).

The Pearson results with the subgroups were statistically significant but very moderate in comparison with the California datasets. However, the results are somewhat consistent with the findings from the California datasets with regards to the positive and negative correlations with the specific subgroup variables. Also, since none of the schools in the dataset met NAEP scale proficiency at the proficient level, a comparison with the NAEP proficient level variable was not possible. Refer to the tables in Appendix X for the descriptive statistics and output matrix for the Pearson results.

Michigan 2005 8th Grade

An analysis of the 8th grade 2005 Michigan dataset produced a very moderate positive correlation between the two variables for the proficient level NAEP scale and state AYP results (r = 0.071, significant at the .01 level). There was a moderate positive correlation for the basic level NAEP scale and state AYP results ($r = 0.294$, $p < .01$). These results establish a weaker relationship vs. the results that were found in the state of California. However, although there was a weaker relationship overall vs. California, the pattern found in California was somewhat consistent. There was a weaker relationship between state AYP and the proficient level NAEP scale and a stronger relationship with the basic level NAEP scale in both Michigan and California.

The correlation results with state AYP and the subgroups included a moderate positive correlation with WHITEPER ($r = .278$, $p < .01$) and ASIANPER ($r = .169$, $p < .01$). A minor positive correlation was found with NATAMPER ($r = .052$, $p < .01$), HISPANICPER ($r = .002$, $p < .01$), ELLPER ($r = .047$, $p < .01$), and SPECIALEDPER ($r = .015$, $p < .01$). There was a moderate negative correlation with EDPER ($r = -.223, p < .01$), BLACKPER ($r = -.299, p < .01$), and TOTALREVENUE ($r = -.114, p < .01$).

The results are somewhat consistent with the Michigan 2005 4th grade dataset. The positive correlation with WHITEPER and ASIANPER along with the negative correlation with EDPER is consistent with the California datasets although the results were more moderate in comparison. The negative correlation with the BLACKPER is consistent with the 2005 4th grade Michigan dataset but not with the California datasets. This suggests that BLACKPER might have a greater negative relationship in Michigan vs. California. Also, the SPECIALEDPER and ELLPER did not produce much of a relationship while the California results with those variables showed the SPECIALEDPER was inconsistent but that ELLPER had a larger negative correlation.

It is interesting to point out that some of the correlation results with the basic level NAEP scale and the subgroups were much larger vs. the correlation results with the state AYP results. For example, the correlation with the basic level NAEP scale and EDPER produced an r value = -.740 ($p < .01$) while the correlation with the basic level NAEP scale and WHITEPER established an r value = .776 ($p < .01$). This suggests that the relationship between these variables and the basic level NAEP scale results was much stronger than the relationship with state AYP and the two above mentioned subgroups. It is also interesting to point out that while the relationship with the TOTALREVENUE variable had a negative correlation when compared to state AYP ($r = -.114, p < .01$) and the basic level NAEP scale ($r = -.283, p < .01$), there was a positive correlation with the proficient level NAEP scale ($r = .163, p < .01$). It was predicted that the TOTALREVENUE variable would have a positive influence on school results. However, the results with all the datasets in this study are consistent because there was either a minor negative or positive correlation. The tables in Appendix X provide descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

Michigan 2007 4th Grade

For the 2007 dataset at the 4th grade level, there was a very moderate positive correlation between the two variables for the proficient level NAEP scale and state AYP results (r = 0.083, significant at the .01 level). There was a large positive correlation for the basic level NAEP scale and state AYP results (r = 0.400, p < .01). These results are not surprising and they are consistent with the other datasets in this study.

The correlation results with state AYP and the subgroups included a moderate positive correlation with WHITEPER (r = .301, p < .01) and a very moderate positive correlation with ASIANPER (r = .094, p < .01), HISPANICPER (r = .098, p < .01), and ELLPER (r = -.091, p < .01). A minor positive correlation was found with SPECIALEDPER (r = .014, p < .01). There was a moderate negative correlation with EDPER (r = -.321, p < .01), BLACKPER (r = -.304, p < .01), and TOTALREVENUE (r = -.180, p < .01). The Michigan results are fairly consistent with all of the datasets. The BLACKPER variable along with the EDPER variable provide for a moderate negative correlation while the WHITEPER variable shows a moderate positive correlation. The SPECIALEDPER, ELLPER, HISPANICPER and NATAMPER variables are not consistent and are moderate among the Michigan datasets. The inconsistency in the relationship with these variables is similar to what was found in California.

Similar to the Michigan 2005 8th grade dataset, the correlation results with the basic level NAEP scale and some of the subgroups were much larger vs. the correlation results with the state

AYP results. For example, the correlation with the basic level NAEP scale and EDPER produced an r value = -.684 ($p < .01$) while the correlation with the basic level NAEP scale and WHITPER had an r value = .729 ($p < .01$). This further supports the finding that the relationship between these variables and the basic level NAEP scale results was stronger than the relationship with the state AYP and the two above mentioned subgroups.

Similar to the Michigan 2005 8th grade dataset, the relationship with the TOTALREVENUE variable had a negative correlation when compared to state AYP ($r = -.180$, $p < .01$) and the basic level NAEP scale ($r = -.371$, $p < .01$), however, there was a positive correlation with the proficient level NAEP scale ($r = .164$, $p < .01$). This suggests that revenue might have a greater relationship with the results on the proficient level NAEP scale which is a much more rigorous standard as this study identifies. The tables in Appendix X report the descriptive statistics and the correlation matrix for this dataset.

Michigan 2007 8th Grade

For the 2007 dataset at the 8th grade level, there was a very minor positive correlation between the two variables for the proficient level NAEP scale and state AYP results ($r = 0.055$, significant at the .01 level). There was a stronger positive correlation for the basic level NAEP scale and state AYP results ($r = 0.429$, $p < .01$). All of the Michigan datasets support the predicted hypothesis that there would not be a strong relationship between the proficient level NAEP scale and the state accountability assessment. An interesting finding is that the basic level NAEP scale had a stronger relationship than the proficient level scale in Michigan and California.

A Pearson Product-Moment Correlation was also conducted with state AYP as the dependent variable and in comparison with the subgroup categories and school resources. There was a sizable positive correlation with WHITEPER ($r = .522$, $p < .01$). There was a moderate positive correlation with the ASIANPER ($r = 0.139$, $p < .01$), and a minor positive correlation with NATAMPER ($r = .064$, $p < .01$) and HISPANICPER ($r = .027$, $p < .01$). There was a sizable statistically significant negative correlation with EDPER ($r= -.396$, $p < .01$) and BLACKPER ($r = -.627$, $p < .01$). There was also a statistically significant moderate negative correlation with TOTALREVENUE ($r = -.155$, $p < .01$). There was a very minor negative correlation with ELLPER ($r = -.023$, $p < .01$) and SPECIALEDPER ($r = -.057$, $p < .01$). The BLACKPER variable had a stronger relationship among the Michigan datasets with state AYP vs. the results in California. One of the reasons for this could be due to the larger BLACKPER subgroup population in the state of Michigan. The output tables can be found in Appendix X.

North Carolina 2005 4th Grade

An analysis of the 2005 North Carolina 4th grade dataset, established that there was a moderate positive correlation between the proficient level NAEP scale and state AYP results ($r = 0.163$, respectively, significant at the .01 level). There was also a moderate positive correlation with the basic level NAEP scale and state AYP results ($r = 0.248$, $p < .01$). The stronger relationship between the basic level NAEP scale and the state AYP results has been consistent among the previously reported datasets in Michigan and California.

The correlation results with state AYP and the subgroups included a moderate positive correlation with WHITEPER ($r = .316$, $p < .01$), and a minor positive correlation with ASIANPER ($r = .058$, $p < .01$) and TOTALREVENUE ($r = .025$, $p < .01$). There was a moderate negative correlation with EDPER ($r = -.342$, $p < .01$), NATAMPER ($r = -195$, $p < .01$), BLACKPER ($r = -.246$, $p < .01$), HISPANICPER ($r = -.127$, $p < .01$), and ELLPER ($r = -.147$, $p < .01$). There was a minor negative correlation with SPECIALEDPER ($r = -.095$, $p < .01$). The WHITEPER and the ASIANPER variables were positively associated with state AYP in a consistent manner with all of the datasets in this study.

The negative correlation with the EDPER variable and the state AYP variable is a consistent finding with the other datasets in this study. The TOTALREVENUE variable had a minor positive association with state AYP which is further consistent with some of the variability that has been found among the different datasets in this study, although Michigan and California produced very moderate or minor negative correlations. The results with the other subgroup variables are not necessarily inconsistent with the results in the other datasets in this study. The values have been inconsistent with the correlations between state AYP and the following variables; BLACKPER, HISPANICPER, NATAMPER, ELLPER, and SPECIALEDPER. However, there were two exceptions as Michigan produced a consistent negative correlation with BLACKPER and California produced a consistent negative correlation with ELLPER. Refer to Appendix Y for descriptive statistics and correlation matrix for the North Carolina datasets.

North Carolina 2005 8th Grade

An analysis of the 2005 dataset at the 8th grade level, produced a positive correlation between the proficient level NAEP scale and state AYP results ($r = 0.327$, respectively, significant at the .01 level). There was also a moderate positive correlation with the basic level NAEP scale and state AYP results ($r = 0.245$, $p < .01$). This is one of the few datasets where there was a stronger relationship between the state AYP variable and proficient level NAEP scale vs. the state AYP variable and the basic level NAEP scale variable. This suggests that the NC state AYP accountability standards at the 8th grade level had a stronger relationship with the proficient level NAEP scale vs. the other datasets in this study. It also implies that North Carolina has a more rigorous assessment standard in comparison to the other states in this study.

The correlation results with state AYP and the subgroups included a moderate positive correlation with WHITEPER ($r = .320$, $p < .01$) and TOTALREVENUE ($r = .197$, $p < .01$). There was an minor positive correlation with NATAMPER ($r = .098$, $p < .01$) and ASIANPER ($r = .096$, $p < .01$). There was a moderate negative correlation with EDPER ($r = -.388$, $p < .01$), BLACKPER ($r = -.337$, $p < .01$), HISPANICPER ($r = -.181$, $p < .01$), SPECIALEDPER ($r = -.162$, $p < .01$), and ELLPER ($r = -.132$, $p < .01$).

The results are somewhat consistent with the 2005 4th grade North Carolina dataset with the exception of the TOTALREVENUE variable with a stronger positive correlation and the NATAMPER variable with a minor positive correlation. The WHITEPER variable had a positive association and the EDPER variable had a negative association which is consistent throughout all of the datasets.

It is interesting to point out that with the two 2005 North Carolina datasets, the correlation with state AYP and the subgroup variables were similar in value. This pattern was consistent with the correlation between the NAEPPROFbasic level variable and subgroup variables along with the NAEPPROFprof level variable and the subgroup variables. However, this finding is inconsistent with the California datasets and most of the Michigan datasets. For example, the correlation between state AYP and the WHITEPER variable produced an r value of .320 while the NAEPPROFbasic level variable and WHITEPER had an r value .471 and the correlation between NAEPPROFprof level variable and WHITEPER value had an r value of

.112. This pattern was similar with correlation between the state AYP and the EDPER variable which produced an r value of -.388 while the NAEPPROFbasic level variable and EDPER had an r value -.445 and the correlation between NAEPPROFprof level variable and EDPER value had an r value of -.316. This suggests a variation with the North Carolina dataset as the subgroups appear to have a similar association with the state AYP, NAEPPROFbasic level, and NAEPPROFprof level. The tables in Appendix Y provide descriptive statistics and the correlation matrix.

North Carolina 2007 4th Grade

An analysis of the 2007 NC dataset at the 4th grade level, produced a positive correlation between the proficient level NAEP scale variable and state AYP results ($r = 0.418$, respectively, significant at the .01 level). There was also a more moderate positive correlation for the basic level NAEP scale and state AYP results ($r = 0.336$, $p < .01$). This is the second dataset from the state of North Carolina where there was a stronger relationship between the state AYP standards and proficient level NAEP scale vs. the state AYP standards and the basic level NAEP scale.

The correlation results with state AYP and the subgroups included a moderate positive correlation with WHITEPER ($r = .475$, $p < .01$) and a minor positive correlation ASIANPER ($r = .021$, $p < .01$). There was a sizable negative correlation with EDPER ($r = -.445$, $p < .01$) and BLACKPER ($r = -.426$, $p < .01$). There was a moderate negative correlation with NATAMPER ($r = -.124$, $p < .01$), HISPANICPER ($r = -.215$, $p < .01$), ELLPER ($r = -.255$, $p < .01$), and SPECIALEDPER ($r = -.204$, $p < .01$). There was a minor negative correlation with TOTALREVENUE ($r = -.048$, $p < .01$).

The results are consistent with the 2005 North Carolina datasets in regard to the WHITEPER variable although the correlation is a little larger. The minor positive correlation with ASIANPER is also consistent. The results with the other variables are also consistent with the 2005 North Carolina datasets as the associations are similar in value. The EDPER variable had a negative association which is consistent throughout all of the datasets. Also, consistent with the 2005 North Carolina datasets, the correlation between the subgroups variables and the following three variables state AYP, the NAEPPROFbasic level variable, and the NAEPPROFprof level variable were similar in value. This finding is consistent with the North Carolina datasets but in opposition to the findings from the California datasets and most of the Michigan datasets. For example, the correlation between state AYP and the WHITEPER variable produced an r value of .475 while the NAEPPROFbasic level variable and WHITEPER had an r value .475, and the correlation between NAEPPROFprof level variable and WHITEPER value had an r value of .432. This was similar with correlation between the state AYP and the EDPER variable which produced an r value of -.445, while the NAEPPROFbasic level variable and WHITEPER had an r value -.548, and the correlation between NAEPPROFprof level variable and WHITEPER value had an r value of -.581. This further suggests a variation with the North Carolina dataset in comparison to Michigan and California and the possibility of a closer relationship with the NC state accountability assessment and the NAEP assessments in relationship to the subgroups. The relationships in California and most of the Michigan datasets between the subgroups with both the NAEPPROFprof level and NAEPPROFbasic level were stronger vs. the relationship between the subgroups and the state AYP results. Refer to Appendix Y for the descriptive statistics table and the correlation matrix.

North Carolina 2007 8th Grade

With the analysis of the 2007 North Carolina dataset at the 8th grade level, there was a positive correlation between the proficient level NAEP scale and state AYP results (r = 0.379, respectively, significant at the .01 level). There was also a very weak positive correlation for the basic level NAEP scale and state AYP results (r = 0.039, p < .01). The stronger correlation with the proficient level NAEP scale and state AYP has been consistent with three of the NC datasets and it provides additional support for the suggestion that NC has a more rigorous accountability standard that is more closely aligned with the NAEP proficiency standards. In fact, the Pearson results in the state of North Carolina provided data that established a stronger relationship with the NAEP proficient level and state AYP accountability vs. the results in all of the other state datasets including Texas which is reported later in this chapter.

The correlation results with state AYP and the subgroups included a moderate positive correlation with WHITEPER (r = .333, p < .01), TOTALREVENUE (r= .119, p <.01), and ASIANPER (r = .150, p < .01). There was a moderate negative correlation with EDPER (r = -.358, p < .01), BLACKPER (r = -.276, p < .01), HISPANICPER (r = -.220, p < .01), ELLPER (r = -.206, p < .01), and SPECIALEDPER (r = -.105, p < .01). There was a minor negative correlation with NATAMPER (r = -.092, p < .01). The results are consistent with the other North Carolina datasets as the correlation results are similar in value. The WHITEPER variable had a positive association and the EDPER variable had a negative association which is consistent throughout all of the datasets in this study.

Consistent with the 2005 North Carolina datasets, the correlation between the subgroup variables and the following three variables; state AYP, the NAEPPROFbasic level variable, and the NAEPPROFprof level variable, were similar in value. The correlation between state AYP and the EDPER variable produced an r value of -.358 while the NAEPPROFbasic level variable and EDPER had an r value -.193. The correlation between NAEPPROFprof level variable and EDPER value had an r value of -.446. This gives further support to the earlier finding that the subgroups appear to have a similar association with state AYP, NAEPPROFbasic level, and NAEPPROFprof level in the NC datasets. However, there was one exception, the correlation between state AYP and the WHITEPER variable produced an r value of .333, while the NAEPPROFbasic level variable and WHITEPER had an r value .192, and the correlation between NAEPPROFprof level variable and WHITEPER produced a negative r value of -.126. The correlation was close in range but it was interesting that the WHITEPER variable produced a negative correlation with NAEPPROFprof which was not aligned with the findings in the other datasets. Most of the other datasets in the study had established a positive r value with the analysis of the WHITEPER variable in comparison to the NAEPPROFprof variable.

It is also interesting to point out that the TOTALREVENUE variable had a positive correlation which was larger with the NAEPPROFprof level variable with and r value of .302 ($p < .01$). Thus revenue might have a greater relationship with the proficient level NAEP scale in the state of North Carolina vs. Michigan and California. This finding is interesting because some of the datasets in California and Michigan produced a positive association with the proficient level NAEP scale even though there was a minor negative association with state AYP and the basic level NAEP scale. The tables in Appendix Y provide descriptive statistics and the correlation matrix.

Texas 2005 4th Grade

An analysis of the 2005 Texas dataset at the 4th grade level, produced a very weak negative correlation between the proficient level NAEP scale and state AYP results, (r = -0.043, significant at the .01 level). There was also a very weak negative correlation for the basic level NAEP scale and state AYP results (r = - 0.036, p < .01). This suggests that there is almost no relationship between the Texas state accountability assessment results and the NAEP results in this dataset. As the descriptive statistics in Table Z1 show, over 98% of the schools in this dataset met state AYP, while just over 65% met proficiency on the basic level NAEP scale, and only 10% of the schools met proficiency on the proficient level NAEP scale.

The correlation results with state AYP and the subgroups included a minor positive correlation with TOTALREVENUE (r = .054, p < .01), ASIANPER (r= .043, p <.01), HISPANICPER (r = .087, p < .01), and ELLPER (r = .094, p < .01). There was a minor negative correlation with EDPER (r = -.038, p < .01), BLACKPER (r = -.091, p < .01), HISPANICPER (r = -.087, p < .01), WHITEPER (r = -.038, p < .01), and SPECIALEDPER (r = -.050, p < .01). These results are extremely moderate and they are not consistent with the results from the other state datasets.

It is interesting to note that the correlation with the NAEPPROFbasic level variable and the subgroup variables, along with the correlation with the NAEPPROFprof level and the subgroup variables, produced a more sizable statistically significant results vs. the relationship with the state AYP results. For example, the correlation with EDPER and NAEPPROFbasic was (r = -.457, p < .01) and the correlation with EDPER and NAEPPROFprof was (r = -.509, p < .01) while the correlation with EDPER and state AYP was (r = -.038, p < .01). Similar results were found with the correlation between WHITEPER and the NAEPPROFbasic level variable (r = .544, p < .01), and the correlation with EDPER and NAEPPROFprof (r = .515, p < .01), while the correlation with WHITPER and state AYP was (r = -.038, p < .01). These results suggest that the subgroups had no relationship with the state results and a larger relationship with the NAEP results.

The findings with the relationship between the school level NAEP results and the Texas subgroups are consistent with the findings in the other states. However, the findings with Texas AYP and the subgroups are unusual and this could suggest that the Texas state assessment standards are not well aligned with the NAEP as they have very little in common with the state accountability assessments in the other states (at least in terms of school level performance outcome scores). The output tables can be found in Appendix Z with descriptive statistics and the correlation matrix.

Texas 2005 8th Grade

With the analysis of the 2005 Texas dataset at the 8th grade level, there was a moderate positive correlation between the proficient level NAEP scale and state AYP results (r = 0.137, significant at the .01 level). There was a stronger positive correlation with the basic level NAEP scale and state AYP results (r = 0.393, p < .01). This suggests that there is a stronger relationship between the 2005 8th grade Texas state AYP accountability results and the NAEP assessment vs. the 4th grade 2005 datasets. The 2005 8th grade results are more consistent with the findings from the datasets in the state of California and Michigan.

The Pearson results with state AYP and the subgroups included a sizable positive correlation with WHITEPER (r = .409, p < .01), a moderate positive correlation with

ASIANPER (r= .215, p <.01), and a minor positive correlation with NATAMPER (r = .042, p < .01). There was a sizable negative correlation with EDPER (r = -.451, p < .01) and a moderate negative correlation with BLACKPER (r = -.236, p < .01), HISPANICPER (r = -.267, p < .01), ELLPER (r = -.343, p < .01), and SPECIALEDPER (r = -.285, p < .01). There was a minor negative correlation with TOTALREVENUE (r = -.020, p < .01). These results are not consistent with the 2005 4th grade Texas dataset, however, they are aligned with the results from the other state datasets. For example, the WHITEPER variable had a positive association and the EDPER variable had a negative association which is consistent throughout the datasets in this study.

The correlation with state AYP and the subgroup variables along with the correlation between the NAEPPROFbasic level variable and subgroup variables, are more consistent in value vs. the correlation with the NAEPPROFprof level variable and the subgroup variables. This is similar to the findings from the California and Michigan datasets in this study. The correlation between the state AYP and the WHITEPER variable produced an r value of .409, while the NAEPPROFbasic level variable and WHITEPER had an r value .519, and the correlation between NAEPPROFprof level variable and WHITEPER value had an r value of .280. The correlation the between state AYP and the EDPER variable produced an r value of -.451, while the NAEPPROFbasic level variable and EDPER had an r value -.484, and the correlation between NAEPPROFprof level variable and EDPER value had an r value of -.326. These results suggest that the 2005 8th grade Texas dataset is more aligned with the Michigan and California datasets vs. the North Carolina dataset or the 2005 4th Grade Texas datasets. Over 72% of the schools met state AYP standards in Texas, while over 77% of the schools met proficiency on the basic level NAEP scale, and less than 5% met proficiency on the proficient level NAEP scale. This supports the finding that this dataset is more closely aligned with the basic level NAEP scale in alignment with California and most of the Michigan datasets. Refer to Appendix Z for descriptive statistics and the correlation matrix.

Texas 2007 4th Grade

An analysis of the 2007 dataset at the 4th grade level produced results indicating a very weak positive correlation between the proficient level NAPE scale and state AYP results, (r = .032, significant at the .01 level). There was a moderate negative correlation with the basic level NAEP scale and state AYP results ($r = -0.115$, $p < .01$). This supports the suggestion that there is a stronger relationship between the 2005 8th grade Texas state accountability assessment results and the NAEP vs. the 4th grade assessment results which is similar to the findings with the 2005 Texas datasets. It also suggests that the Texas AYP system at the 4th grade level, has an extremely mild relationship (the lowest of all four states in the sample) with the proficient level NAEP scale and basic level NAEP scale. In fact, similar in value to the 2005 4th grade Texas results, over 99% of the schools met state AYP, just over 61% of the schools met basic level NAEP scale proficiency and just over 11% of the schools met proficiency standards at the proficient level NAEP scale.

The correlation results with state AYP and the subgroups included a minor positive correlation with NATAMPER ($r = .062$, $p < .01$), ASIANPER ($r= .057$, $p <.01$), and WHITEPER ($r = .076$, $p < .01$). There was a moderate negative correlation with EDPER ($r = -.130$, $p < .01$) and SPECIALEDPER ($r = -.304$, $p < .01$). There was a minor negative correlation with TOTALREVENUE ($r = -.093$, $p < .01$), BLACKPER ($r = -.084$, $p < .01$), HISPANICPER ($r = -.026$, $p < .01$), and ELLPER ($r = -.052$, $p < .01$).

These results are somewhat consistent with the 2005 Texas 4th grade results. Both 4th grade Texas datasets produced some of the most unusual results in comparison with the other datasets. It is interesting to note that the correlation with the NAEPPROFbasic level and the subgroup variables, along with the correlation with the NAEPPROFprof level and the subgroup variables, produced a more sizable statistically significant result vs. the correlation with state AYP and the subgroups. For example, the correlation with EDPER and NAEPPROFbasic was ($r = -.422$, $p < .01$), and the correlation with EDPER and NAEPPROFprof was ($r = -.512$, $p < .01$) while the correlation with EDPER and state AYP was ($r = -.130$, $p < .01$). Similar results were found with the correlation between WHITEPER and NAEPPROFbasic level ($r = .546$, $p < .01$) and the correlation with EDPER and NAEPPROFprof level at ($r = .492$, $p < .01$), while the correlation with WHITPER and state AYP was ($r = .076$, $p < .01$).

These results suggest that the subgroups had no relationship with the state AYP results in contrast to a larger statistically significant relationship with the NAEP results. The findings with the relationship between the school level NAEP results and the Texas subgroups are consistent with the findings in the other states. However, the relationship with state AYP and the subgroup variables in both of the 4th grade Texas datasets (2005 and 2007) were very weak. This further supports the finding that the Texas state assessment standards at the 4th grade level had a very low correlation with the NAEP and the other state assessments that are used in this sample. The output tables including descriptive statistics and the correlation matrix can be found in Appendix Z.

Texas 2007 8th Grade

For the 2007 dataset at the 8th grade level, there was a moderate positive correlation between the two variables for the proficient level NAEP scale and state AYP results, ($r = .137$, significant at the .01 level). There was also a moderate positive correlation for the basic level NAEP scale and state AYP results ($r = 0.224$, $p < .01$). These results support the finding that there is a stronger relationship between the 8th grade Texas state accountability assessment results and the NAEP results, vs. the 4th grade Texas results (in the 2005 and 2007 datasets) and the NAEP results.

The Texas findings also support the conclusion that the state of Texas has a less rigorous AYP accountability standard (especially at the 4th grade level) vs. the other states in the sample in terms of school level proficiency. As the descriptive statistics indicate (refer to Appendix Z) both of the 4th grade datasets in Texas had over 98% of the schools meeting the state AYP requirements while less than 12% met NAEP scale proficiency at the proficient level. A lower percentage of schools met state AYP proficiency in the 8th grade which produced results that were more closely aligned to the proficiency on the basic level NAEP scale. Thus the 8th grade Texas dataset produced results similar to California and most of the Michigan datasets. The Texas 4th grade datasets were not similar to any other dataset in this study establishing a very weak relationship with the NAEP.

The correlation results with state AYP and the subgroups included a moderate positive correlation with WHITEPER ($r = .256$, $p < .01$), and ASIANPER ($r= .167$, $p <.01$), and a minor positive correlation with TOTALREVENUE ($r = .019$, $p < .01$). There was a moderate negative correlation with EDPER ($r = -.325$, $p < .01$), BLACKPER ($r = -.117$, $p < .01$), HISPANICPER ($r = -.192$, $p < .01$), and SPECIALEDPER ($r = -.119$, $p < .01$). There was a minor negative correlation with NATAMPER ($r = -.049$, $p < .01$) and ELLPER ($r = -.050$, $p < .01$). These results are somewhat consistent with the 2005 8th grade Texas dataset and the results from the Michigan and California datasets. However, they are not consistent with the findings from the 4th grade Texas dataset. Refer to Appendix Z for the descriptive statistics and the correlation matrix.

Summary for Research Question One

As discussed throughout this section, the relationship between the basic level NAEP scale proficiency and state AYP results was more consistent than the relationship between the proficient level NAEP scale results in the state of California, Michigan, the 4th grade 2005 North Carolina dataset, and the 8th grade datasets in Texas. The other North Carolina datasets produced a closer association with the proficient level NAEP scale results and state AYP vs. the basic level results. The 4th grade Texas datasets were unusual and had an extremely mild relationship with state AYP and the NAEP results at both the basic level and proficient level.

With regard to the subgroups, the EDPER variable consistently produced a statistically significant negative association with all of the datasets. The range was from moderate to sizable but it was generally one of the larger associations among all of the datasets and state AYP. There was a similar finding with the WHITEPER variable as it was positively associated to state AYP at either the moderate to sizable level in all datasets. The ASIANPER variable produced a minor to moderate positive relationship with all datasets. The results with the other variables were inconsistent among the datasets. However, the HISPANICPER variable and ELLPER variable in California were negatively correlated at a moderate to sizable level. The

BLACKPER variable was negatively associated to state AYP in a moderate to sizable level among all of the Michigan datasets. The TOTALREVENUE variable had a tendency to be negatively correlated at either a minor to moderate level with the exception of North Carolina which produced minor to moderate positive associations. It was interesting that TOTALREVENUE appeared to have a greater positive impact on the NAEP scale proficient results in North Carolina. The rest of the independent variables in this study were not consistent with their results, although the NATAMPER and SPECIALEDPER variables were minor to moderate in a negative correlation with most of the datasets.

Results for Research Question Two

2. Is there a significant correlation between the schools that meet proficiency standards on the NAEP and schools that meet proficiency standards on the state accountability assessments (STAR, MEAP, End of Grade Test, and TAKS) in the sample states? Is there a combination of factors that best predicts proficiency status on the NAEP and state accountability assessments?

The hypothesis for research question two (as stated in chapter three) is that there would not be a strong relationship between school level achievement on the NAEP and the state accountability assessment results in mathematics and reading. Similar to research question one, it was predicted that there would be a relationship between state accountability assessment results and the demographic characteristic (categorical variables) along with school funding.

In order to answer research question two and test the predicted hypothesis, the statistical tests that were used include Pearson's Product-Moment Correlation and Logistic Regression.

The Logistic Regression equation for research question two based on the independent and dependent relationships is the following:

Logit (Y, of meeting proficiency standards) = α *(constant)* + *(β_1 * DUMMYREV)* + *(β_2 * EDPER)* + *(β_3 * NATAMPER)* + *(β_4 * ASIANPER)* + *(β_5 * BLACKPER)* + *(β_6 * HISPANICPER)* + *(β_7 * WHITEPER)* + *(β_8 * ELLPER)* + *(β_9 * SPECIALEDPER)*+ *(β_{10} *NAEPPROFprof)*[17] +*(β_{11} * NAEPPROFbasic)*[18]

The above equation is used for both of the dependent variables which include school level state accountability assessment results in reading (or ELA) and mathematics. Thus for each year and state assessed, the equation is used two times with the different dependent variables that were mentioned above. The Omnibus Test of Model Coefficients was analyzed with each logistic regression in order to test for statistical significance and goodness of fit. A Cox & Snell R along with a Nagelkerke R Square was calculated in order to determine the proportion of variance in the dependent variable that is explained by variance in the independent variables. A sample of the logistic regression output tables are included in the first section for the 2005 California 4th grade mathematics results. The rest of the tables for the remainder of the study were placed in the Appendices. These tables include:

1. The observed and predicted frequencies with the constant model.
2. The observed and predicted frequencies with the predicted model.
3. Omnibus Test of Model Coefficients assessing goodness of fit.

---

[17] The NAEP Proficient level was used twice (for mathematics and reading).
[18] The NAEP basic level was used twice (for mathematics and reading).

4. Logistic Regression model summary assessing goodness of fit (Cox & Snell R Square along with the Nagelkerke R Square). [19]
5. The Logistic Regression Results Output Table.
6. The Logistic Regression Observed and Predicted Probabilities Figure.

California 2005 4th grade mathematics

A Pearson Correlation was used to analyze the 2005 dataset at the 4th grade level, and there was a moderate positive correlation between the two variables for the proficient level NAEP mathematics scale and the mathematics state accountability assessment, ($r = 0.114$, $p < .01$). There was a stronger positive correlation for the basic level NAEP scale in mathematics and the mathematics state accountability assessment ($r = 0.428$, p < ,01). Refer to Appendix AA for the Pearson Correlation matrix and descriptive statistics.

A logistic regression was also used to analyze the California 2005 4th grade dataset with the mathematics state accountability assessment as the dependent variable. Although the results of the logistic regression were statistically significant, the predicted probability of the constant(null) model was 92% accurate vs. the logistic model at 92.2 % accuracy. The reason for this had to do with the high number of schools that met proficiency status in mathematics vs. a low number that did not. Over 90% of schools met proficiency status on the mathematics state accountability assessment, while less than 6% met the proficient level NAEP scale proficiency,

[19] There is no widely accepted direct analog to OLS Regression's R square. However, the Cox & Snell Square along with the Nagelkerke Square attempts to explain some of the variance. The Nagelkerke Squared is normally higher than the Cox & Snell Square. A problem with this figure is that it attempts to make a statement about percentage of variance but the variance in a dichotomous dependent variable depends on the frequency distribution of the variable. (Garison, 2010).

and less than 37 % met the basic level NAEP scale proficiency. Please refer to Table 11 for the constant model and Table 12 for the predicted model frequencies.

As Table 13 indicates, the Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model. Table 14 also indicates goodness of fit as the Cox & Snell R value of .169 and the Nagelkerke R Square value of .395 accounts for some of the variance within the model.

Table 12

Logistic regression observed and predicted frequencies with the constant model[20] in the California 2005 4th grade dataset with the mathematics state accountability assessment as the dependent variable.

Observed			Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	21113	.0
		met proficiency status	0	242473	100.0
	Overall Percentage				92.0

a. The cut value is .500

[20] The constant model is calculated based on which response had the higher percentage of values with the dependent variable. Thus in this case 92% of the schools or 243473 weighted schools met proficiency and thus the constant model is 92% accurate. If there was a higher percentage that did not meet proficiency then the constant model would accept that value in contrast. Independent variables are not used to calculate the constant model.

Table 13

Logistic regression observed and predicted frequencies with the predicted model[21] with the California 2005 4th grade mathematics state accountability assessment results

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	4957	16156	23.5
		met proficiency status	2620	239853	98.9
		Overall Percentage			92.9

a. The cut value is .500

Table 14

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 4th Grade California mathematics assessment as the dependent variable

		dimension		
		Chi-square	df	Sig.
Step 1	Step	562.697	1	.000
	Block	562.697	1	.000
	Model	48733.688	11	.000

[21] The predicted model is calculated base don the predicted results of the dependent variable when the independent variables are factored into the model.

Table 15

Logistic regression model summary assessing goodness of fit with the California 4th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	98353.273[a]	.169	.395

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

The logistic equation for the California 2005 4th grade dataset is: Logit (Y, of meeting proficiency on the mathematics state accountability assessment) = 26.255(constant) + (.276 * DUMMYREV) + (1.570 * EDPER) + (-40.919 * NATAMPER) + (-21.681 * ASIANPER) + (-32.669 * BLACKPER) + (-23.717 * HISPANICPER) + (-24.814 * WHITEPER) + (-4.188 * ELLPER) + (.170 * SPECIALEDPER) + (2.154 * NAEPMATHPROFbasic) + (16.531 * NAEPMATHPROFprof).

As Table 15 indicates, the Wald value was significant in the model for all of the independent variables ($p < .05$) other than NAEPMATHPROFprof ($p > .05$) . Since the model is not very useful due to the low level of increase between the predicted model and constant model, the results must be interpreted with caution. For example, the EDPER variable provides a positive influence on the probability that a school meets the proficiency standards. However, this is in contrast to the Pearson Correlation results which show a negative relationship between the EDPER variable and state mathematics accountability assessment ($r= -.193$, $p < .01$), the

EDPER and NAEP basic level scale (r= -.388, p < .01) and the EDPER and NAEP proficient level scale (r= -.574, p < .01).[22]

As previously mentioned, *it is important to emphasize that the logistic regression is a more powerful tool of analysis for predicting the probability that a school will meet proficiency status.* The positive Wald square value in this dataset is also in contrast to almost all of the other results with the EDPER variable throughout the different states in this study. Thus consideration must be given to the fact that the logistic regression results in this dataset might not provide much of an explanation into a school's ability to meet mathematics proficiency. As mentioned, the high percentage of over 90% of schools in this dataset that met state mathematics proficiency might provide a better explanation as to the type of standard implemented in the state at the 4th grade level as they contrast the NAEP mathematics results from the same dataset with just over 6% proficiency at the proficient level NAEP scale. This is an interesting finding as the percentage variance in proficiency between the NAEP assessment and state accountability assessment is not aligned.

---

[22] Refer to Table AA2 in the Appendix for the Pearson matrix that includes subgroup correlations.

Table 16

Logistic regression analysis for the 2005 California 4th grade data with the mathematics assessment as the dependent variable

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
DUMMYREV	.276	.015	345.437	1	.000	1.318	1.280	1.357
EDPER	1.570	.046	1148.478	1	.000	4.808	4.390	5.265
NATAMPER	-40.919	1.175	1213.049	1	.000	.000	.000	.000
ASIANPER	-21.681	1.066	413.796	1	.000	.000	.000	.000
BLACKPER	-32.669	1.053	962.524	1	.000	.000	.000	.000
HISPANICPER	-23.717	1.043	516.840	1	.000	.000	.000	.000
WHITEPER	-24.814	1.060	548.188	1	.000	.000	.000	.000
ELLPER	-4.188	.070	3594.132	1	.000	.015	.013	.017
SPECIALEDPER	.170	.073	5.462	1	.019	1.186	1.028	1.367
NAEPMATHPROFbasic	2.154	.022	9892.924	1	.000	8.620	8.261	8.993
NAEPMATHPROFprof	16.531	214.259	.006	1	.939	1.5117	.000	3.601189
Constant	26.255	1.039	638.764	1	.000	2.52711		

Figure 22 represents the predicated probability of the schools' mathematics proficiency status for the 4th Grade 2005 California dataset. Please note the high percentage that met proficiency on the mathematics state accountability assessment which contributed to the minor increase in the constant model vs. the predicted model.

Figure 22. Logistic regression for the California 2005 4th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

   80000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                   pI
F        I                                                                                                   pI
R  60000 +                                                                                                   p+
E        I                                                                                                   pI
Q        I                                                                                                   pI
U        I                                                                                                   pI
E  40000 +                                                                                                   p+
N        I                                                                                                   pI
C        I                                                                                                  ppI
Y        I                                                                                             p    ppI
   20000 +                                                                                             p   ppp+
         I                                                                                             p  ppppI
         I                                                                                       p   p pp ppppI
         I                                                                          pp    ppppppppppppppppppppI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9        1
  Group:  nnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnnpppppppppppppppppppppppppppppppppppppppppppppppppp

