DISTORTIONS IN STATE LEVEL PERFORMANCE OUTCOMES ON HIGH STAKES ASSESSMENTS

By

Joseph E. Hornback

B.S., University of Kansas, 1995

M.A., University of Kansas, 1996

M.A., University of Kansas, 2000

Submitted to the Department of Educational Leadership and Policy Studies and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Education.

Committee:

____

Chairperson

____

____

____

____

____

The Dissertation Committee for Joseph E. Hornback certifies this is the approved version of the following dissertation:

DISTORTIONS IN STATE LEVEL PERFORMANCE OUTCOMES ON HIGH STAKES ASSESSMENTS

Committee:

____
Chairperson*

____

____

____

____

Date approved: ________________________

ABSTRACT

This dissertation addresses two research questions:

1. Do states misrepresent their progress on their own state assessments?
2. If states do distort their progress, are their predictors to suggest why this distortion occurs?

The first research question requires that distortion be defined. For the purposes of this dissertation I calculated the growth from 2003 to 2005, 2005 to 2007, and 2007 to 2009 on each state's individual state assessment and the NAEP. To calculate the growth I used a modified growth equation that subtracts the two scores and divides that by the maximum score on that test in that year, from the first score. This calculation produces a Practical Normed Growth (PNG) for the state assessment as well as the NAEP. To determine the distortion index I subtract the NAEP's PNG from the state assessment's PNG. A positive distortion index indicates the state assessment's growth was greater than the NAEP's growth and the state distorted their progress. A negative distortion index indicates the NAEP's PNG was greater than the state assessment's PNG and the state did not misrepresent their progress. This analysis was done on the elementary reading assessment. This assessment includes three growth periods to compare, creates three observations of the 50 states, or 150 data points possible for distortion.

The first research question, do states distort their progress? The answer is yes. On the elementary reading assessments the states had a positive distortion index 76 times out of a possible 150, or 51%. The observed distortions came from three basic models. First, the state assessment scores went up, but the NAEP scores went down or stayed the same. Second, the state assessment scores stayed the same, but the NAEP scores went down. Third, the state assessment scores went down, but the NAEP scores went down more. In each of these possible scenarios the states have misrepresented the education progress of their state to their stakeholders. In the first scenario, if the scores on the state assessments go up while the NAEP assessment scores go down indicates a narrowed curriculum and an overemphasis on the state assessment. In the second scenario, if the scores on the state assessment stayed the same while the NAEP scores dropped indicates a less effective focus on the state assessment, at the expense of the NAEP. The third scenario, both the state assessment and NAEP scores fall, with the NAEP scores falling faster indicates a state struggling to do anything well.

The second research question, are there predictors to suggest why this distortion occurs? The answer is yes. On the elementary reading assessment comparison, two predictors were statistically significant in the final model, state population and African-American status. The coefficients of each of these predictors indicate that states with lower populations and higher numbers of African-Americans distort their progress more on the elementary reading assessments.

ACKNOWLEDGEMENTS

As a lifelong educator I have never viewed the process of educating a child as either a sprint or a marathon, but rather a relay. To this end, I have many people to thank. First, and most important, is my wife Melissa. Her support and belief in me has made all of this worth the time and effort required to do it well. Next I would like to thank my dissertation chair, and mentor, Dr. Argun Saatcioglu. His high expectations and incredible intellect have motivated me to reach higher than I thought possible. In addition, I would like to thank the faculty members of the University of Kansas School of Education I have had the privilege of working within my current program and as part of the three degrees I have earned previously. Specifically, I would like to thank Dr. John Rury, Dr. Mickey Imber, and Dr. Jerry Bailey, each of whom I admire greatly and value the time and effort they provided to my development as a person, student, and as a professional. Finally, I would like to thank three very special people; Dr. Perry Perkins, Mr. John Hilton, and Mr. Rick Moulin. Dr. Perkins has been a constant source of encouragement and guidance from the day I met him and is a testament to the power of leadership in a school and district. Mr. Hilton and Mr. Moulin, my two assistant principals, have made it possible for me to earn this degree. I appreciate their willingness to assist with my professional responsibilities and provide encouragement along the way more than I could ever express in words.

DEDICATION

This dissertation and degree was motivated by and dedicated to my four children; Grace, Helen, Noah, and John. I hope my effort and dedication to this degree provide an example for you to follow in your lives, illustrating that grit, determination, and a stubbornness to believe in yourself will compensate for any shortcomings in talent or intelligence.