      Predicted Probability is of Membership for proficient
      The Cut Value is .50
      Symbols: n - not proficient
               p - proficient
      Each Symbol Represents 5000 Cases.
```

California 2005 4th grade ELA/reading

For the year 2005 at the 4th grade level, there was a moderate positive correlation between the ELA state accountability assessment and the proficient level NAEP scale reading assessment ($r = 0.110$, $p < .01$). There was also a moderate positive correlation with the basic level NAEP scale reading assessment and the ELA state accountability assessment ($r = 0.364$, p

$< .01$). These findings are consistent with the findings in the other datasets throughout this study. Refer to Appendix AA for the Pearson correlation matrix and the descriptive statistics.

A logistic regression model was also used to analyze the 4th grade 2005 California dataset with the ELA state accountability assessment as the dependent variable. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, the NAEPELAPROFbasic level variable and NAEPELAPROFprof level variables were analyzed separately because the Omnibus Model of Coefficients test was not statistically significant ($p > .05$) indicating that the model did not provide a goodness of fit. When the NAEPELAPROFprof level variable was not used for the logistic regression model, the Omnibus Model of Coefficients test was statistically significant ($p < .05$) indicating goodness of fit with the model. Thus the rest of this analysis provides the logistic regression results without the inclusion of the NAEPELAPROFprof level variable.

The constant (null) model was correct 82.6% of the time vs. the predicted model which was accurate 87.7% of the time. There was a very low increase in accuracy for the predicted model vs. the constant model. A consideration with the analysis of this dataset is that over 80% of the schools met state proficiency standards on the state ELA assessment while just over 5% met the proficient level NAEP scale proficiency. Thus the logistic regression model was analyzed with the consideration that the predicted model provided a low increase in accuracy. This was similar to the 2005 4th grade mathematics results in California. Refer to Table BB1 and BB2 for the observed and predicted frequencies of the constant model vs. the predicted model. These tables and the rest of the tables and figures from this dataset were placed in Appendix BB.

The Cox & Snell R value was .329 and the Nagelkerke R Square was .546 explaining some of the variance within the model. The logistic equation for the California 2005 4th grade dataset is: Logit (Y, of meeting state ELA proficiency) = -12.989 (constant) + (.057 * DUMMYREV) + (-2.889 * EDPER) + (-10.345 * NATAMPER) + (34.899 * ASIANPER) + (13.495 * BLACKPER) + (-18.651 * HISPANICPER) + (24.148 * WHITEPER) + (-4.151 * ELLPER) + (-1.339 * SPECIALEDPER) + (16.905 * NAEPREADPROFbasic) + (-2.693 * NAEPREADPROFprof).

As table BB5 in Appendix BB indicates, the Wald value was significant in the model for most of the independent variables ($p< .05$) other than NAEPREADPROFbasic. The statistically significant variables that were positively associated with the odds that a school met the state ELA proficiency standards were DUMMYREV (B = .057, Exp (B) = 1.059), ASIANPER (B = 34.899, Exp (B) = 1.433), BLACKPER (B = 13.495, Exp (B) = 725940.193), HISPANICPER (B = 18.651, Exp (B) = 1.259), and WHITPER (B = 24.148 - , Exp (B) = 3.071). The statistically significant variables that had negative impact on the state ELA assessment include EDPER (B = -2.9889, Exp (B) = .056), NATAMPER (B = -10.345 , Exp (B) = .000), ELLPER, (B =- 4.151, Exp (B) = .016), and SPECIALEDPER (B = - 1.339, Exp (B) = .262).

It was not surprising that the EDPER variable was negatively associated with the probability that a school met the state ELA proficiency as this finding is consistent with findings in the other datasets and in contrast to the outlier finding with the 2005 California 4th grade mathematics dataset. The DUMMYREV variable produced a mild positive association with the probability that a school met the ELA proficiency status. It was also found that the ASIANPER was positively associated with the chance that a school met the state ELA proficiency which is consistent with the findings in other datasets in this study.

The negative impact of the SPECIALEDPER and ELLPER variable in predicting school success is not surprising as the hypothesis in this study predicts that there would be a negative impact. However, this finding is inconsistent when analyzing the other datasets in this study. Listed in Appendix BB is Figure BB1 that represents the predicated probability of school level AYP proficiency status.

California 2005 8th grade mathematics

For the 2005 8th Grade California dataset with the state mathematics assessment[23] as the dependent variable, there was a moderate positive correlation between the mathematics proficient level NAEP scale and the mathematics state accountability assessment, ($r = 0.174$, significant at the .01 level). There was a strong positive correlation with the mathematics basic level NAEP scale and the mathematics state accountability assessment ($r = 0.522$, $p < .01$). Refer to Appendix AA for the descriptive statistics and correlation matrix.

The mathematics state accountability assessment was used as the dependent variable for a logistic regression to analyze the California schools 2005 8th grade dataset. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. The constant (null) model was correct 55.7% of the time vs. the predicted model which was accurate 79.8% of the time. There was a 24% increase with the predicted model vs. the constant model. This signifies a positive influence by the predicted model. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model. The Cox & Snell R value was 0.404 and the Nagelkerke R Square was 0.541 explaining some of the variance within the model. The output tables for the logistic regression can be found in Appendix BB.

The logistic regression equation for the schools in the 2005 California 8th grade dataset with mathematics as the dependent variable is the following: Logit (Y, of meeting state

[23] The state of California uses an algebra test as their 8th grade state accountability assessment.

accountability proficiency in mathematics) = 21.201(constant) + (-0.613 * DUMMYREV) + (-2.416 * EDPER) + (-26.810 * NATAMPER) + (-15.457 * ASIANPER) + (-24.802 * BLACKPER) + (-20.646 * HISPANICPER) + (-18.284 * WHITEPER) + (1.513 * ELLPER) + (6.461 * SPECIALEDPER) + (0.903 * NAEPMATHPROFbasic) + (0.336 * NAEPMATHPROFprof).

As Table BB10 indicates, the Wald value was significant in the model for all of the independent variables ($p < .05$). Several of the variables appear to have a strong negative influence on a school's ability to meet proficiency which include EDPER (B = - 2.416 , Exp (B) = .089), NATAMPER (B = -26.810 , Exp (B) = .000), ASIANPER (B = -15.457 , Exp (B) =0.00), BLACKPER (B = -24.082 , Exp (B) =0.00), HISPANICPER (B = -24.082 , Exp (B) =0.00), and WHITEPER (B = -18.284, Exp (B) =0.00). The NATAMPER, ASIANPER, BLACKPER, HISPANICPER, and WHITEPER variables must be looked at with caution because they produced extremely low odds ratio results. This indicates that the estimating coefficient with these variables are unstable.

The negative influence with the EDPER variable is not surprising and is consistent with findings from other datasets. The minor negative impact of the DUMMYREV variable, while somewhat surprising, is consistent with the findings in other datasets. The DUMMYREV variable has provided either a mild negative or positive impact with the datasets in this study. The positive impact with the SPECIALEDPER variable (B = 6.416 , Exp (B) = 639.712) and the ELL variable (B = 1.513 , Exp (B) = 4.538) is somewhat surprising but it is consistent with the findings in the other datasets since these variables have produced inconsistent results among the different datasets. The NAEPPROFprof variable (B = .336 , Exp (B) = 1.400) and the

NAEPPROFbasic variable (B = .903 , Exp (B) = 2.446) had a positive influence on the school's ability to meet mathematics proficiency.

California 2005 8th grade ELA/Reading

An analysis of the 2005 8th Grade California dataset, provided a very moderate positive correlation between the proficient level NAEP reading scale and the state ELA accountability assessment, ($r = 0.057$, significant at the .01 level). There was as a moderate positive correlation with the basic level NAEP scale in reading and the ELA state accountability assessment ($r = 0.251$, $p < .01$). Please refer to Appendix AA for the descriptive statistics and the correlation matrix.

The ELA state accountability assessment was used as the dependent variable for a logistic regression to analyze the California schools 2005 8th grade dataset. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, the NAEPREADPROFbasic level variable and the NAEPREADPROFprof level variables were analyzed separately. The reason for this was because the Omnibus Model of Coefficients test was not statistically significant ($p > .05$) indicating that the model did not provide a goodness of fit. However, when the NAEPELAPROFprof level variable was not used for the logistic regression model, the Omnibus Model of Coefficients test was statistically significant ($p < .05$) indicating goodness of fit with the model. Thus the rest of this analysis provides the logistic regression results without the NAEPELAPROFprof level variable included in the output.

The constant (null) model was correct 62.7% of the time vs. the predicted model which was accurate 91.5% of the time. There was a 28.8% increase with the predicted model vs. the constant model. This signifies a strong positive influence with the predicted model. Refer to Table BB11 and BB12 in Appendix BB for the observed and predicted frequencies with the constant model vs. the predicted model. The Cox & Snell R value was 0.582 and the Nagelkerke R Square was 0.793 explaining some of the variance within the model. The output tables can be found in Appendix BB.

The logistic regression equation for the schools in the 2005 California 8th grade dataset with the state ELA assessment as the dependent variable is the following: Logit (Y, of meeting state accountability proficiency in ELA) = -10.989 (constant) + (0.64 * DUMMYREV) + (-8.936 * EDPER) + (6.483 * NATAMPER) + (26.503 * ASIANPER) + (11.846 * BLACKPER) + (17.035 * HISPANICPER) + (24.831 * WHITEPER) + (-2.940 *ELLPER) + (-.112 * SPECIALEDPER) + (.046 * NAEPMATHPROFbasic).

As Table BB15 indicates, the Wald value was significant in the model for all of the independent variables ($p< .05$). The variables that had a negative influence on a school's ability to meet proficiency include EDPER (B = - 8.936 , Exp (B) = .000), ELLPER (B = -2.940, Exp (B) = .000), and SPECIALEDPER (B = -.112 , Exp (B) =0.104). The EDPER variable has provided a consistent negative impact on the odds that a school met proficiency status in almost all of the datasets throughout this study including the mathematics assessments, reading (or ELA) assessments, and AYP results. However, the negative impact of the SPECIALEDPER variable and the ELLPER variable is in opposition to the finding when the 2005 8th grade mathematics assessment was the dependent variable.

The following variables had a positive influence on the probability of a school meeting proficiency on the 8th grade ELA assessment, DUMMYREV (B = .064 , Exp (B) = 1.066), NATAMPER (B = 6.483, Exp (B) = 653.924), ASIANPER (B = 26.503, Exp (B) = 3.236), BLACKPER variable (B = 11.846 , Exp (B) = 139526.2), HISPANICPER (B = 17.035 , Exp (B) = 2.50), WHITEPER (B = 24.831 , Exp (B) = 6.083), and NAEPREADPROFbasic (B = .046 , Exp (B) = 1.047). The NATAMPER variable must be looked at with caution because it produced extremely high odds ratio results. The reason for this could be due to the low level of NATAMPER population. Thus there was very little impact with this variable. The BLACKPER variable also produced high odds ratio results indicating that those values should be interpreted with caution.

It is interesting that other than the EDPER variable and NAEPREADPROFbasic variable, all of the other variables had an opposite influence on the results vs. the 8th grade California 2005 mathematics results. For example DUMMYREV, NATAMPER, ASIANPER, BLACKPER, HISPANICPER, and WHITEPER all had a positive influence with the state ELA assessment as the dependent variable. However, the same variables had a negative impact when the state mathematics assessment was the dependent variable. The opposite was true with the ELLPER and SPECIALPER variables as there was a positive impact with the mathematics assessment results and a negative impact with the ELA assessment results. As mentioned, a portion of the explanation has to due with the unstable odds ratios with some of the variables.

California 2007 4th grade mathematics

For the mathematics 2007 4th Grade California dataset, there was a very moderate positive correlation between the proficient level NAEP scale in mathematics and the state mathematics accountability assessment, ($r = 0.073$, significant at the .01 level). There was also a very moderate positive correlation for the mathematics basic level NAEP scale and state mathematics accountability assessment ($r = 0.083$, $p < .01$). The tables in Appendix AA provide descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

The state mathematics accountability assessment was used as the dependent variable for a logistic regression to analyze the California schools 2005 4th grade dataset. Although the results of the logistic regression were statistically significant, the predicted probability of the constant (null) model was 96% accurate vs. the logistic model at 95.9% accurate. The explanation for the negative impact on the model had to do with the high number of schools that met proficiency status in mathematics vs. the low number that did not. Over 95% of the schools met proficiency status on the mathematics state accountability assessment, while under 79% met the basic level NAEP scale proficiency and 12% of the schools met the proficient level NAEP scale proficiency. Thus although the logistic model was statistically significant, the model was not very useful due to the high level of accuracy on the null model which is predicted based on chance. The output tables from the logistic regression can be found in Appendix BB.

California 2007 4th grade ELA/reading

With the analysis of the reading 2007 4th grade California dataset, there was a very moderate positive correlation between the proficient level NAEP scale reading assessment and the state reading accountability assessment, ($r = 0.089$ at the .01 level). There was also a moderate positive correlation between the basic level NAEP scale reading assessment and the California ELA accountability assessment, ($r = 0.257$ $p < .01$). Refer to Appendix AA for descriptive statistics and the correlation matrix.

A logistic model was also used to analyze the 4th grade California 2007 dataset with the ELA state accountability assessment as the dependent variable. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, similar to some of the other datasets, the NAEPREADPROFbasic variable and NAEPREADPROFprof variables were analyzed separately from the other variables. This was done because the Omnibus Model of Coefficients test was not statistically significant ($p > .05$) indicating that the model did not provide a goodness of fit. However, when the test was analyzed without the NAEPREADPROFprof variable, the Omnibus Model of Coefficients test was statistically significant ($p < .05$) indicating goodness of fit with the model. Thus the logistic results are reported without the NAEPELAPROFprof level variable.

The constant (null) model was correct 89.2% of the time vs. the predicted model which was accurate 92.2% of the time. There was a mild increase in accuracy for the predicted model vs. the constant model. A problem with this dataset is that over 80% of the schools met state proficiency standards on the state ELA assessment while just over 5% met NAEP proficiency at the proficient level. The data analysis with the logistic regression model should be interpreted while considering that factor. Refer to Appendix BB for tables relevant to the logistic regression for the 2007 4th Grade California dataset.

California 2007 8th grade mathematics

A Pearson correlation analysis of the 2007 8th Grade California dataset showed that there was a moderate positive correlation between the proficient level NAEP scale mathematics assessment and the mathematics state accountability assessment, ($r = 0.242$, significant at the .01 level). There was also a strong positive correlation with the basic level NAEP scale in mathematics and the state mathematics accountability assessment ($r = 0.567$, $p < .01$). Appendix AA shows the output tables including descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and the financial variable. These correlation results are higher than the ones found in the other California datasets indicating that there is a closer relationship with the California state accountability mathematics assessments and the NAEP assessment (especially at the basic level NAEP scale) vs. the other California datasets with the analysis of research question two.

In order to provide an analysis of the 2007 8th grade California dataset, the mathematics state accountability assessment was used as the dependent variable for a logistic regression. The constant (null) model was correct 60.2% of the time vs. the predicted model which was accurate 80.4% of the time. Thus there was a large increase with the predicted model vs. the constant model at just over 20%. This signifies a positive influence on the regression model. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model. The Cox & Snell R value was .413 and the Nagelkerke R Square was 0.559. The output tables can be found in Appendix BB.

The logistic regression equation for the schools in the 2007 California 8th grade dataset with mathematics as the dependent variable is the following: Logit (Y, predicted probability of meeting mathematics state accountability assessment proficiency) = 20.344 (constant) + (0.072 * DUMMYREV) + (-2.671 * EDPER) + (-36.668 * NATAMPER) + (-16.289 * ASIANPER) + (-24.471 * BLACKPER) + (-19.300 * HISPANICPER) + (-17.430 * WHITEPER) + (2.049 * ELLPER) + (-8.676 * SPECIALEDPER) + (1.853 * NAEPMATHPROFbasic) + (17.014 * NAEPMATHPROFprof).

The only independent variable that was not statistically significant in the regression model was the NAEPMATHPROFprof ($p > .05$). Three of the statistically significant variables ($p < .05$) had a positive influence on the schools ability to meet proficiency status: DUMMYREV

(B =.072, Exp (B) = 1.074), ELLPER (B = 2.049 , Exp (B) = 7.764), and NAEPMATHPROFbasic (B = 1.853, Exp (B) = 6.832). This indicates that financial resources had a mild positive influence on a schools ability to meet the mathematics proficiency status while the ELLPER variable had a greater positive impact. If a student met the basic level NAEP scale proficiency status they also had a greater odds of meeting state accountability proficiency status.

The statistically significant variables ($p < .05$) that were negatively associated with a school meeting the 4th grade mathematics proficiency include: EDPER (B = -2.671, Exp (B) =0.069), NATAMPER variable (B = - 36.668 , Exp (B) = 0.00), ASIANPER (B = -16.289, Exp (B) =0.00), BLACKPER (B = -24.471, Exp (B) =0.00), HISPANICPER (B = -19.300, Exp (B) =0.00), WHITEPER (B = -17.430, Exp (B) =0.00), and SPECIALEDPER (B = -8.676 , Exp (B) =0.00). It is surprising that the WHITEPER variable, and ASIANPER variable had a negative influence on the results. However, as previously mentioned, view the results of variables with low odds ratios with caution as they indicate an unstable impact by the independent variable. The other variables that were negatively associated with the probably of meeting proficiency standards have not been consistent between the datasets with the exception of the EDPER variable. The EDPER variable has produced results showing a consistent negative influence on the predicted probably of meeting proficiency status in almost all of the datasets throughout this study.

It is also important to consider that the state accountability assessment in mathematics at the 8th grade level is an algebra test in California. This might help to explain why there was a stronger correlation between the NAEP assessment and the state accountability assessment. The rigor of the California mathematics assessment could have a greater relationship with the NAEP vs. some of the other assessments that are analyzed with the California datasets in this study. This might also help to explain why there was a large increase in the predicated probably of a school meeting proficiency status with this dataset.

California 2007 8th grade ELA/reading

An analysis of the 2007 8th grade California dataset showed that there was a very moderate positive correlation between the proficient level NAEP scale reading assessment and the state accountability assessment in reading, ($r = 0.112$ at the .01 level). There was a much stronger positive correlation between the basic level NAEP scale in reading and the California ELA accountability assessment, ($r= 0.581$ $p < .01$). This supports earlier findings that the basic level NAEP scale proficiency is more closely aligned with the state accountability assessments with the 2007 California dataset. Refer to Appendix AA for descriptive statistics and the correlation matrix.

The California ELA state accountability assessment was used as the dependent variable for a logistic regression to analyze the 2007 8th grade dataset. The constant (null) model was correct 67.7% of the time vs. the predicted model which was accurate 89.4% of the time. The result was a significant increase with the predicted model vs. the constant model of over 22%. This signifies a positive influence and a useful regression model. The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model. The Cox & Snell R value was .523 and the Nagelkerke R Square was 0.730. The output tables can be found in Appendix BB.

The logistic regression equation for the schools in the 2007 8th grade California dataset with the ELA assessment as the dependent variable is the following: Logit (Y, predicted probability of meeting state ELA proficiency status) = -14.432 (constant) + (-0.297 * DUMMYREV) + (-3.812 * EDPER) + (-4.260 * NATAMPER) + (22.903 * ASIANPER) + (11.347 * BLACKPER) + (16.689 * HISPANICPER) + (33.777 * WHITEPER) + (0.891 *ELLPER) + (-1.026 * SPECIALEDPER) + (0.880 * NAEPMATHPROFbasic) + (12.407 * NAEPMATHPROFprof).

Several of the statistically significant variables ($p < .05$) had a positive influence on the schools' ability to meet proficiency status; ASIANPER (B = 22.903 , Exp (B) = 8.840), BLACKPER (B = 11.347, Exp (B) = 84705.098), HISPANICPER (B = 16.689, Exp (B) = 1.77), WHITEPER (B = 33.777 , Exp (B) = 4.667), ELLPER (B = 0.0891 , Exp (B) = 2.437), and NAEPREADPROFbasic (B = 0.880, Exp (B) = 2.411). This indicates that schools with higher percentages of Asian students, White students, Hispanic students, and ELL students have a greater change of meeting math proficiency standards. However, extremely high odds ratios must be interpreted with caution and the BLACKPER variable produced a high odds ratio variable. The Hispanic student variable results might be a little surprising since it has a positive association and the output values have not been consistent among the different datasets.

The statistically significant variables ($p < .05$) that were negatively associated with a school meeting the 4th grade mathematics proficiency include: DUMMYREV (B = -0.297 - , Exp (B) =0.743), EDPER(B = -3.812 - , Exp (B) =0.22), NATAMPER (B = - 4.260 , Exp (B) = 0.014), and SPECIALEDPER (B = -1.026 , Exp (B) =0.358). The EDPER variable has produced results showing a consistent negative impact on the probability that a school meets proficiency status. The DUMMYREV variable had a slight negative influence and the mild fluctuation with the influence of that variable has been consistent among the other datasets. The results with the SPECIALPER variable are aligned with the predicted hypothesis in this study but the results have not been consistent among the different datasets.

Michigan 2005 4th grade mathematics

An analysis of the 2005 Michigan dataset with mathematics as the dependent variable, produced results with a moderate positive correlation between the proficient level NAEP scale in mathematics and the mathematics state accountability assessment, ($r = 0.254$, significant at the .01 level). There was a stronger positive correlation for the mathematics basic level NAEP scale and the state mathematics accountability assessment ($r = 0.392$, $p < .01$). The tables in Appendix CC provide descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

The Michigan 4th grade mathematics state accountability assessment was used as the dependent variable for a logistic regression to analyze the 2005 dataset. The constant (null) model was correct 86.5% of the time vs. the predicted model which was accurate 91.7% of the time. There was a moderate increase for the predicted model vs. the constant model which was just over 5%. This signifies a slight positive influence on the predicted model. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model. The Cox & Snell R value was 0.316 and the Nagelkerke R Square was 0.578. The output tables can be found in Appendix DD.

As table DD5 indicates, the Wald value was significant in the model for most of the independent variables ($p < .05$). However, BLACKPER and the NAEPMATHPROFprof variables were not statistically significant ($p > .05$). The logistic regression equation for the schools in the 2005 Michigan 4th grade dataset with mathematics as the dependent variable is the following: Logit (Y, predicted probability of meeting proficiency on the mathematics state accountability assessment) = -1.822 (constant) + (-0.469 * DUMMYED) + (-1.968 * EDPER) + (-9.762 * NATAMPER) + (7.338 * ASIANPER) + (1.082 * BLACKPER) + (11.799 * HISPANICPER) + (4.554 * WHITEPER) + (48.537 *ELLPER) + (-10.630 * SPECIALEDPER) + (-.192 * NAEPMATHPROFbasic) + (19.069 * NAEPMATHPROFprof).

Several of the statistically significant variables ($p < .05$) had a positive influence on the schools ability to meet proficiency status: DUMMYREV (B = .0469, Exp (B) = 1.598), ASIANPER (B = 7.338 , Exp (B) = 1537.087), HISPANICPER (B = 11.799, Exp (B) = 133160.543), WHITEPER (B = 4.554 , Exp (B) = 95.016), and ELLPER (B = 48.537 , Exp (B) = 1.200). This indicates that financial resources and schools with higher percentages of Asian students, White students, Hispanic students, and ELL students have a greater chance of meeting math proficiency standards. The Hispanic student variable might be a little surprising since it was negatively associated to meeting proficiency status with the results in many of the other state datasets. However, the odds ratio variable was extremely high with the Hispanic variable, indicating an unusual result. The Hispanic population is not that large in Michigan in comparison to other states which might explain the unusual results with this variable. Similar results were found with the ELLPER variable.

The statistically significant variables ($p < .05$) that were negatively associated with a school meeting the 4^{th} grade mathematics proficiency include: EDPER(B = -1.968, Exp (B) =0.14), NATAMPER (B = - 9.762 , Exp (B) = 0.00), SPECIALEDPER (B = -10.360 , Exp (B) =0.00), and NAEPMATHPROFbasic (B = -0.192 , Exp (B) =0.825). It is a little surprising that schools which meet NAEP proficiency in mathematics have a mild negative impact on the schools' chances of meeting proficiency on the math state accountability assessment. The impact is only slightly negative and it is in alignment with the predicted hypothesis that there would not be a strong relationship between the NAEP assessment results and state accountability assessment results.

Michigan 2005 4th grade ELA/reading

An analysis of the 2005 Michigan reading dataset at the 4th grade level, produced a very moderate positive correlation between the proficient level NAEP scale reading assessment the state ELA accountability assessment, ($r = 0.047$ at the .01 level). There was a very moderate negative correlation between the basic level NAEP scale reading assessment and the Michigan ELA accountability assessment, ($r = -0.093$ $p < .01$). A high number of schools met ELA proficiency in Michigan at over 97%, as opposed to a low number of schools that met the proficient level NAEP scale proficiency at under 8%. This provides some of the explanation for the extremely moderate correlation between the two assessments. Refer to the tables in Appendix CC for descriptive statistics and the correlation matrix.

The Michigan 4th grade ELA state accountability assessment was used as the dependent variable for a logistic regression to analyze the 2005 dataset. The logistic regression was implemented and the constant (null) model was correct 96.5% of the time. As a result, the analysis of a logistic regression output table with the state ELA accountability assessment as the dependent variable was not possible. Less than 3% of the schools failed to meet proficiency on the state ELA assessment providing for a low variance among the dependent variable. However, the constant (null) model computed results which can be found in Appendix DD.

Michigan 2005 8th grade mathematics

A Pearson correlation was used to analyze the 8th grade 2005 mathematics dataset, a moderate positive correlation was found between the mathematics proficient level NAEP scale and the mathematics state accountability assessment, ($r = 0.150$, significant at the .01 level). There was a strong positive correlation for the mathematics basic level NAEP scale and the mathematics accountability assessment ($r = 0.640$, $p < .01$). These results provide support to the finding that the basic level NAEP scale is more closely aligned with the state accountability assessments in the state of Michigan. It is also in alignment with literature in the field regarding the closer alignment with the basic level NAEP scale in states across the country (Bandeira de Mello, et al. 2009; Fuller et al. 2007; Gewertz, 2010). Refer to Appendix CC for descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and the financial variable.

The Michigan 8th grade mathematics state accountability assessment was used as the dependent variable for a logistic regression to analyze the 2005 dataset. However, the Omnibus Test of Model Coefficients was only statistically significant ($p < .05$) with the absence of the NAEPMATHPROFprof variable. Thus the NAEPMATHPROFprof variable is not included in the reported results. The constant (null) model was correct 87.3% of the time vs. the predicted model which was accurate 96.3% of the time. There was a moderate increase with the predicted model vs. the constant model which was 9%. This signifies a slight positive influence on the predicted model. The Cox & Snell R Squared value was 0.418 and the Nagelkerke R Square was 0.784 explaining some of the variance within the model. The output tables can be found in Appendix DD.

As Table DD11 in Appendix DD indicates, the Wald Squared value was significant in the model for all of the independent variables (p< .05). However, caution must be given to the influence of some of the variables as they produced abnormally high results. This could be due to the high number of schools that met state proficiency (over 86%) on the mathematics assessment establishing a lower level of variance among the dependent variable.

The logistic regression equation for the schools in the 2005 Michigan 8th grade dataset with mathematics as the dependent variable is the following: Logit (Y, predicted probability of meeting proficiency on the mathematics state accountability assessment) = 142.044 (constant) + (3.183 * DUMMYED) + (-10.205 * EDPER) + (-135.661 * NATAMPER) + (-103.474 * ASIANPER) + (-146.628 * BLACKPER) + (-1652.567 * HISPANICPER) + (-141.129 * WHITEPER) + (-4.129 *ELLPER) + (5.451 * SPECIALEDPER) + (2.719 * NAEPMATHPROFbasic).

The statistically significant variables (p < .05) that had a positive influence on the schools' ability to meet proficiency status include: DUMMYREV (B = 3.183, Exp (B) = 24.115), SPECIALEDPER (B = 5.451, Exp (B) = 232.966), and NAEPMATHPROFbasic (B = 2.719, Exp (B) = 15.166). This indicates that increased financial resources and meeting the NAEP proficiency at the basic level create a greater chance of meeting state proficiency. It is worth noting that the EDPER variable (B = -10.205, Exp (B) = .000) continues to impact a schools chances to meet proficiency in a negative manner. All of the other variables produced high or abnormal predicated probability ratios indicating that the outputs were not very useful. Less than 14% of the schools in the sample did not meet proficiency. Refer to Appendix DD for the output tables.

Michigan 2005 8th grade ELA/reading

With an analysis of 2005 8th grade reading dataset in Michigan, there was a very moderate positive correlation between the proficient level NAEP scale reading assessment and the ELA state accountability assessment, ($r = 0.085$ at the .01 level). There was a much stronger positive correlation between the basic level NAEP scale reading assessment and the ELA state accountability assessment, ($r= 0.618 \quad p < .01$). The correlations once again show that the Michigan state accountability assessment has a close relationship with the basic level NAEP scale proficiency. Refer to Appendix CC for descriptive statistics and the correlation matrix.

The Michigan 8th grade ELA state accountability assessment was used as the dependent variable for a logistic regression to analyze the 2005 dataset. The logistic regression was implemented and the constant (null) model was correct 95.6% of the time. The result was that a logistic regression output with the state ELA accountability assessment as the dependent variable was not possible since such a high percentage of schools met proficiency status on the ELA assessment. Only 5% of the schools in the sample did not meet ELA proficiency status. However, the constant model provided an output reading and it can be found in Appendix DD.

Michigan 2007 4th grade mathematics

For the mathematics 2007 4th grade Michigan dataset, there was a moderate positive correlation between the mathematics proficient level NAEP scale and the mathematics state accountability assessment, ($r = 0.142$, significant at the .01 level). There was a very strong positive correlation for the mathematics basic level NAEP scale and the mathematics state accountability assessment ($r = 0.725$, $p < .01$). This provides more evidence from multiple years and grades of assessment indicating that there is a strong relationship between the basic level

NAEP scale proficiency and the state accountability assessments in Michigan. Refer to Appendix CC for the descriptive statistics and correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

A logistic model was also used to analyze the 4th grade 2007 mathematics state accountability assessment results. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, the NAEPREADPROFbasic level variable and NAEPELAPROFprof level variables were analyzed separately because the Omnibus Model of Coefficients test was not statistically significant ($p < .05$). However, when the NAEPMATHPROFprof level variable and NAEPMATHPROFbasic variables were not used for the logistic regression model, the Omnibus Model of Coefficients test was statistically significant ($p < .05$) indicating goodness of fit with the model.

The constant (null) model in the logistic analysis was correct 95.3% of the time vs. the predicted model which was accurate 100% of the time. Thus there was a very low increase in the accuracy for the predicted model vs. the constant model and thus very little usefulness with this model. Similar to the other Michigan datasets, a problem with this dataset has to do with the fact that such a high number of schools met mathematics proficiency (over 95%) on the state accountability assessment in 2007. This explains why none of the independent variables in the model were statistically significant. Refer to Appendix DD for the SPSS output tables.

Michigan 2007 4th grade ELA/reading

For the 2007 4th grade Michigan dataset in reading/ELA, there was a very moderate positive correlation between the proficient level NAEP scale reading assessment and the ELA state accountability assessment, ($r = 0.117$ at the .01 level). There was a much stronger positive correlation between the basic level NAEP scale in reading and the Michigan ELA accountability assessment, ($r = 0.664$ $p < .01$). Thus the Pearson results continue to show support to the finding that the Michigan state accountability assessment has a strong relationship with the NAEP scale assessment at the basic level. Refer to Table CC for descriptive statistics and the correlation matrix.

A logistic model was used to analyze the 2007 4th grade ELA dataset with the state accountability assessment as the dependent variable. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, similar to some of the other datasets, the NAEPREADPROFbasic level variable and NAEPELAPROFprof level variables were analyzed separately using the stair step option in SPSS. The reason for doing this was because the Omnibus Model of Coefficients test was not statistically significant ($p > .05$) indicating that the model did not provide a goodness of fit. However, when the NAEPELAPROFprof level variable was not used for the logistic regression model, the Omnibus Model of Coefficient test was statistically significant ($p < .05$) indicating goodness of fit with the model. Thus the rest of this analysis provides the logistic regression results without the NAEPELAPROFprof level variable.

The constant (null) model was correct 94.3% of the time vs. the predicted model which was accurate 100% of the time. There was a moderate increase in accuracy for the predicted model vs. the constant model. Also, none of the independent variables were found to be statistically significant. Similar to the other Michigan datasets, a problem with this dataset has to do with the fact that a high number of schools at over 93% met state proficiency standards on the state ELA assessment while just over 5% met NAEP proficiency at the proficient level. Thus the data analysis with the logistic regression model must be analyzed while keeping in mind that the predicted model provided a low increase in the accuracy of the results and statistically significant results were absent. Refer to Appendix DD for the observed predicted frequencies of the constant model vs. the predicted model. This table and the rest of the tables and figures can be found in Appendix DD.

Michigan 2007 8th grade mathematics

A Pearson Correlation analysis of the mathematics 2007 8th grade Michigan dataset produced a moderate positive correlation between the NAEP scale mathematics assessment at the proficient level and the state mathematics accountability assessment, ($r = 0.142$, significant at the .01 level). There was a very strong positive correlation with the mathematics basic level NAEP scale and state mathematics accountability assessment ($r = 0.725$, $p < .01$). The strong positive correlation with the basic level NAEP scale is consistent with the state of Michigan datasets. Refer to Appendix CC for descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

A logistic model was also used to analyze the 8th grade 2007 Michigan dataset with the math state accountability assessment as the dependent variable. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, the NAEPREADPROFbasic variable and NAEPELAPROFprof variables were analyzed separately using the stair step method in SPSS. Similar to previously analyzed datasets in Michigan, the Omnibus Model of Coefficients was not statistically significant ($p > .05$) indicating that when the NAEP variables were included, goodness of fit was not found. However, when the NAEPMATHPROFprof level variable and NAEPMATHPROFbasic variables were not used for the logistic regression model, the Omnibus Model of Coefficients test was statistically significant ($p < .05$) indicating goodness of fit with the model.

The logistic regression showed that the constant (null) model was correct 88% of the time vs. the predicted model which was accurate 100% of the time. There was a 12% increase in the accuracy for the predicted model vs. the constant model. A problem with this dataset which is similar to the other Michigan datasets, has to do with the fact that such a high number of schools met math proficiency on the state accountability assessment in 2007 at over 88%. This helps to explain why none of the independent variables in the model were statistically significant. Refer to Appendix DD for the SPSS output tables with the analysis of the logistic regression.

Michigan 2007 8th grade ELA/reading

For the ELA/reading 2007 8th grade Michigan dataset, there was a very moderate positive correlation between the two variables for the proficient level NAEP scale and the ELA state accountability assessment, ($r = 0.095$ at the .01 level). There was a stronger positive correlation between the basic level NAEP scale and the Michigan ELA accountability assessment, ($r= 0.573$, $p < .01$). Refer to Appendix CC for descriptive statistics and the correlation matrix.

A logistic model was also used to analyze the 8th grade 2007 dataset with the ELA state accountability assessment as the dependent variable. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, the NAEPELAPROFbasic level variable and NAEPELAPROFprof level variables were analyzed separately because of the lack of goodness of fit with the model which has been consistent with the 2007 Michigan datasets. When the NAEPREADPROFprof level variable was not used for the logistic regression model, the Omnibus Model of Coefficients test was statistically significant

($p < .05$) indicating goodness of fit with the model. Thus the rest of this analysis provides the logistic regression results without the NAEPREADPROFprof variable.

The logistic regression showed that the constant (null) model was correct 89.6% of the time vs. the predicted model which was accurate 99.6% of the time. There was a moderate increase in accuracy for the predicted model vs. the constant model of 10%. Only 9% of the schools in the sample did not meet state proficiency standards on the state ELA assessment while only 7% of the schools in the sample met proficiency on the NAEP reading assessment at the proficient level. The data analysis with the logistic regression model must be analyzed while keeping in mind that fact that the predicted model provided a moderate increase in the accuracy of the results. The Cox & Snell R Squared value was .441 and the Nagelkerke R Square was 0.904. Refer to Appendix DD for the SPSS output tables.

As Table DD32 in Appendix DD indicates, the Wald Squared value was significant in the model for all of the independent variables ($p< .05$). The logistic regression equation for the schools in the 2007 Michigan 8th grade dataset with ELA as the dependent variable is the following: Logit (Y, predicted probability of meeting proficiency on the state ELA proficiency assessment) = -33.732 (constant) + (2.870 * DUMMYED) + (-8.989 * EDPER) + (478.465 * NATAMPER) + (74.79 * ASIANPER) + (33.73 * BLACKPER) + (30.202 * HISPANICPER) + (64.567 * WHITEPER) + (-41.385 * ELLPER) + (-96.048 * SPECIALEDPER) + (-3.355 * NAEPMATHPROFbasic).

Several of the independent variables had a positive influence on the ability of the school to meet proficiency status: DUMMYREV (B =2.870, Exp (B) = 17.628), NATAMPER (B = 478.465 , Exp (B) = 6.233), ASIANPER (B = 74.790 , Exp (B) = 3.027), BLACKPER (B = 33.73, Exp (B) = 4.454), HISPANICPER (B = 30.202, Exp (B) = 1.308), and WHITEPER (B

= 64.567 , Exp (B) = 1.099). This indicates that financial resources and schools with higher percentages of Asian students, White students, Black students, and Hispanic students, have a greater change of meeting math proficiency standards. While the NATAMPER was a large positive value, the percentage of Native American students in the Michigan schools were very low and the result is a low impact on the regression model.

The independent variables ($p < .05$) that were negatively associated with a school meeting the 8th grade ELA proficiency include: EDPER (B = -8.989, Exp (B) = 0.00), ELLPER variable (B = - 41.385 , Exp (B) = 0.00), SPECIALEDPER (B = -99.048 , Exp (B) =0.00), and NAEPMATHPROFbasic (B = -3.355, Exp (B) =0.035). The negative association with the NAEP results is in alignment with the hypothesis in this study that there would not be a positive relationship with success on the NAEP and state accountability assessment. The negative association with the EDPER variable is consistent with the results in the other datasets in this study. The negative association with ELLPER and SPECIALEDPER is not surprising based on the predicted hypothesis. Michigan has a level of identification of special education students at 14% of the student population throughout the state based on the 2007-08 statistics (USDOE, NCES, 2008). The national average is just over 13%.

North Carolina 2005 4th grade mathematics

A Pearson Correlation analysis of the 2005 4th grade North Carolina dataset produced a moderate positive correlation between the proficient level NAEP scale mathematics assessment and the mathematics state accountability assessment, (r = 0. 170, significant at the .01 level). There was also a strong positive correlation for the mathematics basic level NAEP scale and the mathematics state accountability assessment ($r = 0.435$, $p < .01$). Refer to Appendix EE for the correlation matrix and descriptive statistics.

The mathematics state accountability assessment was used as the dependent variable for a logistic regression to analyze the North Carolina schools 2005 4th grade dataset. Although the results of the logistic regression were statistically significant ($p < .05$), the predicted probability of the constant (null) model was 93.3% accurate vs. the logistic model at 96.1% accuracy. The low increase in the predicted model can be explained by the high number of schools that met state proficiency status in mathematics vs. a low number that did not. The number of schools in the sample that did not meet state proficiency status was just over 6%.

This logistic model was not very useful due to the high level of accuracy on the null model which is predicted based on chance. Please refer to Appendix FF for the observed and predicted frequencies of the constant model and predicted model. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model as it was also indicated by the Cox & Snell R value of .229 and the Nagelkerke R Square value of .591 which accounts for some of the variance within the model.

The logistic equation for the North Carolina 2005 4th grade dataset is: Logit (Y, of meeting mathematics state proficiency) = -8.573 (constant) + (1.007 * DUMMYREV) + (-3.785 * EDPER) + (-4.610 * NATAMPER) + (10.461 * ASIANPER) + (-8.472 * BLACKPER) + (-19.450 * HISPANICPER) + (-.570 * DUMMYWHITE) + (9.003 * ELLPER) + (-19.664 * SPECIALEDPER) + (23.516 * NAEPMATHPROFbasic) + (14.694 * NAEPMATHPROFprof).

As Table FF5 in Appendix FF indicates, the Wald value was statistically significant in the model for most of the independent variables ($p < .05$) other than NAEPMATHPROFbasic and NAEPMATHPROFprof ($p > .05$). The statistically significant variables that were positively associated with the odds that a school met AYP proficiency standards were DUMMYREV (B = 1.007, Exp (B) = 2.737), ASIANPER (B = 10.461, Exp (B) = 34909.835), and ELLPER (B = 9.003, Exp (B) = 8130.512). The variables that had a negative impact on the mathematics state accountability assessment outcome include EDPER (B = -3.785, Exp (B) = .023), NATAMPER (B = -4.610 , Exp (B) = .010), BLACKPER, (B = -8.472, Exp (B) = .000), HISPANICPER (B = -19.450, Exp (B) = .000), DUMMYWHITE (B = - .570, Exp (B) = .565), and SPECIALEDPER (B = - 19.664, Exp (B) = .000).

It was not surprising that the EDPER variable was negatively associated with the probability that a school met the mathematics proficiency as this finding has been consistent throughout the study in all of the statistically reliable datasets. The DUMMYREV variable was positively associated with the probability that a school met mathematics proficiency which is something that is not surprising and generally consistent throughout this study especially with the datasets in the state of Michigan (some of the datasets produced results that were slightly negative in association). The finding that the ASIANPER was positively associated with the chance that a school met mathematics proficiency corresponds to an earlier finding in this study. However, please exert caution when interpreting the results of variables that produced extremely high or extremely low odds ratio values. The negative impact of the SPECIALEDPER and the positive impact with the ELLPER variable were not consistent between the different states and years of assessment with predicting school level success. Each of the datasets established different findings with respect to these two variables.

Since the NAEPMATHPROFbasic and the NAEPMATHPROFprof variables were not statistically significant ($p > .05$), it is difficult to determine the impact that these variables had on the state mathematics accountability assessment. An explanation why the variables were not statistically significant could have to do with the low number of schools that failed to meet mathematics proficiency. Throughout the study various datasets have produced results similar in value due the low number of schools that failed to meet state proficiency. However, the Pearson test showed that there was a moderate positive correlation with the NAEPMATHPROFprof variable ($r = 0.170$, $p < .01$) and a more sizable positive correlation with the NAEPMATHPROFbasic variable ($r = .453$, $p < .01$) indicating that there was a statistically significant relationship. This finding is somewhat consistent with the Michigan findings although the positive correlation with the Michigan datasets and the basic level NAEP scale were greater in value.

North Carolina 2005 4th grade reading

An analysis of the 2005 4th grade North Carolina reading dataset produced a moderate positive correlation between the proficient level NAEP scale in reading and the state accountability assessment in reading, ($r = 0.210$, at the .01 level). There was a much stronger positive correlation between the basic level NAEP scale in reading and the North Carolina reading state accountability assessment, ($r= 0.568$, $p < .01$). Refer to Appendix EE for the descriptive statistics and the correlation matrix. These results are consistent with the North Carolina 2005 4th grade mathematics analysis and most of the Michigan datasets.

The state accountability assessment in reading was used as the dependent variable for a logistic regression to analyze the North Carolina schools 2005 4th grade dataset. The predicted probability of the constant (null) model was 74.7% accurate vs. the logistic model at 85.3% accurate. This signifies a moderate increase of over 10% for the predicted model vs. the constant model. Please refer to Table FF6 and FF7 in Appendix FF for the observed and predicted frequencies of the predicted model. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model. It was also supported by both the Cox & Snell R value of .395 and the Nagelkerke R Square value of .583 which accounts for some of the variance within the model. Refer to Appendix FF for the output tables.

The logistic equation for the North Carolina 2005 4th grade reading dataset is: Logit (Y, of meeting the reading state accountability assessment proficiency) = 6.594 (constant) + (-.468 * DUMMYREV) + (-4.763 * EDPER) + (-19.529 * NATAMPER) + (4.434 * ASIANPER) + (-2.871 * BLACKPER) + (-1.857 * HISPANICPER) + (-.179 * DUMMYWHITE) + (-.990 * ELLPER) + (-2.513 * SPECIALEDPER) + (1.335 * NAEPREADPROFbasic) + (16.899 * NAEPREADPROFprof).

As Table FF10 in Appendix FF indicates, the Wald value was statistically significant in the model for most of the independent variables ($p< .05$) other than NAEPREADPROFprof ($p > .05$). The only statistically significant variable that was positively associated with the odds that a school met reading proficiency standards was NAEPREADPROFbasic (B = 1.335, Exp (B) = 3.799). While the NAEPREADPROFprof variable had a positive impact on the probability of meeting the reading proficiency standard, the result was not statistically significant ($p > .05$).

As displayed in Table FF10 in Appendix FF, most of the variables had a negative association with the probability that a school met the reading proficiency standard, DUMMYREV (B = -.468, Exp (B) = .627), EDPER (B = -4.763, Exp (B) = .009), NATAMPER (B = -19.629, Exp (B) = .000), ASIANPER (B = -4.434, Exp (B) = .012), BLACKPER (B = -2.871, Exp (B) = .057), HISPANICPER (B = -1.857, Exp (B) = .156), DUMMYWHITE (B = -.990, Exp (B) = .836), ELLPER (B = - 2.513, Exp (B) = .372) and SPECIALEDPER (B = -2.513, Exp (B) = .081).

It was not surprising that the EDPER variable was negatively associated with the probability that a school met the mathematics proficiency as this finding is consistent throughout the study. The DUMMYREV variable had a mild negative association with the probability that a school met reading proficiency which is in contrast to earlier findings with a positive impact in other datasets. It was surprising that the ASIANPER was negatively associated with the chance that a school met reading proficiency as this variable has been positively associated with proficiency status in other datasets in this study. However, the negative impact with the SPECIALEDPER and the ELLPER variable is not surprising and the finding has not been consistent from state to state and year of assessment with predicting school level success. Please refer to Appendix FF for the logistic regression output tables.

North Carolina 2005 8th grade mathematics

An analysis of the mathematics 2005 8th grade North Carolina dataset showed that there was a moderate positive correlation between proficient level NAEP scale mathematics assessment and the mathematics state accountability assessment ($r = 0.239$, significant at the .01 level). There was a stronger positive correlation with the mathematics basic level NAEP scale and the mathematics state accountability assessment ($r = 0.414$, $p < .01$). The result is consistent with earlier findings in North Carolina and somewhat consistent with findings from Michigan and California. Refer to Appendix EE for descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

The North Carolina 8th grade mathematics state accountability assessment was used as the dependent variable for a logistic regression to analyze the 2005 dataset. The constant (null) model was correct 70.1% of the time vs. the predicted model which was accurate 89.9% of the time. Thus there was an increase of over 19% for the predicted model vs. the constant model. This signifies a positive influence on the predicted model. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model as it was also supported by the Cox & Snell R value of 0.515 and the Nagelkerke R Square was 0.731 explaining some of the variance within model. The output tables can be found in Appendix FF.

As table FF15 in Appendix FF indicates, the Wald value was significant in the model for most of the independent variables ($p < .05$) other than the NAEPMATHPROFprof variable and the NAEPMATHPROFbasic ($p > .05$). The logistic regression equation for the schools in the 2005 North Carolina 8th grade dataset with mathematics as the dependent variable is the following: Logit (Y, predicted probability of meeting the state mathematics proficiency status) = -6.200 (constant) + (0.797 * DUMMYREV) + (-6.638 * EDPER) + (-15.537 * NATAMPER) + (44.504 * ASIANPER) + (-11.101 * BLACKPER) + (-6.538 * HISPANICPER) + (-.903 * DUMMYWHITE) + (-10.689 *ELLPER) + (-2.480 * SPECIALEDPER) + (19.292 * NAEPMATHPROFbasic) + (15.960 * NAEPMATHPROFprof).

As displayed in Table FF15 in Appendix FF, the only statistically significant variables that were positively associated with the odds that a school met reading proficiency standards were DUMMYREV (B = .797, Exp (B) = 2.220) and ASIANPER (B = 44.504, Exp (B) 2.127). Both of the variables have generally had a positive association with the probability that a school meets proficiency status with most of the datasets in this study.

Most of the variables had a negative association with the probability that a school met the reading proficiency standard, EDPER (B = -6.638, Exp (B) = .001), NATAMPER (B = -15.537, Exp (B) = .000), BLACKPER (B = -11.101 , Exp (B) = .000), HISPANICPER, (B = - 6.538, Exp (B) = .001), DUMMYWHITE (B = -.903, Exp (B) = .406), ELLPER (B = - 10.689, Exp (B) = .000), and SPECIALEDPER (B = - 2.480, Exp (B) = .084). The EDPER variable has been negatively associated with the probability that a school meets proficiency for most of the datasets in this study. The negative impact of the SPECIALEDPER and the ELLPER variables has not been consistent from state to state and year of assessment with predicting school level success. Please refer to Appendix FF for the logistic regression output tables.

North Carolina 2005 8th grade reading

A Pearson Correlation test was used to analyze the 2005 8th grade North Carolina dataset, there was a very moderate positive correlation between the proficient level NAEP scale reading assessment and the reading state accountability assessment, ($r = 0.068$ at the .01 level). There was a stronger positive correlation between the basic reading NAEP scale and the North Carolina reading accountability assessment, ($r= 0.439$, $p < .01$). The Pearson results in North Carolina are consistent with all of the 2005 datasets with the analysis of research question two as the basic level NAEP scale has a stronger correlation vs. the proficient level NAEP scale. Refer to Appendix EE for descriptive statistics and the correlation matrix.

The North Carolina 8th grade reading state accountability assessment was used as the dependent variable for a logistic regression to analyze the 2005 dataset. The constant (null) model was correct 90.2% of the time vs. the predicted model which was accurate 99.3% of the time. There was an increase of over 9% for the predicted model vs. the constant model. This signifies a moderate positive influence on the predicted model. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model as it was also supported by the Cox & Snell R value of 0.442 and the Nagelkerke R

Square was 0.893 explaining some of the variance within model. The output tables can be found in Appendix FF.

As table FF20 in Appendix FF indicates, the Wald value was significant in the model for most of the independent variables (p< .05) other than the NAEPMATHPROFprof variable and the NAEPMATHPROFbasic. However, as indicated in the next paragraph the output data was somewhat erratic with this dataset due to the extremely high logistic coefficient values and inconsistent odds ratios. Over 90% of the schools in the sample met proficiency status in reading which explains some of the unusual results. Thus caution must be used when analyzing these results as they are somewhat limited in answering the second research question due to the unusual high outputs.

The logistic regression equation for the schools in the 2005 North Carolina 8th grade dataset with reading as the dependent variable is the following: Logit (Y, predicted probability of meeting the reading proficiency status) = 109.443 (constant) + (-13.292 * DUMMYREV) + (-79.681 * EDPER) + (-74.223 * NATAMPER) + (116.143 * ASIANPER) + (-31.216 * BLACKPER) + (19.003 * HISPANICPER) + (- 1.693 * DUMMYWHITE) + (-172.547 *ELLPER) + (5.194 * SPECIALEDPER) + (-29.894 * NAEPMATHPROFbasic + (47.861 * NAEPMATHPROFprof).

As Table FF30 in Appendix FF indicates, the Wald value was statistically significant in the model for most of the independent variables (p< .05) other than NAEPREADPROFbasic and NAEPREADPROFprof (p > .05). Although it was not statistically significant, the NAEPREADPROFbasic variable had a negative association while the NAEPREADPROFprof variable had a positive association with the probability of meeting the reading proficiency standard. However, the Pearson test shows that there was a very moderate positive correlation with NAEPREADPROFprof ($r = 0.068$ at the .01 level) and NAEPREADPROFbasic variable ($r= 0.439$, $p < .01$).

The statistically significant variables that were positively associated with the odds that a school met reading proficiency standards were ASIANPER (B = 116.143, Exp (B) = 2.755), HISPANICPER (B = 19.003, Exp (B) 1.789), and SPECIALEDPER(B = 5.194, Exp (B) 180.183). While the ASIANPER has generally had a positive association with the probability that a school meets proficiency status, the HISPANICPER variable is somewhat surprising as the results have not been consistent in this study. As displayed in Table FF30 in Appendix FF, most of the variables had a negative association with the probability that a school met the reading proficiency standard, DUMMYREV (B = -13.292, Exp (B) .000), EDPER (B = -79,681, Exp (B) = .000), NATAMPER (B = -74.223, Exp (B) = .000), BLACKPER (B = -31.216 , Exp (B) = .000), DUMMYWHITE (B = -1.693, Exp (B) = .000), and ELLPER (B = - 172.547, Exp (B) = .000).

North Carolina 2007 4th grade mathematics

For the mathematics 2007 4th grade North Carolina dataset, there was a strong positive correlation between the NAEP scale proficient level and the mathematics state accountability assessment, ($r = 0.596$, significant at the .01 level). There was a weak positive correlation for the basic level NAEP scale mathematics assessment and state mathematics accountability assessment ($r = 0.079$, $p < .01$). This is the first dataset in the research question two analysis which established a stronger relationship with the proficient level NAEP scale and the state accountability assessment vs. the basic level NAEP scale and the state accountability assessment. This suggests a closer association between the North Carolina 2007 4th grade mathematics assessment and the NAEP. Refer to Appendix EE for the descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

A logistic model was also used to analyze the 4^{th} grade NC dataset with the mathematics state accountability assessment as the dependent variable. The constant (null) model was correct 73.6% of the time vs. the predicted model which was accurate at a rate of 92.3%. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model. It was also supported by the Cox & Snell R value of 0.506 and the Nagelkerke R Square was 0.738 explaining some of the variance within model. Refer to Table Appendix FF for the observed and predicted frequencies of the constant model vs. the predicted model. This table and the rest of the tables and figures in this section of the analysis are in the Appendix FF.

The logistic regression equation for the schools in the 2007 North Carolina 4th grade dataset with mathematics as the dependent variable is the following: Logit (Y, predicted probability of meeting the mathematics proficiency status) = -24.457 (constant) + (1.314 * DUMMYREV) + (-9.461 * EDPER) + (15.782 * NATAMPER) + (36.833 * ASIANPER) + (-4.540 * BLACKPER) + (10.850 * HISPANICPER) + (13.419* WHITEPER) + (6.355 *ELLPER) + (-5.825 * SPECIALEDPER) + (12.064 * NAEPMATHPROFbasic) + (1.144 * NAEPMATHPROFprof).

As Table FF25 in Appendix FF indicates, the Wald value was statistically significant in the model for most of the independent variables ($p< .05$) other than NAEPMATHPROFbasic ($p > .05$). This finding is interesting as the NAEPMATHPROFprof was statistically significant. With the other datasets, the tendency was for the basic level NAEP scale to register a statistically significant output when at times the proficient level NAEP scale was not. However, this dataset provided an opposite result and the finding is also in alignment with the Pearson analysis showing a greater relationship between the proficient level NAEP scale and the state accountability assessment vs. the basic level NAEP scale and the state accountability assessment.

The statistically significant variables that were positively associated with the odds that a school met reading proficiency standards were: DUMMYREV (B = 1.314, Exp (B) = 3.721), NATAMPER (B = 15.782, Exp (B) 7148837.83), ASIANPER(B = 36.833, Exp (B) 9.913), BLACKPER(B = 4.540, Exp (B) 93.682), HISPANICPER(B = 10.850, Exp (B) 51541.625), WHITEPER(B = 13.419, Exp (B) 672812.255), ELLPER(B = 6.355, Exp (B) 575.175), and NAEPMATHPROFprof (B = 1.144, Exp (B) 3.138).

The positive impact that the DUMMYREV variable had is consistent with most of the results in this study. While most of the variables had a positive impact, several of the odds ratios were unusual (either large of small) thus they must be interpreted with caution. The variables that had a negative impact included EDPER (B = -9.461, Exp (B) .000) and SPECIALEDPER (B = -5.825, Exp (B) = .003). The negative impact with the EDPER variable continues to be consistent with all of the results in this study. The negative influence with the SPECIALEDPER variable is not surprising but not consistent in this study.

North Carolina 2007 4th grade reading

It was not possible to conduct a Pearson Correlation test or a logistic regression with the reading state accountability assessment as the dependent variable for the 2007 4th grade dataset. All of the schools in the sample met proficiency status on the North Carolina state accountability assessment in reading. North Carolina changed their reading assessment in 2007 which also changed the line of trajectory that was needed to meet proficiency on the state assessment[24]. This is an example of where the states have a lot of flexibility to determine how many schools meet proficiency by changing the cut score or modifying the line of trajectory for proficiency. It is interesting to note that with the schools in the North Carolina 2007 4th grade sample, 14% of the schools met the proficient level NAEP scale proficiency and 76% met the basic level NAEP scale proficiency in contrast to the 100% proficiency on the state accountability assessment. Even though a statistical analysis was not possible, the fact that all schools met proficiency due to a change in the state cut scores and proficiency level is an interesting finding.

Although a Pearson correlation with the state accountability assessment in reading was not possible, there was a strong negative correlation between the NAEPREADPROFprof and EDPER ($r = 0.599$, $p < .01$) and the NAEPREADPROFbasic and EDPER ($r = 0.548$, $p < .01$). The results continue to establish the negative association between the EDPER variable and school level proficiency. Refer to Appendix EE for the correlation matrix.

North Carolina 2007 8th grade mathematics

For the mathematics 2007 8th grade North Carolina dataset, there was a strong positive correlation between the proficient level mathematics NAEP scale and the state mathematics accountability assessment, ($r = 0.553$, significant at the .01 level). There was a moderate correlation for the mathematics basic level NAEP scale and state mathematics accountability assessment ($r = 0.175$, $p < .01$). This is in alignment with the NC 2007 4th grade mathematics results as there was a stronger correlation with the NAEP proficiency level vs. the basic level NAEP scale. This shows that the NC mathematics assessments for the 2007 assessments are

[24] For detailed information refer to the beginning of chapter 4 for a description of the NC state accountability process.

more closely associated with the proficient level NAEP scale. This is a unique finding in this study in comparison to the other state datasets. Refer to Appendix EE for descriptive statistics and the correlation matrix.

The North Carolina 2007 8th grade state accountability assessment in mathematics was used as the dependent variable for a logistic regression to analyze the 2007 dataset. The constant (null) model was correct 72.8% of the time vs. the predicted model which was accurate 87.9% of the time. There was an increase of over 15% between the predicted model vs. the constant model. This signifies a positive influence on the predicted model. The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model. The Cox & Snell R value was 0.426 and the Nagelkerke R Square was 0.617 explaining some of the variance within the model. Refer to Appendix FF for the output tables.

As table FF30 in Appendix FF indicates, the Wald value was significant in the model for some of the independent variables ($p< .05$). However, BLACKPER, HISPANICPER, WHITEPER, and the NAEPMATHPROFbasic variables were not statistically significant ($p> .05$). The logistic regression equation for schools in the 2007 North Carolina grade dataset with mathematics as the dependent variable is the following: Logit (Y, predicted probability of meeting mathematics proficiency) = - 12.126 (constant) + (0.230 * DUMMYREV) + (-11.291 * EDPER) + (-109.863 * NATAMPER) + (-17.278 * ASIANPER) + (-0.403 * BLACKPER) + (.102 * HISPANICPER) + (1.137 * WHITEPER) + (6.850 *ELLPER) + (-5.027 * SPECIALEDPER) + (15.220 * NAEPMATHPROFbasic) + (1.156 * NAEPMATHPROFprof).

Several of the statistically significant variables ($p < .05$) had a positive influence on the schools' ability to meet proficiency status: DUMMYREV (B = .230, Exp (B) = 1.259), ELLPER (B = 6.850, Exp (B) = 943.624), and NAEPMATHPROFprof (B = 1.156, Exp (B) = 3.176).

This indicates that financial resources had a mild positive impact on the odds of a school meeting proficiency status, while a larger ELL population also had a positive impact. If a school met NAEP proficiency at the proficient level, they had a greater chance of meeting state accountability proficiency. This finding is interesting and in alignment with the Pearson correlation results showing that the 2007 mathematics state accountability assessment had a closer relationship with the proficient level NAEP scale vs. the basic level NAEP scale.

The statistically significant variables ($p < .05$) that were negatively associated with a school meeting the 8th grade mathematics proficiency include: EDPER(B = -11.291, Exp (B) =0.00), NATAMPER variable (B = -109.863, Exp (B) = 0.00), ASIANPER (B = -17.278, Exp (B) =0.00), and SPECIALEDPER (B = -5.027, Exp (B) =0.007). The negative impact of the EDPER variable on a school's ability to meet proficiency has been consistent with the datasets in this study. The large negative influence of the NATAMPER must be interpreted with caution since the value was large while the NATAMPER population in NC is relatively low. The negative impact on the odds of a school meeting proficiency with the ASIANPER variable is somewhat surprising but once again the ASIANPER population in NC is relatively low in comparison to a state like California.

North Carolina 2007 8th grade reading

For the 2007 8th grade North Carolina reading dataset, there was a very weak positive correlation between the two variables for the proficient level NAEP scale in reading and the state reading accountability assessment, ($r = 0.016$ at the .01 level). There was a moderate positive correlation between the basic level NAEP scale in reading and the state accountability assessment in reading, ($r= 0.139$, $p < .01$). These results are much lower and in contrast to the findings in most of the datasets in this study including the NC mathematics results. It is evident that the change in the 2007 reading assessment in NC along with the change in the cut score trajectory had a large positive impact on the state proficiency results. It is also one of the reasons why the NC 2007 mathematics assessment has a closer relationship with the NAEP proficient level scale vs. the reading assessment. Refer to Appendix EE or descriptive statistics and for the correlation matrix.

The North Carolina 2007 8th grade reading state accountability assessment was used as the dependent variable for a logistic regression to analyze the 2007 dataset. The constant (null) model was correct 99.6% of the time vs. the predicted model which was accurate 100% of the time. The logistic model was not very useful in analyzing this dataset due to the fact that such a high number of schools met the reading proficiency standard. As a result, none of the independent variables were statistically significant ($p > .05$). The Omnibus Test of Model Coefficients was not statistically significant. This indicates that the model did not meet the goodness of fit standard. Refer to Appendix FF for the SPSS output tables for the North Carolina 8th grade reading dataset.

The change in the North Carolina state accountability assessment cut score in reading had an impact on the inability to acquire useful logistic regression results. However, this is a key finding in itself showing the level of statistical manipulation that is possible as states change or manipulate assessment cut scores. The finding is also in alignment with the literature in the field showing the variance and statistical manipulation with the cut scores among states across the nation (Darling-Hammond, 2007a; Darling-Hammond, 2007b; Guilfoyle, 2006; Harris, 2007; Ravitch, 2010; Sunderman et al. 2005).

Texas 2005 4th grade mathematics

For the mathematics 2005 4th grade Texas dataset, there was a very moderate positive correlation between the two variables for the mathematics proficient level NAEP scale and the mathematics state accountability assessment ($r = 0.095$, significant at the .01 level). There was also a moderate correlation for the mathematics basic level NAEP scale and the state mathematics accountability assessment ($r = 0.179$, $p < .01$). This signifies one of the lower associations among the datasets. Refer to Appendix GG for the descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

A logistic regression was used to analyze the Texas schools in the 2005 4th grade dataset with the mathematics assessment as the dependent variable. The constant (null) model was correct 97.3% of the time vs. the predicted model which was accurate at the same level. Thus there was no increase and this model was not useful with predicting the outcome of the schools on the mathematics state accountability assessment. Less than 3% of the schools in the sample failed to meet the mathematics proficiency status which had an impact on the usefulness of the dataset.

Although the Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model, the Cox & Snell R Squared value of .083 and the Nagelkerke R Square value of .376 were both very moderate which further suggests that there was a very low impact on the variance within this model. This finding is also supported by the Pearson test results which showed a very moderate positive correlation between the state proficiency assessment and the NAEP assessment at both the basic and proficient NAEP levels. Less than 22% of the schools in the sample met proficient level NAEP scale proficiency while over 97% of the schools met the state accountability proficiency status. Over 95% of the schools met the proficiency status on the basic level NAEP scale.

Although the model is not very useful, several of the variables were statistically significant ($p < .05$). It is important to use caution when interpreting the logistic equation for this model due to the low number of schools that did not meet proficiency. The variables that were not statistically significant ($p > .05$) include NATAMPER, ELLPER, SPECIALEDPER, and NAEPMATHPROFprof . The logistic equation for the Texas 2005 4th grade dataset is: Logit (Y, of meeting mathematics proficiency) = 18.997 (constant) + (-1.529 * DUMMYREV) + (-9.203 * EDPER) + (14520.795 * NATAMPER) + (-16.416 * ASIANPER) + (- 6.421 * BLACKPER) + (-2.397 * HISPANICPER) + (-.855 * DUMMYWHITE) + (.056* ELLPER) + (-.145 SPECIALEDPER), (-.655 * NAEPMATHPROFbasic) + (14.767 * NAEPMATHPROFprof). Listed in the Appendix HH are all of the output tables from the logistic regression for the 2005 Texas 4th grade dataset.

Texas 2005 4th grade reading

An analysis of the 2005 4th grade Texas dataset in reading produced a very moderate positive correlation between the proficient level NAEP scale reading assessment and the state reading accountability assessment, ($r = 0.073$, at the .01 level). There was a moderate positive correlation between the basic level NAEP scale in reading and the Texas reading accountability assessment, ($r= 0.293 \quad p < .01$). These results are consistent with the Texas 2005 4th grade mathematics analysis showing a weak relationship with the NAEP among most of the datasets in Texas. Refer Appendix GG for descriptive statistics and the correlation matrix.

A logistic model was also used to analyze the 2005 Texas 4th grade dataset with the reading state accountability assessment dependent variable. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, the NAEPELAPROFprof level variable was analyzed separately using the stair step method in SPSS. The reason for this was due to the Omnibus Model of Coefficients test which was not statistically significant ($p > .05$), indicating that the model did not provide a goodness of fit. However, when the NAEPELAPROFprof level variable was eliminated from the logistic regression model, the Omnibus Model of Coefficients test was statistically significant ($p < .05$) indicating goodness of fit with the model. Thus the rest of this analysis provides the logistic regression results with without the NAEPELAPROFprof level variable.

The logistic regression results showed that the constant (null) model was correct 96.3% of the time vs. the predicted model which was accurate 96.6%. Thus there was an extremely moderate increase and this model was not useful for predicting the outcome of the school level mathematics proficiency status. Much of this had to do with the fact that less than 6% of the schools in the sample failed to meet the mathematics proficiency status. Although the Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model as it was supported by the Cox & Snell R value of .124 and the Nagelkerke R Square value of .459, the results were not very useful. Similar to the Texas mathematics dataset at the 4th grade level, the values were moderate which further suggest that there was a very low impact on the variance within this model. This finding is also supported by the Pearson test results which showed a very moderate positive correlation between the state proficiency assessment and the NAEP assessment at both the basic level and proficient levels. Less than 12% of the schools in the sample met NAEP proficiency status at the proficient level while over 93%

of the schools met the state accountability proficiency status. Over 60% of the schools met the proficiency status on the NAEP reading assessment when measured against the basic level scale.

Although the model is not very useful, several of the variables were statistically significant (< .05). Similar to the 2005 4th grade Texas mathematics dataset, it is important to use caution when interpreting the logistic equation for this model due to the low number of schools that did not meet proficiency. The variables that were not statistically significant (p > .05) include NATAMPER, ELLPER, and NAEPMATHPROFbasic.

The logistic equation for the Texas 2005 4th grade dataset is: Logit (Y, of meeting state reading proficiency) = 3.634 (constant) + (-.081 * DUMMYREV) + (1.327 * EDPER) + (11802.002 * NATAMPER) + (-16.110 * ASIANPER) + (- 4.256 * BLACKPER) + (-1.198 * HISPANICPER) + (-.415 * DUMMYWHITE) + (.103* ELLPER) + (-.265 SPECIALEDPER) (20.007 * NAEPMATHPROFbasic). Listed in the Appendix HH are all of the output tables from the logistic regression for the 2005 Texas 4th grade reading dataset.

Texas 2005 8th grade mathematics

For the mathematics 2005 8th grade Texas dataset, there was a moderate positive correlation between the proficient level NAEP scale and the state mathematics accountability assessment, ($r = 0.180$, significant at the .01 level). There was also a strong positive correlation for the mathematics basic level NAEP scale and state mathematics accountability assessment ($r = 0.630$, $p < .01$). The results from the 8th grade Texas dataset in mathematics are aligned more closely with the results that were found in California, Michigan and the 2005 North Carolina datasets. They are in contrast to the results from the Texas 4th grade datasets. Refer to Appendix

GG for descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

The Texas 8th grade mathematics state accountability assessment was used as the dependent variable for a logistic regression to analyze the 2005 dataset. The constant (null) model was correct 81.7% of the time vs. the predicted model which was accurate 92.4% of the time. Thus there was an increase of over 10% for predicted model vs. the constant model. This signifies a moderate positive influence on the predicted model. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model as it was also supported by the Cox & Snell R value of 0.436 and the Nagelkerke R Square was 0.709 explaining some of the variance within model. The output tables can be found in Appendix HH.

The logistic regression equation for the schools in the 2005 Texas 8th grade dataset with mathematics as the dependent variable is the following: Logit (Y, predicted probability of meeting the mathematics proficiency status) = 14.264 (constant) + (-.520 * DUMMYREV) + (-5.465 * EDPER) + (-241.700 * NATAMPER) + (41.962 * ASIANPER) + (-9.927 * BLACKPER) + (-7.649 * HISPANICPER) + (-.577 * DUMMYWHITE) + (-.814 * ELLPER) + (-9.152 * SPECIALEDPER) + (2.047 * NAEPMATHPROFbasic) + (14.952 * NAEPMATHPROFprof).

As Table HH15 in Appendix HH indicates, the Wald value was statistically significant in the model for most of the independent variables ($p< .05$) other than NAEPMATHPROFprof ($p > .05$). Even though the NAEPMATHPROFprof variable was not statistically significant both of the NAEP assessment variables had a positive impact on the probability of meeting the mathematics proficiency standard. As mentioned, the Pearson test showed that there was a positive correlation with NAEPREADPROFprof (r = 0.180 at the .01 level) and the NAEPREADPROFbasic variable (r= 0.630 p < .01). These significant correlations provide evidence that the Texas 2005 8th grade state accountability assessment in mathematics and the NAEP assessment had a positive relationship. However, the relationship with the basic level NAEP scale was much stronger indicating that the Texas 8th grade assessment in mathematics is more closely associated with the basic level NAEP scale.

The statistically significant variables that were positively associated with the odds that a school met mathematics proficiency standards were NAEPMATHPROFbasic (B = 2.047, Exp (B) = 7.743) and ASIANPER (B = 41.962, Exp (B) 1.1675). Both of the variables had a positive association with the probability that a school meets proficiency status with most of the datasets in this study.

As displayed in Table HH15 in Appendix HH, most of the variables had a negative association with the probability that a school met the reading proficiency standard, DUMMYREV (B = -.520, Exp (B) = .595), EDPER (B = -5.465, Exp (B) = .004), NATAMPER (B = -241.700, Exp (B) = .000), BLACKPER (B = -9.927, Exp (B) = .000), HISPANICPER, (B = - 7.649, Exp (B) = .001), DUMMYWHITE (B = -.577, Exp (B) = .561), ELLPER (B = -.814, Exp (B) = .443), and SPECIALEDPER (B = - 9.152, Exp (B) = .000). The EDPER variable has been negatively associated with the probability that a school meets proficiency with almost all of the datasets in this study. The moderate negative impact with the DUMMYREV variable is somewhat surprising but also aligned with some of the findings from the other datasets in this study. The negative impact of the SPECIALEDPER and the ELLPER variables are not surprising, however, this finding been inconsistent from state to state and year of assessment with predicting school level success. The DUMMYWHITE variable had a moderate negative relationship which has been somewhat consistent in other datasets and the HISPANICPER variable and the BLACKPER variable had a larger negative impact on the probability that a school would meet the state accountability standards in mathematics. However, all three of those variables had a tendency to fluctuate between the different datasets. Please refer to Appendix HH for the logistic regression output tables.

Texas 2005 8th grade reading

A Pearson Correlation analysis of the reading 2005 8th grade Texas dataset showed that there was a very moderate positive correlation between the two variables for the proficient level NAEP scale in reading and the reading state accountability assessment, ($r = 0.028$ at the .01 level). There was also a moderate positive correlation between the basic level NAEP scale in reading and the Texas accountability assessment in reading, ($r= 0.218$, $p < .01$). This result is not aligned with the other datasets in the study as the association with the NAEP was at a mild level vs. other datasets. However, it is aligned with the findings from the 4th grade 2005 Texas datasets. Refer to Appendix GG for descriptive statistics and the correlation matrix.

The logistic model was used to analyze the 2005 Texas 8th grade dataset with the reading state accountability assessment as the dependent variable. Similar to other datasets in Texas, the NAEPELAPROFbasic level variable and NAEPELAPROFprof level variables were analyzed separately because the Omnibus Model of Coefficients test was not statistically significant ($p > .05$) indicating that the model did not provide a goodness of fit. When the NAEPMATHPROFprof and NAEPMATHPROFbasic variables were eliminated from the logistic regression, the Omnibus Model of Coefficients test was statistically significant ($p < .05$)

indicating goodness of fit with the model. However, there were other problems with this dataset which are discussed below.

The logistic regression results showed that the constant (null) model was correct 98.8% of the time vs. the predicted model which was accurate 100%. Just over 1% of the schools failed to meet the reading proficiency standard which was part of the problem. As a result, the Cox & Snell R value of .124 and the Nagelkerke R Square value of 1.00 were unusual. Similar to the Texas 2005 datasets in mathematics and reading at the 4th grade level, the proficiency standard appears to be weakly associated with the NAEP assessment.

As mentioned in the previous chapter, the Pearson test results also establish a very weak relationship between the NAEP assessment and Texas state accountability assessments which further supports the finding that the Texas assessments (with the exception of the Texas 2005 mathematics results) might not be as rigorous as the NAEP or other state accountability assessments. Less than 7% of the schools in the sample met NAEP proficiency status at the proficient level and 81% met the basic level proficiency on the NAEP which is in contrast to the 98% of the schools that met the state accountability proficiency standards.

The only two variables that had a statistically significant Wald values were DUMMYREV and DUMMYWHITE, however, the numbers were very high and unusual indicating that this model was not very useful. Listed in Appendix HH are all of the output tables from the logistic regression for the 2005 Texas 8th grade reading dataset.

Texas 2007 4th grade mathematics

An analysis of the 2007 4th grade Texas dataset produced results showing a very weak positive correlation between the proficient level NAEP scale mathematics assessment and the mathematics state accountability assessment, ($r = 0.046$, significant at the .01 level). There was an extremely minor negative correlation for the mathematics basic level NAEP scale and state mathematics accountability assessment ($r = -0.008$, $p < .01$). These results are somewhat consistent with the other Texas 4th grade datasets showing a very weak relationship with the NAEP. Refer to Appendix GG for descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

A logistic model was also used to analyze the 2007 Texas 4th grade dataset with the reading state accountability assessment as the dependent variable. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, the NAEPELAPROFbasic level variable and NAEPELAPROFprof level variables were analyzed separately. Like many of the other datasets, this was done because the Omnibus Model of Coefficients test was not statistically significant ($p > .05$) indicating that the model did not provide a goodness of fit. However, when the NAEPMATHPROFprof and NAEPMATHPROFbasic variables were eliminated from the logistic regression model, the Omnibus Model of Coefficients test was statistically significant ($p < .05$). The logistic regression results showed that the constant (null) model was correct 99.4% of the time vs. the predicted model which was accurate 100%. Less than 1% of the schools in the sample failed to meet the mathematics proficiency status on the state accountability assessment. As a result, the Cox & Snell R value of .071 and the Nagelkerke R Square value of 1.00 were unusual. Similar to the Texas 2005 datasets in mathematics and reading at the 4th grade level and the Texas 8th grade reading dataset, the proficiency standard appears to be extremely weak in relationship with the NAEP assessment and other state accountability assessments.

Less than 27% of the schools in the sample met NAEP proficiency status at the proficient level while over 99% of the schools met the state accountability proficiency status. Over 98% of the schools did meet the proficiency status on the NAEP mathematics when measured against the basic level NAEP scale. None of the independent variables had a statistically significant Wald value. Listed in Appendix HH are all of the output tables from the logistic regression for the 2007 Texas 4th grade mathematics dataset.

Texas 2007 4th grade reading

For the 2007 reading 4th grade Texas dataset, there was a very moderate positive correlation between the proficient level NAEP scale reading assessment and the reading state accountability assessment, ($r = 0.075$ at the .01 level). There was a moderate positive correlation between the basic level reading NAEP scale and the Texas reading state accountability assessment, ($r = 0.169 \quad p < .01$). This finding is in alignment with most of the other Texas datasets showing a weak relationship with the NAEP. The only exception to this finding was with the Texas 8th grade 2005 state accountability assessment. Refer to Appendix GG for descriptive statistics and the correlation matrix.

A logistic regression was used to analyze the 4th grade Texas schools in the 2007 dataset with the reading assessment as the dependent variable. The constant (null) model was correct 95.9% of the time vs. the predicted model which was accurate at the same level. Thus there was no increase and the model was not useful for predicting the outcome of the schools reading proficiency status. The fact that less than 5% of the schools in the sample failed to meet the reading proficiency status had an impact on the usefulness of the model. Although the Omnibus Test of Model Coefficients was statistically significant (p < ,05) indicating goodness of fit with the logistic model, the Cox & Snell R value of .078 and the Nagelkerke R Square value of .268 were both very moderate which further suggests that there was a very low impact on the variance within this model. Refer to Appendix HH for the output tables.

Less than 13% of the schools in the sample met NAEP proficiency status at the proficient level and over 66% of the schools met the NAEP proficiency level at the basic scale. This is in contrast to over 95% of the schools met the state accountability proficiency status on the 4th grade reading assessment in Texas. This finding is also supported by the Pearson test results which showed an minor positive correlation between the state proficiency assessment and the NAEP assessment at both the basic level and proficient level NAEP scale.

Although the model is not very useful and some of the values were fairly erratic in the output table, several of the variables were statistically significant (< .05). It is critical to use caution when interpreting the logistic equation for this model due to the low number of schools that did not meet proficiency. The variables that were not statistically significant (p > .05) include DUMMYWHITE, ELLPER, SPECIALEDPER, and NAEPMATHPROFprof .

The logistic equation for the Texas 2007 4th grade dataset is: Logit (Y, of meeting state reading proficiency) = 18.997 (constant) + (-.940 * DUMMYREV) + (-.756 * EDPER) + (717.507 * NATAMPER) + (-16.131 * ASIANPER) + (- 13.919 * BLACKPER) + (-12.288 * HISPANICPER) + (-.032 * DUMMYWHITE) + (.058* ELLPER) + (-.117 SPECIALEDPER), (.304 * NAEPMATHPROFbasic) + (14.146 * NAEPMATHPROFprof). Listed in Appendix HH are all of the output tables from the logistic regression for the 2007 Texas 4th grade dataset in reading.

Texas 2007 8th grade mathematics

A Pearson correlation with the mathematics 2007 8th grade Texas dataset showed that there was a moderate positive correlation between the two variables for the proficient level NAEP scale mathematics assessment and the state mathematics accountability assessment, ($r = 0.119$, significant at the .01 level). There was also a moderate positive correlation for the mathematics basic level NAEP scale and the mathematics state accountability assessment ($r = 0.109$, $p < .01$). This result is not consistent with the other state datasets but it is consistent with most of the other datasets from Texas. Refer to Appendix GG for descriptive statistics and the correlation matrix for the listed proficiency variables including the demographic characteristics and financial variable.

A logistic model was also used to analyze the 2007 Texas 8th grade dataset with the mathematics state accountability assessment dependent variable. The logistic regression was implemented with all of the variables that were described in chapter three as independent variables. However, the NAEPELAPROFbasic level variable and NAEPELAPROFprof level variables were analyzed separately. Similar to many of the other datasets, the Omnibus Model of Coefficients test was not statistically significant ($p > .05$) indicating that the model did not provide for goodness of fit. However, when the NAEPELAPROFprof level variable was eliminated from the logistic regression model, the Omnibus Model of Coefficients test was statistically significant ($p < .05$) indicating goodness of fit with the model. Thus the rest of this analysis provides the logistic regression results with without the NAEPELAPROFprof level variable.

The logistic regression results showed that the constant (null) model was correct 93.4% of the time vs. the predicted model which was accurate 93.5%. Most of the logistic models among the Texas datasets have not been valuable at predicting the odds that a school would meet state proficiency status. However, this consistent result among the Texas datasets further supports the finding that the Texas state accountability assessment is less rigorous vs. the NAEP assessment and other states in the sample. Thus the fact that there has been a lack of useful statistical results with the logistic models in Texas, this becomes an interesting finding in itself. The Pearson results also support those findings which show a very weak relationship with the NAEP assessment.

Although the Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model as was supported by the Cox & Snell R value of .165 and the Nagelkerke R Square value of .429, the statistical usefulness of this model is limited. Less than 17% of the schools in the sample met NAEP proficiency status at the proficient level while over 92% of the schools met the state accountability proficiency status. Over 91% of the schools met the proficiency status on the NAEP mathematics when measured against the basic level NAEP scale. This could suggest that the basic level NAEP scale is more closely aligned with the Texas state accountability assessment but the problem is that the low correlation with the Pearson test suggests that the same schools were not consistent at meeting proficiency with the basic level NAEP scale and state accountability assessment. Another problem has to do with the low number of schools that failed to meet proficiency which makes it much more difficult to make confident conclusions.

Although the logistic model is not statistically useful, all of the variables were statistically significant ($< .05$). Similar to many of the other Texas datasets, it is important to use caution when interpreting the logistic equation for this model due to the low number of schools that did not meet proficiency. The logistic equation for the Texas 2007 8^{th} grade mathematics dataset is: Logit (Y, of meeting state mathematics proficiency) = 35.742 (constant) + (-.413 * DUMMYREV) + (-1.167 * EDPER) + (-85.062 * NATAMPER) + (38.943 * ASIANPER) + (-30.602 * BLACKPER) + (-29.326 * HISPANICPER) + (-1.557 * DUMMYWHITE) + (-2.003* ELLPER) + (-6.877 SPECIALEDPER), (-.745 * NAEPMATHPROFbasic). Listed in Appendix HH are all of the output tables from the logistic regression for the 2007 Texas 8th grade mathematics dataset.

Texas 2007 8th grade reading

With the 2007 8th grade Texas dataset in reading, it was not possible to complete a correlation between the NAEP and state accountability assessment in reading. The reason is that the dependent variable output was constant as all of the schools in the sample met the state proficiency status in reading. This also made it impossible to complete a logistic regression analysis. This again supports the earlier conclusion that the Texas state accountability assessments might not be as rigorous as the NAEP assessment and some of the other state assessments in this sample.

While 100% of the schools in the sample met the Texas state proficiency standard in reading, less than 7% of the schools in the sample met the NAEP proficiency status at the proficient level, and just over 69% of the schools met the proficiency status on the basic level NAEP scale reading assessment. This might suggest that the basic level NAEP scale might be more closely aligned with the Texas state accountability assessment vs. the proficient level NAEP scale. The problem is that this can not be verified via statistical analysis with the Texas 2007 8th grade reading dataset. There is no doubt that Texas has the weakest relationship among the four states in comparison to the NAEP assessment.

Summary for research question two

There were several consistent findings with the analysis of the data to answer research question two. The Pearson Correlation analysis consistently produced a strong negative correlation with the EDPER variable and state accountability proficiency status. The logistic regression showed that the EDPER variable provided a negative influence on the probability that a school met proficiency which was consistent with almost all of the datasets. There were a few exceptions including the Michigan 2007 4^{th} grade mathematics and the Michigan 4^{th} grade reading datasets. However, the output values produced extremely high odds ratio values indicating instability and the data was not very useful. The same was true for the 2007 8^{th} grade NC reading dataset. However, the output values did not produce statistically significant results ($p > .05$) with that dataset.

The DUMMYREV variable had either a positive influence on the odds that a school met proficiency or in the case of a few of the datasets, there was a mild negative influence. Most of the other variables within the analysis of this research question were not consistent as the results fluctuated between the datasets. However, the ASIANPER and WHITEPER variables generally had a positive impact on the state accountability assessment results. The ELLPER and SPECIALEDPER variables were inconsistent with their influence among the datasets. The NATAMPER, BLACKPER, and HISPANICPER variables also produced inconsistent results between the different state datasets. However, some of the variables were consistent within a given state but not between the states. For example, the BLACKPER variable had a strong negative correlation with school level proficiency in Michigan and it generally had a negative impact on the probability of a school meeting proficiency standards with the logistic regression.

This could be explained by the different sizes in subgroup populations among the states. Michigan had a higher BLACKPER vs. the other states in this study. Other than DUMMYREV and EDPER, most of the results were either inconsistent or at times produced unusual odds ratio values. The NAEP variables were often left out of the analysis because goodness of fit was not found with the analysis of many of the models when the NAEP variable was included.

With the California datasets, the results were generally aligned with the Michigan results and most of the North Carolina datasets showing a closer association between the state accountability assessment and the basic level NAEP scale vs. the proficient level NAEP scale and state accountability assessment. Most of the logistic regression analysis in the California datasets produced the greatest increase between the constant (null) model and the predicted model. The reason for this could be do to the greater balance among the schools that met and did not meet proficiency. However, sample size could also have an impact as the state of California had the largest school samples.

It is noteworthy that the sample from California and Texas were very similar with regards to the total number of schools analyzed but the statistical analysis produced inconsistent results between the two states. The California regression results were much more statistically useful than the Texas results. This established a finding that the Texas proficiency assessments had the weakest relationship with the NAEP among the four states in the sample. Based on the analysis of the datasets, it is probable that the Texas state accountability assessment was less rigorous than the other three states in the sample.

The Michigan statistical analysis with the mathematics and ELA assessments as the dependent variables produced results that were consistent throughout the state. There were a high number of schools that met state accountability assessment proficiency in each of the Michigan datasets. The logistic regression analysis had the potential to provide rich and robust results; however, due to the high number of schools that met proficiency, the logistic regression results were less useful vs. the findings in California and North Carolina.

The logistic results in Michigan produced consistent findings with the EDPER variable which had a negative impact on the chance that a school would meet proficiency standards on the state accountability assessments. The logistic regression also showed that educational resources had a positive impact on school level ability to meet proficiency in Michigan. The Pearson analysis was a very valuable method for analyzing the data in Michigan as it produced results that are aligned with literature in the field which shows that the basic level NAEP scale proficiency is more closely associated with results on the state accountability assessments (Bandeira de Mello, et al. 2009; Fuller et al. 2007; Gewertz, 2010). This finding was consistent with all Michigan datasets as a very strong relationship was found between the two variables. The results between the basic level NAEP scale and the state accountability assessment in Michigan were the strongest among the states in the sample.

The North Carolina analysis produced some of the most intriguing results with the analysis of the second research question. All of the 2005 NC datasets were aligned with the results from the Michigan datasets showing a greater positive correlation with the basic level NAEP scale and the state accountability assessment vs. the proficient level NAEP scale and the state accountability assessment.[25] The DUMMYREV variable tended to be positively associated with the probability that a school met state accountability proficiency with the logistic regression analysis.

---

[25] The correlation relationship with the basic level NAEP was slightly higher in Michigan but still at a consistent level with the NC 2005 datasets.

With the 2007 results, the NC mathematics assessments at both the 4th and 8th grade provided the closest correlation with the proficient level NAEP scale. In fact, these were the only datasets that produced a greater positive relationship with the proficient level NAEP scale vs. the basic level NAEP scale. The logistic results also showed that the NAEPMATHPROFprof variable in 2007 at both the 4th and 8th grade level produced statistically significant results with a positive impact on the probability that a school met state accountability assessment proficiency. In contrast, the NAEPMATHPROFbasic variable did not produce statistically significant results in either grade in NC. This is an interesting finding suggesting a greater level of rigor with the 2007 NC mathematics assessments.

Another interesting finding is that NC changed the reading assessment and the cut score proficiency trajectory in 2007 resulting in a massive increase in the amount of schools that met reading proficiency at both the 4th and 8th grade during the 2007 school year. The result was that most of the schools met proficiency and the Pearson results showed that there was a very weak relationship. This example provides support for the literature review findings which established that by changing the cut score, the state can give a false impression of either growth or decline with student achievement.

Many states are now looking to change the cut scores across the nation to be aligned with what they call college readiness. The state of Michigan is planning to change the cut scores for state accountability proficiency on the MEAP during the 2011 school year. However, they are not going to change the assessment or make any other modifications to the test that would assess college readiness. As a result, it can be predicted that many more schools will fail to meet AYP giving the false impression of a decrease in academic achievement when the results are published to the public. The manner in which the states are playing with politics and changing cut scores has the potential for great damage to the public education system in the long run if no other reforms are included which promote an increased student achievement through actual gains vs. statistical manipulation.

The Texas state accountability assessment in all datasets (other than the 2005 8th grade mathematics dataset) established a very weak correlation between the NAEP and the state accountability assessments. The 8th grade 2005 mathematics assessment was the only one that produced results which were aligned with the other datasets in this study. Statistical analysis with many of the Texas datasets were not possible because either all of the schools or most of the schools met state proficiency. Thus neither the Pearson correlation nor the logistic regression was very helpful in statistically analyzing the datasets. However, the weak relationship with the NAEP supported a finding that the Texas state accountability assessments might not be as rigorous as either the NAEP and/or other state datasets in this study.

Research Question Three

3. Are the demographic (categorical) characteristics and educational resources significant predictors of the schools that make AYP and fail to make AYP in the sample states?

The hypothesis for research question three (as stated in chapter three) is that there would be a relationship between school level AYP proficiency status and the higher levels of demographic subgroup populations. It was also predicted that school district revenue would have an impact on the AYP proficiency status of schools. A logistic regression analysis was used in order to answer research question three and test the predicted hypothesis. The AYP status of schools is the dependent variable for all tests that were analyzed in this section.

The logistic regression analysis was employed to predict the probability that a school would meet the AYP proficiency status. As identified in chapter three, the predictor variables were the overall per pupil funding that the school received, and the percentage of students in the following subgroups within the school; ED, ELL, special education, Asian/Pacific, Black/African American, Hispanic, Native American and White. All of the school samples were weighted by the school enrollment.

The logistic regression equation for research question three is the following: Logit (Y, of meeting AYP proficiency) = α (constant) + (β_1 * DUMMYREV) + (β_2 * EDPER) + (β_3 * NATAMPER) + (β_4 * ASIANPER) + (β_5 * BLACKPER) + (β_6 * HISPANICPER) + (β_7 * WHITEPER) + (β_8 * ELLPER) + (β_9 * SPECIALEDPER). This equation was used for each year and state assessed. The Omnibus Test of Model Coefficients was run on each logistic regression in order to test for goodness of fit along with the Cox & Snell R and the Nagelkerke R Square in order to test the variance within the model.

California 2005 4th grade

The logistic regression results from the 2005 4th grade sample showed that the constant (null) model was correct 53.7% of the time vs. the predicted model which was accurate 80% of the time. There was a very large increase of over 27% accuracy for the predicted model vs. the constant model. This is a very good indication that the model was valuable in determining a schools' probability of meeting AYP. The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model as indicated by the Cox & Snell R value of .456 and the Nagelkerke R Square value of .610, which accounts for some of the variance within the model. Please refer to Appendix II for the logistic regression output tables.

The logistic equation for the California 2005 4th grade dataset is: Logit (Y, of meeting AYP proficiency) = 3.388 (constant) + (.489 * DUMMYREV) + (-2.900 * EDPER) + (-17.560 * NATAMPER) + (7.068 * ASIANPER) + (-5.089 * BLACKPER) + (-2.877 * HISPANICPER) + (.446 * WHITEPER) + (-1.858 * ELLPER) + (2.060 * SPECIALEDPER).