TABLE OF CONTENTS

Chapter 1

Introduction

1.1 Purpose for this Study

In 1965 Lyndon Johnson declared a "War on Poverty" which began the high stakes testing era. By providing schools with additional funds to combat the effects of poverty, and the states accepting those "title" funds, the groundwork was established for a culture built on standardized assessments. In 1969 the reauthorization The Elementary and Secondary Education Act (ESEA) established the National Assessment of Educational Progress (NAEP) to assess the impact of the title funds distributed in 1965, starting a snowball that would culminate 47 years later in the 2002 reauthorization of ESEA, better known as No Child Left Behind.

In 1976, social science researcher Donald Campbell developed what is known as Campbell's Law which states, "the more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it was intended to monitor" (Nichols & Berliner, 2007). In President Johnson's "War on Poverty" the social decision was made to provide funding for students that need additional resources, coupled with the No Child Left Behind policy of assigning the quantitative measure of test scores to remain eligible for those funds, has led to resource dependency, institutionalism, and a high stakes culture. School districts are resource dependent and controlled externally, making the ceremonial adoption of policies and procedures necessary to maintain their legitimacy and funding. Deviations from this new norm threaten their existence and further raise the importance of the state assessments perpetuating a high stakes culture that links resources to results by any means necessary, which has manifested itself in the need for some schools, districts, and states to distort their progress.

In this dissertation I will examine the distortions that occurs at the state level as a result of a high stakes testing culture. Distortions, for the purposes of this dissertation, will be defined as the magnitude of the discrepancy between the reported levels of growth the state reports on its own high stakes state assessment, compared to the rate of growth on a low stakes assessment, the National Assessment of Educational Progress, NAEP.

This study seeks to provide some predictors to suggest possible explanations to why these distortions occur using census data obtained through the Integrated Public Use Microdata Series (IPUMS). Other studies have compared state assessment scores and NAEP scores in snapshots to illustrate the rigor of standards used for the state assessment (McLaughlin, Bandeira de Mello, Blankenship, Chaney, Esra, Hikawa, Rojas, William, and Wolman, 2008) and to highlight the achievement gap that occurs between groups of students (Lee and Reeves, 2012), but none have examined the inflated levels of proficiency reported over time with the integration of census data to try to tell the story as to why it occurs. This pattern of inflated reported gains in educational progress, intended to boost confidence in the process and protect resources, could hinder real progress from occurring. The accurate collection and reporting of achievement data influences the decisions policymakers need to make to improve quality of education in this country. When that data is not reflective of the current state of education it potentially creates a false sense of security to allow existing practices to continue. The purpose of using data in the decision making process is to allow for informed judgment by those leaders charged with making policies. When the data is manipulated or misrepresented, it inhibits true progress from occurring.

1.2 Research Questions

This dissertation addresses two research questions:

1. Do states misrepresent their progress on their own state assessments?
2. If states do distort their progress, are their predictors to suggest why this distortion occurs?

The first research question requires that distortion be defined. For the purposes of this dissertation I calculated the growth from 2003 to 2005, 2005 to 2007, and 2007 to 2009 on each state's individual state assessment and the National Assessment of Educational Progress, NAEP. To calculate the growth I used a modified growth equation that subtracts the two scores and divides that by the maximum score on that test in that year, from the first score. This calculation produces a Practical Normed Growth (PNG) for the state assessment as well as the NAEP. To determine the distortion index I subtract the NAEP's PNG from the state assessment's PNG. A positive distortion index indicates the state assessment's growth was greater than the NAEP's growth and the state cheated. A negative distortions index indicates the NAEP's PNG was greater than the state assessment's PNG and the state did not distort their progress. This analysis was done on the elementary reading assessment. This assessments had three growth periods to compare, creates three observations of the 50 states, or 150 data points possible for cheating.

The second research question of this dissertation is to suggest some predictors to suggest why these distortions occur. Using state level data from the United States Census and IPUMS, a fixed effects time variant panel regression model was created to determine if any of the following eight variables were statistically significant when compared to the calculated distortion index; per pupil expenditure spent on education, state population, manufacturing level of the state, parent's level of education, poverty status of the home, father's location in the home, African-American status of the state, and born in the United States to suggest possible relationships.

This study will examine distortion indexes for elementary reading from 2003 to 2009 to answer the two research questions. The amount of distortion that occurs on each test will be examined, and well as the predictors that proved to be statistically significant.