As table II5 in Appendix II indicates, the Wald value was significant in the model for all of the independent variables ($p < .05$) other than WHITEPER ($p > .05$). The statistically significant variables that were positively associated with the odds that a school met AYP proficiency standards were DUMMYREV (B = .489, Exp (B) = 1.631), ASIANPER (B = 7.068, Exp (B) = 1173.966), and SPECIALEDPER (B = 2.060, Exp (B) = 7.850). The variables that had a negative impact on the AYP outcome include: EDPER (B = -2.900, Exp (B) = .055), NATAMPER (B = -17.560, Exp (B) = .000), BLACKPER, (B =- 5.089, Exp (B) = .006), HISPANICPER (B = -2.877, Exp (B) = .056), and ELLPER (B = - 1.858, Exp (B) = .156).

It was not surprising that the EDPER variable was negatively associated with the probability that a school met AYP proficiency as this finding has been consistent with the findings in the previous sections of this study. The DUMMYREV variable was mildly positive in association with the probability that a school met AYP, however, this is in contrast to the negative correlation that occurred when the Pearson Correlation test [26] was used in research question one, TOTALREVENUE ($r = -.187$, $p < .01$). As previously mentioned, while the Pearson correlation test provides information on *whether a relationship exists between existing variables by offering a correlation coefficient it does not indicate the direction of the correlation.* The logistic regression is a more powerful tool of analysis for predicting the probability that a school will meet the AYP proficiency status. In addition, whereas the actual dollar amount was used for the Pearson test, a dummy variable was used to represent a specific range of dollar amount[27] for the logistic regression allowing for a deeper understanding of the influence of income levels on proficiency results. Regardless, the output values in each case provided a moderate influence vs. a substantial influence.

It was also found that the ASIANPER variable was positively associated with the chance that a school met AYP which corresponds to an earlier finding in this study. However, the positive impact of the SPECIALEDPER variable and the negative impact with the ELLPER

---

[26] The Pearson test was used in research question one and it measures the same dependent and independent subgroup variables from research question three. The difference is that research question one also analyzed the NAEP as an independent variable while research question three focuses only on AYP in relation to the subgroup independent variables. Refer to Appendix W for the Pearson Correlation results in California.

[27] As it was described in chapter 3, for the logistic regression, this variable was changed to DUMMYREV which is a range of revenue allocated per student. Please refer to that chapter for detailed information regarding this variable.

variable have not been consistent from state to state and year of assessment with predicting school level success on the state accountability assessments. Each of the dataset results established different findings with respect to these two variables. Listed in the Appendix II is Figure II1 that represents the predicated probability of the school AYP proficiency status.

California 2005 8th Grade

A logistic regression was used to evaluate the California schools that met AYP proficiency status in the 2005 8th grade sample. The constant (null) model was correct 71.1% of the time vs. the predicted model which was accurate 86.4% of the time. There was a large increase of more than 15% accuracy for the predicted model vs. the constant model. Refer to Appendix II for the observed and predicted frequencies of the constant model and the predicted model. The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit ($p < ,05$) with the logistic model as indicated by the Cox & Snell R value of .438 and the Nagelkerke R Square value of .626, which accounts for some of the variance within the model.

The logistic equation for the California 2005 8th grade dataset is: Logit (Y, of meeting AYP proficiency) = -7.14 (constant) + (-.093 * DUMMYREV) + (-4.357 * EDPER) + (8.610 * NATAMPER) + (6.401 * ASIANPER) + (-.007 * BLACKPER) + (1.057 * HISPANICPER) + (4.946 * WHITEPER) + (-.273 * ELLPER) + (-7.710 * SPECIALEDPER).

As table II6 in Appendix II indicates, the Wald value was significant in the model for all of the independent variables ($p < .05$) other than BLACKPER ($p > .05$). The statistically significant variables that were positively associated with the odds that a school met AYP proficiency standards were NATAMPER (B = 8.610, Exp (B) = 5485.244), ASIANPER (B = 6.401, Exp (B) = 602.216), HISPANICPER (B = 1.057, Exp (B) = 2.877), and WHITEPER (B =

4.946, Exp (B) = 140.582). The variables that had negative impact on the AYP outcome include DUMMYREV (B = -.093 , Exp (B) = .911), EDPER (B = -4.357, Exp (B) = .013), ELLPER(B = -.273, Exp (B) = .761), and SPECIALEDPER (B = - 7.710, Exp (B) = .000).

The Exp (B) value on the NATAMPER was extremely large, indicating that the results were unstable and the variable required further analysis.[28] The California sample has very few schools with high percentages of Native American students. The population mean among the schools in the sample was less than 1%. Thus the NATAMPER variable had very little impact on the AYP results among the 2005 8th Grade California dataset due to the small NATAMPER population within the sample.

The negative impact of the EDPER variable was predictable as was the positive influence of the ASIANPER and WHITEPER variables. It was surprising that the HISPANICPER variable had a positive impact on the probability that a school met AYP proficiency as this was in contrast to the 4th grade 2005 California dataset. The SPECIALEDPER variable had a negative impact in contrast to the 4th grade dataset. The ELLPER variable had a consistent negative impact among both of the 2005 datasets. The DUMMYREV had a negative impact but that impact was extremely mild. Listed in Appendix II is figure II2 that represents the predicated probability of the schools AYP proficiency status.

California 2007 4th Grade

A logistic regression was used to evaluate the California schools in the 2007 4th grade sample that met AYP proficiency status. The constant (null) model was correct 68.3% of the time vs. the predicted model which was accurate 80.3% of the time. Thus a large increase was found at over 12% accuracy for the predicted model vs. the constant model. The Omnibus Test

---

[28] Extremely high odds ratio values can be a result of a low number within the independent variable.

of Model Coefficients was statistically significant indicating goodness of fit with the logistic model. This is supported by the Cox & Snell R value of .287 and the Nagelkerke R Square value of .402, which accounts for some of the variance within the model. Refer to Appendix II for the regression output tables.

The logistic equation for the California 2007 4th grade dataset is: Logit (Y, of meeting AYP proficiency) = 8.526 (constant) + (.807 * DUMMYREV) + (.644* EDPER) + (-111.738 * NATAMPER) + (.069 * ASIANPER) + (-10.653 * BLACKPER) + (-9.012 * HISPANICPER) + (-6.528 * WHITEPER) + (-4.068 * ELLPER) + (-3.548 * SPECIALEDPER).

As table II15 in Appendix II indicates, the Wald value was significant in the model for all of the independent variables($p < .05$) other than ASIANPER. The statistically significant variables that were positively associated with the odds that a school met AYP proficiency standards were DUMMYREV (B = .807, Exp (B) = 2.241), and EDPER (B = .644, Exp (B) = 1.904). The variables that had negative impact on the AYP outcome include NATAMPER (B = -111.738 , Exp (B) = .000), BLACKPER (B = -10.653, Exp (B) = .000), HISPANICPER (B = -9.012, Exp (B) = .761), WHITEPER (B = -6.528, Exp (B) = .001), ELLPER (B = -4.068, Exp (B) = .017), and SPECIALEDPER (B = - 3.548, Exp (B) = .029).

The mild positive impact that the EDPER variable had on the results was unusual in relation to the other findings in this study. In fact, this is only the 3rd test among all of the analysis in all research questions where there EDPER variable had a positive impact on school proficiency. However, the positive impact with those datasets produced unusual and unstable reading outputs.[29]

---

[29] The EDPER variable had a consistent negative impact with the exception of the California 2007 4th grade dataset with an extremely moderate positive impact along with the 2005 8th Grade Michigan results and the 2007 8th Grade

The results from all other variables were also somewhat unusual and in contrast to the other datasets from California. The negative impact of the ELLPER is consistent with the other two 2005 California datasets. The results from this dataset should be viewed with caution due to the inconsistencies that were found with the variables. Refer to Appendix II for the tables and figure from the logistic regression.

California 2007 8th Grade

A logistic regression was also used to evaluate the California schools in the 2007 8th grade sample that met AYP proficiency status. The constant (null) model was correct 73.6% of the time vs. the predicted model which was accurate 85.4% of the time. Thus a large increase was found at over 11% accuracy for the predicted model vs. the constant model. The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model as indicated by the Cox & Snell R value of .349 and the Nagelkerke R Square value of .510, which accounts for some of the variance within the model. Refer to Appendix II for the logistic regression output tables.

The logistic equation for the California 2007 8th grade dataset is: Logit (Y, of meeting AYP proficiency) = -2.294 (constant) + (-.445 * DUMMYREV) + (-1.784* EDPER) + (10.808 * NATAMPER) + (7.488 * ASIANPER) + (-1.455 * BLACKPER) + (2.765 * HISPANICPER) + (6.423 * WHITEPER) + (.180 * ELLPER) + (-6.010 * SPECIALEDPER).

As table II20 in Appendix II indicates, the Wald value was significant in the model for all of the independent variables (p< .05). The statistically significant variables that were positively associated with the odds that a school met AYP proficiency standards were:

Michigan results. The results from Michigan in the two identified datasets produced extremely high odds ratio values and a low increase in the predicted probability.

NATAMPER (B = 10.808, Exp (B) = 49401.423), ASIANPER (B = 7.488, Exp (B) = 1787.022), HISPANICPER (B = 2.765, Exp (B) = 15.876), WHITEPER (B = 6.423, Exp (B) = 616.049), and ELLPER (B = .180, Exp (B) = 1.197).

The variables that had a negative impact on the AYP outcome include: DUMMYREV (B = -.445, Exp (B) = .641), EDPER (B = -1.784, Exp (B) = .168), BLACKPER (B = -1.455, Exp (B) = .233), and SPECIALEDPER (B = -6.010, Exp (B) = .002). The negative impact with the EDPER variable was predictable and it also supports the theory that the mild positive impact with the EDPER variable in the 2005 dataset was likely an anomaly. The mild negative impact with the DUMMYREV variable is somewhat surprising but not necessarily inconsistent with other results in this study. The DUMMYREV variable has generally had either a positive association with the Pearson correlation test or a positive impact with the logistic regression test. However, the mild negative impact has been found with some of the other datasets and tests in this study.

Although there was a positive impact with the NATAMPER variable, the Exp (B) value was extremely large, indicating unstable results. As previously mentioned, the California schools have very few schools with high percentages of Native American Students. The population mean among the schools in the sample was less than 1%. Thus the NATAMPER variable had very little impact on the AYP results among the 2007 8th Grade California dataset due to the small NATAMPER population within the sample.[30] The positive impact with the ELLPER variable and the negative impact with the SPECIALEDPER variable were not consistent among the California datasets. However, this finding was in support of previous conclusions that both variables had an inconsistent impact on school level proficiency.

---

[30] Extremely high odds ratio values can be a result of a low number within the independent variable.

It is interesting that the logistic regression with all of the California datasets produced an increase in the constant model accuracy vs. the predicated model accuracy that was either large or very large (between 10% to 27%). This indicates that the logistic models in California were useful in predicting school level success as measured by the state accountability assessments. However, when analyzing all four datasets, it was interesting that very few of the subgroup variables were consistent with their impact on the school AYP status. The EDPER variable and DUMMYREV variable provided the most consistent results. The ELLPER variable was consistent in 3 of the 4 datasets in California. Listed in Appendix II is a table that represents the predicated probability of the schools AYP proficiency status.

Michigan 2005 4th grade

A logistic regression was used to analyze the Michigan 2005 4th grade dataset with the state AYP status as the dependent variable. The model was not statistically useful because none of the independent variables were statistically significant and there was no difference between the predicted probabilities of the constant (null) model which was 99.4% accurate vs. the logistic model at 100% accuracy. The reason for this had to do with the high number of schools that met AYP proficiency status. Less than 1% of the schools in the sample did not meet AYP proficiency in Michigan during the 2005 school year. Thus an analysis of the impact of the independent variables (subgroup populations) was not possible. However, since the sample of schools was selected randomly through stratified random sampling procedures that were implemented by NCES, this finding suggests that the 2005 AYP standard in Michigan was not very rigorous. This is not surprising due to the line of trajectory in Michigan which was reported earlier in the chapter. The 2005 percentage of students that need to meet proficiency in order for the school to meet AYP was 56% in mathematics and 48% in ELA. Please refer to Appendix JJ for the logistic regression output tables.

The logistic equation for the Michigan 2005 4^{th} grade dataset is: Logit (Y, of meeting AYP proficiency) = -4993.928(constant) + (91.817 * DUMMYREV) + (-302.215 * EDPER) + (4670.965 * NATAMPER) + (3998.331 * ASIANPER) + (5036.939 * BLACKPER) + (10730.701 * HISPANICPER) + (5314.615 * WHITEPER) + (-609.437 * ELLPER) + (-185.594 * SPECIALEDPER).

Michigan 2005 8^{th} Grade

The Michigan 2005 8^{th} grade dataset was analyzed using a logistic regression with state AYP status as the dependent variable. Similar to the 4^{th} grade 2005 Michigan sample, the model was not statistically useful because only three of the independent variables; EDPER, ELLPER, and SPECIALEDPER, were statistically significant ($p < .05$). However, even the statistically significant variables did not do a good job at explaining the impact on the AYP status of schools in the database. The constant (null) model was 93.5% accurate producing a higher level of accuracy vs. the logistic model at 91.9% accuracy. There were a high number of schools that met AYP proficiency status similar to the 4^{th} grade 2005 dataset in Michigan. Less than 7% of the schools in the sample failed to meet AYP proficiency. An analysis of the impact of the independent variables (subgroup populations) was not possible nor was it statistically useful. This further suggests that the 2005 AYP standard in Michigan was not very rigorous in comparison to the California dataset. Please refer to Appendix JJ for the output tables.

The logistic equation for the Michigan 2005 4^{th} grade dataset is: Logit (Y, of meeting AYP proficiency) = 6334.298 (constant) + (.087 * DUMMYREV) + (5.997 * EDPER) + (-

6307.690 * NATAMPER) + (-5783.507 * ASIANPER) + (-6338.654 * BLACKPER) + (-6379.861 * HISPANICPER) + (-6333.128 * WHITEPER) + (3.121 * ELLPER) + (-2.495 * SPECIALEDPER).

Michigan 2007 4th Grade

A logistic regression was used to analyze the Michigan 2007 4th grade dataset with the state AYP status as the dependent variable. While all of the independent variables were statistically significant ($p < .05$), this model was once again not statistically useful with analysis of the sample because a high percentage of schools met the state AYP proficiency status. Less than 6% of the schools in the sample did not meet AYP proficiency in Michigan during the 2007 school year in the 4th grade sample. There was an extremely moderate increase from the constant (null) model which was 94.1% accurate vs. the predicted model at 95.4% accuracy. Thus an analysis of the impact of the independent variables (subgroup populations) was not very useful when analyzing all of the logistic regression output tables. The results from the 2007 4th grade Michigan dataset are consistent with the results from the 2005 datasets at the 4th and 8th grade level suggesting a less rigorous AYP standard vs. the state of California.

The logistic equation for the Michigan 2007 4th grade dataset is: Logit (Y, of meeting AYP proficiency) = 84.724 (constant) + (-1.356 * DUMMYREV) + (-9.701 * EDPER) + (-80.485 * NATAMPER) + (-67.645 * ASIANPER) + (-71.802 * BLACKPER) + (-44.632 * HISPANICPER) + (-75.542 * WHITEPER) + (-24.244 * ELLPER) + (8.650 * SPECIALEDPER).

Michigan 2007 8th Grade

The Michigan 2007 8th grade dataset was analyzed using a logistic regression with state AYP status as the dependent variable. However, once again, while all of the independent variables were statistically significant ($p < .05$), this model was very limited in analyzing the impact of the independent variables on state AYP. Less than 7% of the schools in the sample did not meet AYP proficiency in the 2007 8th grade Michigan dataset. There was an extremely moderate increase from the constant (null) model which was 93.2% accurate was a higher level of accuracy vs. the logistic model at 97% accuracy. Thus an analysis of the impact of the independent variables (subgroup populations) was not statistically useful when analyzing all of the logistic regression output tables. Please refer to Appendix JJ for the output tables.

The logistic equation for the Michigan 2007 8th grade dataset is: Logit (Y, of meeting AYP proficiency) = 53.964 (constant) + (.555 * DUMMYREV) + (1.264 * EDPER) + (233.332 * NATAMPER) + (65.618 * ASIANPER) + (-58.474 * BLACKPER) + (-56.374 * HISPANICPER) + (-50.767 * WHITEPER) + (-7.751* ELLPER) + (-18.583 * SPECIALEDPER).

North Carolina 2005 4th grade

A logistic regression was used to analyze the North Carolina schools in the 2005 4th grade sample. The constant (null) model was correct 61.9% of the time vs. the predicted model which was accurate 70.9% of the time. Thus a moderate increase of more than 9% accuracy for the predicted model vs. the constant model occurred. This is an indication that the model was somewhat valuable in determining a schools' probability of meeting AYP. The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model as indicated by the Cox & Snell R value of .229 and the Nagelkerke R Square value of .312 which accounts for some of the variance within the model. Please refer to Appendix KK for the SPSS output tables for the logistic regression.

The logistic equation for the North Carolina 2005 4th grade dataset is: Logit (Y, of meeting AYP proficiency) = -6.077 (constant) + (-.041 * DUMMYREV) + (-.382 * EDPER) + (-59.362 * NATAMPER) + (12.149 * ASIANPER) + (4.416 * BLACKPER) + (6.833 * HISPANICPER) + (1.196 * DUMMYWHITE) + (-4.180 * ELLPER) + (-1.101 * SPECIALEDPER).

As table KK5 in Appendix KK indicates, the Wald value was significant in the model for all of the independent variables ($p < .05$). The statistically significant variables that were positively associated with the odds that a school met AYP proficiency standards were: ASIANPER (B = 12.149, Exp (B) = 1888978.973), (BLACKPER (B = 4.416, Exp (B) = 82.748), HISPANICPER (B = 6.833 , Exp (B) = 928.177), and DUMMYWHITE (B = 1.196, Exp (B) = 3.307). The variables that had negative impact on the AYP outcome include: DUMMYREV (B = -.041 , Exp (B) = .959), EDPER (B = -.382 - , Exp (B) = .682), NATAMPER (B = -59.362 , Exp (B) = .000), ELLPER (B =- 4.180, Exp (B) = .015), and SPECIALEDPER (B = - 1.101, Exp (B) = .332).

It was not very surprising that the EDPER variable had a negative impact on the probability that a school met AYP proficiency as this finding has been consistent with the findings in the previous sections of this study. The DUMMYREV variable had an extremely moderate negative impact on the probability that a school met AYP which is somewhat consistent with results in other datasets. It was also found that the ASIANPER variable had a positive impact on the odds that a school met AYP which corresponds to an earlier finding in this study. However, the Exp (B) value was extremely high indicating a problem with this variable which can be explained by the extremely low percentage of Asian students in North Carolina which is under 3% in total.

The negative impact that the NATAMPER had on AYP is consistent with other findings in this study. However, the positive impact of the BLACKPER and HISPANICPER variables with predicting school level proficiency was inconsistent with all datasets and according to all research questions in this study. Listed in Appendix KK are the output tables and figures from the logistic regression analysis.

North Carolina 2005 8th grade

A logistic regression was used to analyze the North Carolina schools in the 2005 8th grade sample. The constant (null) model was correct 75.5% of the time vs. the predicted model which was accurate 83.9% of the time. Thus a moderate 8.4% increase in the accuracy for the predicted model vs. the constant model. This is an indication that the model had a slight value in determining a schools' probability of meeting AYP. The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model as indicated by the Cox & Snell R value of .209 and the Nagelkerke R Square value of .311 which accounts for some of the variance within the model. These values are consistent with the moderate increase in the accuracy level of the predicted model vs. the null model. Please refer to Appendix KK for the logistic regression output tables.

The logistic equation for the North Carolina 2005 8th grade dataset is: Logit (Y, of meeting AYP proficiency) = 2.051(constant) + (.890 * DUMMYREV) + (-2.988 * EDPER) + (5.821 * NATAMPER) + (6.902 * ASIANPER) + (-3.458 * BLACKPER) + (-3.492 *

HISPANICPER) + (-.249 * DUMMYWHITE) + (.263 * ELLPER) + (-6.133 * SPECIALEDPER).

As Table KK10 in Appendix KK indicates, the Wald value was statistically significant in the model for all of the independent variables ($p < .05$) other than ELLPER ($p > .05$). The statistically significant variables that were positively associated with the odds that a school met AYP proficiency standards were: DUMMYREV (B = .890, Exp (B) = 2.435), NATAMPER (B = 5.821, Exp (B) = 337.433), ASIANPER (B = 6.902, Exp (B) = 994.336), and SPECIALEDPER (B = 2.060, Exp (B) = 7.850). The variables that had a negative impact on the AYP outcome include EDPER (B = -2.988, Exp (B) = .050), BLACKPER (B =- 3.458, Exp (B) = .031), HISPANICPER (B = -3.492, Exp (B) = .030), DUMMYWHITE (B = -.249, Exp (B) = .780), and SPECIALEDPER (B = - 6.133, Exp (B) = .002).

It was not surprising that the EDPER variable had a negative impact on the probability that a school met AYP proficiency as this finding continues to be consistent with the findings in the previous sections of this study. The DUMMYREV variable was positively associated with the probability that a school met AYP along with NATAMPER and ASIANPER. The NATAMPER variable was inconsistent in predicting the probability of a schools AYP status and it tends to fluctuate among the different datasets in this study. The low number of NATAMPER students among the states in this sample influences the inconsistent findings.

The positive association with the ASIANPER has been consistent in NC. The SPECIALPER variable had a negative association, which is in contract to the positive association in the 2005 4th grade North Carolina dataset. This variable continues to be unpredictable among the different datasets in this study. The BLACKPER, HISPANICPER, and WHITEPER variables had a negative impact, which is in contrast to the positive impact with the

2005 4th grade North Carolina dataset. The WHITEPER variable was moderately negative in value and thus not a substantial impact on the AYP outcome. These variables also tend to fluctuate from positive to negative between the different datasets. Listed in Appendix KK is Figure KK2 that represents the predicated probability of the school AYP proficiency status.

North Carolina 2007 4th grade

A logistic regression was used to evaluate the North Carolina schools in the 2007 4th grade sample that met AYP proficiency status. The constant (null) model was correct 51.6% of the time vs. the predicted model which was accurate 72.6% of the time. This represents a large increase of 21% accuracy for the predicted model vs. the constant model. The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model. It is supported by the Cox & Snell R value of .269 and the Nagelkerke R Square value of .359 which accounts for some of the variance within the model. These values are consistent with the moderate increase in the accuracy level of the predicted model vs. the null model.

The logistic equation for the North Carolina 2007 4th grade dataset is: Logit (Y, of meeting AYP proficiency) = 5.966 (constant) + (-.206 * DUMMYREV) + (-2.248* EDPER) + (-4.564 * NATAMPER) + (.009 * ASIANPER) + (-6.228 * BLACKPER) + (.678 * HISPANICPER) + (-3.084 * WHITEPER) + (-6.259 * ELLPER) + (-4.193 * SPECIALEDPER).

As Table KK15 in Appendix KK indicates, the Wald value was significant in the model for most of the independent variables ($p < .05$) other than, HISAPANICPER and ASIANPER ($p > .05$). There was an absence of statistically significant variables that were positively associated with the odds that a school met AYP proficiency standards. The variables that had a negative impact on the AYP outcome include: DUMMYREV (B = -.206 , Exp (B) = .814), EDPER (B = -2.248, Exp (B) = .106), NATAMPER (B = -4.564, Exp (B) = .010), BLACKPER (B = -6.228, Exp (B) = .002), WHITEPER (B = -3.084, Exp (B) = .046), ELLPER (B = -6.259, Exp (B) = .002), and SPECIALEDPER (B = -4.193, Exp (B) = .015). The negative impact of the EDPER variable is both predictable and consistent. The slight negative impact of the DUMMYREV variable is also consistent with the other datasets which establish either a positive or slightly negative impact. The negative impact of the other variables produced results showing a lack of consistency among the other datasets in this study.

North Carolina 2007 8th grade

A logistic regression was used to evaluate the North Carolina schools in the 2007 8th grade sample that met AYP proficiency status. The constant (null) model was correct 82.2% of the time vs. the predicted model which was accurate 84.4% of the time. Thus an extremely moderate increase was found of over 2% accuracy for the predicted model vs. the constant model. When looking at the logistic formula and the statistically significant variables, a major consideration is the low level of increase with the predicted model and thus the impact of the independent variables was not as great as in other state datasets. The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model as indicated by the Cox & Snell R value of .165 and the Nagelkerke R Square value of .272 which accounts for some of the variance within the model. These values are consistent with the moderate increase in the accuracy level of the predicted model vs. the null model. Refer to Appendix KK for the logistic regression output tables.

The logistic equation for the North Carolina 2007 8th grade dataset is: Logit (Y, of meeting AYP proficiency) = -2.411 (constant) + (.309 * DUMMYREV) + (-1.114* EDPER) +

(-5.560 * NATAMPER) + (11.399 * ASIANPER) + (-5.560 * BLACKPER) + (-10.367 * HISPANICPER) + (-2.588 * WHITEPER) + (-7.247 * ELLPER) + (-1.894 * SPECIALEDPER).

As Table KK20 in Appendix KK indicates, the Wald value was significant in the model for all of the independent variables ($p < .05$). The statistically significant variables that were positively associated with the odds that a school met AYP proficiency standards were DUMMYREV (B = .309, Exp (B) = 1.362), and ASIANPER (B = 11.399, Exp (B) = 89196.438). The variables that had negative impact on the AYP outcome include: EDPER (B = -1.114 , Exp (B) = .328), NATAMPER (B = -5.560, Exp (B) = .004), BLACKPER (B = -5.560, Exp (B) = .004), HISPANICPER (B = -10.367, Exp (B) = .000), WHITPER (B = -2.588, Exp (B) = .075), ELLPER (B = -7.247, Exp (B) = .001), and SPECIALEDPER (B = - 1.894, Exp (B) = .112).

Since the predicted probably showed a low increase over the constant model, the logistic regression from this dataset was not as statistically useful as the results from the other NC datasets when there was a greater increase in the predicted probability. However, it is noteworthy that the EDPER variable had a negative impact, the DUMMYREV had a moderately positive impact, and the ASIANPER produced a positive impact. All three of those variables were consistent with the results in all of the NC datasets. Listed in Figure KK4 in Appendix KK is the observed and predicted probability of the logistic model.

Texas 2005 4th grade

A logistic regression was used to analyze the Texas schools in the 2005 4th grade dataset. The constant (null) model was correct 98.2% of the time vs. the predicted model which was accurate at the same level. Thus there was no increase and this model is not useful with predicting the outcome of school level AYP status. Much of this had to do with the fact that less than 2% of the schools in the sample failed to meet AYP. This suggests that the Texas AYP proficiency status is not as rigorous as North Carolina or California for the 2005 school year. This finding is fairly consistent with the overall number of schools in the state that only 10.3% of all schools (elementary, middle, and high schools) in Texas did not meet the AYP proficiency status in the 2005 school year (Texas Education Agency).

The Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model as indicated by the Cox & Snell R value of .034 and the Nagelkerke R Square value of .209 which further suggests that there was a very low impact on the variance within this model. Refer to Appendix LL for the SPSS output tables for the logistic regression.

Although the model is not statistically useful, all of the variables were statistically significant ($p < .05$) other than SPECIALEDPER ($p > .05$). The logistic equation for the Texas 2005 4th grade dataset is: Logit (Y, of meeting AYP proficiency) = 11.425 (constant) + (-.072 * DUMMYREV) + (-4.035 * EDPER) + (-30.263 * NATAMPER) + (5.348 * ASIANPER) + (--5.346 * BLACKPER) + (-2.652 * HISPANICPER) + (-.972 * DUMMYWHITE) + (4.068 * ELLPER) + (-.253 * SPECIALEDPER). Listed in Appendix LL are all of the output tables from the logistic regression for the 2005 Texas 4th grade dataset.

Texas 2005 8th grade

The Texas schools in the 2005 8th grade sample were analyzed with a logistic regression. The constant (null) model was correct 72.5% of the time vs. the predicted model which was accurate 77.2% of the time. The result is a mild increase in the accuracy for the predicted model vs. the constant model at 4.7%. The model has limited statistical usefulness due to the mild increase but it is more valuable than the 2005 4th grade Texas logistic model which showed no increase in the predicted model. The Omnibus Test of Model Coefficients was statistically significant ($p < .05$) indicating goodness of fit with the logistic model as indicated by the Cox & Snell R value of .277 and the Nagelkerke R Square value of .410 which accounts for some of the variance within the model. These values are consistent with the moderate increase in the accuracy level of the predicted model vs. the null model. Refer to Appendix LL for the logistic regression output tables.

The logistic equation for the Texas 2005 8th grade dataset is: Logit (Y, of meeting AYP proficiency) = 9.278 (constant) + (-.324 * DUMMYREV) + (-1.873 * EDPER) + (-159.361 * NATAMPER) + (4.796 * ASIANPER) + (-5.775 * BLACKPER) + (-5.158 * HISPANICPER) + (-.395 * DUMMYWHITE) + (-3.014 * ELLPER) + (-6.536 * SPECIALEDPER).

As Table LL10 in Appendix LL indicates, the Wald value was significant in the model for all of the independent variables ($p < .05$). The only variable that had a positive impact on the odds that a school met AYP proficiency standards was ASIANPER (B = 44.796 , Exp (B) = 121.007). All of the other variables had a negative impact on the AYP outcome DUMMYREV (B = -.324, Exp (B) = .724), EDPER (B = -1.873, Exp (B) = .154) NATAMPER (B = -159.361 - , Exp (B) = .000), BLACKPER (B = -5.775, Exp (B) = .003), HISPANICPER, (B = - 5.158, Exp (B) = .006), DUMMYWHITE (B = -.395, Exp (B) = .673), ELLPER (B = - 3.014, Exp (B) = .049), and SPECIALEDPER (B = -6.536, Exp (B) = .049). When analyzing these results it is important to remember that this model was limited in statistical usefulness due to the very mild increase in the predicted probability vs. the constant model.

The positive association with the ASIANPER variable is consistent with the findings in the other datasets in this study. It was not surprising that the EDPER variable was negatively associated with the probability that a school met AYP proficiency as there was also a sizable negative correlation (r = -.451) with the analysis using Pearson test (as listed in Appendix W under research question one). The DUMMYREV variable had a mild negative association with the probability that a school met AYP. The rest of the variables also had a negative impact on the probability of meeting AYP which was not surprising and not consistent with the other datasets in this study. The results with many of the variables have been inconsistent between the different datasets. Listed in the Appendix LL are all of the output tables for the logistic regression analysis.

Texas 2007 4th grade

A logistic regression was used to analyze the Texas schools in the 2007 4th grade sample. The constant (null) model was correct 99.1% of the time vs. the predicted model which was accurate 100% of the time. Thus there was an extremely low increase and this model is not statistically useful for predicting the outcome of the schools' AYP status. Similar to the 2005 4th grade Texas logistic results, much of this had to do with the fact that a very low number (less than 1%) of schools failed to meet AYP standards in Texas. This further supports the theory that the Texas AYP proficiency status is not as rigorous as North Carolina or California. The low level of increase in the predicted probability vs. the constant model is consistent with the results in Michigan which also had an extremely large percentage of schools meeting AYP in all of the state samples. Although the Omnibus Test of Model Coefficients was statistically significant indicating goodness of fit with the logistic model, the abnormal Cox & Snell R value of .094 and the Nagelkerke R Square value of 1.000 further support the finding that this dataset was not statistically useful in predicting the impact that the independent variables had on the dependent variable.

None of the variables were statistically significant (p > .05) other than DUMMYWHITE (p < .05). The logistic equation for the Texas 2007 4th grade dataset is: Logit (Y, of meeting AYP proficiency) = 1678.632 (constant) + (1.848 * DUMMYREV) + (-870.963 * EDPER) + (14790.617 * NATAMPER) + (4610.234 * ASIANPER) + (-890.412 * BLACKPER) + (-601.592 * HISPANICPER) + (-.136.588 * DUMMYWHITE) + (-110.274* ELLPER) + (33.556 * SPECIALEDPER). Most of the odds ratio values in the equation were very unusual as listed in Appendix LL.

Texas 2007 8th Grade

The Texas schools in the 2007 8th grade dataset were analyzed with a logistic regression. The constant (null) model was correct 81.1% of the time vs. the predicted model which was accurate 82.3% of the time. The result is an extremely mild increase in the accuracy for the predicted model vs. the constant model of just over 1 %. Therefore, the model has limited statistical usefulness due to the low increase which has been consistent with the Texas datasets in this study. The Omnibus Test of Model Coefficients was statistically significant (p < .05) indicating goodness of fit with the logistic model as the Cox & Snell R value was .200 and the Nagelkerke R Square value of .321 which accounts for variance within the model. Refer to Appendix LL for the output tables.

The logistic equation for the Texas 2007 8th grade dataset is: Logit (Y, of meeting AYP proficiency) = 15.031 (constant) + (-.096 * DUMMYREV) + (-2.426 * EDPER) + (-209.454 *

NATAMPER) + (6.155 * ASIANPER) + (-11.830 * BLACKPER) + (-11.182 * HISPANICPER) + (-1.116 * DUMMYWHITE) + (1.599 * ELLPER) + (-1.560 * SPECIALEDPER).

Due to the low level of increase in the predicted probability, the following results must be interpreted with caution. As table LL20 in Appendix LL indicates, the Wald value was significant in the model for all of the independent variables ($p < .05$). The only variables that were positively associated with the odds that a school met AYP proficiency standards were ASIANPER (B = 6.155 , Exp (B) = 471.173) and ELLPER (B = 1.599, Exp (B) = 4.946). The ASIANPER variable has been positively associated with the odds that a school meets AYP in most of the datasets in this study. The ELLPER variable has not been consistent with either a negative or positive association at different levels among the different datasets.

All of the other variables were negatively associated with impact on the AYP status: DUMMYREV (B = -.096, Exp (B) = .909), EDPER (B = -2.426, Exp (B) = .088) NATAMPER (B = -209.454 - , Exp (B) = .000), BLACKPER (B = -11.830, Exp (B) = .003), HISPANICPER, (B = - 11.182, Exp (B) = .000), DUMMYWHITE (B = -1.116, Exp (B) = .328), and SPECIALEDPER (B = -1.560, Exp (B) = .210). It was not surprising that the EDPER variable was negatively associated with the probability that a school met AYP proficiency as there was also a negative correlation ($r = -.325$) with the analysis using the Pearson Correlation test in the first research question (Refer to Appendix Z for the Pearson results).

The EDPER variable produced consistent results showing a negative impact with most datasets in this study. The DUMMYREV had an extremely mild negative association with the probability that a school met AYP. However, the unstable values must be taken into consideration. The rest of the variables also had a negative association which was not surprising and consistent with the other datasets in this study. Listed in the Appendix LL are all of the output tables for the logistic regression analysis.

Summary for Research Question Three

The results from research question three produce a good measure of school level proficiency vs. the results in research question two because they are using AYP as the calculation. As mentioned earlier in the chapter, the AYP calculation includes the mathematics and reading (or ELA in Michigan and California) state accountability assessments, plus the other indicator that is used to calculate AYP by each state. This is opposed to the one to one analysis of the relationship between individual state accountability assessment and the impact that the subgroups are having on those results with the analysis of research question two. It is important to note that research question two used the NAEP as an independent variable but this variable is not part of the analysis with research question three. The major focus of research question three was an analysis of the relationship between the subgroups and AYP.

The BLACKPER variable had a consistent negative impact on the probability that a school would meet AYP in California while the ASIANPER variable produced a consistent positive impact in all of the California datasets. The DUMMYREV variable had either a positive impact or a slightly negative impact on the AYP status of schools in California. The EDPER variable had a negative impact on three of the California datasets with the exception of the 2007 4^{th} grade dataset where there was a moderate positive impact. The EDPER variable has produced results showing a negative impact in almost all of the datasets throughout this study. Thus the finding from the 2007 California 4^{th} grade dataset is interesting and possibly an anomaly with the impact of that variable. Most of the other variables in California produced inconsistent results on the probability that a school would meet AYP in the state of California.

The results from the state of Michigan showed that the AYP standard in Michigan with all four datasets in this study was not very rigorous (in comparison to California or NC) as very few schools failed to meet AYP. This limited the usefulness of the logistic regression and thus an analysis of the impact that the independent variables were having on state AYP proficiency was not very effective. These results suggest that the statistical manipulation methods (as mentioned in the literature review) had an impact on the school level results in Michigan. The line of trajectory in 2005 and 2007 most likely had a major influence on the AYP results. As the line of trajectory increases a lower number of schools in Michigan are predicted to meet AYP proficiency. The results from the sample are in alignment with the overall state AYP proficiency status within the entire state. Over 88% of the schools in Michigan met AYP in 2005 and over 93% of the schools met AYP proficiency in 2007.

The North Carolina logistic regression analysis was much more statistically useful than the Michigan analysis. The results were consistent with the California datasets showing that the DUMMYREV variable produced a positive or slightly negative impact on the AYP status. The EDPER variable had a consistent negative impact on the dependent variable with the exception of a few datasets that produced counter intuitive positive results. They included the California 2007 4th grade dataset (that produced unstable output values), the 2005 8th Grade Michigan dataset, and the 2007 8th Grade Michigan dataset (the results from Michigan in both of those datasets produced extremely high odds ratio values and a low increase in the predicted probability[31]). The ASIANPER variable had a consistent positive impact or was not statistically significant with a couple of the datasets in the sample.

Both the California and North Carolina datasets produced an increase in the predicted probability of the predicted model vs. the constant model which was moderate to large with the exception of the NC 2007 8th grade dataset. This further supports the theory that California and NC had a more rigorous AYP standard vs. Michigan and Texas. The Texas analysis produced results that were similar to the Michigan results with the low percentage of increase in the predicted model vs. the constant model thus limiting the statistical usefulness of the model in predicting the impact that the independent variables has on the dependent variable. The results suggest that the Texas AYP standards are not as rigorous as the California or NC AYP standards. The Texas line of trajectory (similar to the one found in Michigan) establishing lower standards in the earlier years and rapid increases in the later years until the year 2014 had an impact on the high percentage of schools in the state that met AYP in both 2005 and 2007. It is also probable that the state accountability assessment has a less rigorous standard in Texas which is a conclusion that is consistent with the findings in Texas according to research question two.

Overall, the logistic regression analysis was a valuable method to measure the impact that the independent subgroup variables were having on school level AYP proficiency status in California and North Carolina. It was not as statistically useful of a measure in Michigan and Texas. However, the analysis in all states was useful in comparing the results from the different states in the sample, as the lack of statistical usefulness in Michigan in Texas established important findings about the standards and relationship to the NAEP. This finding further

[31] This indicates that the results must be analyzed with caution as the extreme values and low level of increase between the constant and predicted model suggest bias with the results.

supports the literature review that states are implementing an inconsistent standard across the country.

Future studies that analyze the impact of the subgroup variables might choose to have the NAEP school level proficiency as the dependent variable and thus measure the impact of the subgroups in accordance with the NAEP results. This could in turn provide for a comparative analysis with the results from this study with the impact that the subgroups have on AYP. The magnitude of this study did not allow for that type of analysis.

Qualitative Results and Analysis

A qualitative process was included in this study in order to understand the impact that the NCLB accountability reform is having on teachers and administrators at the school level. It was also implemented to see how schools are responding to NCLB under the different AYP sanctions that are outlined in the reform. If the accountability reform is successful, then it can be assumed that the legislation is having a positive impact at the school level. The quantitative data in this study only provides information regarding how well schools are doing on the state accountability assessments and the NAEP. It also evaluates the impact and/or the relationship between the subgroup populations and school finance with achievement on the NAEP and the state accountability assessments. While the quantitative portion of this study provides information about the demographic characteristics of schools that produce successful and failing formulas under the NCLB legislation, it does not provide information in regard to the cause or reasons for success or failures.

The qualitative portion of this study provides rich findings that include the type of strategies and/or school improvement plans that schools have implemented in response to NCLB with a specific focus on the AYP. It provides data and information, which explains some of the advantages, disadvantages, limitations, and unintended consequences with the implementation of AYP.

Description of Sample Schools

All of the schools were coded with pseudonyms according to the following colors, Blue, Red, Orange, and Purple. This was done to protect the identity of the schools and individuals in this study. A very general description of the school characteristics, which include demographics, and other pertinent quantitative data, is described in a generic manner in order to protect the identity of the schools and the identities of participants in this portion of the study.

General information is presented on the percentage of students in each school according to the different demographics in the school. The general identifiers are extremely low, low, medium, high or extremely high. The specific state accountability achievement data was also described using the above levels. General enrollment data is provided as either, low, medium, or high comparatively. The financial information includes describing the school as receiving a high, medium, or a low level of funding comparatively. This description technique was done to protect the identity of the schools and the participants in this portion of the study.

Blue School Description

Blue School is located in a suburban affluent school district and it is an elementary school. The enrollment at Blue school is at a medium level. The subgroup populations include an extremely high level of White/Caucasian and an extremely low level of the other demographic subgroups according to ethnic orientation. There are an extremely low number of students who are Economically Disadvantaged (ED). Based on school district statistics from 2007, Blue school received a medium level of funding.

The school scored extremely high on all state accountability assessments in reading, mathematics, and science. The school was above proficiency at an extremely high level in all tested areas and tested grades (three though five). The school district also scored extremely high on the assessments in all content areas and in all grades (three through five) for the 2008 to 2010 assessment years. Blue school was selected from the United States sample and they have successfully met the AYP requirements every year since the inception of NCLB in 2003 which includes a consistently high state report card grade. The school district also does extremely well on the state accountability assessments achieving very high scores in all content areas and grades assessed at the elementary level. Blue School has a special identity as a Green school leader in the district and state.

Red School Description

Red School is located in a suburban district but it is geographically close to an urban school district. The school district population is much less affluent than the school district where Blue School is located. Red school is a middle school and the enrollment at Red School is medium to high. The subgroup populations include a high number of White/Caucasian students and an extremely low number of the other ethnic subgroups. There is a high percentage of the students that are economically disadvantaged and a medium level of students that are ELL. Based on school district statistics from 2007, Red school received a medium to low level of funding.

The school scored at either a high or medium-high level on the state accountability assessments. Red School was selected from the United States sample and the school has met AYP proficiency for several consecutive years. Red School scored a state report card grade that was medium during the earlier years of NCLB and has received a consistent grade of high since that time.

Orange School Description

Orange School is located in a suburban district that is geographically close to an urban school district and has an overall district population that is less affluent than Blue school. Orange School is a middle school and the enrollment at Orange school is at a medium to high level. The subgroup populations include a medium level of White/Caucasian students, and an extremely low to low level of the of the other subgroup populations. There is a medium to high percentage of the students that are economically disadvantaged and a medium to high percentage of the students are ELL. Based on school district statistics from 2007, Orange School received a high level of funding comparatively within the sample.

The school scored at either high to medium levels of proficiency on the state accountability assessments. Orange School was selected from the United States sample in this study and the school has met AYP proficiency for several years from 2003 until 2009. Orange School scored a state report card grade of medium to high in most years since the inception of NCLB. Orange School did not meet AYP in a previous year and they anticipate that they will not meet AYP again in an upcoming school year.

Purple School Description

Purple School is located in a small urban district that borders a large urban school district and it encompasses all of the social challenges that exist in a large urban setting. Purple School is a middle school and the enrollment is at a medium level. The subgroup populations include a low to medium level that are White/Caucasian, a low to medium level that are Black, and a low level of the other ethnic subgroups. There is a very mixed level of race/ethnicity in this school. There are an extremely high percentage of students that are economically disadvantaged and a medium to low percentage of the students are ELL. Based on school district statistics from 2007, Purple School received a medium to low level of funding.

The school has scored medium to low on the state accountability assessments. Purple School was selected from the 2007 United States sample in this study and the school has met AYP proficiency for several years. Purple School has scored a state report card grade of either medium or low every year since the inception of NCLB. However, Purple School had problems meeting the AYP requirements in previous years. They have experienced the restructuring process under NCLB sanctions. Several of the faculty that were interviewed specifically reference their experiences with the restructuring process in NCLB.

Semi-Structured Interview Findings

As described in chapter three, the data analysis involved finding emerging categories or domains that were discovered during the interviews in relation to school achievement and the AYP status of schools. Interviews and documents were the two sources that provided analysis of multiple perspectives. They were then compared to other incidents in the data, and comparisons were made that led to tentative categories. As the categories got refined and subcategories were created, a framework for patterns and relationships among the coded categories began to develop. The evaluation of discrepant cases and triangulation were among the techniques used to assist the development of patterns (Gawlik, 2007).

The cultural domains or categories were organized according to similar characteristics within the sample of schools. NVIVO 9.0 provided support for the coding of the different domains (or nodes as organized in NVIVO). The cultural domains were further divided into subcategories through a taxonomic domain analysis. Most of the findings within the cultural domains from the four sample schools were verified for consistency via the triangulation methodology. However, there were some distinct differences that were identified through a componential domain analysis which was partially influenced by the independent variables that were analyzed in the quantitative results section. This included a coded analysis to break down the differences that were found among the respondents from the different schools in the sample.

The quantitative datasets along with the literature review were used in the analysis to identify the different domains and to provide further validity to the findings through a triangulation of the data. Patterns emerged with the responses by the teachers and administrators which helped to provide support and/or called into question previous literature addressing NCLB and school accountability.

The following section is divided into categories or domains that were discovered during the qualitative analysis and review. Table 16 provides a breakdown of the findings according to the coded categories or cultural domains and the specific relationships or findings within the cultural domains via a subcategory analysis of taxonomic domains.

Table 17.

Componential and coded analysis in relationship to the four sample schools in the qualitative dataset

Domain Coded Analysis Finding Categories	**Domain Coded Analysis Sub-Categories**	**Blue School**	**Items of Contrast. Red School**	**Orange School**	**Purple School**
NCLB Impression Overview		**Negative**	**Negative**	**Negative**	**Negative**
	Philosophical Intent Impression	**Positive**	**Positive**	**Positive**	**Positive**
Identification as Failing or success will work		**No**	**No**	**No**	**No**
	Sanctions will work	**No**	**No**	**No**	**No**
	Restructuring Works	**N/A**	**N/A**	**No**	**No**
EDPER Variable	**High Percentage in school**	**No**	**Yes**	**Yes**	**Yes**
	Schools with large ED populations are able to do well on State accountability assessments	**Yes**	**No**	**No**	**No**
	Belief that all children can learn and be successful	**Yes**	**Yes**	**Yes**	**Yes**
	Social Capital Impacts Achievement	**No**	**Yes**	**Yes**	**Yes**

		Blue School	Red School	Orange School	Purple School
	Parent Involvement impacts achievement	**Yes**	**Yes**	**Yes**	**Yes**
School Resources/Funding Variable	**Schools funded at different levels can achieve the same**	**Yes**	**No**	**No**	**No**
	How funding is used matters	**Yes**	**Yes**	**Yes**	**Yes**
	Additional Funding like(i.e. Title One) necessary to support ED populations	**N/A**	**Yes**	**Yes**	**Yes**
AYP Positive Consequences	**Increased Urgency Around Data Analysis**	**Yes**	**Yes**	**Yes**	**Yes**
	Disaggregate Subgroup Data	**Yes**	**Yes**	**Yes**	**Yes**
	Positive Impact/Urgency around the SIP	**No**	**Yes**	**Yes**	**Yes**
	Impact Teaching	**No**	**Yes**	**Yes**	**Yes**
AYP Negative Impact	**Stress and Low Morale**	**No**	**No, but the potential was there**	**Yes**	**Yes**
	Single Measurement Accountability was having a negative impact	**No**	**Yes**	**Yes**	**Yes**
	Narrowing Curriculum	**No**	**Yes**	**Yes**	**Yes**

		Blue School	Red School	Orange School	Purple School
	AYP status impact educator decision to work in the school	**No**	**No, (potential for yes)**	**Yes**	**Yes**
	AYP increased school capacity	**No**	**No**	**No**	**No**
Impact on Administration	**Positive Impact**	**No, other than data analysis**	**No**	**No**	**No**
	Negative Impact	**No**	**Yes**	**Yes**	**No**
Impact on District Operation		**Yes**	**Yes**	**Yes**	**Yes**

While not all participants were not in full agreement at each school, a general consensus was made with regards to the thoughts and ideas from each of the four sample schools. All of the coded categories and subcategories were analyzed using NVIVO 9 and verified through triangulation research techniques.

Overview of Participants Views on the NCLB Accountability reform

There was a mixed review of the NCLB accountability policy among the teachers and administrators that were part of the semi-structured interview sample with a consistent slant towards a negative opinion about the effectiveness of the reform. While there were definitely patterns of responses that were consistent, a variety of viewpoints were discovered among the participants. All participants were familiar with the NCLB reform, however, they had varying levels of knowledge about NCLB and the specific accountability provisions including AYP and the mandated sanctions. Most of the educators received their information on NCLB from journals, media articles, or reports along with a reliance on either central administration in the school district or the building administration to acquire their knowledge on NCLB.

It is important to point out that some of the participants answered the questions by identifying school responses to AYP. However, those responses might have been impacted by other independent variables not addressed in this study. Those include the current state of the budget in the United States, along with other district or school level initiatives that might not be a direct result of NCLB. For example, when teachers at Orange School mention the work on their School Improvement Plan (SIP) Team, they also reference the fact that they had been working on school improvement long before NCLB via the North Central Accreditation process. Prior work might have influenced the work with SIP plans in the other schools as well. However, the findings that are identified in this section clearly show that NCLB has provided for an increased and deliberate focus on data analysis that is a response to AYP.

Most of the participants feel that the overarching purpose behind NCLB is highly regarded. This finding is supported by the literature review in chapter two of this dissertation. There was a consensus in the field and with most of the participants in this study that the purpose of school accountability is a positive outcome and it is embraced so that improvements can be made for the benefit of the students. However, the implementation of NCLB and AYP is where many question the value of the reform. For example, the principal from Orange School provided this statement during his initial response to the questions: "Well, I think the idea behind it is admirable. But the accountability issues are misplaced, misthought, and misguided."
Teacher 4 from Blue School provided the following statement in response to what she thinks about the NCLB accountability policy: " I think that it has a place. I think that some of the process is limiting, and the expectations don't meet the tools that we have." To further support this finding, the principal from Blue School provided the following statement:

> Okay well, the philosophical intent behind No Child Left Behind was so that all students could be successful. But holding schools accountable on one hand is important because we all need to be accountable for our work. But on the other hand, the percentages for AYP: I don't think that's what would lead schools to be successful. Those data points- we know what makes schools successful. We know what that formula is for success. And the formula for success is not necessarily testing kids and getting AYP. So that's how I feel about it. I, you know the thing is that all the research tells us and we know what successful schools do regardless of where they are.

Also, Teacher 1 from Red School provided the following opinion regarding the overall impact of the reform:

> I do believe that it's a wonderful program. And I believe in setting high expectations for students, because we want to close that achievement gap. But I… but there are also some parts of NCLB that I'm glad that were or going to tweak because I believe that we have to consider that there are strengths in every student.

The statements from these educators indicate that they believe in the overarching goals and the philosophical intent of NCLB. However, they seem to believe that there are many flaws in the reform which was a consistent response with most of the 20 educators that were interviewed in the qualitative portion of the study. The participant beliefs are validated via a triangulation with the literature review and the quantitative portion of the study addressing issues with the AYP measurement system.

AYP Identification and Sanctions

The majority of the participants believe that identifying schools as failing will not necessarily lead to school improvement. The consensus was that more support was needed to make effective change. Most of the participants felt that financial resources would also have a big impact. There were a few of the participants that believed the identification of failing schools was effective and that it would become a motivating factor because schools would feel a sense of shame with the failing identification. Teacher 4 from Red School was one of those individuals and she provided the following statement:

> I personally think it's a good idea to have the rewards and the sanction thing going on. You know, I often question how some teachers can stay teaching when they don't produce. Personally because being in it for 22 years, I've seen a lot of teachers whose scores, you know, probably aren't where they should be. And they don't seem to think it's a big deal. And that upsets me, so I think this accountability is really important. No, I don't want to be fired because my own kids, you know, don't produce the scores. But I know I work hard and I know I try very hard. So personally, I think it's an okay thing.

Similar to the consensus among the majority of the participants that the identification of failing schools would not work, there was close to a consensus that the sanction provisions are not effective. The majority of the participants believed that the threat of punishments was not a positive means for stimulating change and improvement in schools. The participant answers are aligned with the literature in the field which shows that the sanction provisions are lacking a research basis indicating that they would lead to success (Meier et al. 2004; Sunderman et al. 2005). The principal form Orange School felt that the implementation of NCLB sanctions are not well founded in research and the removal of the principal was not an effective change in itself. In order to be an effective change, there must be excellent leadership that replaces the principal that was removed. The principal from Orange School stated the following:

> There was a point, I believe it was after year four, where the principal had to be moved. And what's happened is they've discovered that they can't find other people. They can't find principal to put in those places. So that part of the law is, has now been excised. So some of the with the reauthorization of No Child Left Behind, they are making some changes, but---
>
> And it's also interesting because of the changes that they had—the outside tutoring for example, pay for outside tutoring—there's no statistical evidence that proves it works. But it does open up money to some of the private companies. And there are some that believe that that's really the genesis of No Child Left Behind, is that there's a $376 billion market that the private sector can't get to and they would love to.

As mentioned, the statement by the principal from Orange School is in alignment with some of the literature in the field stating that the sanction provisions were created on more of an idealistic view point vs. an effective reform that was well grounded in empirical research. Others believe that for ideological reasons, some individuals may want to get public money into the private sector when it comes to education. The push for more charter schools and the initial attempt by former president Bush and some elements of the Republican party to have vouchers as part of the initial NCLB reform support this viewpoint (Peterson & West, 2003).

Several of the participants believed that sanctions will not lead to effective school improvement. The principal from Red School felt that much more was needed in order to create effective change vs. simply "dropping down the hammer". Here is the statement from the principal at Red School in response to question 4 regarding the implementation of NCLB sanctions:

> I don't believe it's going to lead to school improvement. Just based on what I've seen. I mean, and based on my experience with the people that I'm talking to from a large urban district and you know, I just, I don't believe that, you know, dropping a hammer on a school is what's going to make it change. It—what it takes is a lot of handwork and an understanding by the staff and by the administration of what needs to be done and trying to draw parents in. You know, I don't know if you're familiar with Dufours Model.

In support of the comments made by the principal from Red School, the Blue School principal provided the following response on the motivation for her work and the work of her faculty in the school. She did not believe that AYP identification of schools and sanctions was the motivating factor:

> It also doesn't necessarily mean that they're going to succeed. What they have to do is implement what we know to be good practice, best practice. And also something that needs to be considered is, you know, a piece of it, the leadership piece and the teacher piece, part of AYP might- you can desegregate the scores so that you know who the teachers are.

Restructuring

This sub-section specifically focuses on the impact that restructuring had with Purple school. The participants from Purple School referenced experience they had with restructuring in their school. Three of the teachers worked in that school during the restructuring process and the administrator was hired in response to the restructuring. Only one of the teacher participants was not part of the restructuring but she was aware of the fact that the school did go through that process.

The Purple School principal was assigned to the school following the restructuring which also involved transforming the school and eliminating one grade level from the school. This grade change was part of the restructuring process. It was interesting that upon entering Purple school, there was a large banner above the door stating that the school had successfully met the AYP requirements. Teacher 2 from Purple school actually referenced the banner and stated that it was a sense of pride because of the previous stress that they went through via restructuring. Meeting AYP was obviously a sense of accomplishment and possibly a sense of relief for this urban school that had experience with the sanction provisions that were mandated in NCLB. Here is the response by the principal from Purple School referencing the restructuring that the school went through:

> Yes and we went through that early because when I came here as a first-year principal, this school had not made AYP in six years so we were in like the phase six and we're in different phases. And I went through some of that and it did that help make us a better school? I really don't think so. I think what makes you a better school is finding the communication with the administrator and the teacher and the communication with the parents. I think you got to get the parents involved and buy into what you're trying to do. I don't think just being identified as a failing school that, yeah.

Based on the above response, it is evident that the principal from Purple School did not think that restructuring had anything to do with the level of success they had with AYP after the restructuring phase. While Orange School did not go through the restructuring phase, they are in danger of additional phases and sanctions in NCLB. Here is a reference from the Orange School Principal in regards to his view on the potential success with AYP sanctions:

> I think the punishing effects of being labeled as failing lead to changes, but I'm not sure it leads to improvement. Again, what they are counting as success is a multiple-choice test; speaks nothing to creativity. It speaks nothing to previous background.

The statement by the principal from Orange School supports the views of the principal from Purple School. While changes definitely occurred due to restructuring at Purple school, it is questionable whether or not those restructuring changes led to effective improvement. The principal from Purple School does not think so. It is also possible that the restructuring of Purple School was a form of statistical manipulation that was referenced in the literature review. The school took on a smaller population with smaller subgroup populations. This form of restructuring might have assisted with improvements under the AYP measurement while at the same time it might not have actually led to any real improvements in student achievement and/or learning.

Here is the response from Teacher 1 at Purple School in response to question 4 about the impact of sanction provisions from NCLB. She alludes to the fact that restructuring was a surface level form of manipulation that did not lead to effective change:

> We've actually gone through one of the sanctions. If anything it was a waste of money and we're still where we're at. We, we actually reconfigured, taking the 6th grade out of here. I think it's a lot of paper shuffling to make the government think we're doing something. But it's not really doing anything. I mean we did make AYP for several years, but I think it has nothing to do with the configuration. It was the actual student population we had. I think it was more so the camaraderie we had as a staff, I don't think it had to do with the AYP or the NCLB sanctions. Because ever since I've been here, we've been in some form of corrective action, be it under NCLB or under the state before that. So-We try to embrace the small learning communities, but I think overall we just felt good about one another and ourselves as a building, which led to better teaching. The district's in financial hardship and I don't see that going on now. And I don't think it has anything to do with test scores. I think it has to do with who we are as adults and what our

> perception of our professionalism is. I don't think it has anything to do with the government, honestly.

The comments from Teacher 1 at Purple school credit the hard work and PLC philosophy among the teachers as responsible for success under AYP. It is possible that the team philosophy and positive leadership of the Purple School principal who arrived in response to the sanctions in AYP might have been a motivating factor with the improvement. However, just removing the administrator in itself and replacing him or her is not well founded in research indicating that improvements would occur. However, if positive and effective leaders enter the building this could be a factor that leads to improvement under NCLB as research shows that administration has an impact on student achievement (Marzano, Waters, & McNulty, 2005). Thus restructuring could have a positive impact if conditions were in place for effective leaders to take over struggling schools.

Teacher 1 and Teacher 2 were present during the restructuring at Purple School. They felt that it had no impact on the achievement of the school while Teacher 3 was not sure if the restructuring actually led to improvement. Listed below is a comment from Teacher 3 at Purple school in response to question 4 regarding sanctions:

> They did because we were, we were one of them. We did not make AYP for seven years. We were part of the school reform grant. We had received a school reform grant and a University came in and worked with us and people from the county came in and worked with us. And then we also ended up with a re-structuring of administration. I don't know if any of that really did do the trick, but we did make AYP after that.

It is evident that teacher 3 experienced the restructuring but was not sure if it was the answer. Here is another response from Teacher 3 in regard to question 8 regarding the positive consequences from AYP, she addresses restructuring with her response:

> Presently, none. But when we first, when we first we're going through the seven years of no AYP, I saw a lot of people from the outside wanting to help us. It was a positive in that we

> adopted some strategies but it was also negative because we had three different people with three different views doing three different things. So we were caught in the middle. And everybody had an answer, and no one was listening to what we thought we needed. And, but it did bring in people, who I mean, brought in people from the University level, we adopted a lot of reading strategies and that seemed to help. The problem was is that we were not the ones that had the power to do the change. We were still being administrated from top down. Instead of from us telling them what we needed or what we felt we needed, they were telling us what we should do. So unfortunately, that didn't work.

Based on this response, the restructuring seemed to bring a lot of attention to the school and increased resources. However, Teacher 3 does not necessarily feel that restructuring led to improvements in achievement or that it led to an increase in the internal capacity of the school to make improvements. Based on the example from Purple School and literature in the field, the success level of sanctions and the restructuring process should be questioned during the reauthorization of NCLB as to whether it would and has lead to effective change and improvement with schools.

Economically Disadvantaged Status and School Funding

This section is divided into several subcategories or cultural domains. The Economically Disadvantaged (ED) variable along with the school financial variable is somewhat related and connections were found within the participant responses between each of the subcategories or taxonomic domains.

EDPER Variable

The findings from the quantitative portion of this study established that the EDPER variable had the most consistent impact on school level achievement among the datasets from all four states in the sample. The Pearson Correlation analysis provided a consistent negative correlation with school achievement in regards to the EDPER variable. The logistic regression analysis also identified that EDPER had a negative impact on the probability that a school would meet proficiency on the state accountability assessments and/or meet AYP. Thus question 5 from the semi-structured interview was an important question to analyze in comparison to the findings from the quantitative dataset. The literature review is aligned with the findings from the quantitative portion of this study and the responses from the participants during the interview process. Together this allowed for a more reliable finding via the triangulation of data with the literature review, quantitative findings, and the qualitative findings in this study.

The majority of the participants believe that it was unrealistic to expect schools that serve larger populations of ED students to perform as well as schools with more affluent students on the state accountability assessments. During the interviews many of the participants struggled with answering this question, as it was evident that most or many of them believed that all children can succeed and can learn. This philosophical belief is an optimistic viewpoint that many educators hold and often provides for a purpose driven philosophy among many educators. However, the way the question was framed, provided for a target which specifically addressed achievement on the state accountability assessment in relationship to the EDPER variable.

Most of the participants believe that the state accountability assessment is not a good measure of student learning and/or growth. This belief is also supported by the literature review in this study which provides information contesting both the validity and reliability of a single measurement accountability system (Darling-Hammond, 2007a: Darling-Hammond, 2007b; Elmore, 2002; Harris, 2007; Hess & Petrilli, 2006; Kane, Douglas, & Geppert, 2002; Linn & Haug, 2002; Ravitch, 2010; Sunderman et al. 2005). Teacher 3 from Red School gave the following response which supports many of the above mentioned conclusions:

> Realistic. I have two thoughts on that. My initial reaction is no because of the parental support. When you have more affluent districts you do tend to have a little more parental support. You have parents concerned whether the students are doing homework. Whether they are being involved in school. But then the second part of me, that educator side, jumps in and says, How

> can we say that we can't educate kids? We need to just do what we've got to do and if that means changing the way we do things and not doing sit in your seat and do work if it's hands-on, if it's what, making those adjustments and differentiating for each student that is sitting in your classroom. Then I have a hard time saying no, you know, do what you've got to do. I think that all kids have the capability of learning. It's just going to be finding the way they learn.

Only a few of the participants responded that it would be possible for schools with higher percentages of ED students to perform as well on state accountability assessments as schools with more affluent students. However, they made provision statements indicating that other things must occur including an array of resources and/or funding.

It was interesting that the principal from Blue School responded concisely that it was possible. She referenced the students on free and reduced lunch[32] in her school that met proficiency status on the state accountability assessment. However, it must be noted that the demographics of Blue school show that there is an extremely low number of ED students in the school. There is also research in the field showing that the social capital that students bring can have a positive impact on other students. Fiske and Ladd (2000) found that positive peer or spillover effects from students can have a positive impact on the school and other students. Thus the high scores with ED students in that school could be a result of some of the social capital defined in the literature review (Elmore, 2002; Harris, 2007; Wiley et al. 2005; Maleyko & Gawlik, 2011; Meier et al. 2004; Mathis, 2004a; Reeves, 2006).

Since it was a small number of ED students in Blue School that were assessed on the state accountability assessment, reliability issues come into play with that specific subgroup in the school (Harris, 2007). However, it must also be noted that the principal from Blue School had a lot of experience including over 37 years working in six different states and several districts. Thus it was evident that her response was based on all of that experience.

[32] Free and Reduced Lunch statistics are used to measure Economically Disadvantaged (ED) status.

The principal from Blue School provided a response that was not in alignment with the rest of the participants, the findings from this study, and the literature in the field. However, it was aligned with the previously mentioned purpose driven belief that “all children can learn”. Since Blue School has never had any problems with AYP and their demographics provide for a general formula of school success with state accountability as supported by the quantitative findings (impact of subgroup variables), it was both probable and admirable for this type of statement.

When I asked question 5 to all of the participants, it was easy to read the body language that many wanted to simply say yes, “All children can and will learn”. However, many paused or struggled with their answers before stating their beliefs. The principal from Blue School was the only one who came out and clearly provided a response in a positive manner that EDPER does not make a difference with student achievement. However, it was equally apparent that many did not feel that the AYP system was fair or following the research in the field when it comes to the impact of the EDPER variable. A consensus was found among the participants that the measurement system was not fair based on their responses. This is why I believe that 15 of the participants responded with a clear no while 4 provided a maybe response under certain circumstances and only the principal from Blue School provided a decisive yes. For example, the principal from Orange School decided not to be politically correct and stated the following;

> Well, previous research shows that you can take the state accountability test scores and SES, and you’ll have almost a perfect alignment. And I think a big key here in your question is that students perform as well on state accountability assessments. Because I don’t think that kids with greater need are going to do as well on the accountability assessment. That doesn’t mean they’re not going to be successful later in life. It doesn’t mean they’re not going to be a productive citizen. It doesn’t mean that they’re not going to have great, great, great success.

The above statement from the principal at Orange School supports other findings with the limitations of the state accountability assessments which are aligned with the literature review in this study. This finding is also identified and supported by the subsequent cultural domains/categories that are discussed in this section. It clearly supports a conclusion from the literature review and the findings in the quantitative dataset that we might not be measuring success or school achievement gains through the AYP measurement tool. AYP may be measuring the EDPER variable better than it is actually measuring student achievement, learning, and growth.

The principal from Red School gave an answer indicating that it was possible under certain circumstances for schools with larger populations of ED students to do as well, but certain conditions must be met including many resources or additional support. She also felt that school level revenue and funding is a key component of success with ED students which will be addressed later in this section. When asked the question on the impact of ED status, the principal from Red School gave the following response:

> I do. I do believe that it's realistic. But I believe that—I believe you have to have—I ... at this—I have that kind of school. I'm at a high percent free and reduced lunch. But I believe my teachers need to have a 100 things in their bag of tricks as opposed to 20 that my neighboring school district has. I believe that my teachers need to be over-trained. And I believe that the relationship piece is a key factor with the kids. And you know, we try to really build relationships and I have a 1,000 different interventions here. Not a 1,000 but you know what I mean. I have a multitude of interventions. And ... so I do believe that the students can learn and I do believe that they can make the gains that are needed to be made. But I believe that you have to have a level playing field. And that in order to level that playing field, my teachers have a lot of training.

It is evident that her views are in alignment with the philosophical belief addressed earlier that all children can and will learn. However, it is clear that the principal from Red school believes that additional funding, resources, and a well organized strategic plan or school improvement process is necessary in order to overcome the challenges that come with high percentages of ED students. Since Red School was selected from the sample as a successful outlier school on the state accountability assessment despite their large ED population, it is important to give careful consideration to the thoughts and ideas from the educators in that school. It was clear that the principal along with all of the teacher participants from Red School believed that additional resources were essential to their success with their SIP and student population. It was also clear that the teachers believe that the skillful leadership of the principal was the reason for the acquisition of these resources and successful implementation of the resources in order to benefit their students.

Parent Involvement and Social Capital

Since there was a consensus among the participants that ED status has a huge impact on the state assessment results, it is important to analyze the causes. The literature review synthesized research in the field to identify an aspect related to ED status which was identified as social capital. The definition for social capital can be found in Table 1 which is located in chapter one. The referenced literature shows that school success on state accountability assessments often involves a reliance on social capital and the positive aspects that come from the home.

The quantitative dataset supports the literature review findings by showing the strong negative impact that the EDPER variable had on state accountability achievement. Thus the reality is that the state accountability system is really measuring the social capital of specific communities. In order to increase achievement on state accountability assessments, the best method might be to increase the social capital in the community. The problem with this idea is that schools can not necessarily make those increases in social capital on their own. However, there are some things that can be done to support a positive increase in social capital[33]. In order to increase social capital in communities, there must be an overarching political resolve and initiative to provide greater reform that not only improves learning in schools, but also increases the ability of families and communities to send students to school with a certain level of readiness for achieving at high levels.

Many of the participants identified parent support and home demographics as a key element. They often felt that this was a major contributing factor to school success with AYP and it was a determinant factor regarding the level of success among ED students on the state accountability assessments. At least 12 of the participants in the interviews identified parental support as a major ingredient for school success or failure. Teacher 1 from Orange school has a lot of experience working with large populations of both ED and ELL students, she provided the following statement regarding the impact of ED status and social capital:

> I'm conflicted on that. I know there are examples of high performing schools with SES populations that would, you know, we can link SES population with school success. You know, there are certain populations that are more likely to succeed in a school setting. So we would, you know, I would say "no, it's not fair." Obviously the child's family background impacts their school success and when schools have a large portion of kids who are low SES then, you know, they're going to have a much harder road to educate them. But there are examples of high performing schools, you know, that you can't write off schools.

Teacher 1 from Orange School also provided the following statement about parental support, "You know, parent, parental involvement is such a key to school success." Teacher 2 from Orange school provided the following response which supports her colleague's response:

---

[33] Increased social capital also involves issues like the level of parental education, income levels, social programs and services, adequate health care, and the quality of communities (ex. free from violence, drugs, alcohol, and other detrimental factors that are often associated with low income neighborhoods).

> With the economically disadvantaged, you see, its' a different population, completely different. The home life, the parents, all the circumstances, or the factors that come into play are different. These are different so they would really impact student achievement or the student's performance.

All of the participants from Orange School believe that social capital and parent involvement are critical to success. Orange School had the largest ELL population in the sample so the participants also mentioned ELL status in conjunction with social capital. The findings from the quantitative dataset did not provide a consistent analysis that shows ELL status has a negative impact on achievement[34]. Although the participants where mentioning this in conjunction with social capital, it is probable that the ELL factor had less of a negative impact on state accountability achievement vs. ED status. However, in Orange School, the percentages of ELL and ED students are aligned. Thus it is probable that they might have been incorrect with their interpretation of the ELL variable when in fact they did not have a way to decipher between the two variables because the ELL population and ED populations were both at medium to high levels in the school. Thus there is a clear triangulation of the data in this study between the qualitative data, quantitative data, and the literature review with the negative impact of the EDPER variable. However, the same conclusion can not be made with the ELLPER variable due to the inconsistent findings in the quantitative dataset.

School Funding

Question 6 in the semi-structured interview was also related to school resources and funding as it asked participants to provide their thoughts regarding schools funded at different levels and their ability to achieve at the same level on state accountability assessments. All of the participants in the interviews from Red School, Orange School, and Purple School felt that

[34] The exception is the state of California which showed that the ELLPER variable had a negative impact on achievement on the state accountability assessments and the AYP status of schools in most of the datasets.

school funding was a key factor and schools could not achieve at the same level if they were funded at different levels. Most of them felt that additional funding was needed to compensate for high percentages of ED students in schools to level the playing field. For example, the referenced statement in the subsequent section provided by the principal from Red School, supports the belief that the ED students can make achievement gains if the proper resources are available which starts with funding. Also, teacher 3 from Red School provided a response in alignment with other responses from her building that funding resources are critical:

> I think that definitely plays into whether or not a school can achieve, you know, some of those guidelines set by NCLB. That you know, if you don't have the money to provide some of the programs and especially if you're talking about the economically disadvantaged districts. If they don't have the money or the resources to provide some of the things that these kids don't get at home, that your more affluent kids might be getting at home.

The participants from Red School believe that funding was one of the most important ingredients to success. The participants from Orange School also thought it was critical along with Purple School. However, participants from both Red School and Purple School believe that additional funding through categorical resources like Title One funding were essential to meet the needs of ED students. They reference how it not only was the great equalizer but it also provided additional resources to support the SIP. Many of those resources include targeted interventions for students and/or professional development for faculty with the implementation of research-based best practices. Teacher 1 from Red school provided the following response regarding the resources they use to implement the SIP:

> We have the S S T. We have categorical support teachers that we've added that are, you know, that are certified teachers. We have Title I programs. We have my program which is reading program. We have math support teachers. We have after-school tutoring. We have ... what do we not have here? I'm like we have, we have offered so many things to be sure that we hit AYP. I can't even—you know, we have small groups that we, we are—we have people come in from New Zealand to train us so that we do, you know, the latest and greatest of everything.

This additional response from Teacher 1 indicates that the level of resources that a school receives in conjunction with the positive impact of the leadership with the strategic implementation of the SIP is critical to accelerate achievement among at-risk student populations:

> I believe so because we have a principal that's behind us 100 percent and wants—and is like so into the students. And so she makes sure that she provides with people coming in with our school improvement time just to make sure that we are all, you know, getting funding. So we have resources, we have a literacy library that we can go to, to hit every area, every subject that we're around.

However, the findings from those three schools along with one teacher from Blue School do not necessarily provide a concise triangulation of the data with the quantitative dataset in this study. The quantitative results showed that the TOTALREVENUE variable was either moderately positive in relationship to school success on state accountability assessments or at times, correlated in a very mild negative way with school success on the state accountability assessments. However, when we examine the responses from a few of the participants, a triangulation might be possible with an adaptation to the findings. For example, Teacher 4 from Red School provided the following response:

> That's a hard one too because when I think about the money we get, you know, it all depends on how it's used. I know in our building, I mean, I can only speak for what I know. So in this building, I do know that our title money or whatever money comes in is used in a very positive way, in a way that will support the kids. If it's not used correctly though, obviously I don't see how it can benefit anybody.

Teacher 2 from Blue school provided the following response:

> Yea, that is an interesting one because with this whole debate on school funding especially right now. You probably know our district is not near the top of any of the schools that are funded, and yet our district tends to perform pretty well overall. So, I am not sure that there is, has to be a direct correlation between school funding and student performance. I think a lot of it really depends on how that funding is used obviously. And I think that especially if you have a student population who comes with a lot of background knowledge who are ready to learn, your teaching is supported at home and followed through, you don't necessarily need like a lot of the bells and whistles and a lot of the funding to help those students

Both responses provide support to an interesting finding that is identified in this study which shows that TOTALREVNUE[35] is a two- tiered variable. The amount of funding is important to set the conditions for having an impact. However, step two (or tier two) is how the funding is used. If the funding is not used appropriately to support student achievement and growth, the variable might not have a positive impact. This finding is supported by the literature review (Green et al. 2007; Green, Huerta & Richards, 2007). This might be one of the reasons why the quantitative dataset showed either a positive impact or a mild negative impact with the different datasets. As schools received funding at different levels, that variable in itself did not necessarily produce positive results for the school. However, if used effectively, it could positively support school success on state accountability assessments.

The mild negative impact with some of the quantitative datasets supports both the responses from the participants referenced above along with the referenced literature review showing that there is not a clear consensus on the impact of the TOTALREVENUE variable. This is also supported by the literature review where there was not a clear consensus on the role of funding although it was slanted more towards a positive impact in the literature review. Thus future research might be necessary to analyze the funding levels according to the specific school responses to accelerate achievement. This study provides results that show an interesting dynamic with the TOTALRENEVUE variable which needs to be further analyzed in relationship with school improvement plans and school level responses to the NCLB school accountability reform.

[35] The reference to this variable was defined in the quantitative portion of this study. This variable measures educational resources (school finance). Refer to page 94.

Another interesting finding within the funding category is the response from the majority of the participants from Blue School. Three of the participants from Blue School did not feel that financial resources had much of an impact on school level success while one was less informed on the issue and the other participant disagreed and felt that funding was important. Teacher 3, Teacher 2, and the Principal from Blue School felt that schools that were funded at different levels could achieve at the same level on state accountability assessments. For example, the principal from Blue School provided the following statement in regards to school funding:

> Yes, our district is a good example. If you take our district versus another affluent district lets say, and I don't know what you get in your district but we get much less per pupil here. And that's pretty low on the scale of funding per pupil where some schools get 10, 11, 12 thousand per student. And our scores are among the highest in the state. Our scores, we have very good scores, at least here at school we do.

The ideas from the principal at Blue School and some of the other teachers provide for an interesting conundrum. The demographics within the school district are at a lower level (more affluent) when it comes to the EDPER variable. However, the TOTALREVENUE variable is at a lower level in comparison to other districts in the state. The question is which variable is having a greater impact on school success levels? It is possible to make a conclusion from the participant responses from Blue School in conjunction with the district financial data. By looking at those two factors alone, the conclusion might be that revenue does not have much of an impact and schools that are funded at different levels can achieve at the same level. Since Blue School is achieving at high levels on the state accountability assessments and AYP, one finding is that they are using those resources in an effective manner to accelerate student achievement. Based on the walkthrough tour that I conducted throughout the school, it was evident that best practices were being implemented as there was meaningful instruction going on that appeared to be aligned with the school curriculum and school improvement plan (based on my limited time in the handful of classrooms that I visited). Other artifacts in the form of product assessments were also found throughout the building establishing the fact that student outcomes were being measured and celebrated with a diverse assessment model. I also analyzed the SIP and identified that it was aligned with the responses from the Blue School participants.

Another possible explanation to the success level from Blue School has to do with the EDPER variable. The findings from this study along with the research in the field clearly show that the EDPER variable has the greatest influence on school achievement and AYP status. Since the EDPER variable has more of an impact vs. TOTALREVENUE, it is probable that Blue School was achieving at higher levels due to the more affluent population or social capital that the students in that school community had acquired. This in combination with an effective SIP is a formula for success when measured by the state accountability assessments. As previously mentioned, the literature review also provided support to the fact that AYP and standardized assessments are measuring school demographics and ED status (Elmore, 2002; Harris, 2007; Wiley et al. 2005; Maleyko & Gawlik, 2011; Meier et al. 2004; Mathis, 2004a; Reeves, 2006).

Since the dependent variable in the quantitative portion of this study is achievement on state accountability assessments and the AYP status of schools, the results provide a justification for the need to change the AYP formula in order to measure effective school improvement vs. a measurement of the EDPER variable. The qualitative findings from this study also support this conclusion as the majority of participants indicated that the EDPER variable had a dramatic impact on the results as discussed. This finding was validated via a triangulation of the data with the artifacts collected from the school, the literature review, and the quantitative data in this study.

Positive Consequences

Most of the participants provided valuable information leading to specific findings regarding the positive impact that AYP is having on classroom instruction and school improvement strategies. Those positive impacts were coded and categorized by several sub-categories or attributes with the implementation of a componential analysis. Those subcategories include; data analysis, the School Improvement Plan (SIP), other school level responses, administrative actions, impact on teachers and teaching practices, impact on administrative practices, and the influence with curriculum including the district level responses. For example, one benefit with the creation or enhancement of the SIP process was an increased focus on the development and implementation of research based strategies in the classroom. As mentioned, the results regarding the implementation of strategies must be analyzed with caution because it is possible that the implementation of an effective SIP or instructional strategies might have occurred without the presence of AYP. For example, many schools were involved with the NCA or other accreditation processes long before NCLB and AYP. Thus for those schools, AYP might not have been the driving force as other variables had a greater impact. For example, the principal from Blue School provided the following response:

> But we were driven by what was best for kids. So the same thing about Indianapolis, Indiana. Same thing about Louisville, Kentucky. Same thing with, you know, no matter where I am, and here I am; I happen to be right now is this district, but prior to this I was in district 2, and district 2 has different demographics than my current district. And still there, some schools may be driven by that. But for the most part, it's been driven by what's best for kids. So there's some, there have been differences in each of those. I don't think it's AYP that drives people to do things. I don't think that those practices are necessarily best for kids. But now there are some things that are good for kids. I mean, you know, we'll get to it.

Data Analysis in Schools

The most consistent and positive finding that came from the qualitative interviews was the increased emphasis on the use of data. This includes school level data in the summative sense with the state accountability assessments, along with an analysis of individual student data to guide instruction and adapt instructional techniques. The increased focus on subgroups was another area that had a positive impact on school level practices. Almost all of the participants pointed to the increased an emphasis on data analysis in response to AYP. This finding emerged with Blue School which was the first school where the interviews took place. The principal from Blue School provided an answer stating that there were no benefits from AYP throughout the interview. However, based on her additional responses it was evident that her skill with data analysis had increased under AYP. For example, in response to question 11 regarding the specific things that are being implemented in the school in response to AYP, she responded in the following manner:

> I do believe that, now I'm thinking it's a result of looking at data. And as a result of the law of the law itself: No Child Left Behind. All of us are doing RTI, Response to Intervention. All of us are doing universal screening three times a year. When we look at our data here in our district, we know that our African American population is scoring significantly lower than almost any other ethnicity. And so that has led us into courageous conversations. That's a district-wide initiative.

She continued with the following response indicating that the analysis of data was also leading to the implementation of effective research based strategies:

> So, Courageous Conversations was written by Glen Singleton. And the district has gone through three years of training. First they trained the administrators and now we're training, well, the last two years we've been training, building equity teams. And then next year in conversations, just keeping that conversation about race, in part, as part of our professional work too because there's no reason. There's no reason. In fact, I could share with you my data.

She also added the following statement in continuation of her response:

> NCLB and AYP got us, I think, all of us to be taking a better look at data. And four years ago when that came, we didn't have a database. Now we have a database. Four years ago when I came, we didn't have a universal screening where each grade level were doing the same tests. Now we have grade-level assessments, and grade level assessments are determined by the PLCs.

Since Blue School was where the initial finding occurred on the analysis of data, it is important to continue to share answers from the principal at Blue School. Her responses are interesting because they show that while she believed that there was no positive impact from AYP initially, after the interview, it was identifiable that there was at least one in the form of data analysis. It was evident that she also came to realize this through the interview process as the analysis of data was consistently mentioned throughout her interview. In support of that finding, the principal from Blue School provided the following answer when asked question 13 regarding how her practice as an administrator has been impacted by AYP:

> So, it has forced me--now I have always been able to interpret data. I can analyze data. That I know how to do. But it was creating the database. So, and it was, so anyway. So I think I used to be much better at formative data than I was at our summit data, our state accountability data. And I can always analyze the data, but I've had to learn to disaggregate data. Okay. I've had to learn to look at demographics. I've had to learn how to dig deeper, like if everybody were at a high level in reading, you still have to dig deep. That's how we found our boy writers. So, it's harder than it's been. I've had to learn how to dig deeper.

Similar to the principal from Blue School, while most of the participants responded that more data analysis in schools was occurring because of AYP, they often did not respond to question 8 directly with data analysis as a positive outcome. However, the increased emphasis on data analysis was easy to identify in all four schools through the use of the multiple questions that were designed in the survey. The analysis of data was a consistent and clear finding with almost every interview and it was triangulated by an examination of the SIP, school website, a tour of the school, and with some of the artifacts that were collected from the participants.

The participants frequently responded regarding a positive increase in data according to different questions within the interview protocol. Those included questions in relationship to the influence with the school improvement process in their school or sometimes with their teaching practices, administrative practices, or implementation practices with other educators in the school.[36] This increase in emphasis on data appears to be a school, district, and even a county initiative in response to AYP. Three of the schools specifically mentioned a common data analysis system that was being implemented in their districts. However, each believed that more work and progress needed to occur with the development of that system. Other respondents mentioned data as having a positive impact on the SIP plan. For example, Teacher 1 from Purple School provided the following response, "Well, that's what drives our school improvement plan is the data. What can we do? What different techniques can we use in the classroom to help raise student achievement."

Several of the participants also identified the analysis of subgroup data as another important response to AYP. This was another consistent finding within the qualitative interviews. The principal from Red School provided the following response to question 12 regarding the impact on the SIP:

> Well, we're looking at our minority students. And our special end students as subgroups. We're looking at every subgroup now. And that's very much part of AYP. And our special end students, we're really trying to bring them up. And just kind of looking to see if we're treating them differently in any way. You know. Or if there is something different we can do with them. I spend tons of time on paperwork. Tons. As a result of AYP.

In response to question 8 regarding the positive consequences to AYP, the principal from Red School responded in the following manner:

---

[36] The questions asked during the semi-structured interview protocol provided a great design for eliciting the responses leading to this finding on data analysis. It was evident that multiple questions (sometimes redundant in nature based on the protocol) were needed in order to come to this conclusion.

> Well, my teachers are much more knowledgeable about assessment. They're much more knowledgeable about data collection. You know, they are very aware of their students and where they need to get them. You know, I have some intervention programs that I would not have if it weren't for that.

Teacher 3 from Orange School gave the following response to question 8. The response is supported by the SIP artifact that was received from Orange School which showed that data analysis had occurred in order to develop specific research based strategies. The tour of Orange School in the hallways and several classrooms also supports this statement and finding regarding data analysis through the triangulation methodology.

> I think more people outside of those of us who've been on our steering committee are familiar with data. I don't think that that was necessarily the case, especially when we first started NCA back, I don't even know how many years now. You know when I first took that over as chair, the only people analyzing data were teacher 4 and I. That was it. You know, we would sit here all summer and analyze data, and nobody else knew how to do that. And I think we have a lot more people, class A coming in, where people can attain scores. I think it has been positive. And I kind of directly relate that to AYP in some ways.

Teacher 4 from Orange School gave the following response to question 10 regarding the response that their school has taken because of AYP.

> Oh, I think we have. I mean, if we achieved AYP obviously, we must have. You know, like I mentioned earlier, we do have strategies for math and science--I'm sorry, math and reading and writing, and those are the areas where we really need to be focused on, we're trying to find anyway. So, we do a lot of extra things for those kids, in math especially, and we'd have great extra double dipping help for reading and writing. All our ELL kids are tutored. Again, they have an extra language arts time during the day. Again, we have our NCA goals for math, reading and writing and we do them every quarter. And so we're constantly, you know, scoring, looking at the data, comparing, seeing what needs to be worked on. What can we do to improve this NCA strategy goals and- So, yeah.

This shows that while the data analysis was not directly referenced by the participants as a positive consequence under question 8, it was identifiable at different portions of the interview and with different questions among most of the participants. It also shows that a variety of questions were necessary in order to acquire this valuable information and triangulate the findings regarding the positive impact with data analysis in schools and among the faculty. For example, teacher 3 provided the following response to question 13 regarding how her teaching practice has changed:

> I think that back when I first started: A, the population was different and; B, being a younger teacher that didn't have to worry about AYP or anything like that, you kind of looked at the class as a whole. And I think that over time and with AYP and with scores that I can actually point to and with the looping, I'm looking at the class as individual students. So when I go to create a plan or a lesson, I'm creating it for my on-target students and my at-grade level, but then I'm taking it, "okay but what about this kid, and this kid, and this kid?" And how am I going to get them there also.

Teacher 3 from Red school gave the following response to question 11 regarding the impact on teaching practice:

> I think that a lot of – AYP has probably forced a lot more differentiated instruction within the classroom and really looking at students more individually and what they need specifically in order to be successful.

It is evident that not only is there an increased focus on data in the aggregate (overall school scores), but also with the disaggregate (within the subgroups). AYP has further provided for an increased focus on individual students as supported by the participant comments. Teacher 4 from Purple school gave the following response to question 13 about the impact on her teaching practice.

> I feel like I definitely pay more attention to where I get the data. Maybe before we had no child left behind, I didn't necessarily look at okay these students are all not obviously grasping the concept of summarizing. Now, I use a lot of data to guide my instruction so I am constantly doing data analysis on my assessments and making sure that I am teaching things that directly correlate to what they need to learn.

Based on the responses from all of the participants there is a conclusive finding that data analysis is a positive outcome from AYP. This includes overall school scores and summative data, formative classroom instructional data, and individual student data leading to specific strategies. However, there are two factors that must be considered with the analysis of this finding. The first is that many schools were already engaged in data analysis prior to NCLB.

The AYP provisions created an increased sense of urgency around data analysis and it involved a greater analysis of subgroup populations because AYP measures those subgroups. While the data analysis response to AYP is a positive outcome, the specific responses that schools make in regard to their data analysis varied based on the participant responses. Most thought that they were implementing research based best practices in alignment with their SIP. However, the participants did not feel that AYP necessarily provided them with the additional support or resources that were needed once the data analysis occurred.

Future research might want to include more in-depth questioning techniques regarding what specific response the school made as a result of their data analysis and/or what specific resources or implementation strategies they think are necessary in order to improve student achievement according to their data analysis. While some of that information was found in the interviews, it is an area that could be further studied to determine the successful practices that schools are implementing in response to their data analysis techniques. Some of this information is identified in the following sub-section on school responses and SIP implementation.

School Improvement Plan Positive Initiatives

The triangulation technique was used to analyze the data in order to develop valid findings within the cultural domain under positive impact and the subsection or taxonomic domain analysis coded as SIP positive initiatives. The impact that AYP is having on the SIP is the dependent variable for the qualitative research analysis. A question regarding the impact of the SIP was specifically asked during the semi-structured interview protocol. The majority of the participants (other than a couple) felt that AYP had some sort of an impact on their SIP. However, like some of the other positive findings, the impact with the SIP is something that was discovered in more than just question 12 during the interviews. Thus the coded analysis of this domain was critical. The questions that were asked throughout the interview allowed for the finding of rich information about the positive impacts even if the participants came into the interview with a negative perception of NCLB and AYP, which was the case with many of the participants.

A consistent finding within the positive impact domain included the belief on the part of the participants that data analysis helped to drive their SIP. Thus data analysis, the SIP variable, and implementation of effective instructional strategies are interrelated within the specific domain. Data analysis was mentioned by many of the participants which led to responses about specific strategies that were being implemented through implementation of the SIP. Some of those strategies included: a focus on mathematics and literacy, the implementation of Response to Intervention (RTI) model, the initiation of research-based strategies aligned with Dr. Marzano's findings, writing across the curriculum, technology integration, Bulletin Board Configuration (BBC), interventions with reading and mathematics, professional learning communities (PLC), specific strategies targeting subgroups or gender targets where gaps exist between males and females, an increased focus on assessment to drive instruction, focused professional development for faculty, and targeted interventions in the classroom.

The responses by the participants were supported by the analysis of artifacts in the form of the school improvement plans. The school improvement plan for each of the sample schools provided specific information on research based instructional strategies that were being implemented in the schools. The SIP plans were colleted from the participants and then triangulated with additional data that was found on the school websites. Purple School was the only school that provided a limited amount of SIP artifacts.

It was clear that through the implementation of the SIP plan, there is an increased focus on PLC and/or collaboration among teachers. The SIP plan became the center for collaboration strategies among the sample schools. Teacher 2 from Orange School provided the following response indicating that the pressure to collaborate and work together was increased due to AYP:

> Positive consequences: there's, the school culture is very positive, there's more collaboration. All of us are trying to find ways to target subgroups, maybe we're applying school wide strategies; there are more interventions. So … It's a middle school concept to begin with, I mean, we've always have, we don't have any common team prep anymore. But in the past, we've always had that sense of teaming and working together, but I think yeah, it adds to the tension or we feel you have to do it, it's because of the adequate yearly progress.

The administrator from Blue School felt that they were doing a lot of collaboration. While she did not directly indicate that the reason for this had to do with AYP, based on her responses, it was identifiable that AYP had an impact on the implementation of collaborative strategies:

> And I think that's a good thing. I really do think looking at data is a good thing. So anyways, we've kind of come full circle. I've mentioned RTI. I mentioned universal screening. I've mentioned courageous conversations as a result of our data. And if we take a next step, we do co-work days. We do a lot of PLCs; we do a lot of collaboration, and we look at student work. We look at student data, and then you have to think about the implications for teaching. To me, that's just good teaching, you know. So...It's good practice. This got us to looking at data. NCLB and AYP got us, I think, all of us to be taking a better look at data. And four years ago when that came, we didn't have a database. Now we have a database. Four years ago when I came, we didn't have a universal screening where each grade level were doing the same tests. Now we have grade-level assessments, and grade level assessments are determined by the PLCs. So what the first grade and the second grade testing kids, it won't look like the same thing in fourth and fifth grade. It won't look like the same thing, but everybody is testing kids. And I do believe that that is a result of AYP and No Child Left Behind. And I think that's a good result. But looking at student work and determining the implications for teaching that is also just good practice. Because if you don't do anything with the data, what's the point in collecting the data?

Based on the response provided by the principal from Blue School, it appears that PLC and collaboration have been increased under NCLB resulting in the implementation of specific research based strategies within the school and classrooms. However, once again, caution must be given to this conclusion, because it was possible that collaboration would have occurred with or without NCLB as it was a school based strategy that was being advocated by researchers prior to NCLB (Dufour & Eaker, 1998).

The last comment by the principal from Blue school alludes to that finding where she states that it was just “good practice.” It is clear that the principal from Blue School believes that the increased focus on data was a result of AYP which eventually led to the implementation of specific strategies via Professional Learning Communities (PLC).

Impact on Teaching Strategies

Similar to the findings with the impact on the SIP, it was often necessary to analyze and code responses from several questions from the semi-structured interview protocol to determine the impact on instructional strategies. For most teachers, they did not necessarily feel that there was a direct impact. However, the impact on teaching strategies was answered in response to questions addressing the SIP and/or the strategies that were being promoted by the building administration. It is interesting that all of the teachers from Blue school felt that there was no impact on their teaching.[37]

For example, Teacher 1 from Blue school provided the following response to question 13 about the impact of AYP on their teaching practice. “I don’t think it has. I think each student is a case-by-case student and I just know that I look at that student, and I just, I do whatever is needed to help them.” Teacher 2 from Blue school provided a similar response:

> I can say fortunately, I guess I can fortunately, it hasn’t really been. As a teacher, I haven’t had to implement any new scripted programs or I haven’t been expected to give any whatever assignments or do state accountability assessment practice or anything like that. I think since we are doing fine, it hasn’t really been an issue and people have mostly left us alone. Fortunately, I have been able to just kind of keep doing the same thing. And well we keep getting the same good results.

[37] However, the principal from Blue school alluded to the fact that it forced teachers to look at more data and subgroup data like special education.

With the exception of Blue School, most of the other teachers and administrators in the other three schools in the qualitative sample indicated that AYP was having an impact on their teaching practice. It is probable that Blue School provided a different finding about the impact due to their high test scores on the state accountability assessments and the high level of social capital among students in that school which assisted in the attainment of those scores. This is a componential domain finding as the attributes in Blue School are much different than the attributes in the other three schools which is likely a result of the different ED population and the challenges that face the other three schools in the qualitative sample.

The teachers in Red School felt that AYP was having an impact on how they looked at data and differentiate instruction along with their lesson planning. They had instituted a web based system for planning lessons. A sample was collected as an artifact from Red School. Teacher 4 from Red School felt that AYP had a positive impact on her teaching:

> I think it has made me a better teacher, to be honest, I know a lot of people don't feel that way, but I'm just one of those people though that I want to do good, I want my kids to do good. And if I have to change the way I do things, then I do and I'm fine with that. I think it's good. It's more work but its fine. Well, it is very curriculum driven, I mean I am GLCE driven and I make sure, I make sure that those kids know what GLCEs it is are hitting. And they always say, "Well, why do we have to do this? Why do we have to do this?" I show them all the state says, you know, this is important to seventh grade because you'll need this now so that when you get to eighth grade you'll flow into this. You know and I try to show them that it bridges over and then it'll bridge over again into the high school.

The teachers from Purple School also felt that they were more data driven but this was also given a mixed review:

> I have to keep a lot more documentation. Quarterly data is collected as far as pre-imposed test, teaching the strategies per content area. So it is a lot-There's a lot more paperwork that we have to keep. We have to note in our lesson plans when we're doing the strategy. We have to keep evidence of doing it. So student evidence so there's a lot of paperwork involved. When instead of just "Okay, Yes, I'm doing it; these are good teaching strategies." We have to prove that we're doing them, which is annoying. Yeah, I mean, I guess, I mean in one positive way, I mean, I guess I'm kind of making sure that those strategies are being included. I kind of feel that they're

> just good teaching strategies. So I think it's positive in a way that we have to look at the strategies and make sure we're including them. But then again it would be the additional paperwork.

It is evident that Teacher 3 from Purple School felt that while a positive impact was the implementation of strategies based on data analysis, some of the required paperwork under AYP was distracting to her job as a teacher. The teachers from Orange School also felt that they were looking more at individual students and data as a result of the SIP and AYP. The participants from Orange School indicated that they were also implementing strategic programs to meet the AYP targets in mathematics and reading (the AYP tested areas). For example, Teacher 4 from Orange School taught social studies classes at the 7th and 8th grade level, however, his curriculum was modified to support goals in reading and mathematics. Here is a response from Teacher 4 in Orange School regarding the impact of AYP on his teaching:

> Well as a social studies teacher, I think a lot more, when I'm planning lessons and planning instruction, how to support the reading and the writing that's going on in language arts. And as far as when I'm doing, graphs and charts and things, I try to figure out a way to support. For example, when we're doing proportions—they were going to be doing proportions in math in a month or two--but I was working on data that used proportions associated. So kind of introduce, so that the math teacher doesn't have the whole load or the language art teachers have the whole load. Because you know, social studies not counting for AYP, which, you know, I kind of do everything I can to help with the reading and writing for us all.

Administrative Response and Impact

Through a triangulation of the data, it is evident that all teachers were impacted by the SIP plan. However, the responses and implementation levels along with the SIP response varied among the teachers in the different schools. The impact of leadership and school administration also played a factor with the level of implementation among the teachers in the sample schools. All of the teachers in Red School felt that school administration was having a positive impact on their teaching, the SIP strategies, and the level of impact with AYP in their schools. The positive feelings of respect and admiration by the teachers in Red School towards the principal from Red School, was evident based on the data collected during the interviews. For example, Teacher 1 from Red School provided the following response when asked about the impact on other teachers which was a follow-up question to any administrative responses that were being implemented[38]:

> I believe so because we have a principal that's behind us 100 percent and wants—and is like so into the students. And so she makes sure that she provides with people coming in with our school improvement time just to make sure that we are all, you know, fundings. So we have resources, we have a literacy library that we can go to, to hit every area, every subject that we're around.

Other interviews from the teachers at Red School also alluded to the level of respect that they had for their principal. I made a determination that strong leadership was present at Red School based on my interaction with the principal, interviews with the teachers, tour of the school, and examination of the SIP. As mentioned, Red School was selected because they were somewhat of an outlier school with achievement on the state accountability assessment despite their large ED population. However, it was possible that the responses that were taking place from the Red School teachers and the impact on their SIP might have been due to the building principal and not directly related to AYP. Thus it was critical to really analyze the responses from the principal at Red School because she was definitely implementing positive research based strategies in her school.

The principal from Red School felt that the sanctions and punishment in NCLB were not going to be the things that made true change in achievement within the school. The responses indicated that she felt the state kept changing the requirement and that were things that were occurring that would not make effective change. Based on my interview with the principal from Red School, I believe that she was motivated by an intrinsic value that she can and would make a

---

[38] This response was also referenced earlier in the chapter.

difference with students. NCLB was not a true motivating factor, although there were some requirements by AYP that had to be addressed by her school. The principal from Red School indicated that she felt the teachers were more aware of individual students and individual student data but that beyond that, there was not much benefit to the NCLB legislation. Here is a response that she provided:

> You know, they are very aware of their students and where they need to get them. You know, I have some intervention programs that I owl not have if it weren't for that. And the negative is I believe that in education we have, we put so much in the test and do not focus enough on the student. And it really puts your staff at a great deal of ... at a stress level that is just monumental. I feel that as educators were constantly chasing our tail, you know. We do not need to throw programs out that are not working; we need to tweak what doesn't work and that's what has been a fault of us ours for years and years. I believe you need to concentrate on every child and for that, you know, that we do, do that. But I just think that it has put a lot of stress on kids, on teachers, on administrators. And I believe that some of the stress is unnecessary because, you know, I don't know what you feel about this, Glenn, but in my ... as far as I've been a teacher and as far as I've been an administrator, there never has been, never not has been, there has never not been school improvement. From the day I entered the door, I've had, I've worked on school improvement and been on a committee. And you know, as an educator at the university level, I recognized five years ago that there were schools that never get a school improvement plan. And I just never could figure out why they didn't have a school improvement plan. And that's been like part of my life for my whole career. And I get tired of being painted with the same crayon as our, with the same paintbrush as people who are not doing what they need to be doing, you know. So I think that it's really kind of, for educators who have always worked really hard at educating, I think that it's just another policy. And framework by people who really are not part of it, you know. If you look at the whole policy of No Child Left Behind and who he had on that committee, it was not experienced educators. And you know, that's with race to the top now too. You know, it's just another policy, you know, where we have to change things that we're not broken for us to begin with because we were adhering to what No Child Left Behind was asking of us. And I just, I guess I don't understand and I've said this at the state level too. Like I said this to the person in charge of that the state accountability process.

The response form the principal at Red School identifies an important finding, the motivation is not necessarily due to AYP and NCLB, it is rather do to a greater cause and many of the SIP practices were in place prior to NCLB. The principal from Red School also felt that the identification of schools as failing and sanctions would not have a positive impact on improving the conditions of schooling. For example, she also provided the following response to question 4 about the impact of NCLB sanctions:

> I don't believe it's going to lead to school improvement. Just based on what I've seen. I mean, and based on my experience with the people that I'm talking to from this large urban district and you know, I just, I don't believe that, you know, dropping a hammer on a school is what's going to make it change. It—what it takes is a lot of hard work and an understanding by the staff and by the administration of what needs to be done and trying to draw parents in. You know, I don't know if you're familiar with Dufours Model.

An analysis of the administrative responses occurred with a triangulation of the data between the administrative interviews, teacher interviews, quantitative data from the school, artifacts collected, and tours of the school with field notes. It was evident to me that AYP was having an impact on all of the administrators. However, the greatest impact was on the principal from Orange School because they were potentially entering their second year of not meeting AYP proficiency standards. The teachers from Orange School felt that there were definitely responses occurring by the principal in order to meet the AYP requirements. (Orange school also appeared to have a strong leader as the principal). Some of those included scheduling changes, targeted lesson planning, more interventions, the implementation of the RTI model, and a narrowing of the curriculum to focus on reading and mathematics (the AYP tested areas).

There was also a substantial impact on the principal from Purple School, as that school had previously not made AYP and went through restructuring. There was somewhat of an impact on the principal from Red School. However, because they had made AYP, the impact was much less vs. the two afore mentioned school principals. The impact on the Blue school principal was at the lower level among all of the four sample schools because their state accountability scores were very high and the social capital of the students were at a high level in that school.

All of the school administrators were required to comply with the state requirements regarding their SIP and the accountability processes that were implemented in response to AYP which involved a great deal of paper work. The state SIP plan that requires data analysis is very intensive and numerous administrator and teacher hours were spent with the creation of that plan. However, it was not evident that filing the state required paper work led to any meaningful school improvement efforts on the part of the schools.

Negative Impact or Unintended Consequences

Participants identified several negative outcomes from NCLB and AYP. The findings also identified unintended consequences with the implementation of the NCLB accountability reform which are aligned with findings from the literature review in chapter two. Some of those negative findings and unintended consequences include; high levels of non-beneficial stress and pressure among faculty, a negative climate among faculty and the school community leading to low morale, the implementation of ineffective top down strategies, a narrowing of the curriculum, the implementation of high stakes testing based on a single measure that is not very good at measuring school improvement (this includes a measurement of social capital but not school improvement or growth), the possibility that educators would not want to work in schools that have been identified as failing under AYP, a moving target that is established by the state, and the inability of AYP to use sanctions(including restructuring) and the identification of schools to institute effective school improvement for the benefit of disadvantaged students.

Stress and Low Morale

All of the participants in the interviews mentioned that AYP created an increase in stress on the faculty. Only a couple felt that there was positive stress but they also identified negative stress in the school. Participants from the schools that did not meet AYP indicated that there was a negative impact on school culture and the result was a low level of morale among the faculty. Orange School did not make AYP previously and they were concerned that they might not meet the AYP standards in an upcoming school year. Teacher 3 from Orange School provided the following response to the negative impact that AYP was having on their school:

> I think frustration levels are a lot higher. I think that more and more people are getting so overwhelmed because they feel so helpless that there's almost a backsliding of wanting to try, in some ways. You know, you see this goal up here and you know what you're working with and it's sort of like high, they throw their hands up in the air. They just get more and more frustrated, and I think that your stress levels increase and people aren't maybe as happy coming into work every day. And knowing how hard you are working, I mean, you're jumping through hoops, you're performing like a circus clown up there. You really are doing everything you can. And I just don't think that the scores necessarily reflect everything you are doing for kids.

Along with having a negative impact on the stress levels in the building, it is also evident that Teacher 3 does not feel that AYP is a fair measure of instruction in the school. Based on a tour of the school and visits to the classroom along with an analysis of the SIP, there was evidence of effective research based instruction. Some of that included higher level critical thinking, writing across the curriculum, an intensive amount of technology integration, multiple interventions, and the use of differentiated instruction including product based assessments. Curricular aligned student work was on display throughout the building including the classrooms and hallways. The consistency of strategies was also identifiable based on visitations to classrooms and the tour throughout the school. Technology integration and authentic instruction were thriving in this school. Yet, the school had not met AYP standards in a previous year.

The principal from Orange School was definitely under a lot of stress due to the AYP measure, he was not a proponent of NCLB and felt that it was simply measuring students' ability to do well on a standardized test vs. true improvements in achievement and the acquisition of life long learning skills. Here is his response:

> Well, I'd say again I have no problem with accountability. I think we do need to be accountable. I think that we're passing—I think that we had a time where we could truly evolve what education is, and instead of doing that we're looking at trying to go back to some model that was idealized that didn't really exist 30 years ago. And that's sad to me that we're not taking this opportunity to change the way that we're doing things and what we're doing and how we're doing it. And instead we're looking at a multiple-choice test. The rest of our lives have changed. I can

> remember when my mom bought a 25-foot phone cord so she could sit on the porch and talk on the phone. And now people don't have a home phone. People think nothing of driving down the highway at 70 miles an hour making a phone call. You look at the music industry. The music industry fought against all the changes that were going on, and they suffered the consequences. And one of my fears is that public education is going to do the same thing. Is that we're being told to fight for something that no longer exists. And I think you'll see this with lots of different academy's coming in, lots of different charter schools, you'll see some other things where the parents in the areas that are sophisticated enough to take advantage of some of those opportunities are going to do it. And I fear that public schools are going to be left with those that aren't sophisticated, aren't well connected enough to search out other avenues. And that kind of worries me because I think that we do have some opportunities here, and we're chasing our tail. We're chasing this multiple-choice test. I think our greatest advantage has always been our creativity. It's been out willingness to reinvent things. It's been our willingness to redo things. And instead of going down that trail we're going down this very regimented based on- Young Zoë writes about how he grew up in China, and in China they have all these really, really good test takers. And you become exalted in your area if you're the best test taker for that area, but that's it. They don't have the creativity that's why they have the lower end market because they can provide cheap labor. What they can't do is they don't innovate. They don't create. And we seem to be chasing that model. And ironically they're chasing our model of much more openness, of kids having choices, of being a well-rounded kid instead of kid that can take a test.

It is evident the principal from Orange School feels that NCLB and AYP are measuring the wrong things. They are not measuring nor promoting creativity and higher level critical thinking skills. This finding is supported by the literature review with an increased focus on test taking skills under NCLB (Guilfoyle, 2006; Meier et al. 2004) and less of an emphasis on critical thinking skills (Guilfoyle, 2006; Schoen & Fusarelli, 2008).

A Single Measurement Accountability System

The triangulation of the data from Orange School provides support for the argument that a single high stakes measure might not be the most effective at measuring school success. When the state publishes their list of successful schools and failing schools, not one official from the state has ever stepped foot in the school to collect additional data or implement multiple measures. Yet, the visit to this school along with the data collected during the semi-structured interviews, indicates that the school had embraced the SIP, and was working hard with their data to target the subgroups. However, they were not getting the results based on the AYP measure. As previously identified in this chapter and in the literature review, it is probable that AYP was not measuring the implementation of effective SIP strategies in Orange School. The data indicates that AYP might have been measuring social capital in this school.

Several of the participants indicated that the reliance on a single-measurement accountability system was not fair or effective. Teacher 4 from Purple School provided the following response: "I don't feel like a school should be identified as a school that is making progress based solely upon one factor and that is standardized assessment scores." This finding is supported by the literature review that calls into question the use of a single measurement accountability system (Darling-Hammond, 2007a: Darling-Hammond, 2007b; Elmore, 2002; Harris, 2007; Hess & Petrilli, 2006; Kane, Douglas, & Geppert, 2002; Linn & Haug, 2002; Ravitch, 2010; Sunderman et al. 2005). There is a low level of confidence in the reliability and effectiveness with the implementation of a single measurement accountability system.

Narrowing the Curriculum

Findings form the literature review showed that there was a narrowing of the curriculum with the implementation of NCLB (Borowski & Sneed, 2006; Darling-Hammond, 2007a; Guilfoyle, 2006; Lewis, 2006; Maleyko & Gawlik, 2011; Sunderman et al, 2005). The areas that are being measured by NCLB (mathematics and reading) are the areas where schools are reallocating resources in order to ensure success on the state accountability assessments. This finding was supported by participants in the semi-structured interview protocol, along with an identification of school artifacts in the form of the SIP. All of the strategies that were mentioned by the participants (which are a response to NCLB) focus on mathematics, reading or writing.

There was an absence of the identification of strategies that focused on the other core curricular areas or the extended curriculum including physical education, the arts, or music. In fact, many mentioned that the extended curriculum was being eliminated and those resources were being placed in areas that supported mathematics and reading instruction. While placing additional resources in literacy and mathematics is a major goal of NCLB, an unintended consequence is the reduction or elimination of many of the extended curricular areas.

The principal from Red School in response to question 19 regarding the impact on the curriculum in the district or school responded in the following manner:

> Well, I think it's narrowing the curriculum. And its…. Taking away some of the choices kids have because they have to, you know, have to do all math. It kind of bothers me because I, you know, I think the purpose of education is to also help a child become more well rounded. But they also need to know math.

The principal from Orange School provided the following response to question 9 regarding the negative consequences of AYP:

> Oh, I think you're seeing a narrowing of the curriculum. You see the curriculum become reading and math, and you lose out on some of those creative aspects that really are or should be our strength. You become more and more teaching to this multiple-choice test. Because as the, as we move forward and as they tie this test to teacher, individual teachers, those teachers—I mean it's a very predictable response. Whatever it is that you test is what you're going to see. So you're going to see teachers teaching more and more to the test, instead of what kids really, really need in the long run.

Teacher 4 from Orange School also provided a response to question 9 regarding the negative impact indicating that there was less time to implement authentic learning opportunities and more pressure to cover material including the Grade Level Content Expectations (GLCE).

Potential for Teacher Turnover

The literature review provided empirical findings that indicated teachers would be more likely to transfer when a school was labeled as failing (Mintrop, 2004; Sunderman et al, 2005). Question 15 for both teachers and administrators asked participants if the AYP status of schools would impact their choice to teach in a school or remain in a school. The majority of the participants from Blue School and Red School indicated that it would not impact their decision. However, there were two exceptions, Teacher 2 from Blue School indicated that it would as she had experience teaching in an urban setting where the school had struggled to meet AYP. Teacher 4 from Red school also indicated that she might consider leaving due to the stress related to AYP.

The majority of the participants from Orange School and Purple School indicated that it would impact their decision. Two teachers from Purple School stated that that they would not leave the school but they referenced their seniority and the fact that it would be a financial burden for them to leave. Orange School and Purple School have experienced not making AYP while Red School and Blue School have never had a problem with AYP. Thus it is likely that the participants from Purple school and Orange school provided an answer partially due to the stress and low morale that had been related to not meeting AYP. This was an important finding within the coded componential domain analysis.

It is also likely that similar to question 5 (where participants were asked to give their input on whether schools with large populations of ED could achieve at the same level as schools with more affluent populations) there is an underlying philosophical value that educators can and will make a difference with the most challenging students. Thus it would be anticipated that most would want to say that the AYP status of schools would not impact their choice. However, the educators that felt the direct pressure of not meeting AYP were providing a realistic viewpoint with their answer vs. the more idealistic answer that they would be willing to work in schools that were struggling under the NCLB accountability reform. Listed below is a response by the principal from Orange school:

> It could because again the, unfortunately the reality becomes that somebody's going to tell you you're a failing school because of these numbers. You can be doing the greatest things with kids. You can be pulling kids from a non-reader to reading at the 3rd grade level in one year, and you're going to be called a failing school. That's insane to me. We've always known that there are other districts, there's other schools in this district, which have "better teachers" because their kids do better on tests. It has nothing to do with the teachers; they work as hard as a lot of my teachers do. I'll take a lot of my teachers over any of those teachers, but their kids are getting, doing well on the test so they're a "better teacher."

It is evident that the principal from Orange School has felt the stress and frustration from being labeled as a failing school under NCLB. His comments indicate that he believes they are working very hard to meet the needs of their students but that some of the demographic issues and/or the large ED and ELL population had a negative impact on student achievement on the state accountability assessment. His reference to other schools supports an earlier finding with both the quantitative and qualitative dataset that AYP (in his estimation) is measuring demographics or more specifically the EDPER variable.

Inability of AYP to Support the Internal Capacity of Schools

As it was previously indicated, one of the positive findings from the implementation of AYP was the increased sense of urgency to collect and analyze data including a focus on subgroup populations. It was also evident that the data analysis was assisting with the development of the SIP in schools. However, based on the interviews from participants, there was no indication that AYP provided any support beyond that sense of urgency to analyze data.

Most of the participants believed that the implementation of the SIP might have occurred prior to AYP. Participants did not feel that NCLB and AYP provided them with any direction or support to provide schools with the tools to advance themselves and student achievement. In fact, many felt that NCLB was an unfunded mandate. The Red School principal provided a response indicating that NCLB was not funded properly:

> So that part is a difficult thing for me. Especially because the state accountability test is a criterion-based test and it measures how the student is doing for that particular year and you're measuring all students. It should be a year's education, a year's growth. And I think that that's what we're moving towards in the new testing that's coming about finally, you know. So it was in No Child Left Behind, a lot of mandates were put in place but nothing was really funded. I have a difficult time with that. It brought about, you know for example, one thing here is we have to have foreign language. You have to have students graduating with two years of foreign language but nothing is funded for that. And then, the school district, you know, comes up with the assessment. Once again, everything's not equitable, you know. So, and then of course, all the mandates are being made and more and more funding is being cut from education. So I think that it's trying to hold the schools accountable but where is the accountability from the Federal Government for backing these programs up?

The principal from Red School along with the teachers from Red School provided responses indicating that they required additional resources in order to advance student achievement with their student population. Based on the ED status from Red School (and as previously mentioned), they were an outlier school that was beating the odds on the state accountability assessments. However, the teachers and administrators believed that funding and resources were needed in order to make that happen. Based on the coded analysis and triangulation of the data, it was also evident that strategic leadership and thoughtful research based implementation strategies were also necessary.

There were no answers from any of the participants indicating that they received any specific direction or support from NCLB as a result of AYP. Even the sanction provisions were somewhat questioned with regards to whether or not they could support increasing the capacity of schools. Participants from Orange School stated that the school of choice option would not work. The principal from Orange School provided the following response to the sanctions and/or tutoring options:

> But the sanctions are also changing right now. In part, you know, one of the most significant pieces of the law was that the only person that would lose their job would be the principal. There was a point, I believe it was after year four, where the principal had to be moved. And what's happened is they've discovered that they can't find other people. They can't find principals to put in those places. So that part of the law is, has now been excised. So some of the- with the reauthorization of no child left behind, they are making some changes but- And it's also interesting because some of the changes that they had, that they put in place--the outside tutoring for example, paying for outside tutoring--there's no statistical evidence that proves it works. But it does open up money to some of the private companies. And there are some that believe that that's really the genesis of no child left behind, is that there's a $376 billion market that the private sector can't get to and they would love to.

This response was also referenced earlier in this chapter. However, it was referenced again because it supports literature findings that the sanctions provisions are not supported by empirical research (Meier et al, 2004; Sunderman et al, 2005).

District Impact

The responses from the participants indicated that there was an impact on the district operation. However, the teacher participants were often negative about the direct impact. Many indicated that the district was implementing increased methods for analyzing data including the implementation of a centralized data management system to track student progress. The centralized data management system was referenced by at least one participant in all four of the sample schools. Thus it is evident that they believe there was an increased sense of urgency by the school district to analyze data in response to AYP.

The principal from Blue School also indicated that the district was analyzing subgroup data with the goal to close the achievement gap with the Black/African American population that was prevalent in her district. While she did not have a specific gap in her school, she was aware of the achievement gap on a district level as she provided the following response:

> There's no reason why kids at our district—Here you go-; here you go with black. Look at that, look at the discrepancy in our scores. I mean, holy cow! And this is grade three. There's no reason for that. But that's nationwide. And you can have those. So anyway. Now those are just grade school. That's just the data from the elementaries. But it carries through into the middle school and into the high school. And it’s become, it should be a national conversation but it's become a district conversation.

The principal from Blue School also indicated that the district response includes increased professional development. The majority of the participants in the interviews felt that data analysis and increased professional development were the district responses to AYP. However, the finding with professional development should be analyzed with caution because professional development was part of the district plans prior to NCLB. In the case of the district that Blue School belongs to, professional development could be linked to AYP because they were specifically attempting to close their district level achievement gap with the Black student population. However, based on the coded analysis, answers from other participants can not necessarily provide the same connections.

It is probable that the administrative answers to question 20 regarding the district impact provided informed opinions because they have more of a direct link to central administration vs. the teachers. Those responses included additional professional development, data analysis, and alignment of the curriculum to state accountability assessments. It was interesting that the principal from Purple School believed that there was district decentralization of decisions in response to AYP. However, based on some of the other answers, this response could have also been a reaction to budgetary issues. Teacher responses also included a direct link to the implementation of specific curriculum and the Grade Level Content Expectations (GLCE) in order to produce successful scores on the state accountability assessments.

Summary for Qualitative Findings

The motivation for the implementation of the semi-structured interview protocol was to answer research question four in this study and determine the impact that AYP is having on classroom instruction and school improvement initiatives at the school level. There was a clear indication that a positive impact was the sense of urgency that schools implemented around the analysis of data including disaggregating subgroup populations and the analysis of data that drives classroom instruction. The data analysis techniques that were implemented by the schools in the qualitative sample, led to the creation of research based school improvement strategies. However, it is not conclusive regarding the type of positive impact that AYP had on the SIP because many schools were engaged in the development and implementation of SIP prior to NCLB. It is reasonable to conclude that the SIP became a more intensive tool for schools as a result of AYP which is especially true among the schools that struggled to make AYP.

Similar to the finding in this study regarding educational resources, data analysis in itself did not necessarily lead to the implementation of an effective SIP or effective classroom instruction. The educational resources or funding provides schools with the ability to implement effective strategies. However, the school must then use that funding in an effective manner in order to make gains. That finding was also supported by the quantitative data analysis in this study and the literature review. There was a similar finding with data analysis, it set the parameters for schools to analyze student data and the subgroup populations. However, the second step was left to school to determine how to address disaggregating the data. Thus AYP in itself did not provide a formula for an increase in the internal capacity of schools to implement an effective SIP or effective classroom instruction. However, it did provide the first step which was a sense of urgency around the data. The school then had to determine how to make achievement gains for the benefit of disadvantaged students. This study also found that the data analysis that schools and districts implemented involved an increase in the usage of technology to assist with data collection and analysis. This finding was consistent among the schools in the sample.

Based on participant responses, it was also clear that AYP was creating a lot of stress among the faculty. This is especially true among Orange School and Purple School, both had experience with failing to meet AYP standards. However, a level of stress was also found at Red School in regard to the AYP requirements due to their large ED population. Blue School appeared to have less stress in response to AYP, however, the social capital of the students was at a higher level in that school with a very low percentage of ED students.

It was also evident that most of the participants did not feel that AYP was a fair measure. The triangulation of the data with the participant responses, collection of artifacts, literature review, and quantitative dataset from this study, indicate that the current AYP measure might be doing a fantastic job at measuring the level of ED populations and the amount of social capital in schools. However, it is not effective in measuring effective school improvement plans, classroom instruction, and increased student achievement.

The qualitative data also indicated that the AYP status of schools might impact a teacher's choice to teach or remain in a school. Thus an unintended consequence of AYP is teacher turnover. Another unintended consequence is a narrowing of the curriculum. The participant interviews along with the artifact data in the form of the SIP from the schools clearly indicate that mathematics, reading, and writing are the instructional priorities. That in itself is not necessarily a negative consequence. However, many of the schools are eliminating resources in specific curricular areas resulting in the narrowing of the curriculum. It is also evident based on the data collected from Purple School that AYP sanctions and the restructuring process are not necessarily effective in making positive change and improvements in the school. The restructuring in Purple School and subsequent success, was possibly a result of the statistical manipulation (Similar to a finding that was discussed in the literature review).

CHAPTER V

DISCUSSION AND CONCLUDING REMARKS

Recommendations for AYP reform

The intent of the NCLB reform is to improve the quality of classroom instruction and educational programs for disadvantaged students. The consensus found in the literature suggests that policies aligned with the AYP provisions in NCLB are not completely founded in empirical research given the goals of the reform. They were created on a political or ideological basis instead of a scientific research basis. For example, many of the findings and conclusions from the literature review (Borowski & Sneed, 2006; Elmore, 2002; Harris, 2007; Kane, et al. 2002; Lewis, 2006; Maleyko & Gawlik, 2011; Sunderman, et al. 2005) and findings in this study indicate that using a single once a year high stakes test for an accountability measure is unreliable in evaluating school effectiveness. The participants from the semi-structured interview believe that AYP and state accountability assessments were not a very good measure of school success that lead to increased student achievement.

The inconsistent results among the states with the level of success according to AYP and the state accountability assessment in the quantitative portion of this study support the findings with the referenced literature. Since the NAEP was measured in comparison to the state accountability assessment results in each of the four sample states, there was a consistent measure to analyze the data between the states. The empirical quantitative data from this study suggests that the AYP standards in North Carolina were much more rigorous than the AYP standards in Texas. Texas had the highest percentage of schools meeting AYP, yet the school level achievement on the Texas state accountability assessment provided for a very low correlation with NAEP proficiency status. The quantitative data also showed that the Michigan AYP standards were implemented with less rigor than the North Carolina or California AYP standards. The Texas state AYP and state accountability assessments provided for a very low correlation with the NAEP at both the proficient and basic level proficiency standard[39]. The AYP standards in Michigan showed that there was a strong relationship between the basic level NAEP scale standard and AYP. This finding is supported by empirical research that shows proficiency on most state accountability assessments is aligned with the basic level NAEP scale proficiency.

An important positive finding in this study involved an increased emphasis that schools were placing on data analysis in response to AYP. It was evident that schools were focusing on subgroup populations and disadvantaged students through the implementation of the SIP. However, AYP failed to provide schools with any specific guidance or research based best practices to increase their internal capacity. Each school was left on its own to develop those strategies. Elmore (2002) states that if there is a lack of support given to the internal capacity of schools to make improvements, then the sanction provisions will have little impact on improving classroom instruction as schools are labeled as failing according to AYP. The qualitative research findings in this study are aligned with Elmore's position. The interviews with the participants did not provide information which established a connection between AYP and an increase in the internal capacity of schools to make improvements. AYP also failed to channel additional resources to schools. In fact, many of the participants felt that there were many initiatives through NCLB that lacked proper funding.[40]

This study also found that there was an increased chance that quality educators might leave a school or fail to work in a school due to a failing label under NCLB. Schoen and

[39] Other than the 2007 8th grade dataset Texas

[40] As discussed in chapter 4, some of the participants referred to NCLB as an unfunded mandate.

Fusarelli (2008) interviewed educators and found that the accountability sanctions created pressure on principals who considered the possibility of leaving schools due to sanction provisions. Marzano, Waters, and McNulty (2005) found that school leadership has a substantial impact on student achievement. If high quality principals leave schools due to the pressures of probation, then it is probable that those schools will have less of a chance of making improvements. Improvements that Purple School made following their restructuring process was a result of the principal's leadership and SIP leadership team. As mentioned in chapter 4, the strong and viable leadership from the principal in Red School was a major reason for the school's success with the state accountability assessments despite the fact that they had a high percentage of ED students.

It was also evident that both Blue School and Orange School were led by experienced and skillful administrators. The difference between the two schools was that Orange School was facing the possibility of not meeting AYP standards for their second year and there was a great deal of stress and low morale among the faculty. It was also probable that the skillful principal might choose to leave the school due to the stress associated with the sanction provisions in NCLB. The principal from Blue School was not worried about meeting AYP and much of this had to do with not only the SIP, but also the social capital and affluent student population. It was evident based on visitations to the schools that both administrators were implementing an effective research based SIP. The difference was that Blue School was getting the results as measured by AYP and the state accountability assessments.

The findings associated with Blue School's success with the state accountability assessment and lack of success with the state accountability assessment at Orange School, provide an illustration of the broader findings from the quantitative dataset in this study. They showed that the EDPER variable was the greatest predictor of success as measured by NCLB and AYP. All of the datasets in the quantitative study provided consistent results with the negative impact that higher populations of ED students had on achievement with state accountability assessments and AYP. The findings in this study provide support to the literature review indicating that AYP is a measure of the social capital of students.

As mentioned, the EDPER variable was the only one that produced consistent results in almost all datasets with a negative impact on AYP and state accountability assessment results. The quantitative results from this study do not support the findings from Kane et al. (2002) regarding the racial subgroup populations. The subgroup populations had varying levels of impact on the state accountability assessment results and AYP among the different datasets. The findings from this research study did not indicate that AYP was a measure of the minority status and racial subgroups as the quantitative datasets were inconsistent with the impact that different subgroups had on state accountability assessment results and AYP. There were two exceptions as the BLACKPER in Michigan had a consistent negative association with a school's ability to meet AYP proficiency and the HISPANICPER and ELLPER variables had a consistent negative association in California. It is possible that the above mentioned subgroup populations in those states had a close correlation with the EDPER variable which explains why they produced consistent negative results.

Sunderman et al. (2005) criticizes the sanction provisions that were created in the NCLB legislation. They found that there is a lack of empirical data which shows that the school of choice provisions or supplemental service provisions will produce improvements for students. They point to the fact that the sanctions might actually weaken the conditions for students in schools as funds are diverted. The findings in the qualitative portion of this study support the findings by Sunderman et al. as many of the participants in the semi-structured interview protocol did not believe that the sanctions were effective at leading towards school improvement. The findings in the quantitative portion of the study also support the findings by Sunderman et al. as there was either a positive association with meeting AYP and state proficiency or a mild negative association. A conclusion from this study (as supported by the quantitative findings and the qualitative findings through a triangulation of the data) was that educational resources were a two-tiered variable. The manner in which schools used educational resources (funding) to support their SIP had an impact on school success level.

Borowski and Sneed (2006) believe that the accountability provisions and sanctions in the NCLB legislation are unproven and lack scientifically backed data that supports the implementation of those provisions. In their estimation, the legislation is a bandage job that attempts to balance between implementing federal intervention and allowing for local and state control of education. Thus, they believe the potential exists for grave harm as states have at times lowered standards or manipulated their statistical data in order to avoid the costly penalties and sanctions. The qualitative data from Purple School and their experience with the restructuring sanction support the views presented by Borowski and Sneed. Purple School made changes to their school by removing a grade from the building and thus lowering the number of students in the school. This might have looked like a formula for improvement as presented to the state, however, in the estimation of the interview participants from Purple School, restructuring never really led to improvements in the school. It was simply a method towards manipulation of the school structure to increase the odds of meeting AYP but not necessarily leading to improvements in achievement among the students in the school. The quantitative dataset also supports the arguments by Borowski and Sneed which showed that in 2007 NC changed the reading state accountability assessment which resulted in a dramatic increase in the number of schools that met AYP proficiency status.

Lewis (2006) argues that the there are major problems with the current model of accountability known as AYP. She believes that it needs to be eliminated and that the new model should be grounded in accountability research that will help to improve the conditions of schools. There is a lack of empirical research in the field that shows that the accountability provisions will improve school programs and classroom instruction. The majority of data that is referenced in the field claims improvements have been made as a result of the impact of NCLB exclusively rely on standardized test data. Data is not available on the direct impact that the accountability provisions are having on school improvement efforts and classroom instruction. This study addressed that issue through the collection of qualitative data via the semi-structured interview protocol. The results from the qualitative data show that there were some benefits of AYP including the increased sense of urgency for data analysis. However, AYP did not necessarily lead to increased achievement among the schools or the implementation of effective SIP plans and/or classroom instruction. The successful implementation of the SIP among the sample schools was a result of the leadership and internal capacity of the schools to implement an effective SIP and instructional techniques. AYP did create an increased pressure and sense of urgency for schools to find a way to become successful as measured by the state accountability assessments.

The Use of Growth Data for AYP

A large number of empirical research studies and theoretical articles (Elmore, 2002; Hess & Petrilli, 2006; Koretz, 2008; Peterson & West, 2006; Popham, 2005b; Schoen & Fusarelli, 2008; Sunderman et al. 2005; Wiley et al. 2005) recommend the use of growth data in the AYP formula. Research by Sunderman et al. (2005) found that the state accountability systems that were developed in each of the states were generally in contrast to the accountability provisions in NCLB. Most of the states used a form of growth data in order to evaluate the schools in their states. They did not solely rely on the absolute cut score method that is prevalent in NCLB. This often creates confusion for school stakeholders. For example, the schools in Chicago were subject to three accountability provisions at the national, state, and local levels -- all of which could create different results from the measures that were implemented in each system.

Peterson and West (2006) found that the state and federal accountability system can often create confusion for the public which was evident in their evaluation of the effectiveness of the AYP provisions in the state of Florida. Florida established a school report card system that ranks schools from A through F known as the “A+ Grading System”. The grades from this system are determined by a combination of both achievement data and growth data. At the time of this data collection, Florida had not yet been approved by the US department of Education to use the growth model for determining AYP.

Peterson & West (2006) used the Florida Comprehensive Achievement Tests (FACT) for their study for the 2002-2003 and 2003-2004 school years. They found that schools making AYP had .09 of a standard deviation higher than the amount gained by students at schools not making AYP. This standard deviation statistic was stated to be equivalent to 1/3 of a school year of student growth. The achievement in reading was .07 of a standard deviation higher in the schools that made AYP when compared to those schools that did not make AYP. In schools that made AYP, the standard deviation was .11 of a standard deviation higher for African Americans and .12 of a standard deviation higher for Hispanic students when compared to the same subgroups of students in schools that did not make AYP. There was a .06 of a standard deviation higher in the area of reading for both of those ethnic groups. When they completed a statistical analysis to take into account for the ELL, special education, socioeconomic status, ethnicity, and mobility rates, the average was a .04 of a standard deviation in mathematics and a .02 of a standard deviation higher in reading for schools that met AYP vs. those schools that did not meet AYP.

Schools in Florida that made AYP outperformed the schools that did not make AYP 71 percent of the time according to the analysis of student growth data on the mathematics assessments. This resulted in a 28 percent error rate between the state measurement system and the federal AYP measurement system. Peterson and West (2006) disaggregated the data to further analyze the differences according to growth data in relation to the Florida system that evaluates the schools on a 5 point scale A through F. While the learning gains were at a .07 standard deviation higher in schools that scored A versus those that scored F, the gains between the A schools and the F schools was .25 of a standard deviation or the equivalent of one school year of growth. In 2004, only 47 of the states' 2649 schools were given an F and 184 were given a grade of E. This is in contrast to a total of 75 percent of the schools that did not make AYP including more than half of the schools that received an A grade.

The findings by Peterson and West (2006) have important implications as the reauthorization of NCLB is being debated in Congress. The study shows that the AYP data does not have the ability to measure the growth that is occurring in schools. It brings into question the true goals in NCLB. For example, would a parent want to send their child to a school that is having more success with student gains vs. those that appear to be making the cut scores in NCLB? The Peterson and West study points out that the accountability system in Florida and the

NCLB accountability system creates confusion for Florida residents. Half of the schools that scored an A on the Florida accountability system are labeled as failing under the NCLB accountability system. The data further suggests that AYP has a large error rate when labeling schools as successful in accordance with the growth data.

The quantitative results from this study support the findings from Peterson and West (2004) as all four states in the study implemented different levels of rigor with their state accountability assessments and AYP. This was evident based on the different associations between state AYP (and the state accountability assessments) and the NAEP assessment. The statistical results from this study varied among the four states on both the Pearson Correlation test and the logistic regression analysis. In fact, the logistic regression was not statistically useful in Texas or Michigan in explaining the impact that the independent variables had on the dependent variable due to the high number of schools in the sample that met AYP proficiency. This was in contrast to the findings from North Carolina and California which produced more reliable and statistically useful logistic regression results.

The majority of the participants in the semi-structured interviews believed that AYP was not a fair measurement system. The use of an effective growth model which measures student gains might be an effective alternative to the current cut score model. However, the implementation of growth models must be evaluated with caution as there are different ways to implement a growth model. The current discussion with the reauthorization of NCLB is looking at replacing the common method of measuring AYP through a cut score measurement with a growth model. The US Department of Education has been experimenting with growth models by approving specific states with the implementation of the growth model formula.

The United States Department of Education has approved several states for the growth model formula, including Florida. Even though Florida has been approved, the Peterson and West (2004) study illustrates the inequities, inconsistencies, and unfair assessments that are being used by the AYP provisions in NCLB. Since some states are allowed to use a growth model and other states are not, this raises a concern regarding the consistency of the NCLB accountability standard and the direction that the federal government is looking for with the NCLB reform. The implementation of the growth model is also much different among the states. For example, North Carolina implements a formula that is much different than Michigan's formula. The AYP formula that is used in the growth model states is different from the AYP formula in other states and this further indicates that comparing schools in different states according to their AYP results is not reliable. The results in this study with the comparison of data from the four sample states (California, Michigan, North Carolina, Texas) provide a clear indication of the inconsistencies with the implementation of AYP among the states.

A Call for Multiple Measures

The use of a single measure with the evaluation of school effectiveness is not a reliable source of measurement. Several researchers (Darling-Hammond, 2007a; Elmore, 2002; Guilfoyle, 2006; Kane et al. 2002; Scheon & Fusarelli, 2008; Wiley et al. 2005) support the use of multiple measures in order to evaluate school effectiveness. Kane et al. (2002) examined the impact of the use of a single measure to evaluate school effectiveness in their study. They used data from California to show that schools with racially homogeneous populations have a better statistical chance of winning performance awards under the subgroup accountability provisions in that state. They point out that schools with larger minority subgroup populations have a greater chance of failure and that the NCLB reform might provide an incentive for states to lower standards so that more students meet the proficiency levels. As previously mentioned, the state of North Carolina might be an example of where this occurred. They changed their ELA state accountability assessment during the 2006-07 school year. The result was that a much higher percentage of schools met AYP requirements and state accountability proficiency level requirements in reading with the 2007 dataset vs. the 2005 dataset in North Carolina.

Many states will be making changes to the cut scores in the near future to be aligned with what is being called college readiness standards. The cut score standard for proficiency in Michigan is projected to be at a higher level for a student to be deemed proficient (Feldscher, 2011). However, this is another form of statistical manipulation which will give a false impression to the public that schools are less successful. The North Carolina 2007 dataset example provided for an impression that more schools were successful when the reality was that no changes were made. Now states like Michigan are moving in the other direction. The ironic thing with the so called "college readiness standard" which will start with a measurement on the 3rd grade MEAP in Michigan, is that they are making no changes to the actual assessment. The only thing that is changing is the cut score and a label is being put on it calling it "a college readiness measure". I would argue that the current MEAP assessment system in Michigan starting in grade three does not measure the skills necessary for success in college. The current MEAP assessment does not measure qualities that are needed for a successful college career. I am making this conclusion due to the fact that I am currently working on my fourth college degree and I know that skills like the ability to critically think, write effectively, collaborate with peers and the ability to discuss or argue points of view are much more important in college vs. memorization and/or answering multiple choice assessments in reading or mathematics.

Kane et al. (2002) identifies four different types of accountability systems that were used across the country prior to NCLB. States like Arizona, North Carolina, and Tennessee used a growth model. States such as Texas and Illinois reverted to a cross cohort comparison where different cohorts of students were compared from one year to the next. Other states like California rated their schools based on change in test scores from one year to another. Some states used accountability methods that looked at including a combination of growth and proficiency cut scores from year to year. The authors found that regardless of the system used, multiple years of data are needed in order to reliably evaluate school performance. The measure of cross cohort gains was found to be an unreliable measure especially when data was examined from year to year. This is the reason why two years of data were used to analyze the datasets in this study. The analysis of two years of different quantitative datasets from each of the four sample states allowed for more reliable findings in this study and the result was some interesting conclusions as discussed throughout the study.

Kane et al. (2002) found the size of the school has an impact on the ability of schools to make drastic improvements from year to year. Larger schools have less of a chance at making significant growth according to a one year measure, where small schools have a greater chance of making significant gains or losses from year to year. The authors used data from North Carolina's identification of successful schools and the failing schools to illustrate this point. They concluded that there is less variance in performance with larger schools. The example from Purple school where they lowered their overall student population provides support to this finding because the school met AYP for several consecutive years following the removal of 6^{th} grade from the building and a lowering of the overall student population (along with subgroups).

Standardized Assessments

The sole reliance on standardized assessments is a limitation in NCLB. This study in combination with the literature review has discovered that standardized assessments are not the most reliable form of measure to evaluate school effectiveness. Part of this has to do with the inconsistency and statistical manipulations that occur within the states as this study has highlighted. Standardized assessments are easy to use and calculate and thus they give school officials, parents, the politicians, and the public an easy method for ranking and sorting school performance. If standardized assessments were used to drive instruction in order to make improvements at the school level, they could become an asset to school officials when used in addition to other forms of measurement. However, the problem is that the standardized assessments have become such a high stakes measure and there is an immense amount of pressure to do well on the test. This pressure has led to the unintended consequences as discovered in this study which include teaching to the test at times, a narrowing of the curriculum, and the possibility that effective educators will not want to work in schools with high levels of categorical populations. This study clearly shows that there is a strong relationship between ED status, social capital, and success on state accountability assessments.

Guilfoyle (2006) is critical of the emphasis on the use of one assessment from different subgroups in order to classify schools as successful or failing. She uses a quote from the USDOE in order to emphasize her argument:

> If a single annual test were the only device a teacher used to gauge student performance, it would indeed be inadequate. Effective teachers assess their students in various ways during the school year. (Guilfoyle, 2006, p. 13)

Researchers (Darling-Hammond, 2007a; Elmore, 2002; Guilfoyle, 2006; Scheon & Fusarelli, 2008) argue that authentic learning projects and portfolios should be emphasized at a higher or equal level in NCLB. An emphasis on those types of assessments with multiple measures would promote higher level thinking and critical problem solving approaches to education. Students require real life 21st century skills in order to become successful in society.

The findings from this study support the arguments from the above mentioned researchers. The participants in the semi-structured interviews felt that AYP was an unfair measure that emphasized a single measurement system. The quantitative dataset also supported the literature to a certain degree as it was evident that the EDPER variable had a major impact on school level AYP results. Thus another system of measurement might be needed (including multiple measures) to effectively evaluate schools with high levels of ED students. It is evident that the AYP measurement system is not effective at measuring success levels with schools that have large populations of ED students. Thus additional measurement systems which might include a growth measure are definitely needed in order to improve the AYP accountability system.

Another form of measurement that could be used for the evaluation of school effectiveness is a process that has been implemented with a great deal of success by the Advanced Education Organization, formerly known as North Central Accreditation (NCA). This process involves the evaluation of school improvement plans, the inclusion of growth data, the use of multiple forms of achievement data, and the reliance on school visits to measure school effectiveness to determine accreditation status. While this measurement would be costly, it would be beneficial to explore this option as a part of the AYP measurement with regard to school effectiveness. This form of measurement might be what is needed to effectively evaluate schools with large ED populations.

It is probable that many would oppose the above mentioned process on a national or state level since it would involve subjective data sets which are more difficult to measure vs. standardized test data. However, it is recommended to explore this type of measurement tool along with other forms of multiple measures in the field including standardized test data with the inclusion of absolute cut score achievement data and growth level data. It is hard to believe that a school can be labeled as failing or successful when not a single official has ever stepped foot in the school. I can not imagine an effective evaluation of school, administrator, classroom teacher, or even local business for that matter, without having ever spoken to or meeting the individuals responsible for the performance in that school or business.

Standardized tests as a sole measure are not effective at measuring school performance. They fail to evaluate skills that are necessary for success in society as students leave public education. If a multiple measures approach was taken with the use of standardized assessments that include a growth measure and a school improvement committee evaluation process, there is the possibility for improving the AYP measurement system in NCLB. This could lead to effective measures of school performance vs. the current measurement system which tends to measure social capital as supported by the findings in this study.

National Curriculum and the Use of the NAEP

AYP provisions are not consistent from state to state as different states have different cut scores and their trajectories are often much different (Sunderman et al. 2005). Research in the field along with the quantitative results in this study suggests that even when one state might have higher standards when compared to another state based on the proficiency levels, the proficiency targets could be deceiving because the quality of the assessments used are much different in each state. The findings from this study show that the use of the NAEP assessment is an effective technique to analyze the consistency of AYP among the states. The findings also show that the standards implemented among the sample states in this study are much different. While Texas might be able to claim that their schools are successful, the result is that meeting AYP standards in Texas is much less rigorous than the other states in this sample. The comparison with the NAEP assessment in this study provided the ability to make that conclusion.

Borowski and Sneed (2006) argue that there is a need for national standards that protect the local rights and state rights for implementing educational policies and that minimal national standards are necessary. They emphasize that the arbitrary accountability provisions throughout the AYP process must change in order to equalize the accountability standards among the states in a fair manner. Hoxby (2005) believes that there is a need to create a national benchmark system in order to effectively evaluate school progress across the country.

The implementation of a national curriculum that uses the NAEP is gaining momentum with the discussions regarding the reauthorization of NCLB. Numerous states have recently committed to a Common Core Set of Standards that will be assessed by changes to their state accountability assessments within the next few years (Sloan, 2010). This momentum towards the creation of common core standards among many states provides support for the realization among the federal government regarding the inconsistent implementation of standards by the states. Since the legal authority to implement education falls to state governments, trying to get a common core set of standards approved during the initial implementation of NCLB in 2003 would have been difficult. However, this study along with empirical research in the field has found that states are implementing an inconsistent standard in their schools. The results is additional public and political pressure at the national level (and among many states) for a national curriculum standard. This public pressure is a positive aspect within current political forum to create much needed changes within the implementation of the NCLB reform.

Ravitch (2010) points out that we should establish a national curriculum that “declares our intention to educate all children in the full range of liberal arts and sciences, as well as physical education” (p, 231). However, it is possible that there would be court challenges to any form of national curriculum that is mandated by the federal government. Some researchers like Mathis (2004a) believe that one curriculum at the national level would not be wise. It is also possible that some states might choose to opt out of NCLB and forfeit their federal funding.

The state of South Dakota has recently decided not to participate in NCLB due to what they deem to be unrealistic AYP standards as the proficiency trajectory continues to rise to 100% for the year 2014 (Verges, 2011). The state is essentially protesting that they are not willing to engage in the perception of lower levels of achievement from schools in their state with the rise in the proficiency line of trajectory. They are in essence confronting the issue of statistical manipulation with the AYP measure. The price to the state is to forfeit of their federal Title One funding. They obviously feel that it is worth that price. Other states that receive lower levels of Title One funding are making the same sort of contemplations.

What We Still Need to Know

The findings from this study show that the current accountability provisions in NCLB have not been effective in evaluating school performance and at providing schools with a failing or successful label. It is apparent from both this study and the literature that there are numerous problems with the implementation of AYP in order to motivate effective school improvement efforts. Reforms to the legislation are needed in order to improve the reliability of measuring school effectiveness. This includes the need to standardize the AYP measurement system across the country while implementing multiple forms of measurements which include growth data.

Since the key to school improvement is through the enhancement of the internal capacity of the school while increasing the internal accountability mechanisms, it is critical that further research expands on the findings from this study in order to determine strategies for implementing effective SIP that increase internal capacity. The qualitative findings from this study clearly showed that data analysis that led to implementing a school SIP was a positive result from the implementation of AYP. However, it would be beneficial to conduct research that builds on the findings from this study which uncovered some of the effective practices and SIP strategies that were implemented in schools.

The positive impact of effective building leadership was a key finding in each school within the qualitative portion of this study. The implementation of an effective AYP formula that provides support to the establishment of effective school leadership capacity is an area that should be addressed. Simply removing staff including the principal does not guarantee effective reform within schools. Resources should be used to invest in the enhancement of leadership capacity and the internal accountability practices within schools.

This study provided findings regarding the impact that AYP is having on student learning, classroom instruction, and school improvement. The implementation of qualitative methodologies examined how teachers and administrators are responding to AYP. Moreover, research is still needed to examine the question of whether a school will improve if the staff is replaced. Will a school get better if it is turned into a charter school? Will it improve in measurable ways if the state takes it over? What policies could be implemented to attract and retain high-quality teachers in low performing urban schools? What will happen in the year 2014 when all schools are required to have 100% proficiency levels in order to meet AYP standards. These, and a good many other questions, remain to be answered within the context of NCLB.

APPENDIX A

Requirements for Schools Not Making AYP

in Michigan

(Michigan Department of Education)

The requirements for Title I schools that do not make Adequate Yearly Progress in English language arts or mathematics depend on the number of years for which the school has not made AYP. The requirements are designed to give Title I schools an opportunity to improve their programs with assistance from the school district and outside experts if the district determines that they are needed. If a Title I school continues not to make AYP, the district is required to take specific actions to improve student academic achievement in the school. At the same time as the improvement efforts are undertaken, students attending schools that do not make AYP in English language arts or mathematics are given other educational options. These options vary depending on the number of years the school has failed to make AYP. The specific requirements for Title I schools that do not make AYP in English language arts or mathematics are as follows:

Schools Not Making AYP for One Year. Because a school's MEAP results can change from year to year for many reasons, a school that does not make AYP for one year is given a second year to demonstrate that it can make AYP. There are no actions that the school or school district must take because a school has not made AYP for one year.

Schools Not Making AYP for Two Years. A Title I school that does not make AYP for two consecutive years is identified for improvement and must take the following steps to improve student academic achievement:

- Develop a two-year school improvement plan in consultation with parents, school district staff, and any outside expert who is providing assistance.
- Submit the plan to the district for peer review and district approval.
- Implement the improvement plan by the beginning of the school year following the year the school was identified.
- Spend at least 10 percent of its Title I allocation each year for the next two years on professional development that directly addresses the achievement problems that caused it to be identified.

When a Title I school is identified for improvement, the school district must also do the following:

- Offer students who are enrolled in the school the option to transfer to other schools in the district that are not identified for improvement, on a space-available basis.
- Provide or pay for transportation for students who choose the transfer option, within certain cost limits.
- Give priority to the lowest-achieving students from low-income families if there is not enough

space available in non-identified schools or funds to cover transportation costs. Districts with Title I schools identified for improvement must spend an amount equal to 20 percent of their Title I allocations for transportation and supplemental educational services, unless a smaller amount is needed. This includes 5 percent for transportation, 5 percent for supplemental educational services, and the remaining 10 percent for transportation, supplemental educational services, or both. Districts are not required to spend additional funds to meet the transportation requirement.

Schools Not Making AYP for Three Years. A Title I school that does not make AYP for three consecutive years continues to be identified for improvement. The school must continue to implement its revised school improvement plan and spend at least 10 percent of its Title I allocation on professional development to address the academic problems that caused it to be identified. The school district must continue to offer the transfer option and provide or pay for transportation. In addition, the school district must:

- Offer low-income students attending the school the opportunity to receive supplemental educational services outside of the school day.
- Give parents of eligible students the option of choosing among the state-approved providers in the area, on a space-available basis.
- Pay the costs of the supplemental educational services, within certain cost limits.**
- Give priority to the lowest-achieving students if there are not enough funds to cover the costs of supplemental services for all eligible students.
- **The maximum cost per student is the amount of Title I funds the district receives per low-income student or the actual cost of the supplemental services, whichever is less. The district maximum cost is an amount equal to 5 percent of its Title I allocation, with another 10 percent available for transportation, supplemental educational services, or both.

Schools Not Making AYP for Four Years. A Title I school that does not make AYP for four consecutive years is identified for corrective action. The school district must continue to offer the transfer option and supplemental educational services. The district must also take at least one of the following actions to improve student academic achievement in the school:

- Replace the school staff that are relevant to the failure to make AYP.
- Implement a new research-based curriculum and provide appropriate professional development for all relevant staff.
- Significantly decrease management authority at the school.
- Appoint an outside expert to advise the school on revising its school improvement plan to address the issues underlying its continued achievement problems.
- Extend the school year or the school day.
- Restructure the internal organization of the school. The school district is responsible for selecting the action or actions that are most likely to address the school's academic problems that caused it to continue to be identified. The district must notify parents and the public regarding the corrective action(s) it chooses to take.

Schools Not Making AYP for Five Years. A Title I school that does not make AYP for five consecutive years is identified for restructuring. The school district must continue to offer the transfer option and supplemental educational services. The district must also take at least one of the following actions to make fundamental changes in how the school is operated in order to improve student academic achievement:

- Reopen the school as a charter school.
- Replace all or most of the school staff that are relevant to the failure to make AYP.
- Enter into a contract to have an outside organization with a record of effectiveness to operate the school.
- Turn the operation of the school over to the state, if the state agrees.
- Restructure the school's governance arrangement in another way that makes fundamental reforms.

Before taking any action, the school district must notify parents and teachers that the school has been identified for restructuring and give them an opportunity to participate in the development of the restructuring plan.

Status of Identified Schools That Subsequently Make AYP

A Title I school that has been identified for improvement, corrective action or restructuring will have a change in status and the requirements that apply if it succeeds in making AYP. The specific provisions are as follows:

Schools Making AYP for One Year After They Have Been Identified for School Improvement, Corrective Action or Restructuring. If a Title I school makes AYP for one year after it has been identified for school improvement, corrective action or restructuring, the school district may delay taking any additional action for one year. The school district must allow students who had previously chosen the transfer option to continue to attend their new school and must provide or pay for transportation. One year of making AYP is not considered to be an interruption to the number of "consecutive" years a school has not made AYP or the school's resulting status.

Schools Making AYP for Two Consecutive Years After They Have Been Identified for School Improvement, Corrective Action or Restructuring. If a Title I school makes AYP for two consecutive years after it has been identified for school improvement, corrective action or restructuring, it is no longer identified or subject to the requirements for identified schools. The district must allow students who had previously chosen the transfer option to continue to attend their new school until they complete the highest-grade level in the school. However, the district is not required to provide or pay for transportation once the student's original school is no longer identified for improvement.

APPENDIX B

NCLB Purpose Statement

Title I — Improving The Academic Achievement Of The Disadvantaged

SEC. 101. IMPROVING THE ACADEMIC ACHIEVEMENT OF THE DISADVANTAGED.

Title I of the Elementary and Secondary Education Act of 1965 (20 U.S.C. 6301 et seq.) is amended to read as follows:

TITLE I--IMPROVING THE ACADEMIC ACHIEVEMENT OF THE DISADVANTAGED

SEC. 1001. STATEMENT OF PURPOSE.

The purpose of this title is to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments. This purpose can be accomplished by —

(1) ensuring that high-quality academic assessments, accountability systems, teacher preparation and training, curriculum, and instructional materials are aligned with challenging State academic standards so that students, teachers, parents, and administrators can measure progress against common expectations for student academic achievement;

(2) meeting the educational needs of low-achieving children in our Nation's highest-poverty schools, limited English proficient children, migratory children, children with disabilities, Indian children, neglected or delinquent children, and young children in need of reading assistance;

(3) closing the achievement gap between high- and low-performing children, especially the achievement gaps between minority and nonminority students, and between disadvantaged children and their more advantaged peers;

(4) holding schools, local educational agencies, and States accountable for improving the academic achievement of all students, and identifying and turning around low-performing schools that have failed to provide a high-quality education to their students, while providing alternatives to students in such schools to enable the students to receive a high-quality education;

(5) distributing and targeting resources sufficiently to make a difference to local educational agencies and schools where needs are greatest;

(6) improving and strengthening accountability, teaching, and learning by using State assessment systems designed to ensure that students are meeting challenging State academic achievement and content standards and increasing achievement overall, but especially for the disadvantaged;

(7) providing greater decision making authority and flexibility to schools and teachers in exchange for greater responsibility for student performance;

(8) providing children an enriched and accelerated educational program, including the use of school wide programs or additional services that increase the amount and quality of instructional time;

(9) promoting school wide reform and ensuring the access of children to effective, scientifically based instructional strategies and challenging academic content;

(10) significantly elevating the quality of instruction by providing staff in participating schools with substantial opportunities for professional development;

(11) coordinating services under all parts of this title with each other, with other educational services, and, to the extent feasible, with other agencies providing services to youth, children, and families; and
(12) affording parents substantial and meaningful opportunities to participate in the education of their children.

APPENDIX C

NCES Selection Procedures for Sample States

The Common Core of Data (CCD) file, a comprehensive list of operating public schools in each jurisdiction that is compiled each school year by the National Center for Education Statistics (NCES), is used as the sampling frame for the selection of sample schools. The CCD also contains information about grade span, enrollment, and location of each school. In addition to the CCD list, a set of specially sampled jurisdictions is contacted to determine if there are any newly formed public schools that were not included in the lists used as sampling frames. Considerable effort is expended to increase the survey coverage by locating public schools not included in the most recent CCD file. (NCES, 2008)

As part of the selection process, public schools are combined into groups known as strata on the basis of various school characteristics related to achievement. These characteristics include the physical location of the school, extent of minority enrollment, state-based achievement scores, and median income of the area in which the school is located. Stratification of public schools occurs within each state. Combining schools within strata by such selected characteristics provides a more ordered selection process with improved reliability of the assessment results. (NCES, 2008)

On average, a sample of approximately 100 grade-eligible public schools is selected within each jurisdiction; within each school, about 60 students are selected for assessment. Both of these numbers may vary somewhat, depending on the number and enrollment size of the schools in a jurisdiction, and the scope of the assessment in the particular year. Students are sampled from a roster of individual names, not by whole classrooms. The total number of schools selected is a function of the number of grades to be assessed, the number of subjects to be assessed, and the number of states that elect to participate in the program, as well as the desired precision of the survey estimates, and the average number of students to be selected within a sampled school. (NCES, 2008)

APPENDIX D

NAEP Subgroup Ranges and Median Values used in the study

Demographic Variable Range Definitions: According to NCES Protocol

Group Number Value for Statistical Analysis	Subgroup population range percentages in the school (rounded to the nearest percentile)	Median
0	0%	0
1	1%-5%	3%
2	6%-10%	8%
3	11%-25%	19%
4	26%-50%	38%
5	51%-75%	63%
6	76%-90%	83%
7	Over 90%	95%

APPENDIX E

NAEP Item Response Theory

Item Response Theory (IRT) is a set of statistical models that describe the relationship between assessment items and proficiency in a subject or skill area. Through the application of IRT models, mathematical functions for the probability of a student responding correctly to a specific test question or getting a particular score in case of open-ended response items are derived, given the proficiency level and characteristics of the questions on the test.

In IRT procedures, proficiency, the latent trait, is indirectly measured through students' performances on test items. IRT models assume that student performance on assessment items reflects both characteristics of the items and proficiency of the student.

Item response theory models are based on the assumption of conditional independence: the probability of correct responses is independent across items given the individual's proficiency. Conditional independence implies that a student's ability to answer items correctly is unrelated to the student's background variables. It also implies that the position of the item in the test booklet, content around the item of interest and the conditions of the test administration do not affect students' ability to answer items correctly.

NAEP adopted IRT models in 1984 to estimate students' proficiency on various subscales. NAEP currently uses three main types of IRT models: the two-parameter model (2PL), the three-parameter model (3PL) and the generalized partial credit model. Differential item functioning analyses [hyperlinked] are conducted to ensure that conditional independence was maintained. (USDOE, NCES, 2001)

APPENDIX F

Plausible Values

Plausible values are proficiency estimates for an individual NAEP respondent, drawn at random from a conditional distribution of potential scale scores for all students in the sample who have similar characteristics and identical patterns of item responses. They are based on a large marginal maximum likelihood regression in which the regressors are about 200 principle components that account for about 90 percent of the total variance of the full set of background variables. NAEP usually assigns five plausible values to each respondent. They are not test scores for individuals in the usual sense but are used as intermediary computations for calculating summary statistics for groups of students. The online manual for AM Statistical Software explains plausible values as the following:

Plausible values were developed as a computational approximation to obtain consistent estimates of population characteristics in assessment situations where individuals are administered too few items to allow precise estimates of their ability. Plausible values represent random draws from an empirically derived distribution of proficiency values that are conditional on the observed values of the assessment items and the background variables. The random draws from the distribution represent values from the distribution of scale scores for all adults in the population with similar characteristics and identical response patterns. These random draws or imputations are representative of the score distribution in the population of people who share the background characteristics of the individual with whom the plausible value is associated in the data.

Because each respondent is given relatively few items in a scaling area, the uncertainty associated with his or her ability, is too large to be ignored and thus estimates for ability can be seriously biased. To address this problem, NAEP computes five plausible values on each subscale for each student. (USDOE, NCES, 2001)

APPENDIX G

NCES Restricted Access Data Regulations

Under the National Assessment of Educational Progress Authorization Act (Public Law 107-279 III, section 303), the Commissioner of the National Center for Education Statistics (NCES) is charged with ensuring that NAEP tests do not question test-takers about personal or family beliefs or make information about their personal identity publicly available. After publishing NAEP reports, NCES makes data available to researchers but withholds students' names and other identifying information. The names of all participating students are not allowed to leave the schools after NAEP assessments are administered. Because it might be possible to deduce from data the identities of some NAEP schools, researchers must promise, under penalty of fines and jail terms, to keep these identities confidential (NCES, 2008).

All of the data that is on the website is coded to protect the identity of the schools and students. In order to identify the schools that are assessed by NCES using the NAEP, an application was filled out and affidavits were signed in order to guarantee confidentiality of the restricted use data. Federal law mandates that this data must be kept confidential.

According to the NCES guidelines, once the data is accessed, it is required to be housed in a secure location at Wayne State University and it can only be accessed at this site. The data was kept in a locked cabinet and the electronic data was stored on a secured computer. The only people who had access to this data was the primary researcher and the research advisor. The appropriate NCES approval standards were met and NCES provided a CD with an encryption key that allowed for the identification of the schools through the data sheets that were downloaded from the NCES website.

APPENDIX H

AM Statistical Software

AM is a statistical software package for analyzing data from complex samples, especially largescale assessments such as the National Assessment of Educational Progress (NAEP) and the Third International Mathematics and Science Studies (TIMSS).
From its origin as a specialized tool for analyzing large-scale assessment data, AM has evolved into a more generalized and growing tool for analyzing data from complex samples in general. Originally, AM was developed to estimate regression models through marginal maximum likelihood (MML). Because large-scale assessments are often low-stakes assessments for students, students are usually asked to respond to only a few items; each student sees only part of the whole test. Otherwise, they would be unlikely to expend real effort on any items. As a result, individual test scores are subject to substantial measurement error, which would bias many statistical estimates. Rather than assign each student an error-filled score, MML procedures represent each student's proficiency as a probability distribution over all possible scores. MML procedures use these probability distributions in the estimation process.

Another characteristic of large-scale assessments has led to a wider applicability of AM—they almost always draw a sample from a complex design. AM automatically provides appropriate standard errors for complex samples using a Taylor-series approximation. This happens automatically even when new procedures are added to the software. Over time, the software has grown to offer a set of non-MML statistics, including regression, probit, logit, cross-tabs, and other statistics that are useful for survey data in general.