1.3 Significance of the Study

This study is important because it examines state level distortions in a new way. Other research has been done with similar data to look at the rigor of state standards and achievement gap analysis, but none of that research has addressed why states have allowed the standards to become less rigorous or why groups of students achieve at different levels. Much of the research focuses on what is happening, this study seeks to suggest why it is happening. By coupling the testing data with predictors outside of education it allows for a sociological approach to a societal issue. The issues we have in education cannot be fixed by a different system of education. Until these issues are examined as symptoms of our society, rather than a stand-alone system that operates in isolation, we are doomed to the same frustrations that have plagued educators since the dawn of public schooling.

Chapter 2

Review of Literature

2.1 Research Question

1. Do states misrepresent their progress on their own state assessments?
2. If states do distort their progress, are their predictors to suggest why this distortion occurs?

This dissertation compares each state's individual state assessment to the NAEP to determine a distortion index. The states with similar distortion indexes will be grouped together and census data will be used to explore various state characteristics to suggest possible predictors these distortion levels. To this end, this dissertation relies on a number of key literatures; cheating in education and cheating in formal organizations. The organizational sociology theories of institutionalism and resource dependency provide a framework to review the applicable research to understand why formal organizations; including schools, school districts, and states would ceremonially adopt polies and procedures while distorting their progress to maintain their legitimacy. The study also draws on the discussion of motivations to distort progress in a high stakes versus a low stakes environment. This literature provides a disciplinary framework to address the discrepancies in student performance explored in this dissertation.

The fundamental contribution of this study is to education policy literature on the potential downside of high stakes policy. Although the literature is limited, there is growing interest around policy scholars on the disadvantages of high stakes testing, which is also reviewed below.

2.2 - Distortions in Organizations – Institutionalism and Resource Dependency

Cheating occurs in all formal organizations. Bureaucracies enabled on quantifiable outcomes lead to risk evasion and cheating (Meyer & Rowan, 1977). Any institution, business, or organization has one primary goal, to survive. According to Meyer and Rowan an organization will avoid risks and cheat if needed to survive. Institutions adopt "rationalized myths" to ceremonially adopt practices to appear legitimate in hopes of achieving similar results with competing schools. This ceremonial compliance at the state level manifests itself in the creation of state standards, state assessments, and cut scores that are necessary to meet the timeline set by the federal government. The spirit of complete competency is lost in the loopholes created by confidence intervals and safe harbor. Districts are able to make AYP by either meeting the state benchmark for proficiency or by reducing their number of non-proficient students by ten percent, safe harbor. Mathematically a district could make AYP every year through safe harbor and never reach full proficiency. In addition, states modify the pool of students it uses to calculate proficiency (Haney, Madaus, and Wheelock, 2003) and allow schools to triage students into special education services to boost proficiency levels (Booher-Jennings, 2005). States ceremonially adopt these practices to maintain their legitimacy and appear compliant to the federal mandates. States avoid the risk of failing by making the test easy enough so that most students can pass them, and inflate the progress of their schools to ensure a higher rate of success that is actually occurring (Haney, Madaus, and Wheelock, 2003). The combination of risk evasion and the ceremonial adoption of policies and practices have led to an increase in cheating in schools.

In DiMaggio & Powell (1983) the authors discuss how institutions become more isomorphic through three processes; coercive, mimetic, and normative. Schools and districts will coerce change from within by using formal and informal directives, they will mimic other districts in hopes of achieving the same results, and they will use professional norms to ensure compliance to an industry acceptable standard of performance. Schools have done this with testing practices and interventions to meet the demands of NCLB. Once a testing practice became the preferred professional practice within a school, any school that was not doing it was exposed to one or all of the three conversion methods that DiMaggio & Powell discuss. It became a directive from the central office or building administration (coercive), it was sold on the benefits it provided to the schools it was really working at (mimetic), and you could be considered "unprofessional" if you did not think they were a good idea (normative). States have followed a similar model. States were directed by the federal government to comply or lose title funds (coercive), it was packaged as the solution to closing the achievement gap (mimetic), and the use a foundational education principal that all students can learn implied in the name No Child Left Behind provides ammunition to proponents when the policy is questioned (normative). States quickly institutionalized.

In Tolbert & Zucker (1983) the authors discuss how institutions will adopt policies or programs to the extent they become institutionalized. The adoption curve of includes early adopters and pioneers accepting every new idea and as each practice becomes institutionalized the rate of adoption speeds up until there are only a few left who have not adopted the latest innovation. This was certainly true with increased focus on reading and math that has manifested itself in the narrowing of curriculum to eliminate opportunities for students in the arts and science (Harris, 2012). When early adopters began to eliminate time for science and art in elementary schools in favor or additional time for reading and math it was initially frowned upon by neighboring schools. However, when the school was able to produce increased rates of proficiency, the practice became accepted, and later institutionalized. As this practice became more institutionalized, schools districts become more isomorphic, ensuring all would adopt. At the state level, the rate of adoption was motivated by funding, which leads to the second organizational theory in play, resource dependency.

According to Pfeffer and Salancik (1978) the purpose of an organization is to achieve the same goals shared by its members. The organization's legitimacy is determined by the evaluation of society, many of whom have alternating and competing criteria. As a result, the participants who provide the most resources have the most input, with the most control over the organization. The internal demands of the organization coupled with the external controls create an obstacle to efficiency and effectiveness. The organization becomes controlled externally based on the actor providing the most resources, with all other actors' input proportional to the amount of resources provided. The dependence on these resources drives the direction of the organization. The dependence of public schools on public funds is evident in the amount of control afforded to the state and federal governments at the local level in our schools. The designation of additional funds to meet the higher needs of disadvantaged students, title funds, by the federal government provides both a carrot and a stick to local school districts to comply with mandates set forth or risk losing the resources needed to achieve their internal mission. These resources drive the policies, procedures, and practices of public school teachers, administrators, and school boards to comply with external mandates and as a result be controlled from the outside.

The institutionalism and resource dependency of a state can lead to goal displacement. The ceremonial adoption of policies and procedures, the input afforded to external actors based on resource contribution, coupled with the pressure to maintain legitimacy inherent in high stakes testing leads states away from their stated mission of educating all students to gaming a system to ensure organizational survival. According to Suchman (1995) an organization will conform to gain legitimacy, modify to maintain legitimacy, and normalize if it loses legitimacy. The pressure on states to maintain their legitimacy on a yearly basis has led them to be in a constant state of modification. The goal of No Child Left Behind is complete proficiency in reading and math, however the high stakes associated with achieving that worthy goal has places schools, districts, and states into survival mode, displacing their goal from educating to surviving.

2.3 - Distortions in Education

Since the passage of No Child Left Behind in 2002 there have been multiple incidents of schools cheating to improve the scores on state assessment tests used to measure their Annual Yearly Progress (AYP). Between 2009 and 2012 there have been documented incidents of schools cheating on standardized tests in Pueblo, Washington DC, Houston, Atlanta, Ohio, and Virginia. School district officials have erased wrong answers and substituted the right ones, as they did in Pueblo, Washington DC, Houston, and Atlanta (Auge' & Simpson, 2012). Teachers have created study guide questions that were identical to the test, as they did in Ohio (Toppo, Amoms, Gillum, & Upton, 2011) and they have coached students during the test to help them avoid picking wrong answers, as they did in Houston and Norfolk, Virginia (Gabriel, 2010). Researchers have examined fluctuations and patterns in student test scores to identify teachers suspected of cheating (Jacob & Levitt, 2003) all in an effort to maintain the integrity of assessments given to students to measure their learning.

In blatant cases, such as the ones outlined above, it easy to identify where the cheating has occurred. Other schools have been able to meet the high demands set forth by NCLB by making more subtle adjustments in their organization to improve test scores. Schools have narrowed their curriculum, fine-tuned their testing practices, and modified their testing pools to give their school the best advantage possible to meet the benchmarks set by the state to make AYP. While the gaming of state assessments is not technically cheating, the practices employed by schools and districts to make AYP have been widely criticized. These adjustments and modifications have caused schools to deviate from their stated educational objectives to focus on proficiency levels in reading and math.

States have begun to make corrective measures to make it more difficult for schools and districts to cheat. Some states have abandoned the paper and pencil option, others require multiple proctors be present when students are testing, and most require training take place prior to administering the state assessments. However, the reported cases of schools and districts cheating between 2002 and 2012 have drawn negative attention to education at the district and school level.