The American Institutes for Research is committed to keeping AM available as a free and growing tool for the research community. Visit this web site for further information, updates, and technical support (American Institute for Research, 2010a)

APPENDIX I

NAEP Weighting Procedures

Since each selected school that participates in the assessment effort and each student assessed represent only a portion of the full population of interest, weights are applied to both schools and students. Every sampled school and student received a base weight equal to the reciprocal of its probability of selection. Base weights can vary because selection probabilities can vary. The final (full-sample) student weight is computed as the product of the student base weight, the student nonresponse adjustment, and the student trimming factor. For non-response, NAEP uses two adjustment factors: one to adjust for sessions that were not conducted and another for students who were selected to take the assessment but did not. NAEP also employs a trimming algorithm to decrease weights for students who came from schools that contributed more than the specified proportion.

Sampling weights are needed to make valid inferences about the populations from which they were drawn. NAEP uses differential sampling weights, obtaining larger samples of respondents from certain subgroups, in order to enhance the precision of estimates for those subgroups. Appropriate sampling weights are computed to obtain unbiased estimates of population characteristics.

The weights permit valid inferences to be drawn between the student samples and the respective populations from which they were drawn and, most importantly, ensure that the results of the assessments are fully representative of the target populations. This procedure also permits the preparation of unbiased estimates of standard errors.

APPENDIX J

NAEP Reading Achievement Levels by Grade

Specific definitions of the *Basic*, *Proficient*, and *Advanced* achievement levels for grades 4, and 8 are presented in the tables that follow. The achievement levels are cumulative. Therefore, students performing at the *Proficient* level also display the competencies associated with the *Basic* level, and students at the *Advanced* level also demonstrate the skills and knowledge associated with both the *Basic* and the *Proficient* levels. For each achievement level listed, the scale score that corresponds to the beginning of that level is shown in parentheses.

Grade 4

Basic **(208)**	Fourth-grade students performing at the *Basic* level should demonstrate an understanding of the overall meaning of what they read. When reading text appropriate for fourth-graders, they should be able to make relatively obvious connections between the text and their own experiences and extend the ideas in the text by making simple inferences.
Proficient **(238)**	Fourth-grade students performing at the *Proficient* level should be able to demonstrate an overall understanding of the text, providing inferential as well as literal information. When reading text appropriate to fourth grade, they should be able to extend the ideas in the text by making inferences, drawing conclusions, and making connections to their own experiences. The connection between the text and what the student infers should be clear.
Advanced **(268)**	Fourth-grade students performing at the *Advanced* level should be able to generalize about topics in the reading selection and demonstrate an awareness of how authors compose and use literary devices. When reading text appropriate to fourth grade, they should be able to judge text critically and, in general, to give thorough answers that indicate careful thought.

Top

Grade 8

Basic **(243)**	Eighth-grade students performing at the *Basic* level should demonstrate a literal understanding of what they read and be able to make some interpretations. When reading text appropriate to eighth grade, they should be able to identify specific aspects of the text that reflect overall meaning, extend the ideas in the text by making simple inferences, recognize and relate interpretations and connections among ideas in the text to personal experience, and draw conclusions based on the text.

***Proficient* (281)**	Eighth-grade students performing at the *Proficient* level should be able to show an overall understanding of the text, including inferential as well as literal information. When reading text appropriate to eighth grade, they should be able to extend the ideas in the text by making clear inferences from it, by drawing conclusions, and by making connections to their own experiences—including other reading experiences. *Proficient* eighth-graders should be able to identify some of the devices authors use in composing text.
***Advanced* (323)**	Eighth-grade students performing at the *Advanced* level should be able to describe the more abstract themes and ideas of the overall text. When reading text appropriate to eighth grade, they should be able to analyze both meaning and form and support their analyses explicitly with examples from the text; they should be able to extend text information by relating it to their experiences and to world events. At this level, student responses should be thorough, thoughtful, and extensive.

USDOE, NCES, 2009

APPENDIX K

NAEP Mathematics Achievement Levels by Grade

Specific definitions of the *Basic*, *Proficient*, and *Advanced* achievement levels for grades 4, and 8 are presented in the tables that follow. Because of changes made to the NAEP mathematics framework in 2005, the achievement-level descriptions and cut points indicated below for grade 12 have been updated. To maintain trend, results for grades 4 and 8 are reported on a 0–500 scale.

The achievement levels are cumulative; therefore, students performing at the *Proficient* level also display the competencies associated with the *Basic* level, and students at the *Advanced* level also demonstrate the skills and knowledge associated with both the *Basic* and the *Proficient* levels. The cut score indicating the lower end of the score range for each level is noted in parentheses.

Grade 4

***Basic* (214)**

Fourth-grade students performing at the *Basic* level should show some evidence of understanding the mathematical concepts and procedures in the five NAEP content areas.

Fourth-graders performing at the *Basic* level should be able to estimate and use basic facts to perform simple computations with whole numbers, show some understanding of fractions and decimals, and solve some simple real-world problems in all NAEP content areas. Students at this level should be able to use—though not always accurately—four-function calculators, rulers, and geometric shapes. Their written responses will often be minimal and presented without supporting information.

***Proficient* (249)**

Fourth-grade students performing at the *Proficient* level should consistently apply integrated procedural knowledge and conceptual understanding to problem solving in the five NAEP content areas.

Fourth-graders performing at the *Proficient* level should be able to use whole numbers to estimate, compute, and determine whether results are reasonable. They should have a conceptual understanding of fractions and decimals; be able to solve real-world problems in all NAEP content areas; and use four-function calculators, rulers, and geometric shapes appropriately. Students performing at the *Proficient* level should employ problem-solving strategies such as identifying and using appropriate information. Their written solutions should be organized and presented both with supporting information and explanations of how they were achieved.

***Advanced* (282)** **Fourth-grade students performing at the *Advanced* level should apply integrated procedural knowledge and conceptual understanding to complex and non routine real-world problem solving in the five NAEP content areas.**

Fourth-graders performing at the *Advanced* level should be able to solve complex and non routine real-world problems in all NAEP content areas. They should display mastery in the use of four-function calculators, rulers, and geometric shapes. The students are expected to draw logical conclusions and justify answers and solution processes by explaining why, as well as how, they were achieved. They should go beyond the obvious in their interpretations and be able to communicate their thoughts clearly and concisely.

Back to Top

Grade 8

***Basic* (262)** **Eighth-grade students performing at the *Basic* level should exhibit evidence of conceptual and procedural understanding in the five NAEP content areas. This level of performance signifies an understanding of arithmetic operations—including estimation—on whole numbers, decimals, fractions, and percents.**

Eighth-graders performing at the *Basic* level should complete problems correctly with the help of structural prompts such as diagrams, charts, and graphs. They should be able to solve problems in all NAEP content areas through the appropriate selection and use of strategies and technological tools—including calculators, computers, and geometric shapes. Students at this level also should be able to use fundamental algebraic and informal geometric concepts in problem solving.

As they approach the *Proficient* level, students at the *Basic* level should be able to determine which of the available data are necessary and sufficient for correct solutions and use them in problem solving. However, these eighth-graders show limited skill in communicating mathematically.

***Proficient* (299)** **Eighth-grade students performing at the *Proficient* level should apply mathematical concepts and procedures consistently to complex problems in the five NAEP content areas.**

Eighth-graders performing at the *Proficient* level should be able to conjecture, defend their ideas, and give supporting examples. They should understand the connections between fractions, percents, decimals, and other mathematical topics such as algebra and functions. Students at this level are expected to have a thorough understanding of *Basic* level arithmetic operations—an understanding sufficient for problem solving in practical situations.

Quantity and spatial relationships in problem solving and reasoning should be familiar to them, and they should be able to convey underlying reasoning skills beyond the level of arithmetic. They should be able to compare and contrast mathematical ideas and generate their own examples. These students should make inferences from data and graphs, apply properties of informal geometry, and accurately use the tools of technology. Students at this level should understand the process of gathering and organizing data and be able to calculate, evaluate, and communicate results within the domain of statistics and probability.

Advanced **(333)**	**Eighth-grade students performing at the *Advanced* level should be able to reach beyond the recognition, identification, and application of mathematical rules in order to generalize and synthesize concepts and principles in the five NAEP content areas.** Eighth-graders performing at the *Advanced* level should be able to probe examples and counterexamples in order to shape generalizations from which they can develop models. Eighth-graders performing at the *Advanced* level should use number sense and geometric awareness to consider the reasonableness of an answer. They are expected to use abstract thinking to create unique problem-solving techniques and explain the reasoning processes underlying their conclusions.

USDOE, NCES, 2009

APPENDIX L

NAEP Non Cognitive Items and Questionnaires

In addition to assessing subject area achievement, NAEP collects information from participating students, teachers, and schools about background variables that are related to student achievement. This information serves, in part, to fulfill reporting requirements of federal legislation. Specifically, under the No Child Left Behind Act, NAEP is required to collect information on and report achievement results disaggregated by the following variables, when possible: gender, race and ethnicity, socioeconomic status (SES), disability status, and English language learner (ELL) status. (Note that the term *English language learner* is used in NAEP 2005 reports; the term *limited English proficient* was used before 2005 and was used on all SD/ELL questionnaires administered to schools up to and including 2005.) Information from the background items also serves to give context to NAEP results and/or allow researchers to track factors associated with academic achievement.

Recent History

In early 2002, the National Assessment Governing Board was granted final authority over the background items. The Board adopted a policy to focus NAEP background data on the primary purpose of the National Assessment—to provide sound, timely information on the academic achievement of students in the United States (National Assessment Governing Board, 2003). The Board also initiated a process to prepare a general framework to guide the collection and reporting of background data. The Background Information Framework for the National Assessment of Educational Progress (NAEP), developed in 2003, defines the purpose and scope of NAEP background data, and calls for a long-term plan for continued development. In response to this call, the National Center for Education Statistics (NCES) developed the NCES Plan for NAEP Background Variable Development, which provides a general procedural map for the development and review of each type of background data (National Center for Education Statistics, 2004).

Types of Background Items

There are three types of background data: student reporting categories, other contextual/policy information, and subject-specific information. While there are some differences in the approaches to the development of each type of data, shared principles underlie all three: The Governing Board provides initial guidance on what will be developed; the Board has multiple opportunities to review and provide input; and the overall development process seeks to reduce burden on respondents and ensure data quality while continuing to meet the needs of the NAEP program. Descriptions of the three types of background data are below.

- General Student Reporting Categories
    - Since the first NAEP assessment in 1969, achievement results have been disaggregated by subgroups of the population. Achievement has also been presented for and compared across subgroups. As mentioned earlier, since the

inception of the No Child Left Behind Act, NAEP has collected information on and reported achievement results disaggregated by the following variables: gender, race and ethnicity, socioeconomic status, disability status, and English language learner status.

  - NCES monitors the quality of the data collected using the current measures and will develop new approaches to measuring student reporting variables when warranted.
  - One new approach NCES is currently investigating is the creation of a new and improved measure of socioeconomic status (SES). NCES commissioned a literature review of how SES has been defined and operationalized in other education studies and other fields such as health and marketing. This led to the idea of adopting a two-pronged approach to measuring SES. This approach involves (1) creating an enhanced student background questionnaire with items that probe resources in the home, parents' education level, and parents' employment status, among other variables; and (2) using geocoding software to link students' home addresses to aggregate SES data available from the United States Bureau of the Census. Development of the new SES measure commenced in 2005, with the goal of piloting it in 2009 and possibly implementing it in 2011.
- Contextual/Policy Information
  - In every assessment, NAEP collects data on basic characteristics of the school and student body in the school; teacher background, qualifications, and experience; and several student characteristics. These variables provide a basic context for achievement.
  - In addition to these core variables, timely policy/contextual issues are rotated across assessments. NCES convenes a policy/contextual issues panel when needed to identify policy/contextual issues that NAEP might address in the future, and to outline the relevant constructs and identify data needed to address these issues.
- Subject-Specific Information
  - The subject-specific items in NAEP are focused and limited. A set of key issues within each subject area will be addressed in a focused and in-depth manner across the life of each assessment framework.
  - When a new assessment framework is approved, NCES reviews the recommendations for background data made by the framework committee. Since 2003, NCES then develops an issues paper to reflect those priorities, identifies the data needed to address the issues, and develops a proposed schedule for rotating topics.

Background items associated with the categories described above are placed within student, teacher, school, and/or SD/ELL questionnaires, as appropriate. The placement of items and content of each questionnaire depend on the questionnaire respondent and the specific subject(s) NAEP is assessing in a given year. Often questionnaires measure similar constructs across respondents and/or subjects to provide additional information and, in some cases, to validate findings.

- Student background questionnaires ask respondents to provide information about factors such as race or ethnicity, school attendance, and academic expectations. Responses to items on the questionnaires also provide information about factors associated with academic performance, students' educational settings and experiences, students' effort on the assessment, and the difficulty and importance of the assessment.
- Teacher questionnaires ask respondents to indicate teacher background, training, and instructional practices (completed by teachers at grades 4 and 8. NAEP typically does not collect teacher information at grade 12.)
- School questionnaires ask respondents to provide information on school policies and characteristics (completed by the principal or assistant principal.)
- Questionnaires about students with disabilities or English language learners questionnaires ask respondents to provide information about students selected in the sample who have disabilities or limited English proficiency (completed by a special education teacher, a bilingual education/English-language-learner teacher, or a staff member who is most familiar with the student.)

In 2006 NCES also administered a department head questionnaire for grade 12 economics. Within each participating school, the questionnaire was administered to the chair or lead teacher of every department that offered at least one economics-related course. The questionnaire asked the respondent to provide information about the characteristics of the department's faculty, hiring requirements, and courses offered by the department. There are currently no plans to administer the department head questionnaire in every NAEP assessment.

Background Item Development Process

Background items are developed through a process similar to that used for developing the cognitive items. It includes reviews by external advisory groups and field testing. When developing the items, NAEP ensures that the items do not infringe on respondents' privacy, that they are grounded in educational research, and that the answers can provide information relevant to the subject being assessed. The following is an overview of the development process for background items:

1. The National Assessment Governing Board oversees the development of the content framework and item specifications for the background items. More details about this process are provided in the Background Information Framework (National Assessment Governing Board, 2003).

2. When a new assessment framework is approved, or when new policy issues are identified for NAEP to address, NCES develops an issues paper to reflect the new priorities, identify the data needed to address the issues, and propose an item rotation plan. The development of the issues paper involves convening a panel of experts in the relevant fields to help identify issues and then conducting a literature review to identify recent developments for the respective issues.

3. NAEP contractors that specialize in survey development draft and revise background items based on the recommendations of the issues paper and expert panel. Again, issue-specific working groups and expert panels are convened to provide input on the items.

4. NCES then reviews the background items to ensure fairness and quality so that NAEP's mission of providing a fair and accurate measure of student achievement and achievement trends over time is fulfilled (see NCES Statistical Standards).

5. The items are piloted, and the results are analyzed.

6. Based upon pilot data results, some items are revised.

7. The background items once again undergo reviews by item development contractors and then by NCES.

8. NCES presents items to the Governing Board for its approval, as specified in Education Sciences Reform Act, P.L. 107-279 . The Board has "final authority on the appropriateness of all assessment items" and is required "to take steps to ensure that all items selected for use in the National Assessment are free from racial, cultural, gender, or regional bias and are secular, neutral, and non-ideological."

9. The items are then submitted for clearance by NCES to the Office of Management and Budget, which checks to make sure the items comply with government policies.

10. Once clearance is received, each background item is typeset into the respective student, teacher, school, and/or SD/ELL questionnaires.

Background Data

The purpose of administering background items is to give context to NAEP results and/or to track factors associated with academic achievement. The data are also the basis for NAEP's major reporting groups. Therefore, it is important to note that since NAEP is based on a cross-sectional design, it is not possible to infer cause-and-effect relationships—it cannot prescribe what should be done. Rather its descriptions of the educational circumstances of students at various achievement levels—considered in light of research from other sources—may provide important information for public discussion and policy action (National Assessment Governing Board, 2003). For more information regarding how NAEP data is analyzed and reported refer to the "Results" section of NAEP's Frequently Asked Questions or the Background Information Framework for the National Assessment of Educational Progress (NAEP) developed by the Governing Board. For more information on how you can explore and manipulate NAEP data, go to the NAEP Research e-Center or the NAEP Data Explorer. Please note that in the NAEP Data Explorer, the results of the background questionnaires are sorted into eight broad categories:

- major reporting groups,
- student factors,

- factors beyond school,
- instructional content and practice,
- teacher factors,
- school factors,
- community factors, and

APPENDIX M

NAEP Validity and Reliability

Because NAEP findings have an impact on the public's understanding of student academic achievement, precautions are taken to ensure the reliability of these findings. In its current legislation, as in previous legislative mandates, Congress has called for an ongoing evaluation of the assessment as a whole. In response to these legislative mandates, the National Center for Education Statistics (NCES) has established various panels of technical experts to study NAEP, and panels are formed periodically by NCES or external organizations, such as the National Academy of Sciences, to conduct evaluations. The Buros Center for Testing, in collaboration with the University of Massachusetts/Center for Educational Assessment and the University of Georgia, is currently conducting an external evaluation of NAEP.

NCES understands that ensuring both reliability and validity of the test is essential for the success of the NAEP and the impact that it has on public education. Since the NAEP involves both multiple choice test items and constructed response test items, NCES institutes the following practices to ensure reliability of scoring:

While multiple-choice questions allow students to select an answer from a list of options, constructed-response questions require students to provide their own answers. Qualified and trained raters score constructed-response questions. (NCES, 2008)

Scoring a large number of constructed responses with a high level of reliability and within a limited time frame is essential to NAEP's success. (In a typical year, over three million constructed responses are scored.) To ensure reliable, quick scoring, NAEP takes the following steps:

- develops focused, explicit scoring guides that match the criteria delineated in the assessment frameworks;
- recruits qualified and experienced scorers, trains them, and verifies their ability to score particular questions through qualifying tests;
- employs an image-processing and scoring system that routes images of student responses directly to the scorers so they can focus on scoring rather than paper routing;
- monitors scorer consistency through ongoing reliability checks;
- assesses the quality of scorer decision-making through frequent monitoring by NAEP assessment experts; and
- documents all training, scoring, and quality control procedures in the technical reports. (NCES, 2008)

NAEP assessments generally contain both constructed-response and multiple-choice questions. The constructed responses are scored using the image-processing system, whereas the responses to the multiple-choice questions are scored by scanning the test booklets. (NCES, 2008)

APPENDIX N

California: Standardized Testing and Reporting (STAR) Validity and Reliability

This appendix is a description of the information that is provided on the California Department of Education (CDE) website which supports the measures that are taken in order to ensure the validity of the STAR assessment which will be used in this study. The CDE uses the STAR assessment tool to measure the California content standards in mathematics and English Language Arts.

Content validity refers to the degree to which the content of a test is congruent with the purpose of the testing, as determined by subject matter experts of the highest caliber. CST items were developed to align with the California Content Standards that are representative of the broader content domains: English–language arts, mathematics, science, and history–social science. Thus, the content-related evidence of validity concerns the extent to which the test items represent these specified content domains and cognitive dimensions. (California Department of Education, 2007)

Content validity also provides information about how well an item measures its intended construct. Such validity is determined by a critical review of the items by experts in the field. For the CSTs, these reviews are conducted by a number of experts in their designated areas from both the CDE and ETS, with ETS senior content staff working directly with CDE content consultants. (CDE, 2007)

The CDE content consultants each have extensive experience in K–12 assessments, particularly in their subjects of expertise, and many are former teachers. At a minimum, each CDE content consultant holds a bachelor's degree; most have advanced degrees in their area of expertise. All ETS content and test development staff have extensive experience with K–12 assessments, experience in teaching students, and understanding of the California Content Standards, and each hold, at a minimum, bachelor's degrees. Most have advanced degrees within their areas of expertise. (CDE, 2007)

Along with a thorough review by content area expert staff at CDE, there is a further usage of a content assessment review panel in order to measure the California standards in the following areas; as appropriate for testing and alignment to the California Content Standards, to ensure freedom from bias in relation to age/grade appropriateness, gender, racial/ethnic, and socioeconomic status, and to ensure that the material is appropriate and interesting to students tested at a particular grade/course. (CDE, 2007)

APPENDIX O

Michigan: Michigan Educational Assessment Program (MEAP)

Listed below in this sub-section is the information that is provided on the Michigan Department of Education website which supports the measures that are taken in order to ensure the validity of the MEAP assessment which will be used in this study. The Michigan department of education uses the MEAP to measure content standards in mathematics and English Language Arts.

The MEAP Office looks at data in many ways to assure items are measuring what they are intended to measure. One of the first criteria considered is whether an item appropriately assesses the content. The Bias/Sensitivity and Content Advisory Committee reviews are one of the best ways to determine the validity of an item. However, examining student performance data from field-assessment assists these committees. (Michigan Department of Education, 2007)

For every assessment item, MEAP staff first examine the "p-value," or the percentage of students who correctly answered the item, as well as the percent of students who chose each of the "distracters" (incorrect answers on a multiple-choice assessment). Particular attention is paid when less than 30% of the students select the correct answer. Since all multiple-choice items on MEAP assessments have four options, chance alone says that 25% of the students should mark the correct answer. Even if the content is appropriate, the item may not be measuring well — perhaps the graphic shown on the assessment is somehow misleading, or the question is poorly worded. P-values are not used to make the final decision on an item, but simply to indicate the need for further review. (MDE, 2007)

Differential Item Functioning is a fancy way of saying an item is potentially biased, or that it functions differently for one group than it does for another, according to statistical data from a pilot. If an item is "flagged" as being potentially biased, it is returned to the BSC for review, because human judgment is needed to determine whether an item is truly biased. Sometimes an item is flagged for what is really a curricular or instructional issue; i.e., one group did not do as well as another because they had not been taught the material measured by the item. All unusual patterns in the data are reviewed to consider anything in the context of the item that might have been missed in the first round of reviews. Again, based on BSC and CAC decisions, most items are retained, some may be revised, and some are discarded completely. Changes to an item necessitate that it be pilot tested again before it may appear on an operational assessment. (MDE, 2007)

Discrimination - Item discrimination examines performance between students who score high on the assessment compared to those who score low. If an item discriminates poorly, it means that students who scored poorly on the entire assessment may have done as well or better on an individual item than students who scored well on the entire assessment. This often occurs on very easy items that practically everyone answers correctly. Sometimes an item that discriminates poorly is kept if it measures content that is considered important, that is part of the state Content Standards, but may not have been widely taught. If more low-scoring students do as well or better than high-scoring students on a moderately difficult or difficult item, the item is given a closer look by the MEAP staff and the CAC or BSC. Perhaps there is more than one correct answer, or perhaps something in the knowledge base of the high scoring students is interfering with the way they are answering the question. The committees and MEAP staff also look at the distracters to assure they are not misleading students in unintended ways. (MDE, 2007)

Range - While variety may be "the spice of life," it is also an important part of the assessment. The MEAP staff aggressively seeks a wide range of difficulty in items. There is, however, no "magic formula" for how many "difficult" or how many "easy" questions are used. The MEAP staff does everything they can to help assure that overall differences from one year to the next are small. The most important goal is that each item measures an important part of the curriculum framework, benchmarks, standards, and expectations.

Other Factors - For constructed-response items, the staff examines the percent of students receiving points at each score level. If no one is receiving the top score possible, the staff takes another look at what the question is asking. This occurs more frequently when a type of response is asked for the first time on a state assessment, or in a grade that has not taken MEAP before. The staff also considers consistency among those who score the assessment.

Range finding and Scoring

For every MEAP constructed or written response, scoring begins with a process called "rangefinding" in which a committee establishes the "range" of achievement that defines each potential score on a rubric. Participants, generally classroom teachers, typically score 100 or more actual responses representing a range of possible achievement, as well as the state student population. Every single paper is discussed until a consensus is reached on the score the paper should receive. Some papers are easier to score than others, and require little discussion. Others lead to lengthy, spirited discussions because group members are divided in their opinions of what score to give (for example, a "two" or a "three"). The scoring contractor and Michigan Department of Education staffs participate in these meetings, but the educators make the final decisions. In math, science, and social studies, the scoring rubrics are item-specific and can be adjusted during rangefinding. Sometimes students interpret a prompt in a way that was not intended when the prompt was written. If it is considered to be a valid interpretation of the item, students are given the benefit of the doubt and the response is scored accordingly. In pilot rangefinding, problems with items often lead to improvements in the questions. Independent scorers score all MEAP written responses (constructed or extended). Before being hired, scorers qualify on a set of responses already scored during rangefinding. Additional rangefinding papers are used during scoring for validity purposes, as sort of a "pop quiz" to monitor whether scorers are scoring according to state guidelines. The MEAP staff also studies daily "inter-rater reliability" reports tracking the degree to which each scorer's scores agree exactly with those of a second scorer, are within one point (adjacent), or are non-adjacent (two or more points apart). If scorers disagree by more than one point on a response, it is sent to a third scorer with more training and experience (e.g., scoring director) for resolution. Such situations are rare. Additional data show whether a scorer is scoring low or high compared to others and the number of responses scored daily to track progress. This information is used by MEAP staff and the scoring contractor to monitor and adjust the scoring process over time.

Standard Setting

Right after a new MEAP assessment is administered, a process called "standard setting" is conducted to determine "cut" scores for reporting and categorizing student performance into levels of achievement. Standard setting begins with the selection of a statewide committee representing the geographic and ethnic diversity of our state. While most standard-setting panelists are classroom teachers, the process also includes administrators, curriculum specialists, counselors, parents, and business leaders. Over three days, standard setters rate student work on MEAP assessments against a performance standard. For all current MEAP assessments, the final recommendations for "cut" scores from standard-setting committees were reviewed and approved by the Bias/Sensitivity Committee, Content Advisory Committee, Assessment Advisory Committee, Technical Advisory Committee, and the State Board of Education. In April 2007, the State Board of Education approved the following performance levels for all MEAP content areas and grades.

Level 1: Advanced
Level 2: Proficient
Level 3: Partially Proficient
Level 4: Not Proficient

Reliability and Validity

The MEAP staff often fields questions about two critical technical concepts in measurement: reliability and validity. To assist and advise staff in making decisions about such issues, the MEAP Office contracts and consults with a Technical Advisory Committee comprised of nationally known psychometricians (experts in measuring student achievement). The MEAP staff has always followed, and will continue to follow, current psychometric practice in developing, administering, analyzing, and scoring the Michigan Educational Assessment Program assessments.

For the MEAP assessments, reliability values are determined by using internal consistency formulas, which indicate how homogeneous items are in an assessment, or the degree to which students' responses to each item correlate with their total assessment scores. Generally, Cronbach's Coefficient Alpha has been used as the measure of internal consistency reliability when constructed-response items appear on a MEAP assessment. It can also be used when there are solely multiple-choice items, or when combinations of item types are used. Typically, the longer the assessment, the higher the reliability. Both the reliability of MEAP assessments and the inter-rater reliability of the scoring process meet high technical standards.

Validity addresses the question of whether an assessment measures what it is supposed to measure. It refers to the degree of appropriateness, meaningfulness, and usefulness of the specific inferences made from assessment scores. There are three kinds of validity discussed in Standards for Educational and Psychological Assessment
(AERAAPA-NCME, 1985, updated 1999): criterion validity, construct validity, and content validity.
Psychometricians are often concerned about criterion and construct validity. Criterion validity refers to whether a measure can predict a student's future performance. For example, for the

ACT and SAT, which are used to predict college success, criterion validity is very important. This is not, however, the purpose of the MEAP High School Assessment (HSA).

The dilemma of whether to estimate construct validity on the basis of the total score, or upon strand scores, is one with which psychometricians constantly struggle. Construct validity is concerned with the parts (or dimensions) of an assessment, and whether they relate to the construct under study in a total assessment. A construct validity analysis could show whether questions fit into particular strands; for example, whether all geometry items on an assessment are most strongly related to one another, or if one fits better with data analysis. MEAP results are determined using the total assessment score, not scores from individual strands, dimensions, or assessment components. The Rasch model in Item Response Theory (IRT) is used to equate and scale all MEAP assessments. Item Response Theory assumes that the assessments under study are “unidimensional.” This means that the assessments measure one construct (or one domain) only, such as mathematics. Ongoing research evaluates these assumptions. Because the current MEAP assessments are achievement assessments used to assess what students have learned and should be able to achieve in specific content areas and grades, the most important type of validity of concern is content validity. To verify content validity, assessment items must reflect content defined within the Michigan Curriculum Framework, the basis for the content of all MEAP assessments.

APPENDIX P

North Carolina: End of Grade Test Validity and Reliability

North Carolina tests are curriculum-based tests designed to measure the objectives found in the North Carolina *Standard Course of Study (NCSCS)*. The responsibility of updating the *Standard Course of Study* falls to the North Carolina Department of Public Instruction (NCDPI) Division of Instructional Services. Curriculum specialists, teachers, administrators, university professors, and others assist in the process of updating curricula. Once curricula are adopted or tested objectives are approved (e.g. NC High School Comprehensive Test) by the North Carolina State Board of Education, in areas where statewide tests are required, the test development process begins. (North Carolina Public Schools, 2008)

The *Standard Course of Study* is reviewed for possible revisions every five years; however, test development is continuous. The NCDPI Accountability Services/Testing Section test development staff members begin developing **operational** test forms for the North Carolina Testing Program when the State Board of Education determines that such tests are needed. The need for new tests may result from mandates from the federal government or the North Carolina General Assembly. New tests can also be developed if the Board determines that the development of a new test will enhance the education of North Carolina students (e.g. NC Tests of Computer Skills). The test development process consists of six phases and takes approximately four years. The phases begin with the development of test specifications and end with the reporting of operational test results.

PHASE 1: DEVELOP THE TESTING PLAN

Step 1: Develop the Test Specifications (Blueprint)

Prior to developing test specifications, it is important to outline the purpose of a test and what types of inferences (e.g. diagnostic, curriculum mastery) are to be made from test scores. Millman and Greene (1993, *in* Robert Linn, ed)[1] offer a rationale for delineating the purpose of the test. "A clear statement of the purpose provides the overall framework for test specification, item development, tryout, and review. A clear statement of test purpose also contributes significantly to appropriate test use in practical contexts." Using a test's purpose as the guiding framework, NCDPI curriculum specialists, teachers, NCDPI test development staff, and other content, curriculum, and testing experts establish the test specifications for each of the grade levels and content areas assessed. In general, test specifications include the following:

1. Percentage of questions from higher or lower thinking skills and classification of each test question in the two dimensions of difficulty[2] and thinking skill level[3]
2. Percentage of item types such as graphs, charts, diagrams, political cartoons, analogies, and other specialized constraints
3. Percentage of test questions that measure a specific goal, objective, domain, or category
4. For tests that contain passages, the percentage of types of passages (e.g. literary vs. nonliterary passages, percentage of composition vs. literary analysis, etc.).

PHASE 2: ITEM DEVELOPMENT (ITEM TRYOUTS[4] AND REVIEW)

Step 2: Develop Test Items

While objectives for the new curriculum might not yet be implemented in the field, there are larger ideas that carry over from the previous curriculum cycle. These objectives are known as **common curriculum** objectives. Some examples of common curriculum objectives are historical trends in literature and theorems in geometry. Items can be developed from old test items that are categorized as common curriculum items or they can be developed as new items.

Old test items include those items from the previous curriculum cycle that were developed but not field tested. They can also be items that were field tested but not used in the statewide operational administration. If a curricular match is found for certain items, these items will be retained for further development with the new curriculum and tests. Items may be switched from grade to grade or from course to course to achieve a curriculum match. For example, a mathematics item may be moved from grade 5 to grade 4. If they are moved from grade to grade or course to course, they are considered to be new curriculum objective items. If they remain in the same grade or course, they are considered to be common curriculum items. Any item that has been used in a statewide operational test that matches the new curriculum will be released for training or for teachers to use in the classroom. While additional training may be required for writing new item types, the teachers can begin item development of common curriculum items due to their existing familiarity with the content.

Step 3: Review Items for Tryouts

The review process for items developed from the common curriculum is the same as it would be for the review of newly written items developed for any statewide test. The review process is described in detail in the "Phase 3: Field Test Development" section.

Step 4: Assemble Item Tryout Forms

As time and other resources permit, **item tryouts** are conducted as the first step in producing new tests. Item tryouts are a collection of a limited number of items of a new type, a new format or a new curriculum. Only a few forms are assembled to determine the performance of new items and not all objectives are tested. Conducting item tryouts has several advantages. The most important advantage is that an opportunity exists, during this process, to provide items for field-testing that are known to be psychometrically sound. In addition, it provides an opportunity to identify the need for a particular type of item (e.g. analogies). Having this data prior to field-testing and operational testing informs the item development and the test development process.

Conducting item tryouts will become increasingly important as the state moves to embedded field tests. Item tryouts provide an opportunity to determine the feasibility of and best possible plan for embedding, which can vary by subject or grade. Experimental items or sections can be tried out to determine whether students perceive them to be radically different from other sections. In addition, item tryouts provide an opportunity to examine the impact of the experimental sections on students' performance.

Step 5: Administer Item Tryouts

When item tryouts are administered as a stand-alone item tryout, a limited number of forms are produced, thus minimizing the number of children and schools impacted. Once these items are embedded in operational forms, the types of novel items that can be evaluated are severely constrained.

Step 6: Review Item Tryout Forms

Teachers are recruited to review the item tryout forms for clarity, correctness, potential bias, and curricular appropriateness. The NCDPI staff members, who specialize in the education of children with special needs, also review the forms.

Step 7: Review Item Tryout Statistics

Item statistics are examined to determine items that have a poor curricular match, poor response choices (foils), and confusing language. In addition, bias analyses can be run and the bias committee can review flagged items for revision. During a first-year item tryout, timing data can be collected to determine how long the new tests should be or to determine the amount of time needed for a given number of items. All of this information provides an opportunity to correct any flaws in the items that are to be included in the field tests.

PHASE 3: FIELD TEST DEVELOPMENT

Step 8: Develop New Items

North Carolina educators are recruited and trained as item writers for state tests. The diversity among the item writers and their knowledge of the current NC*SCS* are addressed during recruitment. The use of classroom teachers from across the state as item writers and developers ensures that instructional validity is maintained through the input of professional educators with current classroom experience. In cases where item development is contracted to an external vendor, the vendor is encouraged to use North Carolina educators in addition to professional item writers to generate items for a given project.

Step 9: Review Items for Field Test

Another group of teachers is recruited for reviewing the written test items. Each item reviewer receives training in item writing and reviewing multiple-choice test items. Based on the comments from the reviewers, items are revised and/or rewritten, item-objective matches are re-examined and changed where necessary, and introductions and diagrams for passages are refined. Analyses to verify that there is a valid representation by objectives also occur. Additional items are developed as necessary to ensure sufficiency of the item pool. Test development staff members, with input from curriculum specialists, review each item. Representation for students with special needs is included in the review. This process continues until a specified number of test items are written to each objective, edited, reviewed, edited, and finalized. Test development staff members, with input from the curriculum staff and other content, curriculum, and testing experts, approve each item to be field-tested.

Step 10: Assemble Field Test Forms

Items for each subject/course area are assembled into forms for field-testing. Although these are not the final versions of the tests, the forms are organized according to the specifications for the operational tests (test blueprints). If the items on the field test have been through the item tryout process, the field-test forms are parallel and can also be **quasi-equated** because the item-level statistics are already available for those items. New items or those that have been substantially changed since the item tryouts are analyzed after field testing. The item performance should be markedly better and the item rejection rates much lower for those items that were included in item tryouts. If the items have not been through tryouts (and do not have item statistics) **parallel** forms can be assembled which match test specifications and are parallel in terms of content coverage; however, difficulty of the forms cannot be addressed statistically.

Step 11: Review Field Test Forms

A new group of teachers is recruited to review the field test forms to ensure that clarity, correctness, potential bias, and curricular appropriateness are addressed. The NCDPI staff members from the Limited English Proficient (LEP) and Exceptional Children's Sections also review each field test form. The NCDPI test development staff, curriculum staff, and other content specialists (e.g. exceptional children, LEP) review teacher comments about the items, and necessary changes are made to items in the test. Teacher responses to the field test items are also used to verify the answer keys.

Step 12: Administer Field Tests

For a stand-alone or explicit field test, a stratified random sample of students is selected to take the field test forms. To ensure broad representation, schools are selected from across the state and are representative of the state based on the ethnic/racial characteristics of the student population, geographic location, and scores on previous versions of the tests among other characteristics. (Note that once field tests become embedded in operational tests, there will no longer be a need for stratified random sampling for field tests. The field test "sample" will census the entire population of students with the exception of those students who take the alternate assessments. Periodic stand-alone item tryouts may be necessary for new item types.)

The administration of the field test forms must follow the routine that will mimic the statewide administration of a test. The test administrator's manual for the field test administration includes instructions about the types of data to be collected in addition to student responses to the test items during the test administration. Examples of the types of data collected during field testing are Teacher Test Item Review Form, student demographic information, students' anticipated course grades as recorded by teachers, teachers' judgments of students' achievement level, field test administration time, and/or accommodations used for students with disabilities or identified as Limited English Proficient.

The above process will be modified for embedded field tests. For example, teachers will continue to provide the anticipated course grade and achievement judgments; however, they will no longer be able to complete the Teacher Item Review Form during the test administration since they will no longer be aware of which section is experimental.

Step 13: Review Field Test Statistics

The field test data for all items are analyzed by the NCDPI in conjunction with services contracted at the L. L. Thurstone Psychometric Laboratory, UNC-Chapel Hill and Technical Outreach for Public Schools (TOPS). The classical measurement model and the three-parameter logistic **item response theory** (IRT) model (including **p-value, biserial correlation, foil counts, slope, threshold, asymptote, and Mantel-Haenszel bias statistics**) are used in the analyses. Only the items approved by the NCDPI Division of Accountability Services/Testing Section staff members, with input from staff members from the Division of Instructional Services are sent to the next step. For stand-alone field tests, teacher comments are also reviewed.

Step 14: Conduct Sensitivity/Fairness Reviews

A separate committee conducts sensitivity/fairness reviews to address potential bias in test items. The NCDPI Division of Accountability Services/Testing Section "casts a wide net" when statistically identifying potentially biased test items in order to identify more items for review instead of fewer items. Bias Review Committee members are selected for their diversity, their experience with special needs students, or their knowledge of a specific curriculum area. The NCDPI Division of Instructional Services and additional content specialists review items identified by the field test data as biased. Items are retained for test development only if there is agreement among the content specialists and testing specialists that the item appropriately measures knowledge/skills that every student should know based on the North Carolina *Standard Course of Study*.

PHASE 4: PILOT TEST DEVELOPMENT

Step 15: Assemble Equivalent and Parallel Forms

The final item pool is based on approval by the (1) NCDPI Division of Instructional Services for curriculum purposes and (2) NCDPI Division of Accountability Services/Testing Section for psychometrically sound item performance. To develop **equivalent** forms, the test forms are balanced on P+ (sum of p-values). If the tests have a subsection or exhibit dimensionality, the subsections are equated. Finally, to the extent possible, the sections are balanced on slope. Each test matches the test specifications. The test development staff members, in collaboration with the NCDPI Division of Instructional Services, reviews the timing data to determine the appropriate number of test items. Curriculum content specialists also review the forms to determine if the test specifications have been implemented and to ensure that test forms by grade are parallel in terms of curricular coverage.

Step 16: Review Assembled Tests

A separate group of educators participates in the review of the assembled tests. Representation for students with special needs is included. The group reviews the assembled tests for content validity, responds to test items for an additional answer key check, and addresses the parallel nature of the test forms.

When embedding is fully implemented, teachers will review only the operational portions. At the operational stage, the types of edits allowed are quite limited to avoid invalidating the final item calibration. Should the item be determined to be unusable without the changes, it can be returned to the field test stage for revision and recalibration. The field test or item tryout sections will continue to be reviewed separately, since for those items, major revisions are still allowed.

Step 17: Final Review of Tests

Test development staff members, with input from curriculum staff, other content, curriculum, and testing experts and editors, conduct the final content and grammar check for each test form. If at this point a test item needs to be replaced, the test development staff must rebalance the entire form. If a large number of items are replaced after the series of reviews, the form is no longer considered to be the same form that originally went to review. Therefore the "new" form must go back to a teacher review.

Step 18: Administer Test as Pilot[5]

Because the field test forms are disassembled to form a global item pool from which the final tests are made, a **pilot test** of the final forms will allow any remaining glitches or "bugs" to be caught without negative ramifications for students or schools. The pilot test mimics an administration of the operational test in every way except that the standards are not yet in place. Thus the test can have no stakes for students. If there are stakes for schools they must be delayed until after the standard setting and final test administration data analyses.

Step 19: Score Tests

The NCDPI Division of Accountability Services/Testing Section must complete the following in order to provide local education agencies (LEAs) with the ability to scan multiple-choice answer sheets and report student performance at the local level:

(1) Answer key text files must be keyed with the goal/objective information and then converted to the format used by the WINSCAN/SCANXX program.

(2) A program converts the IRT files containing the item statistics to scale scores and standard errors of measurement. State percentiles must be added to create equating files.

(3) The equating files are created so the appropriate conversions occur: (a) raw score to scale score, (b) scale score to percentile, and (c) raw score to standard error of measurement.

(4) Files that convert scale scores to achievement levels are added.

(5) The test configuration file must be completed next. This file describes the layout of the header/answer sheets, the student survey questions, Special Code instructions, answer keys, and the linkage test scores for WINSCAN/SCANXX.

(6) Using the WINSCAN or the SCANXX program, header and answer sheets are scanned. This consists of selecting the appropriate test configuration file and scanning answer sheets. The program reads the answer key, equating the file and achievement level files. The individual items are compared to the answer keys and the raw score is calculated by summing the number correct. Each multiple-choice test item receives equal weight. Raw scores are then converted to other scores.

As mentioned earlier, when the move to an embedded model is complete for a subject or content area, the student's final score is based solely on performance on the operational sections of the test.

Step 20: Establish Standards

Industry guidelines require that standards be set using data from a pilot test or first year of fully operational. When data are not available from a pilot or first year fully operational test, interim standards are set using model based estimates from field tests. In addition, North Carolina has used the Contrasting Groups Method, a student-based method of standard setting, to determine standards for state tests. This method involves having students categorized into the various achievement levels by expert judges who are knowledgeable of the students' achievement. Teacher judgment of student achievement is compared to actual student performance on the operational tests. Analysis of this data is used in setting performance standards (e.g., achievement levels, cut scores) for the tests. Once the performance standards for a test are determined, typically they are not changed unless a new curriculum, revised test, or a new scale is implemented.

PHASE 5: OPERATIONAL TESTING

Step 21: Administer Tests as Fully Operational

The tests are administered statewide following all policies of the State Board of Education, including the North Carolina *Testing Code of Ethics*. Standardized test administration procedures must be followed to ensure the validity and reliability of test results. Students with disabilities and students identified as Limited English Proficient may use accommodations when taking the tests.

PHASE 6: REPORTING

Step 22: Reporting Test Results

For multiple-choice tests, reports are generated at the local level to depict performance for individual students, classrooms, schools, and LEAs. Results are distributed a week or two after the tests are administered. These data can be disaggregated by subgroups of gender and race/ethnicity as well as other demographic variables collected during the test administration. Demographic data are reported on variables such as free/reduced lunch status, LEP status, migrant status, Title I status, disability status, and parents' levels of education. The results are reported in aggregate at the state level usually at the end of June of each year. The NCDPI uses these data for school accountability and to satisfy other federal requirements (e.g. Annual Yearly Progress (AYP) requirement, No Child Left Behind Act of 2001).

TIMELINE FOR TEST DEVELOPMENT

Phase	**Timeline**
Phase 1: Develop Test Specifications (Blueprint)	4 months
Phase 2: Item Development for Item Tryout	12 months
Phase 3: Field Test Development and Administration	20 months
Phase 4: Pilot Test Development and Administration	4 months for EOC tests (9 months for EOG tests)
Phase 5: Operational Test Development and Administration	4 months
Phase 6: Reporting Operational Test Results	Phase 6 completed as data become available.
Total Time	**44-49 months**

Note: Some phases require action by some other authority than the NCDPI Testing Section (e.g. contractors, field staff). These phases can extend or shorten the total timeline for test development.

APPENDIX Q

Texas: Texas Assessment of Knowledge and Skills (TAKS)

This appendix provides information that was found on the Texas Education Agency website which supports the measures that are taken in order to ensure the validity of the TAKS assessment which will be used in this study. The Texas Education Agency of uses the TAKS assessment tool to measure the Texas content standards in mathematics and Reading.

Validity is a process of collecting evidence to support inferences made from the scoring results of an assessment. In the case of TAKS and SDAA, test results are used to make inferences about students' knowledge and understanding of the TEKS. For RPTE, test results provide a measure of progress, indicating annually where each LEP student is on a continuum of English language development designed for second-language learners. (Texas Education Agency, 2004)

Content Validity Standards-referenced assessments, such as the TAKS, SDAA, and RPTE tests, are based on an extensive definition of the content they assess. Test validity is therefore content based and tied directly to the statewide curriculum. In order to ensure the highest level of content validity, the process of aligning TAKS, SDAA, and RPTE to the curriculum was carefully approached and included numerous committees of Texas educators. When TAKS was designed as the standards-referenced assessment for the TEKS, advisory committees consisting of educators from school districts across the state were formed for each subject area at each grade level. Teachers, test development specialists, and TEA staff members worked together in these committees to identify TEKS student expectations important to assess and to develop test objectives, item development guidelines, and test item types. In addition, committees met starting in 2001–2002 to review and edit TAKS items for content and bias and to review data from field testing. A similar process was conducted for both SDAA and RPTE when they were developed. (TEA, 2004)

Relation to the Statewide Curriculum
The item writers as well as the reviewers for each stage of development verify the alignment of test items with the objectives to ensure that the items measure appropriate content. The sequential stages of item development and item review provide many opportunities for Texas educators to offer suggestions for improving or eliminating items and to offer insights into the interpretation of the statewide curriculum. The nature and specificity of these various review procedures provide additional strong evidence for the content validity of the TAKS, SDAA, and RPTE tests. (TEA, 2004)

Educator Input
Not only do Texas educators provide valued input on the content and the match between the items and the statewide curriculum, but many current and former Texas educators and some educators from other states also work as independent contractors to write items specifically to measure the objectives. This provides for a system of checks and balances for item development and review that reduces single-source bias. In other words, because test items Technical Digest 2003–2004 121 Chapter 14: Validity are written by many different people with different backgrounds, it is less likely that items will suffer from a bias that might occur if items were

written by a single author. The direct input from educators offers additional evidence regarding the content validity of constructed TAKS, SDAA, and RPTE tests. (TEA, 2004)

Test Developer Input

The staff at TEA, as well as professional test developers from Harcourt Educational Measurement, Pearson Educational Measurement, and BETA, Inc., provide a wealth of test building experience, including content expertise. Each internal review of an item by these experts increases the probability of the item being an accurate measure of the intended objective. Hence, these reviews are offered as additional evidence for the content validity of the TAKS, SDAA, and RPTE tests. (TEA, 2004)

Construct Validity

Content and construct validity are traditionally separate categories, but in a discussion of the validity of achievement tests, these distinctions become somewhat obscured. Content validity describes whether the test objectives adequately represent what students should be able to do and whether the items, which are based on test objectives, measure intended responses. Construct validity is the extent to which a test can be said to measure a theoretical construct or trait. In the case of the TAKS, SDAA, and RPTE tests, the construct tested is the academic content required by the statewide curriculum. With curriculum-based achievement tests, both types of validity are intertwined. The construct validity is grounded in the content validity of the test. Further evidence of construct validity for TAKS was provided as part of the Higher Education Readiness Component of TAKS. TEA gathered performance data for the exit level mathematics and ELA tests on a sample of college students at two- and four-year institutions throughout the state on which a contrasting-groups study was performed. The percent of college students at the "Met Standard" and "Commended" levels on these tests was compared to that of high school students (see Chapter 8: Higher Education Readiness Component). (TEA, 2004)

Criterion-Related Validity

Criterion validity indicates the relationship between test performance and performance on some other measure. This other measure can be evaluated concurrently or in the future and is then correlated with the test score. In this way, the test score is compared with a criterion that is thought to be a reasonable estimate of the same construct the original test purports to measure. As part of the TAKS Higher Education Readiness Component, a concurrent validity study was conducted in 2003–2004 to correlate performance on exit level TAKS with performance on national testing programs. (TEA, 2004)

Appendix R

The Number of Test Items in Each Assessment

State/Assessment	Grade	Content Area	Test Items
California	4^{th}	Reading	35
California	4^{th}	Mathematics	65
California	8^{th}	Reading	35
California	8^{th}	Mathematics	65
Michigan/MEAP	4^{th}	Reading	31
Michigan/MEAP	4^{th}	Mathematics	61
Michigan/MEAP	8^{th}	Reading	31
Michigan/MEAP	8^{th}	Mathematics	51
USDOE/NAEP	4^{th}	Reading	20
USDOE/NAEP	4^{th}	Mathematics	36
USDOE/NAEP	8^{th}	Reading	20
USDOE/NAEP	8^{th}	Mathematics	36
North Carolina	4^{th}	Reading	50
North Carolina	4^{th}	Mathematics	50
North Carolina	8^{th}	Reading	53
North Carolina	8^{th}	Mathematics	60
Texas	4^{th}	Reading	32
Texas	4^{th}	Mathematics	34
Texas	8^{th}	Reading	32
Texas	8^{th}	Mathematics	40

The test items are average numbers based on the information that was available via the NCES website and State Education Agency Websites.

APPENDIX S

Documentation of Consent to use the Teachers' Voice Survey

From: Gail Sunderman [mailto:]
Sent: Monday, April 20, 2009 9:13 AM
To: Maleyko, Glenn M
Subject: Re: The No Child Left Behind: The Teachers¹ Voice survey, Sunderman, Tracey, Kim, & Orfield, 2004

Dear Glenn, I apologize if I haven't responded sooner. You are welcome to use the Teachers' Voice Survey. We ask that you acknowledge the Civil Rights Project in any work that you do that uses it. I'm not sure if you are also requesting a copy of the survey. If so, please let me know.

Gail Sunderman

From: **Glenn M Maleyko**
To:
Sent: **Thursday, April 16, 2009 5:57:39 PM**
Subject: **The No Child Left Behind: The Teachers¹ Voice survey, Sunderman, Tracey, Kim, & Orfield, 2004**

Dr. Sunderman,

I am a doctoral student at Wayne State University in Detroit, Michigan and my dissertation research is going to focus on the validity and reliability of the AYP measurement in NCLB. I have five research questions and my fifth one involves the impact of AYP on classroom instruction and school improvement. I am hopeful that I might be able to get permission to use your survey instrument Teacher's Voice Survey from your 2004 publication? I appreciate any assistance that you could provide to me regarding if and how I might be able to get permission to use that survey?

Thank you in advance for your assistance.

Sincerely,

Glenn Maleyko

APPENDIX T

NCLB Teacher Interview Questions

Interview Structure and Questions for Teachers

During the past several years educators have witnessed a shift in accountability policy with the onset of the NCLB act legislated by the federal government, I am trying to understand your views and thoughts concerning the impact that No Child Left Behind (NCLB) and specifically Adequate Yearly Progress (AYP) are having on school improvement initiatives in your school and classroom instruction.

NCLB Background Questions

1. What are your thoughts about the NCLB accountability policy? (Brief information will be provided to the interviewee if questions arise)

2. From which sources (i.e. books, journals, other educators, your school or district, etc) do you receive most of your information about NCLB? Are you familiar with the AYP provisions? (If not then a brief explanation will occur)

3. Do you feel that identifying schools that have not made AYP will lead to school improvement? If so, how? If not, why not?

4. Are you familiar with the NCLB sanctions for schools that fail to make AYP? (If not then a brief explanation will occur). Do you believe that any of the sanctions will lead to school improvement?

5. Do you believe that it is realistic to expect schools that serve large populations of economically disadvantaged (ED) students to perform as well on state accountability assessments as schools that serve large populations of affluent students as it is required in NCLB?

6. Do you believe that it is realistic to expect schools that are financed at different levels to perform at the same level on state accountability assessments as it is required by NCLB?

NCLB School Impact Questions

7. Do you believe that meeting the AYP requirements is an attainable goal for your school? Why or why not?

8. What positive consequences, if any, do you observe in your school as a result of AYP?

9. What negative consequences, if any, do you observe happening as a result of AYP?

School Level Reaction to NCLB

10. Do you believe that your school has responded to the AYP requirements in NCLB, if so, what response has your school taken?

11. What specific things, if any, have been implemented in your school which you believe is a result of AYP?

12. What impact, if any, is AYP having on your school improvement plan (SIP)?

Individual Teacher Impact Questions

13. How has your practice as a teacher been affected as a result of AYP?

14. What impact, if any, is AYP having on your classroom instruction?

15. Would the AYP status of a school impact your choice to teach in that school or remain in that school?

Impact on other educators in the school

16. Is the NCLB policy having an effect on teachers and administrators in your school? If so, what are these effects?

17. What impact, if any, is AYP having on the instruction of other teachers in your school? (tentative pending response to question 15)

18. Have you observed any administrative or policy changes by the building administration which you believe is a response to AYP? What are the most notable changes in the last two years as a result of AYP in your opinion?

Curriculum and District Impact

19. What impact, if any, is AYP having on the curriculum in your school or district? What are your thoughts about that?

20. What impact, if any, is AYP having on the operation of your district?

Closing

21. Is there anything else that you would like to add that would help me to better understand your views and thoughts on the NCLB accountability provisions along with the impact on your job as an educator?

APPENDIX U

NCLB Administrator Interview Questions

During the past several years educators have witnessed a shift in accountability policy with the onset of the NCLB act legislated by the federal government, I am trying to understand your views and thoughts concerning the impact that No Child Left Behind (NCLB) and specifically Adequate Yearly Progress (AYP) are having on school improvement initiatives in your school and classroom instruction.

NCLB Background Questions

1. What are your thoughts about the NCLB accountability policy?

2. From which sources (i.e. books, journals, other educators, your school or district, etc) do you receive most of your information about NCLB and AYP?

3. Do you feel that identifying schools that have not made AYP as failing will lead to school improvement? If so, how? If not, why not?

4. Are you familiar with the NCLB sanctions for schools that fail to make AYP? (If not then a brief explanation will occur). Do you believe that any of the sanctions will lead to school improvement?

5. Do you believe that it is realistic to expect schools that serve large populations of economically disadvantaged (ED) students to perform as well on state accountability assessments as schools that serve large populations of affluent students as it is required in NCLB?

6. Do you believe that it is realistic to expect schools that are financed at different levels to perform at the same level on state accountability assessments as it is required by NCLB?

NCLB School Impact Questions

7. Do you believe that meeting the AYP requirements is an attainable goal for your school? Why or why not?

8. What positive consequences, if any, do you observe in your school as a result of AYP?

9. What negative consequences, if any, do you observe happening as a result of AYP?

School Level Reaction to NCLB

10. Do you believe that your school has responded to the AYP requirements in NCLB, if so, what response has your school taken?

11. What specific things, if any, have been implemented in your school which you believe is a result of AYP? (This question is tentative pending the response the response to question 10).

12. What impact, if any, is AYP having on your school improvement plan?

Individual Administrator Impact Questions

13. How has your practice as an administrator been affected as a result of AYP?

14. How has AYP affected your decision-making as a school administrator?

15. What, if any, administrative or policy changes have you implemented in your school in regards to AYP? What are the most notable changes in the last two years as a result of AYP?

16. Would the AYP status of a school impact your choice to work in that school or remain in that school?

Impact on other educators in the school

17. Is the NCLB policy having an effect on the teachers in your school? If so, what are these effects?

18. How has AYP affected the instructional teaching practices of your teachers? (optional pending response to question 16.)

Curriculum and District Impact (Question 19 and 20 are tentative pending the responses to question 18)

19. What impact, if any, is AYP having on the operation of your district?

20. What administrative or policy changes, if any, have you observed of the school district in regards to AYP? What are the most notable changes in the last two years as a result of AYP?

21. What impact, if any, is AYP having on the curriculum in your school or district? What are your thoughts about that?

Closing Question.

22. Is there anything else that you would like to add that would help me to better understand your views and thoughts on the NCLB accountability provisions along with the impact on your job as an educator?

APPENDIX V

Research Information Sheet

Title of Study: The Impact of No Child Left Behind (NCLB) on School Achievement and Accountability

Principal Investigator (PI): Glenn Maleyko
Education Leadership and Policy Studies Ph. D student
519-969-3476

Purpose:

You are being asked to participate in a research study regarding the impact that No Child Left Behind (NCLB) and specifically Adequate Yearly Progress (AYP) are having on school improvement initiatives in your school and with classroom instruction. During the past several years educators have witnessed a shift in accountability policy with the onset of the NCLB act legislated by the federal government. You have been chosen because you are an educator in this school and you are eligible to participate in this study as a result. This study is being conducted through Wayne State University, and in three other schools in the state of Michigan.

Study Procedures:

If you take part in the study, you will be asked to answer questions that pertain to NCLB and AYP so that I can get a better understanding of the impact that the legislation is having at the school level. The questions that you will be asked will give me an understanding of your views regarding the impact of AYP on your school program. I plan to record the interview for transcription purposes only and I will destroy the interview tape once they are transcribed. At no point will I ask your name or write your name on the interview transcription or interview survey. Your participation will involve approximately 20-30 minutes of your time in order to answer the questions that I ask.

Benefits

As a participant in this research study, there may be no direct benefit for you; however, information from this study may benefit other people now or in the future with providing a better understanding of the Adequate Yearly Progress (AYP) provisions in NCLB.

Risks

There are no known risks at this time to participation in this study.

Costs

There will be no costs to you for participation in this research study.

Compensation

You will not be paid for taking part in this study.

Confidentiality:

All information collected about you during the course of this study will be kept without any identifiers.

Voluntary Participation /Withdrawal:

Taking part in this study is voluntary. You are free to not answer any questions or withdraw at any time. Your decision will not change any present or future relationships with Wayne State University or its affiliates.

Questions:

If you have any questions about this study now or in the future, you may contact Glenn Maleyko at 519-969-3476 or my research advisor Dr. Martyza Gawlik at 313-577-1712. If you have questions or concerns about your rights as a research participant, the Chair of the Human Investigation Committee can be contacted at (313) 577-1628. If you are unable to contact the research staff, or if you want to talk to someone other than the research staff, you may also call (313) 577-1628 to ask questions or voice concerns or complaints.

Participation:

By participating in the interview you are agreeing to participate in this study.

APPENDIX W

California Tables and Figures for Research Question One, Pearson Correlation.

Table W1

Descriptive statistics for the 2005 California 8th grade dataset

	Mean	Std. Deviation	N *
STATEAYP	.28	.448	399570
NAEP BASIC level	.3724	.48345	396958
NAEP PROF level	.0000	.00000	396958
TOTAL REVENUE PER STUDENT	9296.53	1292.792	398351
EDPER	.5623	.26532	408538
NATAMPER	.0059	.01131	408538
ASIANPER	.1079	.13346	408538
BLACKPER	.0929	.10938	408538
HISPANICPER	.5398	.28404	408538
WHITEPER	.2425	.23294	408538
ELLPER	.2831	.20517	393066
SPECIALEDPER	.1156	.09012	375723

* Different weighted sample sizes are present because some of the school only took one of the NAEP assessments(math or reading) resulting in the inability to calculate a school level NAEP proficiency. Also, some of the schools did not have a recorded AYP proficiency score or state accountability assessment data for different reasons. A choice was made not to eliminate the schools with incomplete data because they provided information on the relationships between the subgroup populations and the proficiency assessment results (either NAEP or state accountability assessments) that were recorded in the database. In the places where missing datasets are present, SPSS does not use those sample schools in the calculation of the relevant outputs. This will be consistent among all of the following datasets with the display of the descriptive statistics.

Table W2

Pearson correlation matrix for the California 2005 8th grade listed variables (Demographic, financial, and proficiency status: Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	Total Revenue	EDPER	NATAMPER
STATEAYP	1	.396**	.[a]	-.293**	-.658**	.149**
NAEP BASIC level	.396**	1	.[a]	-.262**	-.579**	.117**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.360**	-.163**	-.609**	.582**	-.481**	-.186**
NAEP BASIC level	.317**	-.144**	-.538**	.509**	-.458**	-.096**

**. Correlation is significant at the 0.01 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

Table W3

Descriptive statistics for the 2007 California 4th grade dataset

	Mean	Std. Deviation	N *
STATEAYP	.6741	.46870	183882
NAEP Basic Level	.3684	.48236	183320
NAEP Proficient Level	.0621	.24137	183320
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	11488.60	2765.230	183932
EDPER	.6501	.30680	183150
NATAMPER	.0042	.00651	183932
ASIANPER	.1001	.13538	183932
BLACKPER	.0785	.10210	183932
HISPANICPER	.6010	.30935	183932
WHITEPER	.1994	.24594	183932
ELLPER	.4394	.27734	179901
SPECIALEDPER	.0917	.07911	175392

Table W4

Pearson correlations matrix for the California 2007 4th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.393**	.178**	-.088**	-.347**	.010**
NAEP BASIC level	.393**	1	.337**	-.187**	-.779**	.178**
NAEP PROF level	.178**	.337**	1	-.014**	-.494**	.001

	ASIANPER	BLACKPER	HISPANICPER	WHITEPER	ELLPER	SPECIALEDPER
STATEAYP	.316**	.024**	-.404**	.300**	-.406**	-.060**
NAEP BASIC level	.476**	.009**	-.801**	.696**	-.712**	.009**
NAEP PROF level	.264**	-.145**	-.434**	.430**	-.336**	.030**

Table W5

Descriptive statistics for the 2007 California 8th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.2752	.44661	302443
NAEP BASIC level	.4943	.49997	302589
NAEP PROF level	.0174	.13071	302589
TOTAL REVENUE PER STUDENT	11419.31	1624.585	301548
EDPER	.5953	.26591	300317
NATAMPER	.0060	.01592	302971
ASIANPER	.1046	.11959	302971
BLACKPER	.0738	.09524	302971
HISPANICPER	.5629	.28786	302971
WHITEPER	.2348	.23061	302971
ELLPER	.2826	.19615	297105
SPECIALEDPER	.1126	.06134	290614

Table W6

Pearson correlations matrix for the California 2007 8th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	**NAEP BASIC level**	**NAEP PROF level**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATEAYP	1	.480**	.216**	-.365**	-.583**	.125**
NAEP BASIC level	.480**	1	.135**	-.255**	-.712**	.087**
NAEP PROF level	.216**	.135**	1	-.016**	-.256**	-.017**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.289**	-.181**	-.542**	.565**	-.370**	-.193**
NAEP BASIC level	.429**	-.106**	-.673**	.616**	-.538**	-.102**
NAEP PROF level	.242**	-.084**	-.213**	.167**	-.152**	-.056**

**. Correlation is significant at the 0.01 level (2-tailed)

APPENDIX X

Michigan Tables and Figures for Research Question One, Pearson Correlation.

Table X1

Descriptive statistics for the 2005 Michigan 4th grade dataset

Descriptive Statistics

	Mean	Std. Deviation	N
STATEAYP	.9944	.07441	46162
NAEP BASIC level	.6779	.46727	43401
NAEP PROF level	.0000	.00000	43401
TOTAL REVENUE PER STUDENT	9756.85	1313.817	47109
EDPER	.3847	.29179	45721
NATAMPER	.0098	.03030	47109
ASIANPER	.0293	.04529	47109
BLACKPER	.2039	.32550	47109
HISPANICPER	.0405	.06353	47109
WHITEPER	.7126	.32180	47109
ELLPER	.0423	.08535	46553
SPECIALEDPER	.1194	.06947	43703

Table X2

Pearson correlations matrix for the Michigan 2005 4th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.112**	.[a]	-.031**	-.157**	.015**
NAEP BASIC level	.112**	1	.[a]	-.108**	-.272**	.115**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.049**	-.182**	.048**	.165**	.011*	.044**
NAEP BASIC level	.134**	-.388**	-.019**	.365**	.047**	-.004

**. Correlation is significant at the 0.01 level (2-tailed).

a. Cannot be computed because at least one of the variables is constant.

Table X3

Descriptive statistics for the 2005 Michigan 8th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.94	.235	62892
NAEP BASIC level	.7840	.41153	64223
NAEP PROF level	.0742	.26217	64223
ELLPER	.0502	.14336	63222
SPECIALEDPER	.1436	.07192	59362
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9939.95	1373.088	64247
EDPER	.3482	.25181	63584
NATAMPER	.0070	.01250	64264
ASIANPER	.0208	.02826	64264
BLACKPER	.2173	.32299	64264
HISPANICPER	.0345	.04684	64264
WHITEPER	.7180	.32500	64264

Table X4

Pearson correlation matrix for the 2005 Michigan 8th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.294**	.071**	-.114**	-.223**	.052**
NAEP BASIC level	.294**	1	.149**	-.283**	-.740**	.086**
NAEP PROF level	.071**	.149**	1	.163**	-.285**	-.083**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.169**	-.299**	.002	.278**	.047**	.015**
NAEP BASIC level	.179**	-.782**	-.122**	.776**	-.095**	.069**
NAEP PROF level	.420**	-.134**	-.063**	.107**	-.041**	-.225**

**. Correlation is significant at the 0.01 level (2-tailed)

Table X5

Descriptive statistics for the 2007 Michigan 4th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.9458	.22643	42631
NAEP Basic Level	.8495	.35755	42469
NAEP Proficient Level	.1076	.30989	42469
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2006-07]	10943.16	2144.399	42631
EDPER	.3345	.25697	42631
NATAMPER	.0182	.06493	42631
ASIANPER	.0284	.04792	42631
BLACKPER	.1665	.30284	42631
HISPANICPER	.0332	.04712	42631
WHITEPER	.7439	.30437	42631
ELLPER	.0294	.04082	41694
SPECIALEDPER	.1178	.08030	40110

Table X6

Pearson correlation matrix for the 2007 Michigan 4th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.400**	.083**	-.180**	-.321**	-.164**
NAEP BASIC level	.400**	1	.146**	-.371**	-.684**	.034**
NAEP PROF level	.083**	.146**	1	.164**	-.382**	-.088**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.094**	-.304**	.098**	.301**	.091**	.014**
NAEP BASIC level	.192**	-.782**	.044**	.729**	.144**	-.027**
NAEP PROF level	.308**	-.167**	-.156**	.161**	-.022**	-.183**

**. Correlation is significant at the 0.01 level (2-tailed)

Table X7

Descriptive statistics for the 2007 Michigan 8th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.9294	.25615	51916
NAEP Basic Level	.8531	.35401	56167
NAEP Proficient Level	.0353	.18442	56167
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2006-07]	10542.11	1763.064	56098
EDPER	.3299	.23657	56167
NATAMPER	.0178	.06074	56167
ASIANPER	.0269	.04506	56167
BLACKPER	.1290	.22747	56167
HISPANICPER	.0369	.05257	56167
WHITEPER	.7786	.24539	56167
ELLPER	.0401	.08356	54723
SPECIALEDPER	.1410	.08563	53198

Table X8

Pearson correlation matrix for the 2007 Michigan 8th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.429**	.055**	-.155**	-.396**	.064**
NAEP BASIC level	.429**	1	.079**	-.261**	-.666**	.074**
NAEP PROF level	.055**	.079**	1	.093**	-.202**	-.045**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.139**	-.627**	.027**	.522**	-.023**	-.057**
NAEP BASIC level	.126**	-.717**	-.256**	.665**	-.343**	-.256**
NAEP PROF level	.053**	-.068**	-.055**	.082**	-.048**	-.058**

**. Correlation is significant at the 0.01 level (2-tailed).

APPENDIX Y

North Carolina Tables and Figures for Research Question One, Pearson Correlation.

Table Y1

Descriptive statistics for the 2005 North Carolina 4th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.6143	.48676	96984
NAEP BASIC level	.7698	.42099	97486
NAEP PROF level	.0739	.26169	97486
TOTAL REVENUE PER STUDENT	8123.00	972.097	97486
EDPER	.4944	.25485	96690
NATAMPER	.0204	.10162	97486
ASIANPER	.0287	.02925	97486
BLACKPER	.3327	.25075	97486
HISPANICPER	.1044	.10028	97486
WHITEPER	.5138	.29281	97486
ELLPER	.1060	.11666	97486
SPECIALEDPER	.1390	.06191	96230

Table Y2

Pearson correlations matrix for the 2005 North Carolina 4th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.248**	.163**	.025**	-.342**	-.195**
NAEP BASIC level	.248**	1	.155**	.080**	-.573**	-.253**
NAEP PROF level	.163**	.155**	1	-.026**	-.427**	-.045**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.058**	-.246**	-.127**	.316**	-.147**	-.095**
NAEP BASIC level	.116**	-.397**	-.127**	.460**	-.052**	-.026**
NAEP PROF level	.200**	-.237**	-.192**	.265**	-.178**	-.050**

**. Correlation is significant at the 0.01 level (2-tailed)

Table Y3

Descriptive statistics for the 2005 North Carolina 8th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.2771	.44755	90765
NAEP BASIC level	.8659	.34072	91676
NAEP PROF level	.0389	.19343	91676
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	8129.13	919.078	91731
EDPER	.4809	.20551	91081
NATAMPER	.0097	.04644	91731
ASIANPER	.0221	.02773	91731
BLACKPER	.3399	.23201	91731
HISPANICPER	.0726	.06048	91731
WHITEPER	.5557	.25538	91731
ELLPER	.0654	.06162	88360
SPECIALEDPER	.1630	.08715	83538

Table Y4

Pearson correlations matrix for the 2005 North Carolina 8th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.245**	.327**	.197**	-.388**	.098**
NAEP BASIC level	.245**	1	.079**	.028**	-.445**	.038**
NAEP PROF level	.327**	.079**	1	.056**	-.316**	-.032**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.096**	-.337**	-.181**	.320**	-.132**	-.162**
NAEP BASIC level	-.042**	-.484**	-.146**	.471**	-.127**	-.207**
NAEP PROF level	.421**	-.125**	-.161**	.112**	-.125**	-.137**

**. Correlation is significant at the 0.01 level (2-tailed)

Table Y5

Descriptive statistics for the 2007 North Carolina 4th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.4619	.49855	99367
NAEP BASIC level	.7662	.42326	99885
NAEP PROF level	.1300	.33629	99885
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9181.78	1102.643	99885
EDPER	.4722	.23339	99885
NATAMPER	.0477	.16344	99885
ASIANPER	.0288	.02901	99885
BLACKPER	.2929	.21851	99885
HISPANICPER	.1317	.11285	99885
WHITEPER	.4570	.28834	99885
ELLPER	.1196	.11910	98249
SPECIALEDPER	.1350	.09629	95403

Table Y6

Pearson correlation matrix for the 2007 North Carolina 4th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.336**	.418**	-.048**	-.445**	-.124**
NAEP BASIC level	.336**	1	.214**	.141**	-.548**	-.394**
NAEP PROF level	.418**	.214**	1	.079**	-.581**	-.106**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.021**	-.426**	-.215**	.475**	-.255**	-.204**
NAEP BASIC level	.103**	-.229**	-.238**	.475**	-.242**	-.087**
NAEP PROF level	.264**	-.375**	-.273**	.432**	-.199**	-.103**

**. Correlation is significant at the 0.01 level (2-tailed)

Table Y7

Descriptive statistics for the 2007 North Carolina 8th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.1739	.37905	96480
NAEP Basic Level	.8813	.96626	97123
NAEP Proficient Level	.0598	.23719	97123
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9424.60	1312.122	97123
EDPER	.4672	.20341	97123
NATAMPER	.0390	.13932	97123
ASIANPER	.0225	.02393	97123
BLACKPER	.3101	.23010	97123
HISPANICPER	.0917	.07462	97123
WHITEPER	.5095	.26807	97123
ELLPER	.0660	.06543	94439
SPECIALEDPER	.1530	.07565	88663

Table Y8

Pearson correlations matrix for the 2007 North Carolina 8th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.039**	.379**	.119**	-.358**	-.092**
NAEP BASIC level	.039**	1	.031**	.046**	-.193**	-.128**
NAEP PROF level	.379**	.031**	1	.302**	-.446**	-.066**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.150**	-.276**	-.220**	.333**	-.206**	-.105**
NAEP BASIC level	-.029**	-.127**	-.060**	.192**	-.055**	-.023**
NAEP PROF level	-.066**	.227**	-.223**	-.126**	.241**	-.109**

**. Correlation is significant at the 0.01 level (2-tailed)

APPENDIX Z

Texas Tables and Figures for Research Question One, Pearson Correlation.

Table Z1

Descriptive statistics for the 2005 Texas 4th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.9839	.12591	195738
NAEP Basic Level	.6509	.47668	199774
NAEP Proficient Level	.1044	.30584	199774
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	8774.39	1201.540	198443
EDPER	.5630	.31130	199774
NATAMPER	.0025	.00407	199774
ASIANPER	.0314	.05265	199774
BLACKPER	.1633	.20879	199774
HISPANICPER	.5288	.32308	199774
WHITEPER	.2741	.29269	199774
ELLPER	.2939	.25046	192878
SPECIALEDPER	.1094	.08865	186924

Table Z2

Pearson correlations matrix for the 2005 Texas 4th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	-.036**	.043**	.054**	-.038**	-.073**
NAEP BASIC level	-.036**	1	.250**	-.060**	-.457**	.227**
NAEP PROF level	.043**	.250**	1	.070**	-.509**	.112**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.043**	-.091**	.087**	-.038**	.094**	-.050**
NAEP BASIC level	.293**	-.235**	-.391**	.544**	-.409**	-.011**
NAEP PROF level	.415**	-.172**	-.425**	.515**	-.327**	-.096**

**. Correlation is significant at the 0.01 level (2-tailed)

Table Z3

Descriptive statistics for the 2005 Texas 8th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.7248	.44659	193857
NAEP Basic Level	.7717	.41976	202154
NAEP Proficient Level	.0445	.20628	202154
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	8553.17	1238.803	203550
EDPER	.4904	.26894	203550
NATAMPER	.0026	.00319	203550
ASIANPER	.0318	.04844	203550
BLACKPER	.1847	.20962	203550
HISPANICPER	.4376	.29926	203550
WHITEPER	.3433	.28997	203550
ELLPER	.1179	.14324	195917
SPECIALEDPER	.1565	.09852	189073

Table Z4

Pearson correlation matrix for the 2005 Texas 8th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.393**	.137**	-.020**	-.451**	.042**
NAEP BASIC level	.393**	1	.117**	-.044**	-.484**	.213**
NAEP PROF level	.137**	.117**	1	.209**	-.326**	.125**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.215**	-.236**	-.267**	.409**	-.343**	-.285**
NAEP BASIC level	.245**	-.387**	-.277**	.519**	-.354**	-.232**
NAEP PROF level	.405**	-.124**	-.253**	.280**	-.134**	-.154**

**. Correlation is significant at the 0.01 level (2-tailed)

Table Z5

Descriptive statistics for the 2007 Texas 4th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.9919	.08955	171332
NAEPCOMBPROFbasic	.6171	.48609	172789
NAEPCOMBPROFprof	.1138	.31763	172789
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	10153.03	1701.802	173176
EDPER	.5639	.27784	173176
NATAMPER	.0019	.00285	173176
ASIANPER	.0369	.05567	173176
BLACKPER	.1628	.19216	173176
HISPANICPER	.5537	.30911	173176
WHITEPER	.2447	.27650	173176
ELLPER	.3300	.25868	165976
SPECIALEDPER	.1050	.09972	166949

Table Z6

Pearson correlation matrix for the 2007 Texas 4th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	**NAEP BASIC level**	**NAEP PROF level**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATEAYP	1	.115**	.032**	-.093**	-.130**	.062**
NAEP BASIC level	.115**	1	.282**	-.122**	-.422**	.324**
NAEP PROF level	.032**	.282**	1	.173**	-.512**	.177**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.057**	-.084**	-.026**	.076**	-.052**	-.304**
NAEP BASIC level	.294**	-.235**	-.398**	.546**	-.472**	-.034**
NAEP PROF level	.441**	-.187**	-.405**	.492**	-.310**	-.036**

**. Correlation is significant at the 0.01 level (2-tailed)

Table Z7

Descriptive statistics for the 2007 Texas 8th grade dataset

	Mean	Std. Deviation	N
STATEAYP	.8198	.38434	154567
NAEP Basic Level	.8422	.36458	154573
NAEP Proficient Level	.0783	.26862	154573
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9943.58	2055.118	155218
EDPER	.4965	.24522	153136
NATAMPER	.0026	.00303	155218
ASIANPER	.0356	.04936	155218
BLACKPER	.1766	.19035	155218
HISPANICPER	.4918	.28930	155218
WHITEPER	.2934	.27242	155218
ELLPER	.1411	.16573	150559
SPECIALEDPER	.1443	.09211	152734

Table Z8

Pearson correlation matrix for the 2007 Texas 8th grade listed variables (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATEAYP	NAEP BASIC level	NAEP PROF level	TOTAL REVENUE	EDPER	NATAMPER
STATEAYP	1	.224**	.137**	.019**	-.325**	-.049**
NAEP BASIC level	.224**	1	.126**	-.040**	-.423**	.206**
NAEP PROF level	.137**	.126**	1	.328**	-.393**	.034**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.167**	-.117**	-.192**	.256**	-.050**	-.119**
NAEP BASIC level	.251**	-.266**	-.241**	.394**	-.193**	-.198**
NAEP PROF level	.236**	-.200**	-.280**	.394**	-.198**	-.081**

**. Correlation is significant at the 0.01 level (2-tailed)

APPENDIX AA

California Tables and Figures for Research Question Two, Pearson Correlation.

Table AA1

Descriptive statistics for the 2005 California 4th grade dataset in mathematics

	Mean	Std. Deviation	N *
STATEMATHprofstatus	.9089	.28774	279964
NAEP math met proficiency at the basic level	.8008	.39939	284558
NAEP math met proficiency at the NAEP proficient level	.1137	.31749	284558
TOTAL REVENUE PER STUDENT	$9224.61	$1269.325	280172
EDPER	.6234	.30004	284558
NATAMPER	.0062	.01450	284558
ASIANPER	.0954	.12434	284558
BLACKPER	.0834	.10624	284558
HISPANICPER	.5701	.30628	284558
WHITEPER	.2314	.25221	284558
ELLPER	.4250	.29376	279937
SPECIALEDPER	.0933	.08661	269135

Table AA2

Pearson correlation matrix for the 2005 California 4th grade listed variables (Demographic, financial, and proficiency status) in mathematics: Weighted by school enrollment.

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.428**	.114**	-.145**	-.193**	.029**
NAEP BASIC Math	.428**	1	.179**	-.243**	-.388**	.078**
NAEP PROF Math	.114**	.179**	1	.006**	-.574**	.014**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.130**	-.325**	-.131**	.217**	-.134**	-.023**
NAEP BASIC Math	.250**	-.228**	-.373**	.399**	-.319**	.003
NAEP PROF Math	.320**	-.158**	-.512**	.503**	-.416**	.045**

**. Correlation is significant at the 0.01 level (2-tailed)

Table AA3

Descriptive statistics for the 2005 California 4th grade dataset in ELA and reading

	Mean	Std. Deviation	N
STATEELAprofstatus	.8164	.38713	279964
NAEPREADPROFbasic	.3649	.48142	284558
NAEPREADPROFprof	.0501	.21821	284558
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	$ 9224.61	$ 1269.325	280172
EDPER	.6234	.30004	284558
NATAMPER	.0062	.01450	284558
ASIANPER	.0954	.12434	284558
BLACKPER	.0834	.10624	284558
HISPANICPER	.5701	.30628	284558
WHITEPER	.2314	.25221	284558
ELLPER	.4250	.29376	279937
SPECIALEDPER	.0933	.08661	269135

Table AA4

Pearson correlations matrix for the 2005 California 4th grade listed variables in ELA and reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	NAEP BASIC level Read	NAEP PROF level Read	TOTAL REVENUE	EDPER	NATAMPER
STATE ELA	1	.364**	.110**	-.214**	-.420**	.072**
NAEP BASIC Read	.364*	1	.303**	-.143**	-.729**	.048**
NAEP PROF Read	.110**	.303*	1	.044**	-.405**	-.013**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE ELA	.274**	-.090**	-.419**	.388**	-.438**	.030**
NAEP BASIC Read	.336**	-.124**	-.699**	.680**	-.640**	.078**
NAEP PROF Read	.204**	-.120**	-.349**	.364**	-.266**	-.054**

**. Correlation is significant at the 0.01 level (2-tailed)

Table AA5

Descriptive statistics for the 2005 California 8th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.5662	.49559	374855
NAEP math met proficiency at the basic level	.5117	.49986	408495
NAEP math met proficiency at the NAEP proficient level	.0445	.20612	408495
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	9296.53	1292.792	398351
EDPER	.5623	.26532	408538
NATAMPER	.0059	.01131	408538
ASIANPER	.1079	.13346	408538
BLACKPER	.0929	.10938	408538
HISPANICPER	.5398	.28404	408538
WHITEPER	.2425	.23294	408538
ELLPER	.2831	.20517	393066
SPECIALEDPER	.1156	.09012	375723

Table AA6

Pearson correlations matrix for the 2005 California 8th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	**NAEP BASIC level Math**	**NAEP PROF level Math**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATE Math	1	.522**	.174**	-.282**	-.568**	.119**
NAEP BASIC Math	.522**	1	.211**	-.281**	-.698**	.134**
NAEP PROF Math	.174**	.174**	1	-.067**	-.346**	-.020**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.290**	-.194**	-.518**	.520**	-.383**	.032**
NAEP BASIC Math	.393**	-.144**	-.667**	.622**	-.566**	-.186**
NAEP PROF Math	.368**	-.134**	-.313**	.223**	-.228**	-.052**

**. Correlation is significant at the 0.01 level (2-tailed)

Table AA7

Descriptive statistics for the 2005 California 8th grade dataset in ELA and reading

	Mean	Std. Deviation	N
STATEELAprofstatus	.6174	.48602	401502
NAEPREADPROFbasic	.6491	.47726	397001
NAEPREADPROFprof	.0050	.07071	397001
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	9296.53	1292.792	398351
EDPER	.5623	.26532	408538
NATAMPER	.0059	.01131	408538
ASIANPER	.1079	.13346	408538
BLACKPER	.0929	.10938	408538
HISPANICPER	.5398	.28404	408538
WHITEPER	.2425	.23294	408538
ELLPER	.2831	.20517	393066
SPECIALEDPER	.1156	.09012	375723

Table AA8

Pearson correlations matrix for the 2005 California 8th grade listed variables in ELA and reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	NAEP BASIC level Read	NAEP PROF level Read	TOTAL REVENUE	EDPER	NATAMPER
STATE ELA	1	.251**	.057**	-.326**	-.722**	.199**
NAEP BASIC Read	.251**	1	.052**	-.203**	-.263**	.018**
NAEP PROF Read	.057**	.052**	1	.023**	-.042**	.243**

	ASIANPER	BLACKPER	HISPANICPER	WHITEPER	ELLPER	SPECIALEDPER
STATE ELA	.377**	-.152**	-.684**	.652**	-.631**	-.147**
NAEP BASIC Read	.108**	-.141**	-.195**	.222**	-.130**	.030**
NAEP PROF Read	.120**	.005**	-.098**	.040**	-.047**	-.025**

**. Correlation is significant at the 0.01 level (2-tailed)

Table AA9

Descriptive statistics for the 2007 California 4th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.9590	.19831	183676
NAEP math met proficiency at the basic level	.7849	.41091	183320
NAEP math met proficiency at the NAEP proficient level	.1202	.32514	183320
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	11488.60	2765.230	183932
EDPER	.6501	.30680	183150
NATAMPER	.0042	.00651	183932
ASIANPER	.1001	.13538	183932
BLACKPER	.0785	.10210	183932
HISPANICPER	.6010	.30935	183932
WHITEPER	.1994	.24594	183932
ELLPER	.4394	.27734	179901
SPECIALEDPER	.0917	.07911	175392

Table AA10

Pearson correlations matrix for the 2007 California 4th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.082**	.073**	-.008**	-.144**	.010**
NAEP BASIC Math	.082**	1	.193**	-.098**	-.385**	.111**
NAEP PROF Math	.073**	.193**	1	-.061**	-.622**	.004
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.108**	-.047**	-.129**	.121**	-.187**	.048**
NAEP BASIC Math	.254**	-.064**	-.395**	.354**	-.453**	.010**
NAEP PROF Math	.343**	-.164**	-.570**	.563**	-.424**	.017**

**. Correlation is significant at the 0.01 level (2-tailed)

Table AA11

Descriptive statistics for the 2007 California 4th grade dataset in English Language Arts and reading

	Mean	Std. Deviation	N
STATEELAprofstatus	.8955	.30593	183676
NAEPREADPROFbasic	.3684	.48236	183320
NAEPREADPROFprof	.0652	.24682	183320
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	11488.60	2765.230	183932
EDPER	.6501	.30680	183150
NATAMPER	.0042	.00651	183932
ASIANPER	.1001	.13538	183932
BLACKPER	.0785	.10210	183932
HISPANICPER	.6010	.30935	183932
WHITEPER	.1994	.24594	183932
ELLPER	.4394	.27734	179901
SPECIALEDPER	.0917	.07911	175392

Table AA12

Pearson correlation matrix for the 2007 California 4th grade listed variables in ELA and reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	NAEP BASIC level Read	NAEP PROF level Read	TOTAL REVENUE	EDPER	NATAMPER
STATE ELA	1	.257**	.089**	-.087**	-.272**	.100**
NAEP BASIC Read	.257**	1	.346**	-.187**	-.779**	.178**
NAEP PROF Read	.089**	.346**	1	-.019**	-.504**	.003
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE ELA	.198**	.010**	-.307**	.252**	-.431**	.079**
NAEP BASIC Read	.476**	.009**	-.801**	.696**	-.712**	.009**
NAEP PROF Read	.261**	-.148**	-.444**	.445**	-.344**	.019**

**. Correlation is significant at the 0.01 level (2-tailed)

Table AA13

Descriptive statistics for the 2007 California 8th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.5986	.49018	301358
NAEP math met proficiency at the basic level	.5474	.49774	302628
NAEP math met proficiency at the NAEP proficient level	.0808	.27251	302628
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	11419.31	1624.585	301548
EDPER	.5953	.26591	300317
NATAMPER	.0060	.01592	302971
ASIANPER	.1046	.11959	302971
BLACKPER	.0738	.09524	302971
HISPANICPER	.5629	.28786	302971
WHITEPER	.2348	.23061	302971
ELLPER	.2826	.19615	297105
SPECIALEDPER	.1126	.06134	290614

Table AA14

Pearson correlation matrix for the 2007 California 8th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	**NAEP BASIC level Math**	**NAEP PROF level Math**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATE Math	1	.567**	.242**	-.299**	-.516**	.029**
NAEP BASIC Math	.567**	1	.270**	-.300**	-.688**	.100**
NAEP PROF Math	.242**	.270**	1	-.192**	-.516**	-.004*
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.301**	-.193**	-.463**	.465**	-.275**	-.203**
NAEP BASIC level	.408**	-.131**	-.665**	.626**	-.543**	-.056**
NAEP PROF level	.279**	-.145**	-.443**	.440**	-.278**	-.129**

**. Correlation is significant at the 0.01 level (2-tailed)

Table AA15

Descriptive statistics for the 2007 California 8th grade dataset in English language arts and reading

Descriptive Statistics

	Mean	Std. Deviation	N
STATEELAprofstatus	.6838	.46500	302744
NAEPREADPROFbasic	.6063	.48858	302589
NAEPREADPROFprof	.0265	.16068	302589
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	11419.31	1624.585	301548
EDPER	.5953	.26591	300317
NATAMPER	.0060	.01592	302971
ASIANPER	.1046	.11959	302971
BLACKPER	.0738	.09524	302971
HISPANICPER	.5629	.28786	302971
WHITEPER	.2348	.23061	302971
ELLPER	.2826	.19615	297105
SPECIALEDPER	.1126	.06134	290614

Table AA16

Pearson correlation matrix for the 2007 California 8th grade listed variables in ELA and reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	NAEP BASIC level Reading	NAEP PROF level Reading	TOTAL REVENUE	EDPER	NATAMPER
STATE ELA	1	.581**	.112**	-.394**	-.601**	.149**
NAEP BASIC Reading	.581**	1	.133**	-.253**	-.622**	.114**
NAEP PROF Reading	.112**	.133**	1	.007**	-.278**	-.015**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE ELA	.349**	-.155**	-.599**	.587**	-.457**	.000
NAEP BASIC level	.401**	-.107**	-.575**	.516**	-.530**	-.025**
NAEP PROF level	.276**	-.081**	-.237**	.181**	-.192**	-.077**

**. Correlation is significant at the 0.01 level (2-tailed)

APPENDIX BB

California Tables and Figures for Research Question Two, Logistic Regression.

Table BB1

Logistic regression observed and predicted frequencies with the constant model in the California 2005 4th grade dataset with the ELA state accountability assessment as the dependent variable.

	Observed		Predicted		
			did not meet Proficiency .00	met Proficiency 1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	45739	.0
		met proficiency status	0	217847	100.0
	Overall Percentage				82.6

a. The cut value is .500

Table BB2

Logistic regression observed and predicted frequencies with the predicted model with the California 2005 4th grade ELA state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency .00	met Proficiency 1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	27380	18359	59.9
		met proficiency status	14132	203715	93.5
	Overall Percentage				87.7

a. The cut value is .500

Table BB3

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 4th Grade California ELA assessment as the dependent variable

		Dimension Chi-square	df	Sig.
Step 1	Step	2969.519	1	.000
	Block	2969.519	1	.000
	Model	105131.866	10	.000

Table BB4

Logistic regression model summary assessing goodness of fit with the 2005 California 4th grade ELA assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	138122.783[a]	.329	.546

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table BB5

Logistic Regression Analysis for the 2005 California 4th grade data with the ELA assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	.057	.011	24.979	1	.000	1.059	1.035	1.082
	EDPER	-2.889	.055	2716.584	1	.000	.056	.050	.062
	NATAMPER	-10.345	.598	299.587	1	.000	.000	.000	.000
	ASIANPER	34.899	.492	5040.643	1	.000	1.433	5.469	3.756
	BLACKPER	13.495	.457	870.358	1	.000	725940.193	296162.422	1779392.403
	HISPANICPER	18.651	.460	1644.674	1	.000	1.259	5.110	3.100
	WHITEPER	24.148	.532	2059.919	1	.000	3.071	1.082	8.712
	ELLPER	-4.151	.052	6479.904	1	.000	.016	.014	.017
	SPECIALEDPER	-1.339	.074	325.627	1	.000	.262	.227	.303
	NAEPREADPROFbasic	16.714	97.400	.029	1	.864	1.814	.000	1.463
	Constant	-12.989	.453	821.117	1	.000	.000		

Figure BB1. Logistic Regression for the 2005 California 4th grade dataset with the ELA state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

  160000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R 120000 +                                                                                                   m+
E        I                                                                                                   mI
Q        I                                                                                                   mI
U        I                                                                                                   mI
E  80000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                                                                                   mI
   40000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                   mI
         I                                                                                                  mmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

    Predicted Probability is of Membership for met proficiency status
    The Cut Value is .50
    Symbols: d - did not meet proficiency status
             m - met proficiency status
    Each Symbol Represents 10000 Cases.
```