Schools and districts are not the only ones cheating. States cheat too. As a part of the local control in No Child Left Behind the states are given the authority to determine their own state standards, write and administer their own assessment, determine the cut scores for proficiency, and set the benchmarks schools need to hit starting in 2006 to prepare all students to be proficient by 2014. On other measures used to calculate Annual Yearly Progress (AYP), such as graduation rate, states have had similar flexibility to determine what constituted acceptable progress. According to Carey (2006), for every measure, a significant number of states inflated their progress to appear as though they were advancing. Two examples are graduation rates and proficiency levels on state assessments.

In Hall (2005) the researcher examines how states report their graduation rates. Some states follow a cohort model, tracking students from the day they walk into high school, while others simply look at how many students start their senior year and how many finish. The variation in how graduation rate is calculated make it difficult to compare states, and some states have changed how they calculate graduation rate from year to year, making it difficult to even compare the state against itself. In addition to the calculation of graduation rate, states also have the flexibility to determine what acceptable progress is each year in order for a school to make AYP. The study reports thirty-one states accept any progress at all for a school to make AYP while four different states require at least .1% increase to make AYP, and two more states just require a school to report it to make AYP, not requiring progress. The way a state calculates graduation rates, and the progress needed to make AYP, are both determined at the state level and subject to inflation and the distortion of progress.

In Haney, Madaus, and Wheelock (2003) the researchers discuss how the Department of Education in Massachusetts inflated the proficiency levels it reported.

The state calculated its proficiency rates in reading and math by taking the number of students that passed the test and divided it by the number of students that graduated. This calculation would have some value if the goal was to determine the proficiency level of graduates, but when reporting the proficiency levels for No Child Left Behind, the researchers felt it was not an accurate picture of the proficiency level of all students in Massachusetts. In addition to creative proficiency calculations, states also have the latitude to set the cut scores necessary to be considered proficient. In Kansas the cut score for reading is 68% and for math it is 50%. In science, although not a component of AYP it is still assessed and assigned performance categories, the cut score is 40% to be considered proficient. The arbitrary assignment of cut scores, even within the same state, create opportunities for the state to inflate the number of students it determines is proficient by setting the score needed to be considered proficient so low that a higher percentage of students pass the test. In addition, the lack of consistency within the same state damages the credibility of the measure and calls into question why some tests would require a lower score in order to classified as proficient.

In this dissertation I will examine if states distort their progress in a high stakes testing culture. Distortion, for the purposes of this dissertation, will be defined as the magnitude of the discrepancy between the reported levels of growth the state reports on its own high stakes state assessment, compared to the rate of growth on a low stakes assessment, the NAEP. Schools and districts are not the only entity under pressure as a result of the NCLB legislation. State departments of education are also held accountable for the performance of the schools in their state. The state writes the assessments, sets the cut scores necessary to be proficient in reading and math, and must report the level of growth to the federal government in order to receive the title funds. As a result, some states have demonstrated remarkable growth on their state assessment rates of proficiency, but little or no growth on their NAEP assessment. Based on the definition of distortion provided above, those states have distorted their educational progress, and I will suggest some factors to help understand why this occurred.

Distortion is an emotional word. It is tied to a lack of character or integrity, and viewed as a negative personality flaw. However, distortion as defined above, is a rational response in a high stakes culture. States distort their results because of the over-emphasis on the results tied to testing. Donald Campbell, in 1975 developed Campbell's Law stating, "the more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it was intended to monitor" (Nichols & Berliner, 2007). When resources and legitimacy became tied to state assessment scores it became more likely participants in the system would distort their results. When states were assigned the task of ensuring that all students would pass a reading and math test, or risk being publicly scolded and have to forfeit part of their funding, it is a rational response that states would game the system, lower the bar, and triage the groups to maintain their funding and legitimacy.

Prior to 2002 most states had a proficiency test in reading and math that was given at the elementary level, middle school level, and in high school. Since 1969 schools have administered the National Assessment of Education Progress, NAEP, and the Iowa Test of Basic Skills, (ITBS) with few reported cases of cheating. Schools and districts wanted to do well on these tests, and the data was examined, but no resources were connected to the results so schools cheated less. When resources became tied to results, the amount of distortion increased. The distortion that goes on in schools, districts, and states can be explained within two theories of organizational sociology, institutionalism and resource dependency.

2.4 - What Makes High Stakes Testing High Stakes

Testing became high stakes when resources became tied to results. Prior to NCLB, the Elementary and Secondary Education Act, ESEA, provided addition funds to schools with higher percentages of students it determined were at-risk of failing in school. This legislation was initiated in 1965 as part of President Lyndon Johnson's "War on Poverty." These funds, referred to as title funds, provided money to schools for instructional supplies, professional development, resources to support educational programs, and to promote parental involvement. This legislation must be renewed every five years and was the first to establish federal funding available to all schools.