Table BB6

Logistic regression observed and predicted frequencies with the constant model with the California 2005 8th grade mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	148293	.0
		met proficiency status	0	186618	100.0
	Overall Percentage				55.7

a. The cut value is .500

Table BB7

Logistic regression observed and predicted frequencies with the predicted model with the California 2005 8th grade mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	115866	32427	78.1
		met proficiency status	35209	151409	81.1
	Overall Percentage				79.8

a. The cut value is .500

Table BB8

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 8th Grade California mathematics assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	55.651	1	.000
	Block	55.651	1	.000
	Model	173221.303	11	.000

Table BB9

Logistic regression model summary assessing goodness of fit with the California 2005 8th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	286668.646[a]	.404	.541

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table BB10

Logistic Regression Analysis for the 2005 California 8th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-.613	.007	7446.216	1	.000	.542	.534	.549
	EDPER	-2.416	.043	3134.002	1	.000	.089	.082	.097
	NATAMPER	-26.810	.693	1497.450	1	.000	.000	.000	.000
	ASIANPER	-15.457	.379	1660.765	1	.000	.000	.000	.000
	BLACKPER	-24.082	.385	3917.762	1	.000	.000	.000	.000
	HISPANICPER	-20.646	.377	2998.810	1	.000	.000	.000	.000
	WHITEPER	-18.284	.385	2259.921	1	.000	.000	.000	.000
	ELLPER	1.513	.035	1830.583	1	.000	4.538	4.235	4.864
	SPECIALEDPER	6.461	.080	6596.455	1	.000	639.712	547.357	747.651
	NAEPMATHPROFbasic	.903	.012	5643.557	1	.000	2.466	2.409	2.525
	NAEPMATHPROFprof	.336	.047	51.644	1	.000	1.400	1.277	1.534
	Constant	21.201	.371	3259.492	1	.000	1.613E9		

Figure BB2. Logistic Regression Observed and Predicted Probabilities for in California with the 8th grade mathematics state accountability assessment as the dependent variable

```
Step number: 1

Observed Groups and Predicted Probabilities

   16000 +                                                                                                    +
         I                                                                                                    I
         I                  m                                                                              m  I
F        I                  m                                                                              m  I
R  12000 +                  m                                                                              m m+
E        I                  d                                                                          m   m mI
Q        I         d        d                                                                          m   m mI
U        I         d        d                                                                          mm  mmmI
E   8000 +         d  d     d                                                                          mm  mmm+
N        I         d  d     d      m                      m                                          m mmmmmmmI
C        I         d  d     d    m m                      m                                   m   m  m mmmmmmmI
Y        I        dd dd     d    mmd    m                 m            m          m        m  m   m  m mmmmmmmI
    4000 +  d d  ddd dd     d    dmd   dd        m        m    md      m       m  m m  m  mm  m  mmmmmmmmmmmmm+
         I  d d dddd dd d   d dm dmd  mdd mm  m  dm dd    d  mdmd      m   m   mm m m  m  mm  mm mmmmmmmmmmmmmI
         I  ddd ddddddd d d dmddmdmddmmdd dm dm  dmddd    d mmdmd  m   mm  m  dmd mmmmmdd dmmmmm mmdmmmmmmmmmmI
         I  ddd ddddddddddd dmddmdddddddd ddddm dddddd  mmdddddmdm mmmmdd  dd dmd mmdmdddmddmmdmmdmdmdmmmmmmmdI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

      Predicted Probability is of Membership for met proficiency
      The Cut Value is .50
      Symbols: d - did not meet proficiency
               m - met proficiency
      Each Symbol Represents 1000 Cases.
```

Table BB11

Logistic regression observed and predicted frequencies with the constant model in California 2005 8th grade state accountability assessment ELA results

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	0	130660	.0
		met proficiency status	0	219349	100.0
	Overall Percentage				62.7

a. The cut value is .500

Table BB12

Logistic regression observed and predicted frequencies with the predicted model in California 2005 8th grade ELA state accountability assessment results

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	117557	13103	90.0
		met proficiency status	16570	202779	92.4
	Overall Percentage				91.5

a. The cut value is .500

Table BB13

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 8th Grade California ELA assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	10.726	1	.001
	Block	10.726	1	.001
	Model	305119.896	10	.000

Table BB14

Logistic regression model summary assessing goodness of fit with the California 2005 8th grade ELA assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	157375.762[a]	.582	.793

a. a. Estimation terminated at iteration number 8 because parameter estimates changed by less than .001.

Table BB15

Logistic regression analysis for the 2005 California 8th grade data with the ELA assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	.064	.010	38.351	1	.000	1.066	1.045	1.088
	EDPER	-8.936	.069	16858.006	1	.000	.000	.000	.000
	NATAMPER	6.483	.618	109.890	1	.000	653.924	194.586	2197.567
	ASIANPER	26.503	.458	3348.185	1	.000	3.236E11	1.319E11	7.942E11
	BLACKPER	11.846	.454	681.029	1	.000	139526.203	57315.159	339658.157
	HISPANICPER	17.035	.448	1447.620	1	.000	2.501E7	1.040E7	6.014E7
	WHITEPER	24.831	.474	2741.406	1	.000	6.083E10	2.401E10	1.541E11
	ELLPER	-2.940	.048	3731.592	1	.000	.053	.048	.058
	SPECIALEDPER	-.112	.069	2.637	1	.104	.894	.782	1.023
	NAEPREADPROFbasic	.046	.014	10.737	1	.001	1.047	1.019	1.076
	Constant	-10.989	.444	612.738	1	.000	.000		

Figure BB3. Logistic for the California 2005 8th dataset with ELA as the dependent variable: observed and predicted probabilities

```
               Step number: 1

               Observed Groups and Predicted Probabilities

   160000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R  120000 +                                                                                                   m+
E         I                                                                                                   mI
Q         I                                                                                                   mI
U         I                                                                                                   mI
E   80000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    40000 +                                                                                                   m+
          I                                                                                                   mI
          I  d                                                                                                mI
          Idddddd     dd                   d                                                    m         m m mmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 10000 Cases.
```

Table BB16

Logistic regression observed and predicted frequencies with the constant model in California 2007 4th grade state accountability assessment mathematics results

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	6920	.0
		met proficiency status	0	165291	100.0
	Overall Percentage				96.0

a. The cut value is .500

Table BB17

Logistic regression observed and predicted frequencies with the predicted model in California 2007 4th grade mathematics state accountability assessment results

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	6920	.0
		met proficiency status	225	165066	99.9
	Overall Percentage				95.9

a. The cut value is .500

Table BB18

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 4th Grade California mathematics assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	29.227	1	.000
	Block	29.227	1	.000
	Model	14322.668	11	.000

Table BB19

Logistic regression model summary assessing goodness of fit with the California 2007 4th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	43721.440[a]	.080	.279

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table BB20

Logistic Regression Analysis for the 2007 California 4th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1	DUMMYREV	.794	.019	1817.064	1	.000	2.212	2.133	2.294
	EDPER	-1.733	.142	148.148	1	.000	.177	.134	.234
	NATAMPER	-54.348	1.816	895.949	1	.000	.000	.000	.000
	ASIANPER	40.526	.712	3239.222	1	.000	3.985E17	9.869E16	1.609E18
	BLACKPER	25.272	.652	1503.483	1	.000	9.449E10	2.634E10	3.390E11
	HISPANICPER	31.026	.639	2360.108	1	.000	2.981E13	8.527E12	1.042E14
	WHITEPER	32.142	.721	1989.436	1	.000	9.097E13	2.216E13	3.735E14
	ELLPER	-7.579	.129	3425.311	1	.000	.001	.000	.001
	SPECIALEDPER	1.257	.183	46.983	1	.000	3.516	2.454	5.037
	NAEPMATHPROFbasic	-.328	.029	124.562	1	.000	.721	.680	.763
	NAEPMATHPROFprof	13.553	225.509	.004	1	.952	769195.300	.000	6.914E197
	Constant	-23.656	.584	1639.227	1	.000	.000		

Figure BB4. Logistic for the California 2007 4th dataset with mathematics as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

    160000 +                                                                                                    +
           I                                                                                                    I
           I                                                                                                    I
F          I                                                                                                    I
R   120000 +                                                                                                    +
E          I                                                                                                    I
Q          I                                                                                                    I
U          I                                                                                                    I
E    80000 +                                                                                                   m+
N          I                                                                                                   mI
C          I                                                                                                   mI
Y          I                                                                                                   mI
     40000 +                                                                                                   m+
           I                                                                                                   mI
           I                                                                                                  mmI
           I                                                                                             mmmmmmmI
Predicted  ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:    0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:   ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met proficiency
The Cut Value is .50
Symbols: d - did not meet proficiency
         m - met proficiency
Each Symbol Represents 10000 Cases.
```

Table BB21

Logistic regression observed and predicted frequencies with the constant model in California 2007 4th grade state ELA accountability assessment results

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	0	18585	.0
		met proficiency status	0	153626	100.0
	Overall Percentage				89.2

a. The cut value is .500

Table BB22

Logistic regression observed and predicted frequencies with the predicted model in California 2007 4th grade ELA state accountability assessment results

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	8148	10437	43.8
		met proficiency status	3013	150613	98.0
	Overall Percentage				92.2

a. The cut value is .500

Table BB23

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 4th Grade California ELA assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	1005.193	1	.000
	Block	1005.193	1	.000
	Model	53752.221	10	.000

Table BB24

Logistic regression model summary assessing goodness of fit with the California 2007 4th grade ELA assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	64089.783[a]	.268	.541

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table BB25

Logistic Regression Analysis for the 2007 California 4th grade data with the ELA assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	.717	.017	1875.915	1	.000	2.048	1.983	2.116
	EDPER	-1.435	.096	223.786	1	.000	.238	.197	.287
	NATAMPER	-204.453	3.899	2749.798	1	.000	.000	.000	.000
	ASIANPER	14.363	1.301	121.935	1	.000	1728267.755	135042.470	2.212E7
	BLACKPER	-5.932	1.280	21.478	1	.000	.003	.000	.033
	HISPANICPER	1.946	1.260	2.386	1	.122	6.998	.593	82.635
	WHITEPER	9.506	1.340	50.299	1	.000	13435.264	971.353	185829.760
	ELLPER	-11.276	.107	11050.319	1	.000	.000	.000	.000
	SPECIALEDPER	3.058	.143	454.866	1	.000	21.295	16.077	28.206
	NAEPREADPROFbasic	20.778	79.236	.069	1	.793	1.056E9	.000	2.945E76
	Constant	6.451	1.237	27.190	1	.000	633.339		

Figure BB5. Logistic for the California 2007 4th dataset with ELA as the dependent variable: observed and predicted probabilities

```
               Step number: 1

               Observed Groups and Predicted Probabilities

   160000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R 120000 +                                                                                                     +
E         I                                                                                                    I
Q         I                                                                                                    I
U         I                                                                                                   mI
E   80000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    40000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          I                                                                                           mm   mmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
```

Table BB26

Logistic regression observed and predicted frequencies with the constant model in the California 2007 8th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	113524	.0
		met proficiency status	0	171743	100.0
	Overall Percentage				60.2

a. The cut value is .500

Table BB27

Logistic regression observed and predicted frequencies with the predicted model in the California 8th grade 2007 dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	88319	25205	77.8
		met proficiency status	30828	140915	82.0
	Overall Percentage				80.4

a. The cut value is .500

Table BB28

Logistic regression omnibus test of model coefficients assessing goodness of fit with the California 2007 8th grade mathematics assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	684.066	1	.000
	Block	684.066	1	.000
	Model	152022.326	11	.000

Table BB29

Logistic regression model summary assessing goodness of fit with the California 8th 2007 grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	231476.139[a]	.413	.559

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table BB30

Logistic Regression Analysis for the 2007 California 8th grade data with the mathematics assessment as the dependent variable

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1	DUMMYREV	.072	.007	96.236	1	.000	1.074	1.059	1.090
	EDPER	-2.671	.055	2375.614	1	.000	.069	.062	.077
	NATAMPER	-36.668	.575	4071.790	1	.000	.000	.000	.000
	ASIANPER	-16.289	.423	1481.706	1	.000	.000	.000	.000
	BLACKPER	-24.471	.440	3091.529	1	.000	.000	.000	.000
	HISPANICPER	-19.300	.415	2164.695	1	.000	.000	.000	.000
	WHITEPER	-17.430	.430	1644.296	1	.000	.000	.000	.000
	ELLPER	2.049	.035	3518.224	1	.000	7.764	7.255	8.308
	SPECIALEDPER	-8.676	.097	7972.111	1	.000	.000	.000	.000
	NAEPMATHPROFbasic	1.853	.014	17744.984	1	.000	6.382	6.210	6.558
	NAEPMATHPROFprof	17.014	256.690	.004	1	.947	2.449E7	.000	7.654E225
	Constant	20.344	.411	2450.664	1	.000	6.841E8		

Figure BB6. Logistic Regression for the California 2007 8th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

   32000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                   mI
F        I                                                                                                   mI
R  24000 +                                                                                                   m+
E        I                                                                                                   mI
Q        I                                                                                                   mI
U        I                                                                                                   mI
E  16000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                          m                                                                     m mmI
    8000 +               d          m                         m                           m               mmmm+
         I             m d       m  d  m d                    d     m                m    m     m      m mmmmmI
         I    d    d   d d      dd dd  m d      m m       d   d     m         m    m m    mmm  mm  m mmm mmmmmI
         Idd  dddddddd ddd      ddmddd d dddmd dd dmdmd mddm dd  m  d   m   d mdd mmmmdmmmdmm mmm  d mmm mmmmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met proficiency
The Cut Value is .50
Symbols: d - did not meet proficiency
         m - met proficiency
Each Symbol Represents 2000 Cases.
```

Table BB31

Logistic regression observed and predicted frequencies with the constant model in the California 2007 8th grade dataset with the ELA state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	0	92504	.0
		met proficiency status	0	193806	100.0
	Overall Percentage				67.7

a. The cut value is .500

Table BB32

Logistic regression observed and predicted frequencies with the predicted model in the California 2007 8th grade dataset with the ELA state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	78827	13677	85.2
		met proficiency status	16710	177096	91.4
	Overall Percentage				89.4

a. The cut value is .500

Table BB33

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 8th Grade California ELA assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	4.025	1	.045
	Block	4.025	1	.045
	Model	211793.100	11	.000

Table BB34

Logistic regression model summary assessing goodness of fit with the California 8th grade 2007 ELA assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	148486.219[a]	.523	.730

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table BB35

Logistic Regression Analysis for the 2007 California 8th grade data with the ELA assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-.297	.009	994.066	1	.000	.743	.730	.757
	EDPER	-3.812	.079	2322.196	1	.000	.022	.019	.026
	NATAMPER	-4.260	.815	27.340	1	.000	.014	.003	.070
	ASIANPER	22.903	.503	2069.865	1	.000	8.840E9	3.296E9	2.371E10
	BLACKPER	11.347	.494	527.995	1	.000	84705.098	32179.114	222969.273
	HISPANICPER	16.689	.474	1237.510	1	.000	1.771E7	6986677.504	4.487E7
	WHITEPER	33.777	.534	3996.695	1	.000	4.667E14	1.638E14	1.330E15
	ELLPER	.891	.043	423.126	1	.000	2.437	2.239	2.653
	SPECIALEDPER	-1.026	.102	100.921	1	.000	.358	.293	.438
	NAEPREADPROFbasic	.880	.015	3397.149	1	.000	2.411	2.341	2.484
	NAEPREADPROFprof	12.407	348.162	.001	1	.972	244562.801	.000	5.555E301
	Constant	-14.432	.472	935.004	1	.000	.000		

Figure BB7. Logistic Regression for the California 2007 8th grade dataset with the ELA state accountability assessment as the dependent variable: observed and predicted probabilities

```
           Step number: 1

           Observed Groups and Predicted Probabilities

    160000 +                                                                                                    +
           I                                                                                                    I
           I                                                                                                    I
F          I                                                                                                    I
R   120000 +                                                                                                    +
E          I                                                                                                   mI
Q          I                                                                                                   mI
U          I                                                                                                   mI
E    80000 +                                                                                                   m+
N          I                                                                                                   mI
C          I                                                                                                   mI
Y          I                                                                                                   mI
     40000 +                                                                                                   m+
           I                                                                                                   mI
           I                                                                                                   mI
           I         d       m                                                                 m     mmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  dddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 10000 Cases.
```

APPENDIX CC

Michigan Tables and Figures for Research Question Two, Pearson Correlation.

Table CC1

Descriptive statistics for the 2005 Michigan 4th grade dataset in mathematics

Descriptive Statistics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.8716	.33451	47058
NAEP math met proficiency at the basic level	.8768	.32872	47109
NAEP math met proficiency at the NAEP proficient level	.3045	.46022	47109
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9756.85	1313.817	47109
EDPER	.3847	.29179	45721
NATAM	.0098	.03030	47109
ASIANPER	.0293	.04529	47109
BLACKPER	.2039	.32550	47109
HISPANICPER	.0405	.06353	47109
WHITEPER	.7126	.32180	47109
ELLPER	.0423	.08535	46553
SPECIALEDPER	.1194	.06947	43703

Table CC2

Pearson correlation matrix for the 2005 Michigan 4th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.392**	.254**	-.173**	-.428**	-.140**
NAEP BASIC Math	.392**	1	.248**	-.316**	-.637**	.096**
NAEP PROF Math	.254**	.248**	1	-.091**	-.484**	.025**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.162**	-.424**	.152**	.385**	.147**	.003
NAEP BASIC level	.129**	-.780**	-.039**	.766**	.073**	.100**
NAEP PROF level	.313**	-.345**	-.191**	.340**	.007	-.199**

**. Correlation is significant at the 0.01 level (2-tailed)

Table CC3

Descriptive statistics for the 2005 Michigan 4th grade dataset in English language arts and reading

	Mean	Std. Deviation	N
STATEELAprofstatus	.9718	.16560	47058
NAEPREADPROFbasic	.7848	.41097	43401
NAEPREADPROFprof	.0655	.24734	43401
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9756.85	1313.817	47109
EDPER	.3847	.29179	45721
NATAM	.0098	.03030	47109
ASIANPER	.0293	.04529	47109
BLACKPER	.2039	.32550	47109
HISPANICPER	.0405	.06353	47109
WHITEPER	.7126	.32180	47109
ELLPER	.0423	.08535	46553
SPECIALEDPER	.1194	.06947	43703

Table CC4

Pearson correlation matrix for the 2005 Michigan 4th grade listed variables in ELA and reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	NAEP BASIC level Reading	NAEP PROF level Reading	TOTAL REVENUE	EDPER	NATAMPER
STATE ELA	1	-.093**	.047**	-.160**	-.294**	.051**
NAEP BASIC Reading	-.093**	1	.139**	.111**	.099**	.069**
NAEP PROF Reading	.047**	.139**	1	.103**	.119**	-.051**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE ELA	.034**	-.397**	.101**	.370**	.070**	-.012**
NAEP BASIC level	.052**	.080**	-.014**	-.090**	.027**	-.062**
NAEP PROF level	-.014**	.119**	.123**	-.134**	.069**	-.060**

**. Correlation is significant at the 0.01 level (2-tailed)

Table CC5

Descriptive statistics for the 2005 Michigan 8th grade dataset in mathematics

	Mean	Std. Deviation	N
NAEP math met proficiency at the basic level	.7922	.40570	64240
NAEP math met proficiency at the NAEP proficient level	.1256	.33144	64240
STATEMATHprofstatus	.8653	.34137	63570
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9939.95	1373.088	64247
EDPER	.3482	.25181	63584
NATAMPER	.0070	.01250	64264
ASIANPER	.0208	.02826	64264
BLACKPER	.2173	.32299	64264
HISPANICPER	.0345	.04684	64264
WHITEPER	.7180	.32500	64264
ELLPER	.0502	.14336	63222
SPECIALEDPER	.1436	.07192	59362

Table CC6

Pearson correlation matrix for the 2005 Michigan 8th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	State Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.640**	.150**	-.280**	-.666**	.098**
NAEP BASIC Math	.640**	1	.194**	-.267**	-.705**	.085**
NAEP PROF Math	.150**	.194**	1	.166**	-.355**	-.123**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.187**	-.704**	-.020**	.682**	-.137**	.128**
NAEP BASIC level	.167**	-.760**	-.091**	.749**	-.098**	.074**
NAEP PROF level	.401**	-.177**	-.082**	.155**	-.051**	-.082**

**. Correlation is significant at the 0.01 level (2-tailed)

Table CC7

Descriptive statistics for the 2005 Michigan 8th grade dataset in ELA and reading

	Mean	Std. Deviation	N
NAEPREADPROFbasic	.8648	.34191	64223
NAEPREADPROFprof	.1136	.31737	64223
STATEELAprofstatus	.9491	.21977	61491
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9939.95	1373.088	64247
EDPER	.3482	.25181	63584
NATAMPER	.0070	.01250	64264
ASIANPER	.0208	.02826	64264
BLACKPER	.2173	.32299	64264
HISPANICPER	.0345	.04684	64264
WHITEPER	.7180	.32500	64264
ELLPER	.0502	.14336	63222
SPECIALEDPER	.1436	.07192	59362

Table CC8

Pearson correlation matrix for the 2005 Michigan 8th grade listed variables in ELA and reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	NAEP BASIC level Reading	NAEP PROF level Reading	TOTAL REVENUE	EDPER	NATAMPER
STATE ELA	1	.618**	.085**	-.204**	-.428**	.078**
NAEP BASIC Reading	.618**	1	.142**	-.316**	-.684**	.120**
NAEP PROF Reading	.085**	.142**	1	.069**	-.284**	-.087**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE ELA	.163**	-.511**	.016**	.488**	.054**	.077**
NAEP BASIC level	.167**	-.740**	-.068**	.728**	.065**	.074**
NAEP PROF level	.356**	-.183**	-.115**	.170**	-.066**	-.232**

**. Correlation is significant at the 0.01 level (2-tailed)

Table CC9

Descriptive statistics for the 2007 Michigan 4th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.9461	.22574	42499
NAEP math met proficiency at the basic level	.9023	.29691	42469
NAEP math met proficiency at the NAEP proficient level	.2616	.43954	42469
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2006-07]	10943.16	2144.399	42631
EDPER	.3345	.25697	42631
NATAMPER	.0182	.06493	42631
ASIANPER	.0284	.04792	42631
BLACKPER	.1665	.30284	42631
HISPANIC	.0332	.04712	42631
WHITEPER	.7439	.30437	42631
ELLPER	.0294	.04082	41694
SPECIALEDPER	.1178	.08030	40110

Table CC10

Pearson correlation matrix for the 2007 Michigan 4th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	**NAEP BASIC level Math**	**NAEP PROF level Math**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATE Math	1	.725**	.142**	-.199**	-.422**	.061**
NAEP BASIC Math	.725**	1	.196**	-.254**	-.623**	.080**
NAEP PROF Math	.142**	.196**	1	.117**	-.494**	-.124**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.107**	-.641**	.150**	.578**	.092**	.146**
NAEP BASIC level	.156**	-.803**	.055**	.741**	.130**	.098**
NAEP PROF level	.346**	-.256**	-.131**	.243**	.152**	-.238**

**. Correlation is significant at the 0.01 level (2-tailed)

Table CC11

Descriptive statistics for the 2007 Michigan 4th grade dataset in ELA and reading

	Mean	Std. Deviation	N
STATEELAprofstatus	.9329	.25021	42499
NAEPREADPROFbasic	.8656	.34106	42469
NAEPREADPROFprof	.1685	.37431	42469
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2006-07]	10943.16	2144.399	42631
EDPER	.3345	.25697	42631
NATAMPER	.0182	.06493	42631
ASIANPER	.0284	.04792	42631
BLACKPER	.1665	.30284	42631
HISPANICPER	.0332	.04712	42631
WHITEPER	.7439	.30437	42631
ELLPER	.0294	.04082	41694
SPECIALEDPER	.1178	.08030	40110

Table CC12

Pearson correlation matrix for the 2007 Michigan 4th grade listed variables in ELA and reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	NAEP BASIC level Reading	NAEP PROF level Reading	TOTAL REVENUE	EDPER	NATAMPER
STATE ELA	1	.664**	.117**	-.088**	-.422**	.035**
NAEP BASIC Reading	.664**	1	.177**	-.337**	-.618**	.017**
NAEP PROF Reading	.117**	.177**	1	.024**	-.377**	-.107**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE ELA	.069**	-.482**	.077**	.444**	.034**	-.061**
NAEP BASIC level	.149**	-.695**	.119**	.642**	.103**	-.150**
NAEP PROF level	.206**	-.221**	-.116**	.230**	-.038**	-.206**

**. Correlation is significant at the 0.01 level (2-tailed)

Table CC13

Descriptive statistics for the 2007 Michigan 8th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.8818	.32282	55146
NAEP math met proficiency at the basic level	.8678	.33871	56167
NAEP math met proficiency at the NAEP proficient level	.1567	.36348	56167
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2006-07]	10542.11	1763.064	56098
EDPER	.3299	.23657	56167
NATAMPER	.0178	.06074	56167
ASIANPER	.0269	.04506	56167
BLACKPER	.1290	.22747	56167
HISPANICPER	.0369	.05257	56167
WHITEPER	.7786	.24539	56167
ELLPER	.0401	.08356	54723
SPECIALEDPER	.1410	.08563	53198

Table CC14

Pearson correlation matrix for the 2007 Michigan 8th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	**NAEP BASIC level Math**	**NAEP PROF level Math**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATE Math	1	.725**	.142**	-.199**	-.422**	.061**
NAEP BASIC Math	.725**	1	.196**	-.254**	-.623**	.080**
NAEP PROF Math	.142**	.196**	1	.117**	-.494**	-.124**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.107**	-.641**	.150**	.578**	.092**	.146**
NAEP BASIC level	.156**	-.803**	.055**	.741**	.130**	.098**
NAEP PROF level	.346**	-.256**	-.131**	.243**	.152**	-.238**

**. Correlation is significant at the 0.01 level (2-tailed)

Table CC15

Descriptive statistics for the 2007 Michigan 8th grade dataset in English language arts and reading

	Mean	Std. Deviation	N
STATEELAprofstatus	.9040	.29459	55146
NAEPREADPROFbasic	.8873	.31621	56167
NAEPREADPROFprof	.0767	.26617	56167
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2006-07]	10542.11	1763.064	56098
EDPER	.3299	.23657	56167
NATAMPER	.0178	.06074	56167
ASIANPER	.0269	.04506	56167
BLACKPER	.1290	.22747	56167
HISPANICPER	.0369	.05257	56167
WHITEPER	.7786	.24539	56167
ELLPER	.0401	.08356	54723
SPECIALEDPER	.1410	.08563	53198

Table CC16

Pearson correlation matrix for the 2007 Michigan 8th grade listed variables in ELA and reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	NAEP BASIC level Reading	NAEP PROF level Reading	TOTAL REVENUE	EDPER	NATAMPER
STATE ELA	1	.573**	.095**	-.237**	-.638**	.052**
NAEP BASIC Reading	.573**	1	.103**	-.249**	-.606**	.074**
NAEP PROF Reading	095**	.103**	1	.060**	-.272**	-.053**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE ELA	-.126**	-.679**	-.338**	.705**	-.279**	-.247**
NAEP BASIC level	.108**	-.683**	-.243**	.637**	-.340**	-.166**
NAEP PROF level	-.006	-.113**	-.069**	.143**	-.057**	-.007

**. Correlation is significant at the 0.01 level (2-tailed)

APPENDIX DD

Michigan Tables and Figures for Research Question Two, Logistic Regression.

Table DD1

Logistic regression observed and predicted frequencies with the constant model in the Michigan 2005 4th grade dataset with the mathematics state accountability assessment results as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	5650	.0
		met proficiency status	0	36109	100.0
	Overall Percentage				86.5

a. The cut value is .500

Table DD2

Logistic regression observed and predicted frequencies with the predicted model in the Michigan 2005 4th grade dataset with the mathematics state accountability assessment results as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	3289	2361	58.2
		met proficiency status	1117	34992	96.9
	Overall Percentage				91.7

a. The cut value is .500

Table DD3

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 Michigan 4th grade mathematics assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	1657.236	1	.000
	Block	1657.236	1	.000
	Model	15882.599	11	.000

Table DD4

Logistic regression model summary assessing goodness of fit with the Michigan 2005 4th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	17218.870[a]	.316	.578

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table DD5

Logistic Regression Analysis for the 2005 Michigan 4th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	.469	.063	55.545	1	.000	1.598	1.413	1.807
	EDPER	-1.968	.095	428.525	1	.000	.140	.116	.168
	NATAM	-9.762	2.262	18.633	1	.000	.000	.000	.005
	ASIANPER	7.338	2.896	6.421	1	.011	1537.083	5.271	448270.902
	BLACKPER	1.802	2.128	.717	1	.397	6.062	.094	392.989
	HISPANICPER	11.799	2.334	25.559	1	.000	133160.543	1373.278	1.291E7
	WHITEPER	4.554	2.149	4.492	1	.034	95.016	1.408	6410.191
	ELLPER	48.537	1.458	1107.509	1	.000	1.200E21	6.884E19	2.093E22
	SPECIALEDPER	-10.360	.344	908.506	1	.000	.000	.000	.000
	NAEPMATHPROFbasic	-.192	.090	4.587	1	.032	.825	.692	.984
	NAEPMATHPROFprof	19.069	295.501	.004	1	.949	1.911E8	.000	6.488E259
	Constant	-1.822	2.209	.680	1	.410	.162		

Figure DD1. Logistic Regression for the Michigan 2005 4th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

   20000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                   mI
F        I                                                                                                   mI
R  15000 +                                                                                                   m+
E        I                                                                                                   mI
Q        I                                                                                                   mI
U        I                                                                                                   mI
E  10000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                                                                                   mI
    5000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                 m mI
         I                       d         m                                m         m md         m   mmm mmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9        1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met proficiency status
The Cut Value is .50
Symbols: d - did not meet proficiency status
         m - met proficiency status
Each Symbol Represents 1250 Cases.
```

Table DD6

Logistic regression observed and predicted frequencies for the constant model with the Michigan 2005 4th grade ELA assessment as the dependent variable

	Observed		Predicted		
			did not meet ELAprof	met ELAprof	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency	0	1328	.0
		met proficiency	0	36723	100.0
	Overall Percentage				96.5

a. The cut value is .500

Table DD7

Logistic regression observed and predicted frequencies for the constant model with the Michigan 2005 8th grade mathematics assessment as the dependent variable

	Observed		Predicted		
			did not meet MATHprof	met Mathprof	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency	0	7348	.0
		met proficiency	0	50333	100.0
	Overall Percentage				87.3

a. The cut value is .500

Table DD8

Logistic regression observed and predicted frequencies with the predicted model for the Michigan 2005 8th grade mathematics assessment as the dependent variable

	Observed		Predicted		
			did not meet MATHprof	met MATHprof	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency	6324	1024	86.1
		met proficiency	1114	49219	97.8
	Overall Percentage				96.3

a. The cut value is .500

Table DD9

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the Michigan 2005 8th grade mathematics assessment as the dependent variable

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	734.425	1	.000
			Block	734.425	1	.000
			Model	31241.901	10	.000

Table DD10

Logistic regression model summary assessing goodness of fit with the Michigan 2005 8th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	12756.647[a]	.418	.784

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table DD11

Logistic Regression Analysis with the 2005 Michigan 8th grade mathematics assessment as the dependent variable

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1	DUMMYREV	3.183	.125	652.682	1	.000	24.115	18.890	30.785
	EDPER	-10.205	.193	2787.225	1	.000	.000	.000	.000
	NATAMPER	-135.661	7.774	304.513	1	.000	.000	.000	.000
	ASIANPER	-103.474	7.583	186.208	1	.000	.000	.000	.000
	BLACKPER	-146.628	7.412	391.402	1	.000	.000	.000	.000
	HISPANICPER	-162.567	7.441	477.252	1	.000	.000	.000	.000
	WHITEPER	-141.043	7.405	362.817	1	.000	.000	.000	.000
	ELLPER	-4.129	.414	99.384	1	.000	.016	.007	.036
	SPECIALEDPER	5.451	.435	156.741	1	.000	232.966	99.241	546.884
	NAEPMATHPROFbasic	2.719	.111	602.592	1	.000	15.166	12.207	18.844
	Constant	142.044	7.444	364.146	1	.000	4.887E61		

Figure DD2. Logistic Regression for the Michigan 2005 8th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
          Step number: 1

          Observed Groups and Predicted Probabilities

    80000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R   60000 +                                                                                                    +
E         I                                                                                                    I
Q         I                                                                                                    I
U         I                                                                                                    I
E   40000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    20000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          I                                                                                                   mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 5000 Cases.
```

Table DD12

Logistic regression observed and predicted frequencies with the constant model in Michigan with the 2005 8th grade ELA state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	0	2495	.0
		met proficiency status	0	53686	100.0
	Overall Percentage				95.6

a. The cut value is .500

Table DD13

Logistic regression observed and predicted frequencies with the constant model with the Michigan 2007 4th grade mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	1816	.0
		met proficiency status	0	37225	100.0
	Overall Percentage				95.3

a. The cut value is .500

Table DD14

Logistic regression observed and predicted frequencies for the predicted model with the Michigan 2007 4th grade mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	1816	0	100.0
		met proficiency status	0	37225	100.0
	Overall Percentage				100.0

a. The cut value is .500

Table DD15

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 4th grade Michigan mathematics assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	14689.065	9	.000
	Block	14689.065	9	.000
	Model	14689.065	9	.000

Table DD16

Logistic regression model summary assessing goodness of fit with the Michigan 2007 4th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.012[a]	.314	1.000

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table DD17

Logistic Regression Analysis for the 2007 Michigan 4th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	44.421	116.124	.146	1	.702	1.958E19	.000	1.369E118
	EDPER	276.499	369.899	.559	1	.455	1.207E120	.000	.
	NATAMPER	-3877.322	11104.858	.122	1	.727	.000	.000	.
	ASIANPER	8375.216	15499.943	.292	1	.589	.	.000	.
	BLACKPER	-3661.994	10803.317	.115	1	.735	.000	.000	.
	HISPANIC	-1413.400	13630.313	.011	1	.917	.000	.000	.
	WHITEPER	-2999.614	10589.552	.080	1	.777	.000	.000	.
	ELLPER	-10998.376	12191.642	.814	1	.367	.000	.000	.
	SPECIALEDPER	187.362	376.531	.248	1	.619	2.347E81	.000	.
	Constant	3205.084	10590.985	.092	1	.762	.		

Figure DD3. Logistic Regression for the 2007 Michigan 4th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
              Step number: 1

              Observed Groups and Predicted Probabilities

    40000 +                                                                                                    +
          I                                                                                                   mI
          I                                                                                                   mI
F         I                                                                                                   mI
R   30000 +                                                                                                   m+
E         I                                                                                                   mI
Q         I                                                                                                   mI
U         I                                                                                                   mI
E   20000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    10000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          Id                                                                                                  mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  dddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 2500 Cases.
```

Table DD18

Logistic regression observed and predicted frequencies with the constant model in the 2007 ELA dataset with the 4th grade state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	0	2217	.0
		met proficiency status	0	36824	100.0
	Overall Percentage				94.3

a. The cut value is .500

Table DD19

Logistic regression observed and predicted frequencies with the predicted model for the Michigan 2007 4th grade dataset with the ELA state accountability assessment result as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	2217	0	100.0
		met proficiency status	0	36824	100.0
	Overall Percentage				100.0

a. The cut value is .500

Table DD20

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 4th Grade Michigan ELA assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	7506.241	1	.000
	Block	7506.241	1	.000
	Model	17024.389	10	.000

Table DD21

Logistic regression model summary assessing goodness of fit with the Michigan 2007 4th grade ELA assessment as the dependent variable

Step		dimension1	
	-2 Log likelihood	Cox & Snell R Square	NagelkerkeR Square
1	.003[a]	.353	1.000

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table DD22

Logistic Regression Analysis for the 2007 Michigan 4th grade data with the ELA state accountability assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	132.587	302.382	.192	1	.661	3.819E57	.000	
	EDPER	317.580	1036.575	.094	1	.759	8.382E137	.000	
	NATAMPER	-5943.720	25212.808	.056	1	.814	.000	.000	
	ASIANPER	-5268.270	24630.866	.046	1	.831	.000	.000	
	BLACKPER	-5561.105	24609.922	.051	1	.821	.000	.000	
	HISPANIC	-5918.715	22979.283	.066	1	.797	.000	.000	
	WHITEPER	-5184.593	24457.079	.045	1	.832	.000	.000	
	ELLPER	-1779.894	10174.864	.031	1	.861	.000	.000	
	SPECIALEDPER	288.952	899.547	.103	1	.748	3.093E125	.000	
	NAEPREADPROFbasic	263.262	793.045	.110	1	.740	2.155E114	.000	
	Constant	4729.934	24281.791	.038	1	.846	.		

Figure DD4. Logistic Regression for the Michigan 2007 4th grade dataset with the ELA state accountability assessment as the dependent variable: observed and predicted probabilities

```
              Step number: 1

              Observed Groups and Predicted Probabilities

   40000 +                                                                                                    +
         I                                                                                                   mI
         I                                                                                                   mI
F        I                                                                                                   mI
R  30000 +                                                                                                   m+
E        I                                                                                                   mI
Q        I                                                                                                   mI
U        I                                                                                                   mI
E  20000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                                                                                   mI
   10000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                   mI
         Id                                                                                                  mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 2500 Cases.
```

Table DD23

Logistic regression observed and predicted frequencies with the constant model with the Michigan 2007 8th grade state accountability assessment mathematics results as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	6087	.0
		met proficiency status	0	44646	100.0
	Overall Percentage				88.0

a. The cut value is .500

Table DD24

Logistic regression observed and predicted frequencies with the predicted model in Michigan 2007 8th grade mathematics state accountability assessment results as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	6087	0	100.0
		met proficiency status	0	44646	100.0
	Overall Percentage				100.0

a. The cut value is .500

Table DD25

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 8th grade Michigan mathematics assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	37226.585	9	.000
	Block	37226.585	9	.000
	Model	37226.585	9	.000

Table DD26

Logistic regression model summary assessing goodness of fit with the 2007 Michigan 8th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.001[a]	.520	1.000

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table DD27

Logistic regression analysis for the 2007 Michigan 8th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-29.505	148.123	.040	1	.842	.000	.000	1.858E113
	EDPER	-154.073	1271.831	.015	1	.904	.000	.000	.
	NATAMPER	179.821	8269.339	.000	1	.983	1.246E78	.000	.
	ASIANPER	-110.351	3378.906	.001	1	.974	.000	.000	.
	BLACKPER	9.882	1704.151	.000	1	.995	19579.130	.000	.
	HISPANICPER	19.114	2122.054	.000	1	.993	2.001E8	.000	.
	WHITEPER	66.929	2628.911	.001	1	.980	1.166E29	.000	.
	ELLPER	120.584	1446.012	.007	1	.934	2.338E52	.000	.
	SPECIALEDPER	46.875	3430.351	.000	1	.989	2.279E20	.000	.
	Constant	114.057	3124.808	.001	1	.971	3.423E49		

Figure DD5. Logistic regression for the Michigan 2007 8th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
            Step number: 1

            Observed Groups and Predicted Probabilities

   40000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R  30000 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                    I
U        I                                                                                                   mI
E  20000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                                                                                   mI
   10000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                   mI
         Id                                                                                                  mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 5000 Cases.
```

Table DD28

Logistic regression observed and predicted frequencies with the constant model in the Michigan 2007 8th grade dataset with the ELA state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	0	5294	.0
		met proficiency status	0	45439	100.0
	Overall Percentage				89.6

a. The cut value is .500

Table DD29

Logistic regression observed and predicted frequencies with the predicted model in the Michigan 8th grade dataset with the ELA state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEELAprofstatus	did not meet proficiency status	5069	225	95.7
		met proficiency status	0	45439	100.0
	Overall Percentage				99.6

a. The cut value is .500

Table DD30

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 8th Grade Michigan ELA assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	415.649	1	.000
	Block	415.649	1	.000
	Model	29517.937	10	.000

Table DD31

Logistic regression model summary assessing goodness of fit with the Michigan 8th grade ELA assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	4426.254[a]	.441	.904

a. Estimation terminated at iteration number 14 because parameter estimates changed by less than .001.

Table DD32

Logistic Regression Analysis for the 2007 Michigan 8th grade data with the ELA assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	2.870	.211	184.667	1	.000	17.628	11.654	26.665
	EDPER	-8.989	.576	243.497	1	.000	.000	.000	.000
	NATAMPER	478.465	36.393	172.845	1	.000	6.233E207	6.555E176	5.926E238
	ASIANPER	74.790	4.030	344.472	1	.000	3.027E32	1.124E29	8.147E35
	BLACKPER	33.730	1.299	674.005	1	.000	4.454E14	3.490E13	5.684E15
	HISPANICPER	30.202	1.218	615.054	1	.000	1.308E13	1.202E12	1.423E14
	WHITEPER	64.567	2.280	802.048	1	.000	1.099E28	1.260E26	9.588E29
	ELLPER	-41.385	2.036	413.144	1	.000	.000	.000	.000
	SPECIALEDPER	-96.048	4.897	384.649	1	.000	.000	.000	.000
	NAEPREADPROFbasic	-3.355	.183	335.775	1	.000	.035	.024	.050
	Constant	-33.732	1.572	460.650	1	.000	.000		

Figure DD6. Logistic regression for the Michigan 2007 8th grade dataset with the ELA state accountability assessment as the dependent variable: observed and predicted probabilities

```
            Step number: 1

            Observed Groups and Predicted Probabilities

    40000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R   30000 +                                                                                                    +
E         I                                                                                                    I
Q         I                                                                                                    I
U         I                                                                                                   mI
E   20000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    10000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          Id                                                                                                  mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 5000 Cases.
```

APPENDIX EE

North Carolina Tables and Figures for Research Question Two, Pearson Correlation.

Table EE1

Descriptive statistics for the 2005 North Carolina 4th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.9346	.24718	97486
NAEP math met proficiency at the basic level	.9858	.11822	97486
NAEP math met proficiency at the NAEP proficient level	.2922	.45477	97486
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	8123.00	972.097	97486
EDPER	.4944	.25485	96690
NATAMPER	.0204	.10162	97486
ASIANPER	.0287	.02925	97486
BLACKPER	.3327	.25075	97486
HISPANICPER	.1044	.10028	97486
WHITEPER	.5138	.29281	97486
ELLPER	.1060	.11666	97486
SPECIALEDPER	.1390	.06191	96230

Table EE2

Pearson correlation matrix for the 2005 North Carolina 4th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.453**	.170**	.077**	-.374**	-.060**
NAEP BASIC Math	.453**	1	.077**	.052**	-.223**	-.162**
NAEP PROF Math	.170**	.077**	1	.097**	-.618**	-.108**
	ASIANPER	BLACKPER	HISPANICPER	WHITEPER	ELLPER	SPECIALEDPER
STATE Math	.084**	-.299**	-.151**	.321**	-.114**	-.147**
NAEP BASIC level	.079**	-.170**	.082**	.165**	.067**	.037**
NAEP PROF level	.091**	-.363**	-.234**	.420**	-.201**	-.074**

**. Correlation is significant at the 0.01 level (2-tailed)

Table EE3

Descriptive statistics for the 2005 North Carolina 4th grade dataset in reading

	Mean	Std. Deviation	N
STATEREADprofstatus	.7484	.43396	97486
NAEPREADPROFbasic	.7698	.42099	97486
NAEPREADPROFprof	.1164	.32069	97486
TOTAL REVENUE PER STUDENT	8123.00	972.097	97486
EDPER	.4944	.25485	96690
NATAMPER	.0204	.10162	97486
ASIANPER	.0287	.02925	97486
BLACKPER	.3327	.25075	97486
HISPANICPER	.1044	.10028	97486
WHITEPER	.5138	.29281	97486
ELLPER	.1060	.11666	97486
SPECIALEDPER	.1390	.06191	96230

Table EE4

Pearson correlation matrix for the 2005 North Carolina 4th grade listed variables in reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	**NAEP BASIC level Reading**	**NAEP PROF level Reading**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATE Reading	1	.568**	.210**	.034**	-.596**	-.245**
NAEP BASIC Reading	.568**	1	.198**	.080**	-.573**	-.253**
NAEP PROF Reading	.210**	.198**	1	-.049**	-.481**	-.057**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Reading	.031**	-.421**	-.221**	.518**	-.188**	-.074**
NAEP BASIC level	.116**	-.397**	-.127**	.460**	-.052**	-.026**
NAEP PROF level	.161**	-.280**	-.217**	.317**	-.153**	-.125**

**. Correlation is significant at the 0.01 level (2-tailed)

Table EE5

Descriptive statistics for the 2005 North Carolina 8th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.7139	.45192	91731
NAEP math met proficiency at the basic level	.9357	.24526	91731
NAEP math met proficiency at the NAEP proficient level	.1252	.33091	91731
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	8129.13	919.078	91731
EDPER	.4809	.20551	91081
NATAMPER	.0097	.04644	91731
ASIANPER	.0221	.02773	91731
BLACKPER	.3399	.23201	91731
HISPANICPER	.0726	.06048	91731
WHITEPER	.5557	.25538	91731
ELLPER	.0654	.06162	88360
SPECIALEDPER	.1630	.08715	83538

Table EE6

Pearson correlation matrix for the 2005 North Carolina 8th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.414**	.239**	.093**	-.661**	-.137**
NAEP BASIC Math	414**	1	.099**	.092**	-.388**	.027**
NAEP PROF Math	.239**	.099**	1	.185**	-.456**	-.056**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATEAYP	.025**	-.663**	-.281**	.691**	-.270**	-.168**
NAEP BASIC level	-.078**	-.428**	-.112**	.419**	-.093**	-.238**
NAEP PROF level	.375**	-.252**	-.148**	.233**	-.073**	-.200**

**. Correlation is significant at the 0.01 level (2-tailed)

Table EE7

Descriptive statistics for the 2005 North Carolina 8th grade dataset in reading

	Mean	Std. Deviation	N
STATEREADprofstatus	.8957	.30560	91731
NAEPREADPROFbasic	.8746	.33121	91676
NAEPREADPROFprof	.0389	.19343	91676
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	8129.13	919.078	91731
EDPER	.4809	.20551	91081
NATAMPER	.0097	.04644	91731
ASIANPER	.0221	.02773	91731
BLACKPER	.3399	.23201	91731
HISPANICPER	.0726	.06048	91731
WHITEPER	.5557	.25538	91731
ELLPER	.0654	.06162	88360
SPECIALEDPER	.1630	.08715	83538

Table EE8

Pearson correlation matrix for the 2005 North Carolina 8th grade listed variables in reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	**NAEP BASIC level Reading**	**NAEP PROF level Reading**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATE Reading	1	.439**	.068**	.008*	-.489**	-.158**
NAEP BASIC Reading	.439**	1	.076**	.019**	-.420**	.034**
NAEP PROF Reading	.068**	.076**	1	.056**	-.316**	-.032**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Reading	-.072**	-.470**	-.324**	.540**	-.378**	-.080**
NAEP BASIC level	-.029**	-.464**	-.087**	.439**	-.035**	-.166**
NAEP PROF level	.421**	-.125**	-.161**	.112**	-.125**	-.137**

**. Correlation is significant at the 0.01 level (2-tailed)

Table EE9

Descriptive statistics for the 2007 North Carolina 4th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.2494	.43267	99885
NAEP math met proficiency at the basic level	.9816	.13451	99885
NAEP math met proficiency at the NAEP proficient level	.3068	.46118	99885
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9181.78	1102.643	99885
EDPER	.4722	.23339	99885
NATAMPER	.0477	.16344	99885
ASIANPER	.0288	.02901	99885
BLACKPER	.2929	.21851	99885
HISPANICPER	.1317	.11285	99885
WHITEPER	.4570	.28834	99885
ELLPER	.1196	.11910	98249
SPECIALEDPER	.1350	.09629	95403

Table EE10

Pearson correlation matrix for the 2007 North Carolina 4th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.079**	.596**	.186**	-.635**	-.098**
NAEP BASIC Math	.079**	1	.091**	.016**	-.248**	-.124**
NAEP PROF Math	.596**	.091**	1	.127**	-.679**	-.179**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.156**	-.485**	-.300**	.536**	-.190**	-.163**
NAEP BASIC level	.092**	-.193**	.058**	.177**	.044**	-.018**
NAEP PROF level	.190**	-.441**	-.350**	.563**	-.259**	-.112**

**. Correlation is significant at the 0.01 level (2-tailed)

Table EE11

Descriptive statistics for the 2007 North Carolina 4th grade dataset in reading

	Mean	Std. Deviation	N
STATEREADprofstatus	1.0000	.00000	99885
NAEPREADPROFbasic	.7662	.42326	99885
NAEPREADPROFprof	.1436	.35073	99885
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9181.78	1102.643	99885
EDPER	.4722	.23339	99885
NATAMPER	.0477	.16344	99885
ASIANPER	.0288	.02901	99885
BLACKPER	.2929	.21851	99885
HISPANICPER	.1317	.11285	99885
WHITEPER	.4570	.28834	99885
ELLPER	.1196	.11910	98249
SPECIALEDPER	.1350	.09629	95403

Table EE12

Pearson correlation matrix for the 2007 North Carolina 4th grade listed variables in reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Reading	**NAEP BASIC level Reading**	**NAEP PROF level Reading**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
NAEP BASIC Reading			.226**	.141**	-.548**	-.394**
NAEP PROF Reading		.226**	.[a]	.084**	-.599**	-.112**

	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
NAEP BASIC level	.103**	-.229**	-.238**	.475**	-.242**	-.087**
NAEP PROF level	.275**	-.392**	-.274**	.448**	-.209**	-.134**

**. Correlation is significant at the 0.01 level (2-tailed)

Table EE13

Descriptive statistics for the 2007 North Carolina 8th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.2643	.44097	97123
NAEPMATHPROFbasic	.9223	.26763	96435
NAEPMATHPROFprof	.2000	.39999	97123
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9424.60	1312.122	97123
EDPER	.4672	.20341	97123
NATAMPER	.0390	.13932	97123
ASIANPER	.0225	.02393	97123
BLACKPER	.3101	.23010	97123
HISPANICPER	.0917	.07462	97123
WHITEPER	.5095	.26807	97123
ELLPER	.0660	.06543	94439
SPECIALEDPER	.1530	.07565	88663

Table EE14

Pearson correlation matrix for the 2007 North Carolina 8th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.175**	.553**	.230**	-.589**	-.155**
NAEP BASIC Math	.175**	1	.146**	.014**	-.427**	-.149**
NAEP PROF Math	.553**	.146**	1	.349**	-.571**	-.130**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.155**	-.387**	-.207**	.456**	-.169**	-.168**
NAEP BASIC level	-.051**	-.335**	-.335**	.454**	-.329**	-.131**
NAEP PROF level	.277**	-.296**	-.227**	.354**	-.165**	-.079**

**. Correlation is significant at the 0.01 level (2-tailed)

Table EE15

Descriptive statistics for the 2007 North Carolina 8th grade dataset in reading

	Mean	Std. Deviation	N
STATEREADprofstatus	.9961	.06210	97123
NAEPREADPROFbasic	.8317	.37417	96435
NAEPREADPROFprof	.0603	.23799	96435
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9424.60	1312.122	97123
EDPER	.4672	.20341	97123
NATAMPER	.0390	.13932	97123
ASIANPER	.0225	.02393	97123
BLACKPER	.3101	.23010	97123
HISPANICPER	.0917	.07462	97123
WHITEPER	.5095	.26807	97123
ELLPER	.0660	.06543	94439
SPECIALEDPER	.1530	.07565	88663

Table EE16

Pearson correlation matrix for the 2007 North Carolina 8th grade listed variables in reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Reading	NAEP BASIC level Reading	NAEP PROF level Reading	TOTAL REVENUE	EDPER	NATAMPER
STATE Reading	1	.139**	.016**	.011**	-.036**	.000
NAEP BASIC Reading	.139**	1	.114**	.104**	-.552**	-.291**
NAEP PROF Reading	016**	.114**	1	.308**	-.446**	-.067**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Reading	.059**	-.077**	.050**	.046**	.048**	-.196**
NAEP BASIC level	.107**	-.345**	-.061**	.442**	-.139**	-.097**
NAEP PROF level	.227**	-.222**	-.125**	.241**	-.109**	-.046**

**. Correlation is significant at the 0.01 level (2-tailed)

APPENDIX FF

North Carolina Tables and Figures for Research Question Two, Logistic Regression.

Table FF1

Logistic regression observed and predicted frequencies with the constant model in the North Carolina 2005 4th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency .00	met Proficiency 1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	6373	.0
		met proficiency status	0	89061	100.0
	Overall Percentage				93.3

a. The cut value is .500

Table FF2

Logistic regression observed and predicted frequencies with the predicted model in the North Carolina 4th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency .00	met Proficiency 1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	3121	3252	49.0
		met proficiency status	485	88576	99.5
	Overall Percentage				96.1

a. The cut value is .500

Table FF3

Logistic regression omnibus test of model coefficients assessing goodness of fit with the North Carolina 2005 4th grade mathematics assessment as the dependent variable

		Dimension		
		Chi-square	df	Sig.
Step 1	Step	98.725	1	.000
	Block	98.725	1	.000
	Model	24834.243	11	.000

Table FF4

Logistic regression model summary assessing goodness of fit with the North Carolina 4th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	21971.694[a]	.229	.591

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table FF5

Logistic regression results for the 2005 North Carolina 4th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
1	DUMMYREV	1.007	.052	379.022	1	.000	2.737	2.473	3.029
	EDPER	-3.785	.175	467.699	1	.000	.023	.016	.032
	NATAMPER	-4.610	.446	106.749	1	.000	.010	.004	.024
	ASIANPER	10.461	.965	117.612	1	.000	34909.835	5271.266	231196.166
	BLACKPER	-8.472	.320	701.331	1	.000	.000	.000	.000
	HISPANICPER	-19.450	.475	1676.044	1	.000	.000	.000	.000
	DUMMYWHITE	-.570	.045	161.498	1	.000	.565	.518	.617
	ELLPER	9.003	.332	733.374	1	.000	8130.512	4237.652	15599.493
	SPECIALEDPER	-19.664	.374	2761.211	1	.000	.000	.000	.000
	NAEPMATHPROFbasic	23.516	1031.872	.001	1	.982	1.633E10	.000	.
	NAEPMATHPROFprof	14.694	209.838	.005	1	.944	2408435.387	.000	9.904E184
	Constant	-8.573	1031.872	.000	1	.993	.000		

*Figure FF*1. Logistic Regression for the North Carolina 2005 4th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
            Step number: 1

            Observed Groups and Predicted Probabilities

    80000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R   60000 +                                                                                                   m+
E         I                                                                                                   mI
Q         I                                                                                                   mI
U         I                                                                                                   mI
E   40000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    20000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          I                                                                                                mmmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 5000 Cases.
```

Table FF6

Logistic regression observed and predicted frequencies with the constant model in the North Carolina 2005 4th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	0	24107	.0
		met proficiency status	0	71327	100.0
	Overall Percentage				74.7

a. The cut value is .500

Table FF7

Logistic regression observed and predicted frequencies with the predicted model in the North Carolina 2005 4th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	16112	7995	66.8
		met proficiency status	6056	65271	91.5
	Overall Percentage				85.3

a. The cut value is .500

Table FF8

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 4th Grade North Carolina reading assessment as the dependent variable

		Dimension Chi-square	df	Sig.
Step 1	Step	301.111	1	.000
	Block	301.111	1	.000
	Model	47935.193	11	.000

Table FF9

Logistic regression model summary assessing goodness of fit with the 2005 North Carolina 4th grade readings assessment as the dependent variable

Step	dimension1 -2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	59939.158[a]	.395	.583

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table FF10

Logistic regression analysis for the 2005 North Carolina 4th grade data with the reading assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-.468	.022	469.578	1	.000	.627	.601	.654
	EDPER	-4.763	.088	2938.473	1	.000	.009	.007	.010
	NATAMPER	-19.529	.677	833.316	1	.000	.000	.000	.000
	ASIANPER	-4.434	.377	138.635	1	.000	.012	.006	.025
	BLACKPER	-2.871	.142	410.993	1	.000	.057	.043	.075
	HISPANICPER	-1.857	.220	71.481	1	.000	.156	.101	.240
	DUMMYWHITE	-.179	.025	53.285	1	.000	.836	.797	.877
	ELLPER	-.990	.163	36.844	1	.000	.372	.270	.512
	SPECIALEDPER	-2.513	.167	227.554	1	.000	.081	.058	.112
	NAEPREADPROFbasic	1.335	.024	3101.807	1	.000	3.799	3.624	3.981
	NAEPREADPROFprof	16.899	363.693	.002	1	.963	2.183E7	.000	.
	Constant	6.594	.207	1016.765	1	.000	730.865		

Figure FF2. Logistic Regression for the North Carolina 2005 4th grade dataset with the reading state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

    16000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                   mI
F         I                                                                                                   mI
R   12000 +                                                                                                   m+
E         I                                                                                                   mI
Q         I                                                                                                   mI
U         I                                                                                                   mI
E    8000 +                                                                                                   m+
N         I                                                                                                m mmI
C         I                                                                                               mmmmmI
Y         I                                                                                              mmmmmmI
     4000 +                                                                                            m mmmmmm+
          I                                                                                          m  m mmmmmmI
          Id                                     m                                  m       m m     m  mmmmmmmmI
          Id       d  dd      d d   d d        d ddd md  m  m d m m m   mddm    mm  mm  mmm d m dm m dmmmmmdmmmI
Predicted  ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:    0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:   ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met proficiency
The Cut Value is .50
Symbols: d - did not meet proficiency
         m - met proficiency
Each Symbol Represents 1000 Cases.
```

Table FF11

Logistic regression observed and predicted frequencies with the constant model in the North Carolina 2005 8th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	24041	.0
		met proficiency status	0	56473	100.0
	Overall Percentage				70.1

a. The cut value is .500

Table FF12

Logistic regression observed and predicted frequencies with the predicted model in the North Carolina 8th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	18264	5777	76.0
		met proficiency status	2982	53491	94.7
	Overall Percentage				89.1

a. The cut value is .500

Table FF13

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 8th grade mathematics assessment as the dependent variable

		Dimension Chi-square	df	Sig.
Step 1	Step	128.948	1	.000
	Block	128.948	1	.000
	Model	58290.628	11	.000

Table FF14

Logistic regression model summary assessing goodness of fit with the North Carolina 8th grade mathematics assessment as the dependent variable

Step	dimension1 -2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	39883.042[a]	.515	.731

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table FF15

Logistic regression results for the 2005 North Carolina 8th grade data with the mathematics assessment as the dependent variable

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1	DUMMYREV	.797	.031	676.152	1	.000	2.220	2.090	2.357
	EDPER	-6.638	.147	2042.765	1	.000	.001	.001	.002
	NATAMPER	-15.537	.387	1607.696	1	.000	.000	.000	.000
	ASIANPER	44.504	.824	2920.013	1	.000	2.127E19	4.235E18	1.069E20
	BLACKPER	-11.101	.198	3141.020	1	.000	.000	.000	.000
	HISPANICPER	-6.538	.331	390.921	1	.000	.001	.001	.003
	DUMMYWHITE	-.903	.041	475.658	1	.000	.406	.374	.440
	ELLPER	-10.689	.273	1534.364	1	.000	.000	.000	.000
	SPECIALEDPER	-2.480	.138	324.600	1	.000	.084	.064	.110
	NAEPMATHPROFbasic	19.292	515.240	.001	1	.970	2.390E8	.000	.
	NAEPMATHPROFprof	15.960	354.704	.002	1	.964	8534996.561	.000	.
	Constant	-6.200	515.240	.000	1	.990	.002		

Figure FF3. Logistic regression for the North Carolina 2005 8th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

Step number: 1

Observed Groups and Predicted Probabilities

```
     32000 +                                                                                                    +
           I                                                                                                    I
           I                                                                                                    I
F          I                                                                                                    I
R    24000 +                                                                                                    +
E          I                                                                                                    I
Q          I                                                                                                   mI
U          I                                                                                                   mI
E    16000 +                                                                                                   m+
N          I                                                                                                   mI
C          I                                                                                                   mI
Y          I                                                                                                   mI
      8000 +                                                                                                  mm+
           I                                                                                                  mmI
           Id                                                                                                 mmI
           Id d  d             d      m         d                      m dd      m    m            mm   mmmmmmmmI
Predicted   ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:     0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:    ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm
```

Predicted Probability is of Membership for met proficiency
The Cut Value is .50
Symbols: d - did not meet proficiency
         m - met proficiency
Each Symbol Represents 2000 Cases.

Table FF16

Logistic regression observed and predicted frequencies with the constant model in the North Carolina 2005 8th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	0	7852	.0
		met proficiency status	0	72607	100.0
	Overall Percentage				90.2

a. The cut value is .500

Table FF17

Logistic regression observed and predicted frequencies with the predicted model in the North Carolina 2005 8th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	7339	513	93.5
		met proficiency status	56	72551	99.9
	Overall Percentage				99.3

a. The cut value is .500

Table FF18

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 8th Grade North Carolina reading assessment as the dependent variable

		dimension Chi-square	df	Sig.
Step 1	Step	3945.074	1	.000
	Block	3945.074	1	.000
	Model	44091.321	11	.000

Table FF19

Logistic regression model summary assessing goodness of fit with the North Carolina 8th grade reading assessment as the dependent variable

Step	dimension1 -2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	7363.068[a]	.422	.893

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table FF20

Logistic regression results for the 2005 North Carolina 8th grade data with the reading assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-13.292	.552	580.637	1	.000	.000	.000	.000
	EDPER	-79.681	3.463	529.402	1	.000	.000	.000	.000
	NATAMPER	-74.223	2.638	791.568	1	.000	.000	.000	.000
	ASIANPER	116.143	4.054	820.651	1	.000	2.755E50	9.754E46	7.783E53
	BLACKPER	-31.216	1.323	556.709	1	.000	.000	.000	.000
	HISPANICPER	19.003	3.988	22.702	1	.000	1.789E8	72090.158	4.441E11
	DUMMYWHITE	-1.693	.163	107.467	1	.000	.184	.134	.253
	ELLPER	-172.547	8.140	449.345	1	.000	.000	.000	.000
	SPECIALEDPER	5.194	.420	152.787	1	.000	180.183	79.075	410.572
	NAEPREADPROFbasic	-29.894	1162.318	.001	1	.979	.000	.000	.
	NAEPCOMBPROFbasic	47.861	1162.318	.002	1	.967	6.106E20	.000	.
	Constant	109.443	4.403	617.912	1	.000	3.392E47		

Figure FF4. Logistic Regression for the North Carolina 2005 8th grade dataset with the reading state accountability assessment as the dependent variable: observed and predicted probabilities

```
             Step number: 1

             Observed Groups and Predicted Probabilities

    80000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                   mI
F         I                                                                                                   mI
R   60000 +                                                                                                   m+
E         I                                                                                                   mI
Q         I                                                                                                   mI
U         I                                                                                                   mI
E   40000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    20000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          Id                                                                                                  mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 5000 Cases.
```

Table FF21

Logistic regression observed and predicted frequencies with the constant model in the North Carolina 2007 4th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	69288	0	100.0
		met proficiency status	24911	0	.0
	Overall Percentage				73.6

a. The cut value is .500

Table FF22

Logistic regression observed and predicted frequencies with the predicted model in the North Carolina 4th grade 2007 dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	66008	3280	95.3
		met proficiency status	5613	19298	77.5
	Overall Percentage				90.6

a. The cut value is .500

Table FF23

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 4th grade mathematics assessment as the dependent variable

		dimension		
		Chi-square	df	Sig.
Step 1	Step	1631.547	1	.000
	Block	1631.547	1	.000
	Model	66375.654	11	.000

Table FF24

Logistic regression model summary assessing goodness of fit with the North Carolina 4th grade 2007 mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	42454.525[a]	.506	.738

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table FF25

Logistic regression analysis for the 2007 North Carolina 4th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1	DUMMYREV	1.314	.024	2901.049	1	.000	3.721	3.547	3.903
	EDPER	-9.461	.120	6220.284	1	.000	.000	.000	.000
	NATAMPER	15.782	.858	338.117	1	.000	7148837.835	1329370.490	3.844E7
	ASIANPER	36.833	1.190	957.663	1	.000	9.913E15	9.618E14	1.022E17
	BLACKPER	4.540	.937	23.483	1	.000	93.682	14.935	587.631
	HISPANICPER	10.850	.921	138.735	1	.000	51541.625	8473.310	313518.449
	WHITEPER	13.419	.866	240.022	1	.000	672812.255	123200.249	3674313.448
	ELLPER	6.355	.216	862.377	1	.000	575.175	376.362	879.012
	SPECIALEDPER	-5.825	.188	957.691	1	.000	.003	.002	.004
	NAEPMATHPROFbasic	12.064	871.625	.000	1	.989	173541.145	.000	.
	NAEPMATHPROFprof	1.144	.028	1626.099	1	.000	3.138	2.969	3.318
	Constant	-24.457	871.625	.001	1	.978	.000		

Figure FF5. Logistic regression for the North Carolina 2007 4th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
              Step number: 1

              Observed Groups and Predicted Probabilities

    40000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R   30000 +                                                                                                   m+
E         I                                                                                                   mI
Q         I                                                                                                   mI
U         I                                                                                                   mI
E   20000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    10000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          Id                                                                                               mmmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 5000 Cases.
```

Table FF26

Logistic regression observed and predicted frequencies with the constant model in the North Carolina 2007 8th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	63958	0	100.0
		met proficiency status	23902	0	.0
	Overall Percentage				72.8

a. The cut value is .500

Table FF27

Logistic regression observed and predicted frequencies with the predicted model in the North Carolina 8th grade 2007 dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	61021	2937	95.4
		met proficiency status	7693	16209	67.8
	Overall Percentage				87.9

a. The cut value is .500

Table FF28

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 8th grade mathematics assessment as the dependent variable

		dimension Chi-square	df	Sig.
Step 1	Step	1455.859	1	.000
	Block	1455.859	1	.000
	Model	48728.188	11	.000

Table FF29

Logistic regression model summary assessing goodness of fit with the North Carolina 8th 2007 grade mathematics assessment as the dependent variable

Step	dimension1 -2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	54117.858[a]	.426	.617

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table FF30

Logistic Regression Analysis for the 2007 North Carolina 8th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	.230	.017	183.407	1	.000	1.259	1.218	1.302
	EDPER	-11.291	.132	7347.065	1	.000	.000	.000	.000
	NATAMPER	-109.863	4.492	598.229	1	.000	.000	.000	.000
	ASIANPER	-17.278	.953	328.998	1	.000	.000	.000	.000
	BLACKPER	-.403	.783	.264	1	.607	.668	.144	3.104
	HISPANICPER	.102	.958	.011	1	.915	1.108	.169	7.243
	WHITEPER	1.137	.774	2.160	1	.142	3.118	.684	14.207
	ELLPER	6.850	.367	349.246	1	.000	943.624	460.055	1935.478
	SPECIALEDPER	-5.027	.175	826.840	1	.000	.007	.005	.009
	NAEPMATHPROFbasic	15.220	398.396	.001	1	.970	4072983.007	.000	.
	NAEPMATHPROFprof	1.156	.030	1437.948	1	.000	3.176	2.992	3.372
	Constant	-12.126	398.397	.001	1	.976	.000		

Figure FF6. Logistic Regression for the North Carolina 2007 8th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
               Step number: 1

               Observed Groups and Predicted Probabilities

    32000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R   24000 +                                                                                                    +
E         I                                                                                                    I
Q         Id                                                                                                   I
U         Id                                                                                                   I
E   16000 +d                                                                                                   +
N         Id                                                                                                   I
C         Id                                                                                                   I
Y         Id                                                                                                   I
     8000 +d                                                                                                   +
          Idd                                                                                                  I
          Idd  d d                                                                                             I
          Iddd d dd  d d d   d d  d  dd   d       d    d      m               m  d   m          m   m dm      mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

           Predicted Probability is of Membership for met proficiency
           The Cut Value is .50
           Symbols: d - did not meet proficiency
                    m - met proficiency
           Each Symbol Represents 2000 Cases.
```

Table FF31

Logistic regression observed and predicted frequencies with the constant model in the North Carolina 2007 8th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	0	376	.0
		met proficiency status	0	87484	100.0
	Overall Percentage				99.6

a. The cut value is .500

Table FF32

Logistic regression observed and predicted frequencies with the predicted model in the North Carolina 2007 8th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	376	0	100.0
		met proficiency status	0	87484	100.0
	Overall Percentage				100

a. The cut value is .500

Table FF33

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 8th Grade North Carolina reading assessment as the dependent variable

		dimension		
		Chi-square	df	Sig.
Step 1	Step	.000	1	.993
	Block	.000	1	.993
	Model	4851.728	11	.000

Table FF34

Logistic regression model summary assessing goodness of fit with the North Carolina 8th grade 2007 reading assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.001[a]	.054	1.00

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table FF35

Logistic Regression Analysis for the 2007 North Carolina 8th grade data with the reading assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1	DUMMYREV	14.215	145.286	.010	1	.922	1491590.426	.000	6.944E129
	EDPER	110.381	1928.251	.003	1	.954	8.668E47	.000	.
	NATAMPER	372.830	4271.313	.008	1	.930	8.277E161	.000	.
	ASIANPER	1929.544	9769.083	.039	1	.843	.	.000	.
	BLACKPER	357.490	4314.655	.007	1	.934	1.803E155	.000	.
	HISPANICPER	691.160	4262.685	.026	1	.871	1.468E300	.000	.
	WHITEPER	431.616	5014.117	.007	1	.931	2.808E187	.000	.
	ELLPER	-186.868	2074.244	.008	1	.928	.000	.000	.
	SPECIALEDPER	-27.521	319.160	.007	1	.931	.000	.000	5.209E259
	NAEPREADPROFbasic	35.675	586.116	.004	1	.951	3.114E15	.000	.
	NAEPREADPROFprof	-45.204	579.822	.006	1	.938	.000	.000	.
	Constant	-477.411	5087.526	.009	1	.925	.000		

Figure FF7. Logistic Regression for the North Carolina 2007 8th grade dataset with the reading state accountability assessment as the dependent variable: observed and predicted probabilities

```
              Step number: 1

              Observed Groups and Predicted Probabilities

   160000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R  120000 +                                                                                                    +
E         I                                                                                                    I
Q         I                                                                                                    I
U         I                                                                                                   mI
E   80000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    40000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          I                                                                                                   mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  dddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 10000 Cases.
```

APPENDIX GG

Texas Tables and Figures for Research Question Two, Pearson Correlation.

Table GG1

Descriptive statistics for the 2005 Texas 4th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.9732	.16148	199673
NAEP math met proficiency at the basic level	.9825	.13098	199774
NAEP math met proficiency at the NAEP proficient level	.2460	.43070	199774
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	8774.39	1201.540	198443
EDPER	.5630	.31130	199774
NATAMPER	.0025	.00407	199774
ASIANPER	.0314	.05265	199774
BLACKPER	.1633	.20879	199774
HISPANICPER	.5288	.32308	199774
WHITEPER	.2741	.29269	199774
ELLPER	.2939	.25046	192878
SPECIALEDPER	.1094	.08865	186924

Table GG2

Pearson correlation matrix for the 2005 Texas 4th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.179**	.095**	.006*	-.159**	.100**
NAEP BASIC Math	.179**	1	.076**	.027**	-.119**	.080**
NAEP PROF Math	.095**	.076**	1	.060**	-.621**	.184**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.063**	-.277**	.047**	.133**	-.020**	-.037**
NAEP BASIC level	.072**	-.350**	.101**	.124**	.020**	-.031**
NAEP PROF level	.483**	-.221**	-.528**	.651**	-.444**	.021**

**. Correlation is significant at the 0.01 level (2-tailed)

Table GG3

Descriptive statistics for the 2005 Texas 4th grade dataset in reading

	Mean	Std. Deviation	N
STATEREADprofstatus	.9609	.19378	199673
NAEPREADPROFbasic	.6781	.46721	199774
NAEPREADPROFprof	.1154	.31947	199774
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	8774.39	1201.540	198443
EDPER	.5630	.31130	199774
NATAMPER	.0025	.00407	199774
ASIANPER	.0314	.05265	199774
BLACKPER	.1633	.20879	199774
HISPANICPER	.5288	.32308	199774
WHITEPER	.2741	.29269	199774
ELLPER	.2939	.25046	192878
SPECIALEDPER	.1094	.08865	186924

Table GG4

Pearson correlation matrix for the 2005 Texas 4th grade listed variables in reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Reading	**NAEP BASIC level Reading**	**NAEP PROF level Reading**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATE Reading	1	.293**	.073**	-.030**	-.124**	.115**
NAEP BASIC Reading	.293**	1	.249**	-.073**	-.453**	.208**
NAEP PROF Reading	.073**	.249**	1	.075**	-.502**	.098**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Reading	.093**	-.322**	.046**	.161**	-.026**	-.015**
NAEP BASIC level	.308**	-.266**	-.345**	.512**	-.395**	-.040**
NAEP PROF level	.388**	-.177**	-.428**	.527**	-.334**	-.073**

**. Correlation is significant at the 0.01 level (2-tailed)

Table GG5

Descriptive statistics for the 2005 Texas 8th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.8071	.39456	201950
NAEP math met proficiency at the basic level	.8233	.38144	202154
NAEP math met proficiency at the NAEP proficient level	.1203	.32532	202154
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	8553.17	1238.803	203550
EDPER	.4904	.26894	203550
NATAMPER	.0026	.00319	203550
ASIANPER	.0318	.04844	203550
BLACKPER	.1847	.20962	203550
HISPANICPER	.4376	.29926	203550
WHITEPER	.3433	.28997	203550
ELLPER	.1179	.14324	195917
SPECIALEDPER	.1565	.09852	189073

Table GG6

Pearson correlation matrix for the 2005 Texas 8th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	.630**	.180**	-.026**	-.530**	.120**
NAEP BASIC Math	.630**	1	.171**	-.047**	-.450**	.193**
NAEP PROF Math	.180**	.171**	1	.358**	-.471**	.095**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.235**	-.395**	-.216**	.471**	-.345**	-.337**
NAEP BASIC level	.211**	-.422**	-.206**	.477**	-.255**	-.196**
NAEP PROF level	.497**	-.192**	-.352**	.416**	-.198**	-.188**

**. Correlation is significant at the 0.01 level (2-tailed)

Table GG7

Descriptive statistics for the 2005 Texas 8th grade dataset in reading

	Mean	Std. Deviation	N
STATEREADprofstatus	.9889	.10471	201950
NAEPREADPROFbasic	.8083	.39366	202154
NAEPREADPROFprof	.0661	.24838	202154
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.) [2004-05]	8553.17	1238.803	203550
EDPER	.4904	.26894	203550
NATAMPER	.0026	.00319	203550
ASIANPER	.0318	.04844	203550
BLACKPER	.1847	.20962	203550
HISPANICPER	.4376	.29926	203550
WHITEPER	.3433	.28997	203550
ELLPER	.1179	.14324	195917
SPECIALEDPER	.1565	.09852	189073

Table GG8

Pearson correlation matrix for the 2005 Texas 8th grade listed variables in reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE ELA	**NAEP BASIC level Reading**	**NAEP PROF level Reading**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATE Reading	1	.218**	.028**	-.115**	-.101**	.058**
NAEP BASIC Reading	.218**	1	.130**	-.033**	-.453**	.191**
NAEP PROF Reading	.028**	.130**	1	.189**	-.386**	.127**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Reading	.062**	.049**	-.159**	.119**	-.197**	-.060**
NAEP BASIC level	.225**	-.298**	-.307**	.490**	-.392**	-.239**
NAEP PROF level	.340**	-.159**	-.293**	.358**	-.159**	-.167**

**. Correlation is significant at the 0.01 level (2-tailed)

Table GG9

Descriptive statistics for the 2007 Texas 4th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.9943	.07507	172740
NAEP math met proficiency at the basic level	.9898	.10030	172789
NAEP math met proficiency at the NAEP proficient level	.2655	.44161	172789
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	10153.03	1701.802	173176
EDPER	.5639	.27784	173176
NATAMPER	.0019	.00285	173176
ASIANPER	.0369	.05567	173176
BLACKPER	.1628	.19216	173176
HISPANICPER	.5537	.30911	173176
WHITEPER	.2447	.27650	173176
ELLPER	.3300	.25868	165976
SPECIALEDPER	.1050	.09972	166949

Table GG10

Pearson correlation matrix for the 2007 Texas 4th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	NAEP BASIC level Math	NAEP PROF level Math	TOTAL REVENUE	EDPER	NATAMPER
STATE Math	1	-.008**	.046**	-.069**	-.109**	.052**
NAEP BASIC Math	-.008**	1	.061**	.001	-.001	.069**
NAEP PROF Math	.046**	.061**	1	.005*	-.608**	.220**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.047**	-.048**	-.035**	.062**	-.015**	.043**
NAEP BASIC level	.054**	-.272**	.079**	.090**	.014**	.026**
NAEP PROF level	.506**	-.214**	-.545**	.654**	-.441**	-.037**

**. Correlation is significant at the 0.01 level (2-tailed)

Table GG11

Descriptive statistics for the 2007 Texas 4th grade dataset in reading

	Mean	Std. Deviation	N
STATEREADprofstatus	.9614	.19270	172740
NAEPREADPROFbasic	.6171	.48609	172789
NAEPREADPROFprof	.1232	.32862	172789
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	10153.03	1701.802	173176
EDPER	.5639	.27784	173176
NATAMPER	.0019	.00285	173176
ASIANPER	.0369	.05567	173176
BLACKPER	.1628	.19216	173176
HISPANICPER	.5537	.30911	173176
WHITEPER	.2447	.27650	173176
ELLPER	.3300	.25868	165976
SPECIALEDPER	.1050	.09972	166949

Table GG12

Pearson correlation matrix for the 2007 Texas 4th grade listed variables in reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Reading	NAEP BASIC level Reading	NAEP PROF level Reading	TOTAL REVENUE	EDPER	NATAMPER
STATE Reading	1	.169**	.075**	-.061**	-.115**	.126**
NAEP BASIC Reading	.169**	1	.295**	-.122**	-.422**	.324**
NAEP PROF Reading	.075**	.295**	1	.160**	-.524**	.214**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Reading	.053**	-.148**	-.064**	.163**	-.099**	-.025**
NAEP BASIC level	.294**	-.235**	-.398**	.546**	-.472**	-.034**
NAEP PROF level	.433**	-.193**	-.398**	.490**	-.321**	-.039**

**. Correlation is significant at the 0.01 level (2-tailed)

Table GG13

Descriptive statistics for the 2007 Texas 8th grade dataset in mathematics

	Mean	Std. Deviation	N
STATEMATHprofstatus	.9338	.24867	154299
NAEP math met proficiency at the basic level	.9352	.24619	154594
NAEP math met proficiency at the NAEP proficient level	.1735	.37864	154594
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9943.58	2055.118	155218
EDPER	.4965	.24522	153136
NATAMPER	.0026	.00303	155218
ASIANPER	.0356	.04936	155218
BLACKPER	.1766	.19035	155218
HISPANICPER	.4918	.28930	155218
WHITEPER	.2934	.27242	155218
ELLPER	.1411	.16573	150559
SPECIALEDPER	.1443	.09211	152734

Table GG14

Pearson correlation matrix for the 2007 Texas 8th grade listed variables in mathematics (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Math	**NAEP BASIC level Math**	**NAEP PROF level Math**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
STATE Math	1	.109**	.119**	-.078**	-.222**	.102**
NAEP BASIC Math	.109**	1	.121**	-.047**	-.185**	.202**
NAEP PROF Math	119**	.121**	1	.231**	-.537**	.161**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Math	.163**	-.088**	-.210**	.254**	-.169**	-.192**
NAEP BASIC level	.170**	-.345**	-.056**	.267**	-.087**	-.160**
NAEP PROF level	.329**	-.226**	-.440**	.563**	-.297**	-.162**

**. Correlation is significant at the 0.01 level (2-tailed)

Table GG15

Descriptive statistics for the 2007 Texas 8th grade dataset in reading

	Mean	Std. Deviation	N
STATEREADprofstatus	1.0000	.00000	154306
NAEPREADPROFbasic	.8530	.35414	154573
NAEPREADPROFprof	.0946	.29259	154573
TOTAL REVENUE PER STUDENT (DISTRICT-FIN.)	9943.58	2055.118	155218
EDPER	.4965	.24522	153136
NATAMPER	.0026	.00303	155218
ASIANPER	.0356	.04936	155218
BLACKPER	.1766	.19035	155218
HISPANICPER	.4918	.28930	155218
WHITEPER	.2934	.27242	155218
ELLPER	.1411	.16573	150559
SPECIALEDPER	.1443	.09211	152734

Table GG16

Pearson correlation matrix for the 2007 Texas 8th listed variables in reading (Demographic, financial, and proficiency status): Weighted by school enrollment.

	STATE Reading	**NAEP BASIC level Reading**	**NAEP PROF level Reading**	**TOTAL REVENUE**	**EDPER**	**NATAMPER**
NAEP BASIC Reading		1	.134**	-.038**	-.430**	.185**
NAEP PROF Reading		.134**	1	.283**	-.426**	.063**
	ASIANPER	**BLACKPER**	**HISPANICPER**	**WHITEPER**	**ELLPER**	**SPECIALEDPER**
STATE Reading						
NAEP BASIC level	.239**	-.191**	-.269**	.374**	-.205**	-.176**
NAEP PROF level	.247**	-.199**	-.313**	.426**	-.207**	-.114**

**. Correlation is significant at the 0.01 level (2-tailed)

APPENDIX HH

Texas Tables and Figures for Research Question Two, Logistic Regression.

Table HH1

Logistic regression observed and predicted frequencies with the constant model in the Texas 2005 4th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency .00	met Proficiency 1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	4914	.0
		met proficiency status	0	175551	100.0
	Overall Percentage				97.3

a. The cut value is .500

Table HH2

Logistic regression observed and predicted frequencies with the predicted model in the Texas 4th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency .00	met Proficiency 1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	4914	.0
		met proficiency status	0	175551	100.0
	Overall Percentage				97.3

a. The cut value is .500

Table HH3

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 Texas 4th grade mathematics assessment as the dependent variable

		dimension		
		Chi-square	df	Sig.
Step 1	Step	175.586	1	.000
	Block	175.586	1	.000
	Model	15656.065	11	.000

Table HH4

Logistic regression model summary assessing goodness of fit with the Texas 4th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	29451.590[a]	.083	.376

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table HH5

Logistic Regression Analysis for the 2005 Texas 4th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-1.529	.050	941.339	1	.000	.217	.197	.239
	EDPER	-9.203	.246	1401.827	1	.000	.000	.000	.000
	NATAMPER	14520.795	29208.932	.247	1	.619	.	.000	.
	ASIANPER	-16.416	.734	500.862	1	.000	.000	.000	.000
	BLACKPER	-6.421	.350	337.319	1	.000	.002	.001	.003
	HISPANICPER	-2.397	.341	49.282	1	.000	.091	.047	.178
	DUMMYWHITE	-.855	.041	435.526	1	.000	.425	.392	.461
	ELLPER	.056	.105	.286	1	.593	1.058	.861	1.299
	SPECIALEDPER	-.145	.118	1.516	1	.218	.865	.687	1.090
	NAEPMATHPROFbasic	-.655	.061	113.508	1	.000	.519	.460	.586
	NAEPMATHPROFprof	14.767	126.289	.014	1	.907	2589704.440	.000	8.147113
	Constant	18.997	.482	1554.059	1	.000	1.7808		

Figure HH1. Logistic regression for the Texas 2005 4th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

   160000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R 120000 +                                                                                                    1+
E         I                                                                                                   1I
Q         I                                                                                                   1I
U         I                                                                                                   1I
E  80000 +                                                                                                    1+
N         I                                                                                                   1I
C         I                                                                                                   1I
Y         I                                                                                                   1I
    40000 +                                                                                                   1+
          I                                                                                                   1I
          I                                                                                                   1I
          I                                                                                               11111I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9        1
  Group:  0000000000000000000000000000000000000000000000000011111111111111111111111111111111111111111111111111

Predicted Probability is of Membership for 1.00
The Cut Value is .50
Symbols: 0 - .00
         1 - 1.00
Each Symbol Represents 10000 Cases.
```

Table HH6

Logistic regression observed and predicted frequencies with the constant model in the Texas 2005 4th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	0	6648	.0
		met proficiency status	0	173817	100.0
	Overall Percentage				96.3

a. The cut value is .500

Table HH7

Logistic regression observed and predicted frequencies with the predicted model in the Texas 2005 4th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	562	6086	8.5
		met proficiency status	92	173725	99.9
	Overall Percentage				96.6

a. The cut value is .500

Table HH8

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 4th Grade Texas reading assessment as the dependent variable

		dimension		
		Chi-square	df	Sig.
Step 1	Step	7298.018	1	.000
	Block	7298.018	1	.000
	Model	23958.562	10	.000

Table HH9

Logistic regression model summary assessing goodness of fit with the Texas 4th grade reading assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	32982.506[a]	.124	.459

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table HH10

Logistic regression analysis for the 2005 Texas 4th grade data with the reading assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-.081	.034	5.779	1	.016	.922	.863	.985
	EDPER	1.327	.056	565.997	1	.000	3.771	3.380	4.207
	NATAMPER	11802.002	30680.550	.148	1	.700	.	.000	.
	ASIANPER	-16.110	1.056	232.694	1	.000	.000	.000	.000
	BLACKPER	-4.256	.308	190.287	1	.000	.014	.008	.026
	HISPANICPER	-1.198	.282	18.020	1	.000	.302	.174	.525
	DUMMYWHITE	-.415	.039	114.798	1	.000	.660	.612	.713
	ELLPER	-.103	.088	1.344	1	.246	.903	.759	1.073
	SPECIALEDPER	-.265	.114	5.410	1	.020	.767	.614	.959
	NAEPREADPROFbasic	20.007	84.342	.056	1	.812	4.8868	.000	3.02980
	Constant	3.634	.306	141.188	1	.000	37.850		

Figure HH2. Logistic regression for the Texas 2005 4th grade dataset with the reading state accountability assessment as the dependent variable: observed and predicted probabilities

```
              Step number: 1


              Observed Groups and Predicted Probabilities

   160000 +
          I                                                                                                    I
          I                                                                                                   mI
F         I                                                                                                   mI
R 120000  +                                                                                                   m+
E         I                                                                                                   mI
Q         I                                                                                                   mI
U         I                                                                                                   mI
E  80000  +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
   40000  +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          I                                                                                                   mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 10000 Cases.
```

Table HH11

Logistic regression observed and predicted frequencies with the constant model in the Texas 2005 8th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	33553	.0
		met proficiency status	0	149389	100.0
	Overall Percentage				81.7

a. The cut value is .500

Table HH12

Logistic regression observed and predicted frequencies with the predicted model in the Texas 8th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	25067	8486	74.7
		met proficiency status	5474	143915	96.3
	Overall Percentage				92.4

a. The cut value is .500

Table HH13

Logistic regression omnibus test of model coefficients assessing goodness of fit with the Texas 2005 8th grade mathematics assessment as the dependent variable

		dimension		
		Chi-square	df	Sig.
Step 1	Step	146.411	1	.000
	Block	146.411	1	.000
	Model	104735.891	11	.000

Table HH14

Logistic regression model summary assessing goodness of fit with the Texas 2005 8th grade mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	69615.823[a]	.436	.709

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table HH15

Logistic regression results for the 2005 Texas 8th grade data with the mathematics assessment as the dependent variable

								95% C.I.for EXP(B)	
		B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Step 1	DUMMYREV	-.520	.020	664.487	1	.000	.595	.571	.619
	EDPER	-5.465	.078	4894.648	1	.000	.004	.004	.005
	NATAMPER	-241.700	4.157	3380.619	1	.000	.000	.000	.000
	ASIANPER	41.962	.664	3998.296	1	.000	1.675E18	4.561E17	6.149E18
	BLACKPER	-9.927	.185	2882.300	1	.000	.000	.000	.000
	HISPANICPER	-7.649	.175	1914.221	1	.000	.000	.000	.001
	DUMMYWHITE	-.577	.021	723.933	1	.000	.561	.538	.585
	ELLPER	-.814	.074	121.890	1	.000	.443	.383	.512
	SPECIALEDPER	-9.152	.184	2462.319	1	.000	.000	.000	.000
	NAEPMATHPROFbasic	2.047	.024	7389.787	1	.000	7.743	7.390	8.113
	NAEPMATHPROFprof	14.952	196.483	.006	1	.939	3116975.967	.000	5.496E173
	Constant	14.264	.207	4760.233	1	.000	1565862.806		

Figure HH3. Logistic Regression for the Texas 2005 8th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
              Step number: 1

              Observed Groups and Predicted Probabilities


   160000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R  120000 +                                                                                                    +
E         I                                                                                                    I
Q         I                                                                                                    I
U         I                                                                                                    I
E   80000 +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    40000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          I                                                                                           m   mmmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met proficiency
          The Cut Value is .50
          Symbols: d - did not meet proficiency
                   m - met proficiency
          Each Symbol Represents 10000 Cases.
```

Table HH16

Logistic regression observed and predicted frequencies with the constant model in the Texas 2005 8th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	0	2239	.0
		met proficiency status	0	180703	100.0
	Overall Percentage				98.8

a. The cut value is .500

Table HH17

Logistic regression observed and predicted frequencies with the predicted model in the Texas 2005 8th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	2239	0	100.0
		met proficiency status	0	180703	100.0
	Overall Percentage				100.0

a. The cut value is .500

Table HH18

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2005 8th Grade Texas reading assessment as the dependent variable

		dimension Chi-square	df	Sig.
Step 1	Step	24167.664	9	.000
	Block	24167.664	9	.000
	Model	24167.664	9	.000

Table HH19

Logistic regression model summary assessing goodness of fit with the 2005 Texas 8th grade reading assessment as the dependent variable

Step	dimension1 -2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.081[a]	.124	1.000

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table HH20

Logistic Regression Analysis for the 2005 Texas 8th grade data with the reading assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-191.209	94.672	4.079	1	.043	.000	.000	.003
	EDPER	-174.714	109.233	2.558	1	.110	.000	.000	1.263E17
	NATAMPER	-2503.922	35329.003	.005	1	.943	.000	.000	.
	ASIANPER	1177.887	10857.731	.012	1	.914	.	.000	.
	BLACKPER	1071.798	609.861	3.089	1	.079	.	.000	.
	HISPANICPER	617.042	588.422	1.100	1	.294	9.500E267	.000	.
	DUMMYWHITE	146.649	67.921	4.662	1	.031	4.886E63	748303.382	3.190E121
	ELLPER	-176.862	209.781	.711	1	.399	.000	.000	5.691E101
	SPECIALEDPER	-23.991	254.183	.009	1	.925	.000	.000	8.740E205
	Constant	-41.992	482.231	.008	1	.931	.000		

Figure HH4. Logistic regression for the Texas 2005 8th grade dataset with the reading state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities
```

```
  200000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                   mI
F        I                                                                                                   mI
R 150000 +                                                                                                   m+
E        I                                                                                                   mI
Q        I                                                                                                   mI
U        I                                                                                                   mI
E 100000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                                                                                   mI
   50000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                   mI
         I                                                                                                   mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm
```

```
Predicted Probability is of Membership for met proficiency
The Cut Value is .50
Symbols: d - did not meet proficiency
         m - met proficiency
Each Symbol Represents 12500 Cases.
```

Table HH21

Logistic regression observed and predicted frequencies with the constant model in the Texas 2007 4th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	979	.0
		met proficiency status	0	161201	100.0
	Overall Percentage				99.4

a. The cut value is .500

Table HH22

Logistic regression observed and predicted frequencies with the predicted model in the Texas 4th grade 2007 dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	979	0	100.0
		met proficiency status	0	161201	100.0
	Overall Percentage				100.0

a. The cut value is .500

Table HH23

Logistic regression omnibus test of model coefficients assessing goodness of fit with the Texas 20074th grade mathematics assessment as the dependent variable

		dimension		
		Chi-square	df	Sig.
Step 1	Step	11957.173	9	.000
	Block	11957.173	9	.000
	Model	11957.173	9	.000

Table HH24

Logistic regression model summary assessing goodness of fit with the Texas 4th grade 2007 mathematics assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	.149[a]	.071	1.000

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table HH25

Logistic regression analysis for the 2007 Texas 4th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1	DUMMYREV	-36.847	125.894	.086	1	.770	.000	.000	1.442E91
	EDPER	-1585.395	2656.207	.356	1	.551	.000	.000	.
	NATAMPER	26429.339	99561.684	.070	1	.791	.	.000	.
	ASIANPER	771.301	4501.014	.029	1	.864	.	.000	.
	BLACKPER	888.644	870.958	1.041	1	.308	.	.000	.
	HISPANICPER	769.994	787.426	.956	1	.328	.	.000	.
	DUMMYWHITE	16.397	332.659	.002	1	.961	1.322E7	.000	1.910E290
	ELLPER	164.547	309.252	.283	1	.595	2.895E71	.000	.
	SPECIALEDPER	718.754	531.470	1.829	1	.176	.	.000	.
	Constant	723.546	3583.031	.041	1	.840	.		

Figure HH5. Logistic Regression for the Texas 2007 4th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

   200000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                   mI
R 150000 +                                                                                                    m+
E         I                                                                                                   mI
Q         I                                                                                                   mI
U         I                                                                                                   mI
E 100000 +                                                                                                    m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
    50000 +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          I                                                                                                   mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met proficiency
The Cut Value is .50
Symbols: d - did not meet proficiency
         m - met proficiency
Each Symbol Represents 12500 Cases.
```

Table HH26

Logistic regression observed and predicted frequencies with the constant model in the Texas 2007 4th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	0	6672	.0
		met proficiency status	0	155508	100.0
	Overall Percentage				95.9

a. The cut value is .500

Table HH27

Logistic regression observed and predicted frequencies with the predicted model in the Texas 2007 4th grade dataset with the reading state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEREADprofstatus	did not meet proficiency status	0	6672	.0
		met proficiency status	0	155508	100.0
	Overall Percentage				95.9

a. The cut value is .500

Table HH28

Logistic regression omnibus test of model coefficients assessing goodness of fit with the 2007 4th grade Texas reading assessment as the dependent variable

		dimension		
		Chi-square	df	Sig.
Step 1	Step	76.527	1	.000
	Block	76.527	1	.000
	Model	13143.582	11	.000

Table HH29

Logistic regression model summary assessing goodness of fit with the Texas 4th grade 2007 reading assessment as the dependent variable

Step	dimension1		
	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	42499.956[a]	.078	.268

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table HH30

Logistic Regression Analysis for the 2007 Texas 4th grade data with the reading assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-.940	.048	385.936	1	.000	.391	.356	.429
	EDPER	-.756	.079	90.609	1	.000	.469	.402	.548
	NATAMPER	717.507	26.570	729.250	1	.000	.	9.841E288	.
	ASIANPER	-16.131	.801	405.831	1	.000	.000	.000	.000
	BLACKPER	-13.919	.744	350.111	1	.000	.000	.000	.000
	HISPANICPER	-12.288	.743	273.394	1	.000	.000	.000	.000
	DUMMYWHITE	-.032	.039	.645	1	.422	.969	.897	1.047
	ELLPER	.058	.084	.480	1	.488	1.060	.899	1.249
	SPECIALEDPER	-.117	.089	1.752	1	.186	.889	.748	1.058
	NAEPREADPROFbasic	.304	.033	83.168	1	.000	1.355	1.269	1.446
	NAEPREADPROFprof	14.146	184.007	.006	1	.939	1392103.079	.000	5.900E162
	Constant	17.882	.743	578.446	1	.000	5.833E7		

Figure HH6. Logistic regression for the Texas 2007 4th grade dataset with the reading state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

   160000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R 120000  +                                                                                                    +
E         I                                                                                                    I
Q         I                                                                                                    I
U         I                                                                                                   mI
E  80000  +                                                                                                   m+
N         I                                                                                                   mI
C         I                                                                                                   mI
Y         I                                                                                                   mI
   40000  +                                                                                                   m+
          I                                                                                                   mI
          I                                                                                                   mI
          I                                                                                           mmm m mmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  dddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

     Predicted Probability is of Membership for met proficiency
     The Cut Value is .50
     Symbols: d - did not meet proficiency
              m - met proficiency
     Each Symbol Represents 10000 Cases.
```

Table HH31

Logistic regression observed and predicted frequencies with the constant model in the Texas 2007 8th grade dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	0	9594	.0
		met proficiency status	0	135480	100.0
	Overall Percentage				93.4

a. The cut value is .500

Table HH32

Logistic regression observed and predicted frequencies with the predicted model in the Texas 8th grade 2007 dataset with the mathematics state accountability assessment as the dependent variable

	Observed		Predicted		
			did not meet Proficiency	met Proficiency	
			.00	1.00	
Step 1	STATEMATHprofstatus	did not meet proficiency status	1116	8478	11.6
		met proficiency status	987	134493	99.3
	Overall Percentage				93.5

a. The cut value is .500

Table HH33

Logistic regression omnibus test of model coefficients assessing goodness of fit with the Texas 2007 8th grade mathematics assessment as the dependent variable

		dimension Chi-square	df	Sig.
Step 1	Step	362.101	1	.000
	Block	362.101	1	.000
	Model	26212.122	10	.000

Table HH34

Logistic regression model summary assessing goodness of fit with the Texas 8th 2007 grade mathematics assessment as the dependent variable

Step	dimension1 -2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	44443.585[a]	.165	.429

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table HH35

Logistic Regression Analysis for the 2007 Texas 8th grade data with the mathematics assessment as the dependent variable

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1	DUMMYREV	-.413	.027	226.882	1	.000	.662	.627	.698
	EDPER	-1.167	.054	460.194	1	.000	.311	.280	.346
	NATAMPER	-85.062	6.272	183.951	1	.000	.000	.000	.000
	ASIANPER	38.943	1.874	431.734	1	.000	8.177E16	2.076E15	3.220E18
	BLACKPER	-30.602	.669	2093.578	1	.000	.000	.000	.000
	HISPANICPER	-29.326	.667	1933.769	1	.000	.000	.000	.000
	DUMMYWHITE	-1.557	.046	1137.589	1	.000	.211	.192	.231
	ELLPER	-2.003	.071	789.463	1	.000	.135	.117	.155
	SPECIALEDPER	-6.877	.125	3020.407	1	.000	.001	.001	.001
	NAEPMATHPROFbasic	-.745	.040	348.023	1	.000	.475	.439	.514
	Constant	35.742	.687	2706.044	1	.000	3.331E15		

Figure HH7. Logistic regression for the North Carolina 2007 8th grade dataset with the mathematics state accountability assessment as the dependent variable: observed and predicted probabilities

```
Step number: 1

Observed Groups and Predicted Probabilities

  160000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R 120000 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                    I
U        I                                                                                                    I
E  80000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                                                                                   mI
   40000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                   mI
         I                                                                                           m m    mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

     Predicted Probability is of Membership for met proficiency
     The Cut Value is .50
     Symbols: d - did not meet proficiency
              m - met proficiency
     Each Symbol Represents 10000 Cases.
```

APPENDIX II

California Tables and Figures for Research Question Three, Logistic Regression.

Table II1

Logistic regression observed and predicted frequencies for the constant model with the California 2005 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	122189	.0
		met AYP	0	141575	100.0
	Overall Percentage				53.7

a. The cut value is .500

Table II2

Logistic regression observed and predicted frequencies for the predicted model with the California 2005 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	100490	21699	82.2
		met AYP	31088	110487	78.0
	Overall Percentage				80.0

a. The cut value is .500

Table II3

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the California 2005 4th Grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	160793.030	9	.000
			Block	160793.030	9	.000
			Model	160793.030	9	.000

Table II4

Logistic regression model summary assessing goodness of fit with the California 2005 4th grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	203435.408[a]	.456	.610

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table II5

Logistic Regression Analysis for the 2005 California 4th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	.489	.009	3015.604	1	.000	1.631	1.603	1.660
	EDPER	-2.900	.035	6776.731	1	.000	.055	.051	.059
	NATAMPER	-17.560	.407	1865.035	1	.000	.000	.000	.000
	ASIANPER	7.068	.249	806.336	1	.000	1173.966	720.743	1912.188
	BLACKPER	-5.089	.245	429.940	1	.000	.006	.004	.010
	HISPANICPER	-2.877	.241	142.839	1	.000	.056	.035	.090
	WHITEPER	.446	.251	3.160	1	.075	1.563	.955	2.556
	ELLPER	-1.858	.036	2637.059	1	.000	.156	.145	.167
	SPECIALEDPER	2.060	.058	1250.800	1	.000	7.850	7.002	8.799
	Constant	3.388	.237	203.965	1	.000	29.620		

.

Figure II1. Logistic regression observed and predicted probabilities for the 2005 California 4th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities

    32000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R   24000 +                                                                                                   m+
E         I                                                                                                   mI
Q         I                                                                                                   mI
U         I                                                                                                   mI
E   16000 +                                                                                                   m+
N         I                                                                                                  mmI
C         I                                                                                                  mmI
Y         I         d                                                                                       mmmI
     8000 +        ddm  m  d                                                                               mmmm+
          I       mddd  dm d                                                                             mmmmmmI
          I      dddddddddmdmm     mm       m  m                             m              m     m  m   mmmmmmI
          I     dddddddddddddddd ddddmdddm  ddddm  md m   m  d d  md d m md  ddm   m mm m m md dmmmmmmmmmmmmmmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

      Predicted Probability is of Membership for met AYP
      The Cut Value is .50
      Symbols: d - did not meet AYP
               m - met AYP
      Each Symbol Represents 2000 Cases.
```

Table II6

Logistic regression observed and predicted frequencies for the constant model with the California 2005 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	256045	0	100.0
		met AYP	104059	0	.0
	Overall Percentage				71.1

a. The cut value is .500

Table II7

Logistic regression observed and predicted frequencies for the predicted model with the California 2005 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	233916	22129	91.4
		met AYP	26876	77183	74.2
	Overall Percentage				86.4

a. The cut value is .500

Table II8

Logistic regression omnibus test of model coefficients assessing goodness of fit with the California 2005 8th grade AYP results

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	207387.284	9	.000
			Block	207387.284	9	.000
			Model	207387.284	9	.000

Table II9

Logistic regression model summary assessing goodness of fit with the California 2005 8th Grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	225620.655[a]	.438	.626

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table II10

Logistic Regression Analysis for the 2005 California 8th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-.093	.008	137.183	1	.000	.911	.897	.925
	EDPER	-4.357	.047	8421.143	1	.000	.013	.012	.014
	NATAMPER	8.610	.443	377.066	1	.000	5485.244	2300.289	13080.055
	ASIANPER	6.401	.175	1337.444	1	.000	602.216	427.343	848.650
	BLACKPER	-.007	.188	.001	1	.971	.993	.687	1.437
	HISPANICPER	1.057	.176	36.159	1	.000	2.877	2.039	4.061
	WHITEPER	4.946	.180	755.652	1	.000	140.582	98.806	200.021
	ELLPER	-.273	.054	25.729	1	.000	.761	.685	.846
	SPECIALEDPER	-7.710	.104	5476.414	1	.000	.000	.000	.001
	Constant	-.714	.169	17.873	1	.000	.489		

Figure II2. Logistic regression observed and predicted probabilities for the 2005 California 8th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities

   80000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R  60000 +                                                                                                    +
E        Id                                                                                                   I
Q        Id                                                                                                   I
U        Id                                                                                                   I
E  40000 +dd                                                                                                  +
N        Idd                                                                                                  I
C        Idd                                                                                                  I
Y        Idd                                                                                                  I
   20000 +ddd                                                                                                 +
         Idddd                                                                                                I
         Idddd dm   d        d            m                                                                   I
         Iddddddd ddd dmdm  md  d   m d   d    m m   m    d d   m       d mm      m m m m m  m m  mmmm mmmm   I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met AYP
The Cut Value is .50
Symbols: d - did not meet AYP
         m - met AYP
Each Symbol Represents 5000 Cases.
```

Table II11

Logistic regression observed and predicted frequencies for the constant model with the California 2007 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	54852	.0
		met AYP	0	118177	100.0
	Overall Percentage				68.3

a. The cut value is .500

Table II12

Logistic regression observed and predicted frequencies for the predicted model with the California 2007 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	34188	20664	62.3
		met AYP	13387	104790	88.7
	Overall Percentage				80.3

a. The cut value is .500

Table II13

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the California 2007 AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	58478.159	9	.000
			Block	58478.159	9	.000
			Model	58478.159	9	.000

Table II14

Logistic regression model summary assessing goodness of fit with the California 2007 4th grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	157668.091[a]	.287	.402

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table II15

Logistic Regression Analysis for the 2007 California 4th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	.807	.009	7501.956	1	.000	2.241	2.200	2.282
	EDPER	.644	.048	181.341	1	.000	1.904	1.734	2.092
	NATAMPER	-111.738	1.574	5040.290	1	.000	.000	.000	.000
	ASIANPER	.069	.361	.037	1	.848	1.072	.528	2.173
	BLACKPER	-10.653	.351	920.052	1	.000	.000	.000	.000
	HISPANICPER	-9.012	.344	685.058	1	.000	.000	.000	.000
	WHITEPER	-6.528	.354	339.315	1	.000	.001	.001	.003
	ELLPER	-4.068	.045	8188.868	1	.000	.017	.016	.019
	SPECIALEDPER	-3.548	.090	1546.115	1	.000	.029	.024	.034
	Constant	8.526	.334	650.936	1	.000	5041.981		

Figure II3. Logistic regression observed and predicted probabilities for the 2007 California 4th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities

   20000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                   mI
F        I                                                                                                   mI
R  15000 +                                                                                                   m+
E        I                                                                                                   mI
Q        I                                                                                                   mI
U        I                                                                                                   mI
E  10000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                          m                                                       mmI
    5000 +                             m            m                                                     m mm+
         I                             d d          d  m       m                        m   m   m    mm   mmmmI
         I                            md d    mm    d  d m  m  dm  mdmm m          mmmm mm mmm mmm mmmm mmmmmmI
         I    d     d       d   mdd d ddddd   md  d dd dddd d mdmd mdmmmdmdmmmmmmdmdmdmmmmmdmdmddm mmmmmmmmmmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

     Predicted Probability is of Membership for met AYP
     The Cut Value is .50
     Symbols: d - did not meet AYP
              m - met AYP
     Each Symbol Represents 1250 Cases.
```

Table II16

Logistic regression observed and predicted frequencies for the constant model with the California 2007 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	210682	0	100.0
		met AYP	75750	0	.0
	Overall Percentage				73.6

a. The cut value is .500

Table II17

Logistic regression observed and predicted frequencies for the predicted model with the California 2007 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	196872	13810	93.4
		met AYP	28034	47716	63.0
	Overall Percentage				85.4

a. The cut value is .500

Table II18

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the California 2007 8th grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	123128.116	9	.000
			Block	123128.116	9	.000
			Model	123128.116	9	.000

Table II19

Logistic regression model summary assessing goodness of fit with the California 2007 8th grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	207798.874[a]	.349	.510

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table II20

Logistic regression analysis for the 2007 California 8th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1	DUMMYREV	-.445	.007	4161.871	1	.000	.641	.632	.650
	EDPER	-1.784	.050	1285.062	1	.000	.168	.152	.185
	NATAMPER	10.808	.413	684.925	1	.000	49401.423	21989.891	110982.845
	ASIANPER	7.488	.176	1814.945	1	.000	1787.022	1266.227	2522.017
	BLACKPER	-1.455	.193	56.796	1	.000	.233	.160	.341
	HISPANICPER	2.765	.163	287.185	1	.000	15.876	11.531	21.859
	WHITEPER	6.423	.168	1458.006	1	.000	616.049	443.022	856.654
	ELLPER	.180	.044	16.575	1	.000	1.197	1.098	1.305
	SPECIALEDPER	-6.010	.109	3062.949	1	.000	.002	.002	.003
	Constant	-2.294	.158	211.205	1	.000	.101		

Figure II4. Logistic regression observed and predicted probabilities for the 2007 California 8th grade AYP dataset

```
               Step number: 1

               Observed Groups and Predicted Probabilities

   40000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I   d                                                                                                I
R  30000 +   d                                                                                                +
E        I   d                                                                                                I
Q        I   d                                                                                                I
U        I   d                                                                                                I
E  20000 +   dd                                                                                               +
N        I   ddm                                                                                              I
C        I   ddd                                                                                              I
Y        I ddddd                                                                                              I
   10000 + dddddd m d       m                                                                                 +
         Iddddddddmd d       d                                                                                I
         Iddddddddddmd       d d       d               d                               m                      I
         Iddddddddddd ddmdd d dmmd  d dddd   dd dmd d d   m      dmmm    dm      m m mm m m d  mddmmm         I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met AYP
          The Cut Value is .50
          Symbols: d - did not meet AYP
                   m - met AYP
          Each Symbol Represents 2500 Cases.
```

APPENDIX JJ

Michigan Tables and Figures for Research Question Three, Logistic Regression.

Table JJ1

Logistic regression observed and predicted frequencies for the constant model with the Michigan 2005 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	257	.0
		met AYP	0	41150	100.0
	Overall Percentage				99.4

a. The cut value is .500

Table JJ2

Logistic regression observed and predicted frequencies for the predicted model with the Michigan 2005 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	257	0	100.0
		met AYP	0	41150	100.0
	Overall Percentage				100.0

a. The cut value is .500

Table JJ3

Logistic regression omnibus test of model coefficients assessing goodness of fit with the Michigan 2005 4th grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	3124.591	9	.000
			Block	3124.591	9	.000
			Model	3124.591	9	.000

Table JJ4

Logistic regression model summary assessing goodness of fit with the Michigan 2005 4th grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	.025[a]	.073	1.000

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table JJ5

Logistic Regression Analysis for the 2005 Michigan 4th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	91.817	490.372	.035	1	.851	7.507E39	.000	.
	EDPER	-302.215	1431.238	.045	1	.833	.000	.000	.
	NATAM	4670.965	7130.084	.429	1	.512	.	.000	.
	ASIANPER	3998.331	5912.555	.457	1	.499	.	.000	.
	BLACKPER	5036.939	6604.303	.582	1	.446	.	.000	.
	HISPANICPER	10730.701	12481.172	.739	1	.390	.	.000	.
	WHITEPER	5314.615	6999.523	.577	1	.448	.	.000	.
	ELLPER	-609.437	1034.335	.347	1	.556	.000	.000	.
	SPECIALEDPER	-185.594	497.872	.139	1	.709	.000	.000	.
	Constant	-4993.928	6677.717	.559	1	.455	.000		

Figure JJ1. Logistic regression observed and predicted probabilities for the 2005 Michigan 4th grade AYP dataset

```
                    Step number: 1

                    Observed Groups and Predicted Probabilities

   80000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R  60000 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                    I
U        I                                                                                                    I
E  40000 +                                                                                                   M+
N        I                                                                                                   MI
C        I                                                                                                   MI
Y        I                                                                                                   MI
   20000 +                                                                                                   M+
         I                                                                                                   MI
         I                                                                                                   MI
         I                                                                                                   MI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9        1
  Group:  DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

          Predicted Probability is of Membership for Made AYP
          The Cut Value is .50
          Symbols: D - Did not make AYP
                   M - Made AYP
          Each Symbol Represents 5000 Cases.
```

Table JJ6

Logistic regression observed and predicted frequencies for the constant model with the Michigan 2005 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	3686	.0
		met AYP	0	53317	100.0
	Overall Percentage				93.5

a. The cut value is .500

Table JJ7

Logistic regression observed and predicted frequencies for the predicted model with the Michigan 2005 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	3686	.0
		met AYP	959	52358	98.2
	Overall Percentage				91.9

a. The cut value is .500

Table JJ8

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the Michigan 2005 8th grade state AYP results as the dependent variable

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	12781.160	9	.000
			Block	12781.160	9	.000
			Model	12781.160	9	.000

Table JJ9

Logistic regression model summary assessing goodness of fit with the Michigan 2005 8th grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	14535.860[a]	.201	.528

a. Estimation terminated at iteration number 20 because maximum iterations has been reached. Final solution cannot be found.

Table JJ10

Logistic regression analysis for the 2005 Michigan 8th grade AYP dataset

								95% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1	DUMMYREV	.087	.079	1.203	1	.273	1.091	.934	1.275
	EDPER	5.997	.209	819.809	1	.000	402.248	266.814	606.427
	NATAMPER	-6307.690	41104.960	.024	1	.878	.000	.000	.
	ASIANPER	-5783.507	41104.972	.020	1	.888	.000	.000	.
	BLACKPER	-6338.654	41104.960	.024	1	.877	.000	.000	.
	HISPANICPER	-6379.861	41104.960	.024	1	.877	.000	.000	.
	WHITEPER	-6333.128	41104.960	.024	1	.878	.000	.000	.
	ELLPER	3.121	.218	204.769	1	.000	22.659	14.778	34.743
	SPECIALEDPER	-2.495	.351	50.630	1	.000	.083	.041	.164
	Constant	6334.298	41104.960	.024	1	.878	.		

Figure JJ2. Logistic regression observed and predicted probabilities for the 2005 Michigan 8th grade AYP dataset

```
              Step number: 1

              Observed Groups and Predicted Probabilities


    80000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                                                    I
R   60000 +                                                                                                    +
E         I                                                                                                    I
Q         I                                                                                                    I
U         I                                                                                                    I
E   40000 +                                                                                                   M+
N         I                                                                                                   MI
C         I                                                                                                   MI
Y         I                                                                                                   MI
    20000 +                                                                                                   M+
          I                                                                                                   MI
          I                                                                                                   MI
          I                                                                                                   MI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM

          Predicted Probability is of Membership for Made AYP
          The Cut Value is .50
          Symbols: D - Did not make AYP
                   M - Made AYP
          Each Symbol Represents 5000 Cases.
```

Table JJ11

Logistic regression observed and predicted frequencies for the constant model with the Michigan 2007 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	2311	.0
		met AYP	0	36862	100.0
	Overall Percentage				94.1

a. The cut value is .500

Table JJ12

Logistic regression observed and predicted frequencies for the predicted model with the Michigan 2007 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	872	1439	37.7
		met AYP	345	36517	99.1
	Overall Percentage				95.4

a. The cut value is .500

Table JJ13

Logistic regression omnibus test of model coefficients assessing goodness of fit with the Michigan 2007 4th grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	7552.565	9	.000
			Block	7552.565	9	.000
			Model	7552.565	9	.000

Table JJ14

Logistic regression model summary assessing goodness of fit with the Michigan 2007 4th grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	10012.016[a]	.175	.485

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

Table JJ15

Logistic regression analysis for the 2007 Michigan 4th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-1.356	.055	610.494	1	.000	.258	.232	.287
	EDPER	-9.701	.230	1776.144	1	.000	.000	.000	.000
	NATAMPER	-80.485	5.571	208.698	1	.000	.000	.000	.000
	ASIANPER	-67.645	6.160	120.606	1	.000	.000	.000	.000
	BLACKPER	-71.802	5.471	172.229	1	.000	.000	.000	.000
	HISPANIC	-44.632	5.365	69.207	1	.000	.000	.000	.000
	WHITEPER	-75.542	5.479	190.131	1	.000	.000	.000	.000
	ELLPER	-24.244	1.814	178.675	1	.000	.000	.000	.000
	SPECIALEDPER	8.650	.404	458.341	1	.000	5712.519	2587.579	12611.355
	Constant	84.724	5.484	238.694	1	.000	6.237E36		

Figure JJ3. Logistic regression observed and predicted probabilities for the 2007 Michigan 4th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities

   32000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R  24000 +                                                                                                    +
E        I                                                                                                   mI
Q        I                                                                                                   mI
U        I                                                                                                   mI
E  16000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                                                                                   mI
    8000 +                                                                                                   m+
         I                                                                                                  mmI
         I                                                                                                  mmI
         I                                                                                             mmmm mmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met AYP
The Cut Value is .50
Symbols: D - Did not meet AYP
         m - met AYP
Each Symbol Represents 2000 Cases.
```

Table JJ16

Logistic regression observed and predicted frequencies for the constant model with the Michigan 2007 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	3235	.0
		met AYP	0	44559	100.0
	Overall Percentage				93.2

a. The cut value is .500

Table JJ17

Logistic regression observed and predicted frequencies for the predicted model with the Michigan 2007 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	2117	1118	65.4
		met AYP	302	44257	99.3
	Overall Percentage				97.0

a. The cut value is .500

Table JJ18

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the Michigan 2007 8th Grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	13163.375	9	.000
			Block	13163.375	9	.000
			Model	13163.375	9	.000

Table JJ19

Logistic regression model summary assessing goodness of fit with the Michigan 2007 8th Grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	10505.422[a]	.241	.616

a. Estimation terminated at iteration number 11 because parameter estimates changed by less than .001.

Table JJ20

Logistic regression analysis for the 2007 Michigan 8th grade AYP dataset

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
DUMMYREV	.555	.088	39.825	1	.000	1.742	1.466	2.070
EDPER	1.264	.291	18.862	1	.000	3.538	2.000	6.257
NATAMPER	233.332	10.367	506.617	1	.000	2.162E101	3.242E92	1.442E110
ASIANPER	65.618	6.576	99.556	1	.000	3.146E28	7.941E22	1.246E34
BLACKPER	-58.474	4.912	141.737	1	.000	.000	.000	.000
HISPANICPER	-56.374	4.961	129.127	1	.000	.000	.000	.000
WHITEPER	-50.767	4.869	108.719	1	.000	.000	.000	.000
ELLPER	-7.751	.508	232.840	1	.000	.000	.000	.001
SPECIALEDPER	-18.583	.796	544.847	1	.000	.000	.000	.000
Constant	53.964	4.960	118.356	1	.000	2.732E23		

Figure JJ4. Logistic regression observed and predicted probabilities for the 2007 Michigan 8th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities

   32000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                   mI
F        I                                                                                                   mI
R  24000 +                                                                                                   m+
E        I                                                                                                   mI
Q        I                                                                                                   mI
U        I                                                                                                   mI
E  16000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                                                                                   mI
    8000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                  mmI
         I                                                                             m   m         d m mm mmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met ayp
The Cut Value is .50
Symbols: d - did not meet AYP
         m - met ayp
Each Symbol Represents 2000 Cases.
```

APPENDIX KK

North Carolina Tables and Figures for Research Question Three, Logistic Regression.

Table KK1

Logistic regression observed and predicted frequencies for the constant model with the North Carolina 2005 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	36183	.0
		met AYP	0	58749	100.0
	Overall Percentage				61.9

a. The cut value is .500

Table KK2

Logistic regression observed and predicted frequencies for the predicted model with the North Carolina 2005 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	15514	20669	42.9
		met AYP	6913	51836	88.2
	Overall Percentage				70.9

a. The cut value is .500

Table KK3

Logistic regression omnibus test of model coefficients assessing goodness of fit with the North Carolina 2005 4th grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	24699.734	9	.000
			Block	24699.734	9	.000
			Model	24699.734	9	.000

Table KK4

Logistic regression model summary assessing goodness of fit with the North Carolina 2005 4th Grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
	1	101488.172[a]	.229	.312

a. Estimation terminated at iteration number 8 because maximum iterations has been reached. Final solution cannot be found.

Table KK5

Logistic Regression Results for the 2005 North Carolina 4th grade AYP dataset

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
DUMMYREV	-.041	.015	8.120	1	.004	.959	.933	.987
EDPER	-.382	.057	45.240	1	.000	.682	.610	.763
NATAMPER	-59.362	1.000	3524.242	1	.000	.000	.000	.000
ASIANPER	12.149	.315	1488.024	1	.000	188978.973	101934.837	350351.786
BLACKPER	4.416	.108	1669.752	1	.000	82.748	66.954	102.269
HISPANICPER	6.833	.192	1267.449	1	.000	928.177	637.167	1352.099
DUMMYWHITE	1.196	.020	3658.115	1	.000	3.307	3.181	3.437
ELLPER	-4.180	.137	935.714	1	.000	.015	.012	.020
SPECIALEDPER	-1.101	.135	66.322	1	.000	.332	.255	.433
Constant	-6.077	.142	1822.641	1	.000	.002		

Figure KK1. Logistic regression observed and predicted probabilities for the 2005 NC 4th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities

     8000 +                                                                                                    +
          I                                                                                                    I
          I                                                                                                    I
F         I                                                                              m                     I
R    6000 +                                                                              m                     +
E         I                                                                              m                     I
Q         I                                                                              m                     I
U         I                                                                              m                     I
E    4000 +                                                                              m                     +
N         I                                                               m                m                     I
C         I                                                               m    m     m mm  mm   m   m            I
Y         ID                                                         m  m  mmm     m mm  mm  mm   m            I
     2000 +D                          D                              m mD  mmmm  m mmDm  mmmmmmm mm            +
          ID         D      m         D                 DD        m  D DDD mDmm mm mmDm  Dmmmmmm mm m          I
          ID         D   D  m    D D  D         m       DD    m   D  DDDDD mDmm mDDmmDmm DDmmmmmmmmmmmm        I
          IDD        DDDDD  mmm  DDD  D m m  m mmm     DDDmm  m  mDDDDDDDDDmDDDmmDDmDDDmmDDmmDDmmmmmmmmm       I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  DDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDDmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

        Predicted Probability is of Membership for met AYP
        The Cut Value is .50
        Symbols: D - Did not meet AYP
                 m - met AYP
        Each Symbol Represents 500 Cases.
```

Table KK6

Logistic regression observed and predicted frequencies for the constant model with the North Carolina 2005 8^{th} grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	60063	0	100.0
		met AYP	19485	0	.0
	Overall Percentage				75.5

a. The cut value is .500

Table KK7

Logistic regression observed and predicted frequencies for the predicted model with the North Carolina 2005 8^{th} grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	57551	2512	95.8
		met AYP	10306	9179	47.1
	Overall Percentage				83.9

a. The cut value is .500

Table KK8

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the North Carolina 2005 8th grade state AYP results as the dependent variable

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	18665.023	9	.000
			Block	18665.023	9	.000
			Model	18665.023	9	.000

Table KK9

Logistic regression model summary assessing goodness of fit with the North Carolina 2005 8th Grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	69906.081[a]	.209	.311

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table KK10

Logistic regression results for the 2005 North Carolina 8th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	.890	.018	2440.587	1	.000	2.435	2.351	2.523
	EDPER	-2.988	.079	1439.587	1	.000	.050	.043	.059
	NATAMPER	5.821	.310	353.313	1	.000	337.433	183.894	619.165
	ASIANPER	6.902	.338	418.082	1	.000	994.336	513.101	1926.918
	BLACKPER	-3.458	.149	542.131	1	.000	.031	.024	.042
	HISPANICPER	-3.492	.347	101.471	1	.000	.030	.015	.060
	DUMMYWHITE	-.249	.027	87.829	1	.000	.780	.740	.821
	ELLPER	.263	.291	.817	1	.366	1.301	.735	2.302
	SPECIALEDPER	-6.133	.154	1589.159	1	.000	.002	.002	.003
	Constant	2.051	.185	123.447	1	.000	7.774		

Figure KK2. Logistic regression observed and predicted probabilities for the 2005 NC 8th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities

    8000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R   6000 +                                                                                                    +
E        I    m                                                                                               I
Q        I    m                                                                                               I
U        I d  d                                                                                               I
E   4000 + d  d                                                                                               +
N        I d  d                                                                                               I
C        I d  d   m  d m     m      m  m                                                                      I
Y        Idd  d   d md m  d  m  d   m  m         d                                                            I
    2000 +dddddd  d mddd  d  md d   m  d         d               m                                            +
         Idddddd  d dddd  d  md d   d  d         d       d       m        m                                   I
         Iddddddd d ddddd d  dd dd  ddmdmdd  d  dd   d   d   d m m        mmm     m                           I
         Iddddddddddddddd d  ddddddmddmdmdddddd ddmm dd  d   d m d   d mm dmm     m         m            m    I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9        1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met AYP
The Cut Value is .50
Symbols: d - did not meet AYP
         m - met AYP
Each Symbol Represents 500 Cases.
```

Table KK11

Logistic regression observed and predicted frequencies for the constant model with the North Carolina 2007 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	48342	0	100.0
		met AYP	45339	0	.0
	Overall Percentage				51.6

a. The cut value is .500

Table KK12

Logistic regression observed and predicted frequencies for the predicted model with the North Carolina 2007 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	34294	14048	70.9
		met AYP	11612	33727	74.4
	Overall Percentage				72.6

a. The cut value is .500

Table KK13

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the North Carolina 2007 4th Grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	29391.136	9	.000
			Block	29391.136	9	.000
			Model	29391.136	9	.000

Table KK14

Logistic regression model summary assessing goodness of fit with the North Carolina 2007 4th Grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	100382.026[a]	.269	.359

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

Table KK15

Logistic Regression Analysis for the 2007 North Carolina 4th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-.206	.016	158.450	1	.000	.814	.788	.840
	EDPER	-2.248	.060	1404.512	1	.000	.106	.094	.119
	NATAMPER	-4.564	.355	165.056	1	.000	.010	.005	.021
	ASIANPER	.009	.529	.000	1	.987	1.009	.357	2.847
	BLACKPER	-6.228	.369	284.960	1	.000	.002	.001	.004
	HISPANICPER	.678	.380	3.177	1	.075	1.970	.935	4.151
	WHITEPER	-3.084	.357	74.647	1	.000	.046	.023	.092
	ELLPER	-6.259	.128	2404.799	1	.000	.002	.001	.002
	SPECIALEDPER	-4.193	.108	1510.806	1	.000	.015	.012	.019
	Constant	5.966	.348	294.233	1	.000	390.095		

Figure KK3. Logistic regression observed and predicted probabilities for the 2007 NC 4th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities

    8000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R   6000 +                                                                                                    +
E        I                                                                                                    I
Q        I                                                                                                    I
U        I                                                                                                    I
E   4000 +                                                                m                                   +
N        I                                                                m                                   I
C        I                                                                m                  m                I
Y        I                  m                                 d         m m  m dm m          m      m         I
    2000 +           m     mm    m   dd    m                  d        mm d  m dm m          m    m m         +
         I  d   dd d d    ddmm   m m dd d  d   m              d m      mm d  m dmmm     m    dmmm m m         I
         I  d dddd d d    dddm  dmmmddd d  dd  d     d d m   dd d   d  mm d  m ddmm    mmmm  dmmm mmm         I
         I dd dddd dddmdd dddd  dddddddmddmddd dd ddmd dmmmdmddmdmm d  dd dmmmmdddm   mmmmmm dmmm mmm         I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met AYP
The Cut Value is .50
Symbols: d - did not meet AYP
         m - met AYP
Each Symbol Represents 500 Cases.
```

Table KK16

Logistic regression observed and predicted frequencies for the constant model with the North Carolina 2007 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	71728	0	100.0
		met AYP	15489	0	.0
	Overall Percentage				82.2

a. The cut value is .500

Table KK17

Logistic regression observed and predicted frequencies for the predicted model with the North Carolina 2007 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	70621	1107	98.5
		met AYP	12462	3027	19.5
	Overall Percentage				84.4

a. The cut value is .500

Table KK18

Logistic regression omnibus test of model coefficients assessing goodness of fit with the North Carolina 2007 8th Grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	15728.690	9	.000
			Block	15728.690	9	.000
			Model	15728.690	9	.000

Table KK19

Logistic regression model summary assessing goodness of fit with the North Carolina 2007 8th Grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	65857.874[a]	.165	.272

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Table KK20

Logistic regression analysis for the 2007 North Carolina 8th grade AYP dataset

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
							Lower	Upper
DUMMYREV	.309	.013	544.135	1	.000	1.362	1.327	1.398
EDPER	-1.114	.093	143.432	1	.000	.328	.274	.394
NATAMPER	-5.560	.684	66.034	1	.000	.004	.001	.015
ASIANPER	11.399	.862	174.829	1	.000	89196.438	16464.529	483220.900
BLACKPER	-5.650	.692	66.593	1	.000	.004	.001	.014
HISPANICPER	-10.367	.848	149.542	1	.000	.000	.000	.000
WHITEPER	-2.588	.685	14.251	1	.000	.075	.020	.288
ELLPER	-7.247	.425	290.922	1	.000	.001	.000	.002
SPECIALEDPER	-1.894	.148	162.827	1	.000	.150	.112	.201
Constant	2.411	.678	12.641	1	.000	11.142		

Figure KK4. Logistic regression observed and predicted probabilities for the 2007 NC 8th grade AYP dataset

```
              Step number: 1

              Observed Groups and Predicted Probabilities

    8000 +                                                                                                    +
         I                                                                                                    I
         I     d                                                                                              I
F        Id    d                                                                                              I
R   6000 +d    d    m                                                                                         +
E        Idd   d    d                                                                                         I
Q        Iddd  d    d                                                                                         I
U        Iddd  d    d                                                                                         I
E   4000 +ddd  d   md                                                                                         +
N        Iddd  d  mmd            d                                                                            I
C        Iddd dd  mdd            d                                                                            I
Y        Iddd dd dmdd            d   d             d                                                          I
    2000 +ddd dd dddd     m      d   d       m     d                                                          +
         Idddddd ddddd  ddm      d mmdd      m     dm   m              m                                      I
         Idddddddddddddd dddddd  dd dmddm d d d m dmdm d d    m         m     d                                I
         Iddddddddddddd ddddddm ddddddddmddd ddm dmdm d d    m         m     d                                I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met AYP
          The Cut Value is .50
          Symbols: d - did not meet AYP
                   m - met AYP
          Each Symbol Represents 500 Cases.
```

APPENDIX LL

Texas Tables and Figures for Research Question Three, Logistic Regression.

Table LL1

Logistic regression observed and predicted frequencies for the constant model with the Texas 2005 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	3154	.0
		met AYP	0	173376	100.0
	Overall Percentage				98.2

a. The cut value is .500

Table LL2

Logistic regression observed and predicted frequencies for the predicted model with the Texas 2005 4th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	3154	.0
		met AYP	0	173376	100.0
	Overall Percentage				98.2

a. The cut value is .500

Table LL3

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the Texas 2005 4th grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	6018.967	9	.000
			Block	6018.967	9	.000
			Model	6018.967	9	.000

Table LL4

Logistic regression model summary assessing goodness of fit with the Texas 2005 4th grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	25620.904[a]	.034	.204

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

Table LL5

Logistic regression results for the 2005 Texas 4th grade AYP dataset

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
DUMMYREV	-.072	.029	6.329	1	.012	.930	.879	.984
EDPER	-4.035	.160	635.043	1	.000	.018	.013	.024
NATAMPER	-30.263	3.032	99.591	1	.000	.000	.000	.000
ASIANPER	5.348	.965	30.692	1	.000	210.211	31.691	1394.343
BLACKPER	-5.346	.256	435.009	1	.000	.005	.003	.008
HISPANICPER	-2.652	.225	139.039	1	.000	.071	.045	.110
DUMMYWHITE	-.972	.034	834.234	1	.000	.378	.354	.404
ELLPER	4.068	.200	414.406	1	.000	58.428	39.494	86.439
SPECIALEDPER	-.253	.175	2.086	1	.149	.777	.551	1.094
Constant	11.425	.264	1872.809	1	.000	91586.154		

Figure LL1. Logistic regression observed and predicted probabilities for the 2005 Texas 4th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities

  160000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R 120000 +                                                                                                    +
E        I                                                                                                   mI
Q        I                                                                                                   mI
U        I                                                                                                   mI
E  80000 +                                                                                                   m+
N        I                                                                                                   mI
C        I                                                                                                   mI
Y        I                                                                                                   mI
   40000 +                                                                                                   m+
         I                                                                                                   mI
         I                                                                                                  mmI
         I                                                                                              mmmmmmI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9        1
  Group:  ddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

Predicted Probability is of Membership for met AYP
The Cut Value is .50
Symbols: d - did not meet AYP
         m - met AYP
Each Symbol Represents 10000 Cases.
```

Table LL6

Logistic regression observed and predicted frequencies for the constant model with the Texas 2005 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	48116	.0
		met AYP	0	126661	100.0
	Overall Percentage				72.5

a. The cut value is .500

Table LL7

Logistic regression observed and predicted frequencies for the predicted model with the Texas 2005 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	23699	24417	49.3
		met AYP	15350	111311	87.9
	Overall Percentage				77.2

a. The cut value is .500

Table LL8

Logistic regression omnibus test of model coefficients assessing goodness of fit with the Texas 2005 state AYP results as the dependent variable

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	56749.300	9	.000
			Block	56749.300	9	.000
			Model	56749.300	9	.000

Table LL9

Logistic regression model summary assessing goodness of fit with the Texas 2005 8^{th} grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	148948.818[a]	.277	.401

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table LL10

Logistic regression results for the 2005 Texas 8th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1	DUMMYREV	-.324	.011	800.950	1	.000	.724	.708	.740
	EDPER	-1.873	.036	2692.670	1	.000	.154	.143	.165
	NATAMPER	-159.361	2.353	4587.364	1	.000	.000	.000	.000
	ASIANPER	4.796	.248	372.680	1	.000	121.007	74.362	196.912
	BLACKPER	-5.775	.095	3689.228	1	.000	.003	.003	.004
	HISPANICPER	-5.158	.086	3569.440	1	.000	.006	.005	.007
	DUMMYWHITE	-.395	.012	1096.623	1	.000	.673	.658	.689
	ELLPER	-3.014	.062	2393.937	1	.000	.049	.043	.055
	SPECIALEDPER	-6.536	.117	3121.933	1	.000	.001	.001	.002
	Constant	9.278	.105	7872.524	1	.000	10695.465		

.

Figure LL12. Logistic regression observed and predicted probabilities for the 2005 Texas 8th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities


   16000 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                    I
R  12000 +                                                                                                  1 +
E        I                                                                                                  1 I
Q        I                                                                                                  11I
U        I                                                                                                  11I
E   8000 +                                                                                             1    11+
N        I                                                                                             1    11I
C        I                                                                                             1 1  11I
Y        I                                              1  1             1                         1 1 111  11I
    4000 +                                           0  1  1 1           1               1         1 1 1111 11+
         I0                                      0   0  1  1 1       0   1       1       1 1 1111 1111 1111111I
         I0               0                 1    0   01 1 10 0    0  0  10111   11   1 111111110111111 1111111I
         I0           1 000 0   0      00 1001 100  001 0000 0 1  00 00100000 10101110111110001011111101100111I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  0000000000000000000000000000000000000000000000000011111111111111111111111111111111111111111111111111

Predicted Probability is of Membership for met AYP
The Cut Value is .50
Symbols: 0 - .00
         1 - met AYP
Each Symbol Represents 1000 Cases.
```

Table LL11

Logistic regression observed and predicted frequencies for the constant model with the Texas 2007 4^{th} grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	1385	.0
		met AYP	0	159983	100.0
	Overall Percentage				99.1

a. The cut value is .500

Table LL12

Logistic regression observed and predicted frequencies for the predicted model with the Texas 2007 4^{th} grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	1385	0	100.0
		met AYP	0	159983	100.0
	Overall Percentage				100.0

a. The cut value is .500

Table LL13

Logistic Regression Omnibus Test of Model Coefficients assessing goodness of fit with the Texas 2007 4th grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	15937.592	9	.000
			Block	15937.592	9	.000
			Model	15937.592	9	.000

Table LL14

Logistic regression model summary assessing goodness of fit with the Texas 2007 4th grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	.112[a]	.094	1.000

a. Estimation terminated at iteration number 6 because parameter estimates changed by less than .001.

Table LL15

Logistic regression analysis for the 2007 Texas 4th grade AYP dataset

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1	DUMMYREV	1.848	67.726	.001	1	.978	6.345	.000	2.822E58
	EDPER	-870.963	560.948	2.411	1	.121	.000	.000	1.679E99
	NATAMPER	14790.617	39270.849	.142	1	.706	.	.000	.
	ASIANPER	4610.234	4115.124	1.255	1	.263	.	.000	.
	BLACKPER	-890.412	818.811	1.183	1	.277	.000	.000	.
	HISPANICPER	-601.592	789.938	.580	1	.446	.000	.000	.
	DUMMYWHITE	-136.588	67.069	4.147	1	.042	.000	.000	.006
	ELLPER	-110.274	94.191	1.371	1	.242	.000	.000	1.924E32
	SPECIALEDPER	33.556	43.363	.599	1	.439	3.742E14	.000	3.043E51
	Constant	1678.632	921.827	3.316	1	.069	.		

.

Figure LL3. Logistic regression observed and predicted probabilities for the 2007 Texas 4th grade AYP dataset

```
              Step number: 1

              Observed Groups and Predicted Probabilities

   160000 +                                                                                                    m+
          I                                                                                                    mI
          I                                                                                                    mI
F         I                                                                                                    mI
R  120000 +                                                                                                    m+
E         I                                                                                                    mI
Q         I                                                                                                    mI
U         I                                                                                                    mI
E   80000 +                                                                                                    m+
N         I                                                                                                    mI
C         I                                                                                                    mI
Y         I                                                                                                    mI
    40000 +                                                                                                    m+
          I                                                                                                    mI
          I                                                                                                    mI
          I                                                                                                    mI
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  dddddddddddddddddddddddddddddddddddddddddddddddddddmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmmm

          Predicted Probability is of Membership for met AYP
          The Cut Value is .50
          Symbols: d - did not meet ayp
                   m - met AYP
          Each Symbol Represents 10000 Cases.
```

Table LL16

Logistic regression observed and predicted frequencies for the constant model with the Texas 2007 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	0	27524	.0
		met AYP	0	117818	100.0
	Overall Percentage				81.1

a. The cut value is .500

Table LL17

Logistic regression observed and predicted frequencies for the predicted model with the Texas 2007 8th grade state AYP results

	Observed		Predicted		
			did not meet AYP	met AYP	
			.00	1.00	
Step 1	STATEAYP	did not meet AYP	6540	20984	23.8
		met AYP	4809	113009	95.9
	Overall Percentage				82.3

a. The cut value is .500

Table LL18

Logistic regression omnibus test of model coefficients assessing goodness of fit with the Texas 2007 8th grade AYP dataset

				dimension2		
				Chi-square	df	Sig.
dimension0	Step 1	dimension1	Step	32362.968	9	.000
			Block	32362.968	9	.000
			Model	32362.968	9	.000

Table LL19

Logistic regression model summary assessing goodness of fit with the Texas 2007 8th grade AYP data

Step		dimension1		
		-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
dimension0	1	108710.054[a]	.200	.321

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

Table LL20

Logistic regression analysis for the 2007 Texas 8th grade AYP dataset

							95% C.I.for EXP(B)	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
DUMMYREV	-.096	.016	36.734	1	.000	.909	.881	
EDPER	-2.426	.053	2081.460	1	.000	.088	.080	
NATAMPER	-209.454	3.755	3111.070	1	.000	.000	.000	
ASIANPER	6.155	.313	386.722	1	.000	471.173	255.126	
BLACKPER	-11.830	.135	7665.084	1	.000	.000	.000	
HISPANICPER	-11.182	.128	7672.348	1	.000	.000	.000	
DUMMYWHITE	-1.116	.016	4945.829	1	.000	.328	.318	
ELLPER	1.599	.066	587.699	1	.000	4.946	4.346	
SPECIALEDPER	-1.560	.072	463.967	1	.000	.210	.182	
Constant	15.031	.153	9696.433	1	.000	3372716.601		

Figure LL4. Logistic regression observed and predicted probabilities for the 2007 Texas 8th grade AYP dataset

```
Step number: 1

Observed Groups and Predicted Probabilities


   20000 +                                                                                                     +
         I                                                                                                    1I
         I                                                                                                    1I
F        I                                                                                                    1I
R  15000 +                                                                                                    1+
E        I                                                                                                    1I
Q        I                                                                                                    1I
U        I                                                                                                    1I
E  10000 +                                                                                                 1  1+
N        I                                                                                                 1 11I
C        I                                                                                                 1 11I
Y        I                                                                                                11 11I
    5000 +                                                                                        1       11111+
         I                                                          1 1            1       1  1  11 11 11 11111I
         I                                                 1    1 1 1 1    11   1 11    1 11011  01111 11111111I
         I                             0   0    0     1   000 0 000 0 01 101111101111 010 000101101111111111111I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0        .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  00000000000000000000000000000000000000000000000000111111111111111111111111111111111111111111111111111

     Predicted Probability is of Membership for 1.00
     The Cut Value is .50
     Symbols: 0 - .00
              1 - 1.00
     Each Symbol Represents 1250 Cases.
```

REFERENCES

Abdelmann, C. E., R., Elmore, R., Even, J., Kenyon, S., & Marshall, J. (1999). *When accountability knocks, will anyone answer?* (No. CPRE-RR-42). Philadelphia, PA: Consortium for Policy Research in Education.

Abedi, J. (2004). The no child left behind act and english language learners: Assessment and accountability issues. *Educational Researcher, 33*(1), 4-14.

Abedi, J., & Dietel, R. (2004). Challenges in the no child left behind act for english-language learners. *Phi Delta Kappan, 85*(10), 782.

Abedi, J., Hofstetter, C. H., & Lord, C. (2004). Assessment accommodations for english language learners: Implications for policy-based empirical research. *Review of Educational Research, 74*(1), 1-28.

Addonizio, M. F. (2009). X-efficiency and effective schools: A new look at old theories. *Journal of Education Finance, 35*(1), 1-25.

American Institute for Research (2010a). AM statistical software Retrieved August 15, 2010, from http://www.schooldata.org.

American Institute for Research (2010b). National longitudinal school-level state assessment score database (NLSLSASD) Retrieved August 15, 2010, from http://am.air.org/

Archer, J. (2005). States eyeing expense of hand-scored tests in light of NCLB rules. *Education Week, 24*(38), 1.

Bandeira de Mello, V., Blankenship, C., and McLaughlin, D.H. (2009). *Mapping state proficiency standards onto NAEP scales: 2005-2007 (NCES 2010-456).* National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Boote, D. N., & Beile, P. (2005). Scholars before researchers; on the centralitiy of the dissertation literature review in research preparation. *Educational Researcher, 34*(6), 3-15.

Borowski, J. W., & Sneed, M. (2006). Will NCLB improve or harm public education? *Harvard Educational Review, 76*(4), 457-460.

Brookover, W. B. (1985). Can we make schools effective for minority students? *Journal of Negro Education, 54*(3), 257-268.

Brooks-Gunn, J., & Duncan, G. (1997). The effects of poverty on children. *The Future of Children, 7*(2), 55-71.

Brown, L.I. (2001). *A meta-analysis of research on the influence of leadership on student outcomes.* Unpublished Ph.D., Virginia Polytechnic Institute and State University, VA.

California Department of Education (2007). *The addendum to the california standards tests (cst) technical report spring 2006 administration*. Sacremento, CA.

California Department of Education (2008). *California consolidated state application accountability workbook: For state grants under title ix, part c, section 9302 of the elementary and secondary education act (public law 107-110)*.

California Department of Education (2009a). *California Department of Education* Retrieved April 15, 2009, from http://www.cde.ca.gov/

California Department of Education (2009b). School accountability report card (sarc) Retrieved May, 15, 2009, from http://www.cde.ca.gov/ta/ac/sa/

California State Department of Education (1989). *Technical assistance manual for the California model school accountability report card*. Sacramento, CA.

Camara, W., & Schmidt, E. (1999). *Group differences in standarized testing and social stratification*. New York, NY.

Center on Education Policy (2005). *Hope but no miracle cures: Michigan's early restructuring lessons*.

Center on Education Policy (2007). *What now?: Lessons from Michigan about restructuring schools and next steps under NCLB*. Washington, D.C.: CEP.

Chubb, J. E., Linn, R., Haycock, K., & Wiener, R. (2005). Do we need to repair the monument? Debating the future of no child left behind. *Education Next, 5*(2), 8-19.

Coleman, J. S., & et al. (1966). *Equality of educational opportunity*.

Connell, N. (1996). *Getting of the list: School improvement in New York City*. New York, NY: Education Priorities Panel.

Cronin, J., Dahlin, M., Adkins, D., & Kingsbury, G. G. (2007). *The proficiency illusion*: Thomas R. Fordham Institute and Northwest Evaluation Association.

Darling-Hammond, L. (2007a). Evaluating "No child left behind". *The Nation*(May 21, 2007).

Darling-Hammond, L. (2007b). Race, inequality and educational accountability: The irony of "No child left behind". *Race, Ethnicity and Education, 10*(3), 245-260.

Daro, P., Stancavage, F., Ortega, M., DeStefano, L., & Linn, R. (2007). *Validity study of the naep mathematics assessment: Grades 4 and 8*. Washington, D. C: National Center for Education Statistics.

Davier, M. V., Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful *Ieri monograph series: Issues and methodologies in large-scale assessments*. Hamburg, Germany: IEA-ETS Research Institute.

Dee, T., & Jacob, B. (2009). *The impact of no child left behind on student achievement.* National Bureau of Economic Research working paper no. 15531.

Dufour, R. & Eaker, R. (1998). *Professional learning communities at work: Best practices for enhancing student achievement*. Bloomington, Indiana: Solution Tree.

Duncan, G. J., & Brooks-Gunn, J. (2000). Family povety, welfare reform, and child development. *Child Development, 71*(1), 188-196.

Edmonds, R. (1979a). *A discussion of the literature and issues related to effective schooling.*

Edmonds, R. (1979b). Effective schools for the urban poor. *Educational Leadership, 37*(1), 15-24.

Edmonds, R. (1982). Programs of effective schools:An overview: Universities, state agencies and school districts have established school improvement programs based on effective school research. *Educational Leadership, 40*(3), 4-11.

Education Policy Center (2000). *Strenghtening accountability standards in michigan*. Lansing, MI: Michigan State University.

Elmore, R. F. (2002). Unwarranted intrusion. *Education Next, 2*(1), 30-35.

Elmore, R. F. (2004). *School reform from the inside out; policy, practice and performance.* Cambridge, MA: Harvard Education Press.

Erpenbach, W., Forte-Fast, E., & Potts, A. (2003). *Statewide educational accountability under NCLB*.

Febey, K. S., & Louis, K. S. (2008). Political cultures in education at the state and local level: Views from three states. In B. Cooper, J. Cibulka & L. Fusarell (Eds.), *Handbook of education politics and policy* (pp. 52-69): Routledge.

Feldscher, K. (2011). Michigan board of education raises proficiency scores for meap and mme. *Ann Arbor.com,* (February 8, 2011). Retrieved from http://www.annarbor.com/news/michigan-board-of-education-raises-proficiency-scores-for-meap-and-mme/

Fiske, L. & Ladd, H.F. (2000). *When schools compete: A cautionary tale.* Wasthington, D. C.: Brookings Institute.

Fullan, M., Hill, P., & Crevola, C. (2006). *Breakthrough*. Thousand Oaks, CA Corwin Press.

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

Galvan, J. L. (2006). *Writing literature reviews: A guide for students of the social studies and behavioral sciences* (3rd ed.). Los Angeles: Pryczak Publishing.

Garson, G. David (2010). Logistic Regression from Statnotes: Topics in Logistic Regression Analysis. Retrieved 03/12/2010 from http://faculty.chass.ncsu.edu/garson/pa765/statnote.htm.

Gawlik, M. (2005). *Cutting loose: Autonomy and education in charter schools.* University of California, Berkeley, Berkeley, California.

Gawlik, M. (2007). Beyond the charter schoolhouse door: Teacher-perceived Autonomy. *Education and Urban Society, 39*(4), 524-533.

Gawlik, M. (forthcoming). Beyond the rhetoric: Charter school reform and accountability. *Journal of Educational Research*.

Greene, G. K., Huerta, L. A. & Richards, C. E. (2007). Getting real: A different perspective on the relationship between school resources and student outcomes. *Journal of Education Finance*, 33(1), 49-68.

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). Interpreting research on school resources and student achievement: A rejoinder to Hanusek. *Review of Educational Research, 66*(3), 411-416.

Grubb, W. N., Huerta, L. A. & Goe, L. (2006). Straw into gold, revenues into results: Spinning out the implications of the improved school finance. *Journal of Education Finance, 31*(4), 334-359.

Guilfoyle, C. (2006). NCLB: Is there life beyond testing? *Educational Leadership, 64*(3), 8-

Guo, G., & Harris, M. (2000). The mechanisms mediating the effects of poverty on children's intellectual development. *Demography, 37*(4), 431-447.

Guskey, T. R. (2007). Leadership in the age of accountability. *Educational Horizons, 86*(1), 29-34.

Hakuta, K., Butler, Y. G., & Witt, D. (2000). *How long does it take for english learners to attain proficiency?* Santa Barbara, CA: California University.

Hallinger, P., & Murphy, J. F. (1986). The social context of effective schools. *American Journal of Education*, 94(3), 328-355.

Hanusek, E. A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*(4), 45-62.

Harris, D. N. (2007). High-flying schools, student disadvantage, and the logic of NCLB. *American Journal of Education, 113*(3), 367-394.

Harris, D. N., & Herrington, C. D. (2006). Accountability, standards, and the growing achievement gap: Lessons from the past half-century. *American Journal of Education, 112*(2), 209-238.

Hattie, J. A. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* New York, NY: Routledge.

Haycock, K. (2006). No more invisible kids. *Educational Leadership, 64*(3), 38-42.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). An exchange: Part i: Does money matter? A meta-analysis of studies of the effect of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5-14.

Henchey, N. (2001). *Schools that make a difference: Final Report. Twleve Canadian secondary schools in low income settings*. Kelowna, BC, Canada: Society for the Advancement of Excellence in Education.

Hess, F. M. (2006). Accountability without angst? Public opinion and no child left behind. *Harvard Educational Review, 76*(4), 587-610.

Hess, F. M., & Finn, C. E. (2004). Inflating the life rafts of NCLB: Making public school choice and supplemental services work for students in troubled schools. *Phi Delta Kappan, 86* (1), 34-40 & 54-57.

Hess, F. M., & Petrilli, M. J. (2006). *No child left behind.* New York, NY: Peter Lang Publishing.

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied statistics for the behavioral sciences* (5th ed.). Boston: Houghton Mifflin Company.

Hoff, D. (2007). “Growth” pilot now open to all states. *Education Week, 27*(15), 1 & 20.

Hombro, C. M. (2003). NAEP and NCLB: Technical challenges and practical solutions. *Theory Into Practice, 42*(1), 59-65.

Hoxby, C. M. (2005). Inadequate yearly progress: Unlocking the secrets of NCLB. *Education Next, 2005*(3), 46-51.

Irons, E. J., & Harris, S. (2006). *The challenges of no child left behind: Understanding the issues of excellence, accountability, and choice*.

Jencks, C., & Phillips, M. (1998). *The black-white test score gap*.

Kane, T. J., Douglas, O. S., & Geppert, J. (2002). Randomly accountable. *Education Next, 2*(1), 57-68.

Koretz, D. (2008). A measured approach: Value-added models are a promising improvement, but no one measure can evaluate teacher performance. *American Educator, 32*(3), 18-27 & 39.

Kozol, J. (1991). *Savage inequalities: Children in american schools*. Ann Arbor, MI: Crown Publishing.

Ladd, H. (1996). *Holding schools accountable: Performance based reform in education.* Washington, D. C. : Brookings Institute.

Lamitina, D. (2006). *Quantifying the magnitude by which adequate yearly progress (AYP) achievement targets were missed.* Unpublished Ph.D., University of Nevada, Reno, United States -- Nevada.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational Researcher, 31*(1), 3-12.

Lee, J. (2004). Multiple facets of inequity in racial and ethnic achievement gaps. *Peabody Journal of Education, 79*(2), 51-73.

Levine, D. U., & Lezotte, L. W. (1995). *Effective schools research*.

Lewis, A. C. (2006). Clean up the test mess. *Phi Delta Kappan, 87*(9), 643-645.

Lezotte, L. (1994). The nexus of instructional leadership and effective schools. *School Administrator, 51*(6), 20-23.

Lezotte, L. W. (1992). Learn from effective schools. *Social Policy, 22*(3), 34-36.

Lezotte, L. W., & Jacoby, B. C. (1991). *Effective schools practices that work.*

Lin, A. C., & Harris, D. R. (2008). *The colors of poverty: Why racial and ethnic disparties exist.* New York, NY: Russell Sage Foundation.

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry*. Newbury Park, CA: Sage Publications.

Linn, R. L. (2003). Accountability: Responsibility and reasonable expectations. *Educational Researcher, 32*(7), 3-13.

Linn, R. L. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies, 40*(6), 699-711.

Linn, R. L., & Haug, C. (2002). Stability of school-building accountability scores and gains. *Educational Evaluation and Policy Analysis, 24*(1), 29-36.

Maleyko, G., & Gawlik, M. (2011). No Child Left Behind: what we know and what we need to know. *Education (Chula Vista, Calif.)*, 131(3), 600-24.

Marzano, R. J. (2003). *What works in schools: Translating research into action*. Alexandira, VA: Association for Supervision and Curriculum Development.

Marzano, R. J., Waters, T., & McNulty, B. A. (2005). *School leadership that works: From research to results*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mathis, J. (2003). No Child Left Behind costs and benefits. *Phi Delta Kappan, 84 (2), 679-686.*

Mathis, W. J. (2004a). NCLB and high-stakes accountability: A cure? Or a symptom of the disease? NCLB: Failed schools--or failed law? *Educational Horizons, 82*(2), 143-152.

Mathis, W. J. (2004b). No child left behind and the insomnia plague. *Clearing House, 78*(1), 30.

Mathis, W. J. (2005). The cost of implementing the federal no child left behind act: Different assumptions, different answers. *Peabody Journal of Education, 80*(2), 90-119.

Mathis, W. J. (2006). *The accuracy and effectiveness of adequate yearly progress: NCLB's school evaluation system*. Tempe, AZ: The University of Arizona.

Mays, L. (2008). The cultural divide of discourse: Understanding how english-language learners' primary discourse influences acquisition of literacy. *Reading Teacher, 61*(5), 415-418.

McGee, G. (1997). *What state tests test*. Deerfield, IL: American Educational Research Association.

McMillan, S. (2010, April 26). Ohio drops writing, social studies tests*. Sandusky Register*.

Meier, D., Kohn, A., Darling-Hammond, L., & Sizer, T. R. (2004). *Many children left behind: How the no child left behind act is damaging our children and our schools.* Boston, MA: Beacon Press.

Menken, K. (2008). *English language learners left behind: Standardized testing as language policy.* Clevedon, England: Multilingual Matters.

Meriam, S.B. (1998). *Qualitative Research and Applications in Education.* SanFrancisco: Jossey-Bass Publishers.

Michigan Department of Education (2005). *Michigan consolidated state application accountability workbook: For state grants under title ix, part c, section 9302 of the elementary and secondary education act (public law 107-110)*.

Michigan Department of Education (2007). *Meap: Michigan education assessment program: District and building coordinator handbook for the academic year 2007-08*. Lansing, MI.

Michigan Department of Education (2008a). *Michigan to pilot growth model program for determining adequate yearly progress (AYP).*

Michigan Department of Education (2008b). *What's new in the 2008 Michigan school report cards. Lansing, MI.*

Michigan Department of Education (2009a). *Michigan consolidated state application accountability workbook: For state grants under title ix, part c, section 9302 of the elementary and secondary education act (public law 107-110)*.

Michigan Department of Education (2009b). School report cards Retrieved March 26, 2009, from https://oeaa.state.mi.us/ayp/

Michigan State Board of Education (2003). *Michigan state board of education standards for accreditation education yes! – a yardstick for excellent schools* Lansing, MI.

Miles, K. & Darling-Hammond, L. (1998). Rethinking the allocation of teaching resources: Some lessons from high-performing schools. *Educational Evaluation and Policy Analysis* 20(1)**,** 9-29.

Miles, M.B. & Huberman, A. M. (1994). *Qualitative data analysis* (2nd ed.). London: Sage.

Mintrop, H. (2004). *Schools on probation*. New York, NY: Teachers College Press.

Murnane, R. J. (2007). Improving the education of children living in poverty. *Future of Children, 17*(2), 161-182.

Nelson, S. W., McGhee, M. W., Meno, L. R., & Slater, C. L. (2007). Fullfilling the promise of educational accountability. *Phi Delta Kappan 88*(9), 702-709.

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis,* 26(3), 237-257.

Odden, A., & Busch, C. (1998). *Financing schools for high performance*. San Francisco, CA: John Wiley & Sons, Inc. .

Odden, A., & Clune, W. H. (1998). School finance systems: Aging structures in need of revolution. *Educational Evaluation and Policy Analsysis, 20*(3), 157-177.

Olson, L., & Jacobson, L. (2006). Analysis finds minority NCLB scores widely excluded. *Education Week, 25*(33), 5.

Patton, M. Q. (1990). *Qualitative evaluation methods* (2nd ed.). Thousand Oaks, CA: Sage.

Peterson, P. E., & West, M. R. (2003). No child left behind?: The politics and practice of school accountability. Washington, D.C.: Brookings Institution Press.

Peterson, P. E., & West, M. R. (2006). Is your child's school effective?: Don't rely on NCLB to tell you. *Education Next, 2006*(4), 76-80.

Phelps, J. L., & Addonizio, M. F. (2006). How much do schools and districts matter? A production function approach to school accountability. *Educational Considerations, 33*(2), 51-62.

Popham, W. J. (2005a). AYP wriggle room running out. *Educational Leadership, 63*(1), 85-87.

Popham, W. J. (2005b). Can growth ever be beside the point? *Educational Leadership, 63*(3), 83-84.

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

Pray, L. C., & MacSwan, J. (2002). *Different question, same answer: How long does it take for english learning to acquire proficiency?* . Paper presented at the American Educational Research Association.

Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96 (1), 3-14.

Pope, C., Ziebland, S. & Mays, N. (2000). Qualitative research in health care: Analysing qualitative data. *British Medical Journal, 320,* 114-116.

Public Schools of North Caroliina State Board of Education (2006). *Abcs 2006 accountability report background packet* Raleigh, NC.

Ravitch, D. (2010). *The Death and Life of the Great American School System: How Testing and Choice Are Undermining Education.* New York, NY: Basic Books.

Reeves, D. B. (2004). *Accountability for learning: How teachers and school leaders can take charge*. Alexandria, VA: Association for Supervison and Curriculum Development.

Reeves, D. B. (2006). *The learning leader: How to focus school improvement for better results.* Alexandria, VA: Association for Supervision and Curriculum Development.

Rose, L. C. (2004). No child left behind: The mathematics of guaranteed failure. NCLB: Failed schools--or failed law? *Educational Horizons, 82*(2), 121-130.

Rothstein, R. (2001). Investing in family capital. *American School Board Journal, 188*(2), 18-21.

Rothstein, R. (2004). *Class and schools*. New York, NY: Teacher's College Press.

Rothstein, R. (2008). Whose problem is poverty? *Educational Leadership, 65*(7), 8-13.

Rudestam, K. E., & Newton, R. R. (2007). *Surviving your dissertation: Comprehensive guide to content and process* (3rd ed.). Los Angeles, CA: Sage Publications.

Sanders, W.L. (2000). Value-added assessment from stduent achievement data: Opportunities hurdles. *Journal of Personnel Evaluation in Education*, 14(4), 329-339.

Sanders, W.L., & Rivers, J.C. (1996). Cumulative and residual effects of teachers on future student aceademic achievement: University of Tennesse Value-Added Research and Assessment Center.

Schoen, L., & Fusarelli, L. D. (2008). Innovation, NCLB, and the fear factor: The challenge of leading 21st century schools in an era of accountability *Education Next, 4*(2), 23-29.

Shaul, M. S., & Ganson, H. C. (2005). The no child left behind act of 2001: The federal government's role in strengthening accountability for student performance. *Review of Research in Education, 29*, 151-154.

Sloan, W. (2010). Coming to terms with common core standards. *Information Brief from ASCD, 16*(4).

Spradley, J. P. (1980). *Participant observation*. Forth Worth, TX: Rinehart and Winston, Inc.

Springer, M. G. (2008). Accountability incentives: Do schools practice educational triage? *Education Next, 8*(1), 74-79.

State of North Carolina (2008). *North carolina consolidated state accountability workbook application: For state grants under title ix, part c, section 9302 of the elementary and secondary education act (public law 107-110)*.

Sunderman, G. L., Kim, J. S., & Orfield, G. (2005). *NCLB meets school realities: Lessons from the field*. Thousand Oaks, CA: Corwin Press.

Sunderman, G. L., Tracey, C. A., Kim, J., & Orfield, G. (2004). *Listening to teachers: Classroom realities and no child left behind*. Cambridge, MA: Harvard University.

Taylor, D. L. (2005). Class and schools: Using social, economic, and educational reform to close the black?White achievement gap. *School Effectiveness & School Improvement, 16*(4), 445-449.

Teddlie, C., & Springfield, S. (1993). *Schools make a difference: Lesons learned from a 10 year study of school effects*. New York, NY: Teachers College Press.

Texas Education Agency. History of public education in texas, from http://ritter.tea.state.tx.us/tea/historyoverview.html

Texas Education Agency (2004). *Chapter 15: Validity*. Austin, TX.

Texas Education Agency (2006). *Texas school accountability system*. Austin, TX.

Texas Education Agency (2009). *Texas consolidated state application accountability workbook: For state grants under title ix, part c, section 9302 of the elementary and secondary education act (public law 107-110)*.

The Connecticut Commission on Children (2004). *Children and the long term effects of poverty*.

The National Commission on Education Excellence (1983). *A nation at risk: The imperative for educational reform*.

Tonso, K. (1997). Trustworthiness sample. Wayne State University.

Tyack, D., & Cuban, L. (1995). *Tinkering toward utopia: A century of public school reform*. Cambridge, MA: Harvard University Press.

U.S. Department of Education, National Center for Educational Statistics (2001). NAEP technical documentation Retrieved November 27, 2010, from http://nces.ed.gov/nationsreportcard/pubs/main1998/2001509.asp

U.S. Department of Education, National Center for Education Statistics. (2003). Naep validity studies: An agenda for naep validity research. Washington, D. C.

U.S. Department of Education, National Center for Educational Statistics (2005a). The nation's report card mathematics 2005: National assessment of educational progress.

U.S. Department of Education, National Center for Educational Statistics (2005b). The nation's report card reading 2005: National assessment of educational progress.

U.S. Department of Education, National Center for Educational Statistics (2007a). The nation's report card mathematics 2007: National assessment of educational progress at grades 4 and 8.

U.S. Department of Education, National Center for Educational Statistics (2007b). The nation's report card reading 2007: National assessment of educational progress at grades 4 and 8.

U.S. Department of Education, National Center for Educational Statistics (2008). Number and percentage of children served under individuals with disabilities education act, part b, by age group and state or jurisdiction: Selected years, 1990–91 through 2007–08. *Digest of Education Statistics.* Retrieved June 25, 2011 from http://nces.ed.gov/programs/digest/d09/tables/dt09_052.asp

U.S. Department of Education, National Center for Educational Statistics (2009). The nation's report card: National assessment of educational progress.

U.S. Department of Education (2002). No Child Left Behind Act of 2001, P.L. 107-110.

U.S. Department of Education (2009a). *Secretary spellings approves additional growth model pilots for 2008-2009 school year*. from http://www.ed.gov/news/pressreleases/2009/01/01082009a.html.

U.S. Department of Education (2009b). *Secretary spellings approves additional growth model pilots for 2008-2009 school year*

Verges, J. (2011). S.D. Schools back out of No Child Left Behind. *Argusleader.com*. Retrieved July 12, 2011 from http://www.argusleader.com/article/20110630/NEWS/106300302/S-D-schools-back-out-No-Child-law

Wiley, E. W., Mathis, W. J., & Garcia, D. R. (2005). *The impact of the adequate yearly progress requirement of the federal "No child left behind" Act on schools in the great lakes region.* Tempe, Arizona: Arizona State University.

Wolcott, H. (1990). *Writing up qualitative research*. Newbury Park, CA: Sage Publications.

Zehr, M. A. (2008). Hurdles remain high for english-learners. *Education Week, 27*(39), 1.

Zehr, M. A. (2009). Study finds NCLB law lifted math scores. *Education Week, 29*(13), 8-9.

ABSTRACT

THE IMPACT OF NO CHILD LEFT BEHIND (NCLB) ON SCHOOL ACHIEVEMENT AND ACCOUNTABILITY

by

GLENN MALEYKO

December 2011

Advisor: Dr. Martyza Gawlik

Major: Educational Leadership and Policy Studies

Degree: Doctor of Philosophy

Spurred by the *No Child Left Behind Act* of 2001, virtually every educational reform program now includes an accountability component that requires sound data collection and reporting (NCLB, 2002, section 101). Drawing from empirically based and theoretical literature in the field, this dissertation examines Adequate Yearly Progress (AYP) and the accountability provisions found in Title One of the No Child Left Behind (NCLB) reform. States have the ability to statistically manipulate their AYP implementation, which may give a false impression to the public that AYP is a consistent measure of school effectiveness across the country. The literature review (which was previously published, Maleyko & Gawlik, 2011) identifies the measurement concerns with the implementation of AYP, the benefits of AYP, the unintended consequences, along with the complexities involved with establishing school accountability and the effective and ineffective provisions of the NCLB reform.

One of the most important parts of the NCLB reform is the set of accountability standards for schools, school districts, and states because it is the mechanism in which the framers of NCLB believe that school improvement will occur on a national level. This study examines a portion of the NCLB accountability system in order to measure the impact that the legislation is having on school reform efforts. It further addresses the problem of consistency with the implementation of AYP in different states by comparing the impact that the reform is having on a sample of four states, 1) California, 2) Michigan, 3) North Carolina, and 4) Texas in relationship to the 2005 and 2007 National Assessment of Educational Progress (NAEP) in reading and mathematics. School level comparisons of the NAEP and state accountability assessments are used to measure differences within the sample. The NAEP data was accessed through a restricted level application with the National Center for Educational Statistics(NCES). The NAEP is a complicated measurement tool, this study is the first that could be found which calculates a school level proficiency score in order to compare results in the sample states. Finally, a qualitative methodology was implemented in order to interview teachers and principals to measure the type of responses that schools were implementing as a result of NCLB and AYP.

The findings from this study show that the use of the NAEP assessment is an effective technique to analyze the consistency of AYP among the states. The findings also show that the standards implemented among the sample states in this study are much different. The quantitative data (mathematics and reading assessments in grades 4 and 8) from the years 2005 and 2007 showed that the Michigan AYP standards and state accountability assessment proficiency levels have a close relationship with the basic level NAEP scale proficiency standards. The Texas state AYP and state accountability assessment provided for a very low correlation with the NAEP at the proficient level NAEP scale and basic level proficiency standard. The results showed that North Carolina had the closest relationship with the proficient level NAEP scale and that the standards in North Carolina and California were more rigorous in comparison to Michigan and especially Texas which provided for the lowest rigor with their AYP standard in this study. The findings further show that the Economically Disadvantaged (ED) status variable was the greatest predictor of success as measured by NCLB and AYP. AYP might be doing a better job at measuring ED status vs. actual student achievement. The findings from this research study did not indicate that AYP was a measure of the minority status and racial subgroups as the quantitative datasets were inconsistent with the impact that different subgroups had on state accountability assessment results and AYP between the four sample states.

The findings in this study further indicate that the current accountability provisions in NCLB have not been effective in evaluating school performance. The results from the qualitative data show that there were some benefits of AYP including the increased sense of urgency with data analysis. However, it did not necessarily lead to increased achievement among the schools or the implementation of effective school improvement plans and/or classroom instruction. Finally, this paper concludes with recommended areas of research for policymakers and educators alike who are interested in sustainable reform.

GLENN MALEYKO AUTOBIOGRAPHICAL STATEMENT

I have always been interested in a career in scholarly research. However, my professional ambitions began in Dearborn in 1995 as I started a career in public education. My excitement for research was further enhanced when I started a career in school Administration as I was appointed as the Assistant Principal of Salina in 2002. I became the Principal of the Salina Intermediate School on December 1st, 2003. After serving the Salina Community for twelve years, I was appointed as the Principal at DuVall Elementary in July, 2009 where I served two years. I became the Director of Human Resources for Dearborn Public Schools in July of 2011.

During my years as an educational professional at both DuVall Elementary and Salina Intermediate, I was involved with hosting over 1400 teachers, parents, and administrators to provide them with training on how to implement integrate technology across the curriculum, implement a comprehensive interventions program, integrate the special education co-teaching model, and initiate a comprehensive literacy across the curriculum model.

I have given presentations at multiple conferences over the past 16 years at both the state and national level. Some of those conferences include ICLE, ACTFL, ASCD, MACUL, NMSA, MI-ASCD, the Arkansas Literacy Conference and NCA. I was also named a model school principal by Dr. Bill Dagett's International Center for Leadership in Education(ICLE) in 2009.

I have published at the scholarly level along with educational practitioner publications and bulletins. I have used my knowledge and expertise with my doctoral studies to position myself for a potential career in research and academia while at the same time advancing my professional career in public education.