In 1983, in the landmark report, "A Nation at Risk" reported that "average achievement of most high school students on most standardized tests is now lower than 26 years ago" (Dolezalek, 2009). This report led to a 1989 governor's conference at the urging of President George HW Bush to establish higher standards for schools resulting in "America 2000" that called for "standards for what children should know and be able to do in five core subjects: English, mathematics, science, history, and geography" (Dolezalek, 2009). In addition, it called for testing in grades four, eight, and twelve with a recommendation for a voucher system to allow students additional school choice by allowing them to be used to attend private schools.

In 1994 congress renewed the Elementary and Secondary Education Act (ESEA) and encouraged states to meet high standards by establishing Goals 2000. Included in the eight goals of Goals 2000 is goal number three, "grades 4, 8, 12 competency tests in English, math, science, history, and geography." In 2002, President George W. Bush signs the reauthorization of ESEA with overwhelming bipartisan support, establishing the framework of No Child Left Behind calling for reading and math proficiency testing in grades three, four, five, six, seven, eight, and once in high school by the 2005-06 school year. In addition, it requires teachers to be "highly qualified" requiring a teaching license from the state they teach in and a bachelor's degree.

The defining characteristic of NCLB is the concept of Adequate Yearly Progress, AYP. This established benchmarks for proficiency set by the state to meet in reading and math. In addition, it required schools to test 95% of their students, placing a 1% cap on alternative assessments, and a 2% cap on modified assessments. Schools not making AYP two years in a row were classified as "failing" and are required to offering tutoring or other supplemental services. For schools that continue to not make AYP, they risk losing their title funds, thereby tying resources to results, establishing the high stakes era.

Campbell's Law states, "the more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it was intended to monitor" (Nichols & Berliner, 2007). In President Johnson's "War on Poverty" the social decision was made to provide funding for students that need additional resources, coupled with the No Child Left Behind policy of assigning the quantitative measure of test scores to remain eligible for those funds, has led to resource dependency, institutionalism, and a high stakes culture. School districts are resource dependent and controlled externally, making the ceremonial adoption of policies and procedures necessary to maintain their legitimacy and funding. Deviations from this new norm threaten their existence and further raise the importance of the state assessments perpetuating a high stakes culture that links resources to results by any means necessary, which has manifested itself in the need for some schools, districts, and states to distort their progress.

2.5 - Low Stakes Testing – NAEP

In 1969, with the first reauthorization of Elementary and Secondary Education Act (ESEA), the federal government wanted to measure the impact of the title funds it had begun distributed in 1965. As a result, the Department of Education developed the National Assessment of Educational Progress, NAEP. The goal of the NAEP assessment was to provide a big picture of the educational progress of the country, the "Nation's Report Card." Originally it reported on regions of the country to provide policymakers a big picture of state of education and evaluate the needs of American education on a macro-level. It was not until 1990 that NAEP started to provide state level data, and did so in order to compare state educational programs. On a national level the NAEP provides national trend data in grades four, eighth, and twelve and provides state trend data in grades four and eight. As part of the original "War on Poverty" the states that receive title funds are required to participate in the NAEP assessment when it is given every other year, however no minimum proficiency score is required to ensure continual eligibility for funding.

The NAEP assessment is unique because it uses a survey design. It relies on a stratified, random, multi-stage sample to report data on group characteristics. Originally designed as the "Nations Report Card" it does not provide school or student level data. For example, the pool of reading questions consists of 100-170 items. Each student that takes the test will answer 20-25 questions, with every test item exposed to approximately a quarter of the sample. This requires a large sample of students to ensure that all questions are asked to all the groups of students. The scale scores are then reported by demographic group and grade level for each geographic region of the county and state to determine progress.

Campbell's Law states, "the more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it was intended to monitor" (Nichols & Berliner, 2007). States are required to give the NAEP to receive title funds, as it was initially developed to determine if those funds had any impact. However, the structure of the test, and the sampling needed to represent groups of students, make the unit of analysis at the region or state level. In addition, the performance on the NAEP is not reported at the student or school level, therefore no resources are tied to results of the test, making the NAEP low stakes. The absence of dependence on resources has also minimized the need to ceremonially adopt any preparation, modification, or structural practices at the school level to try to obtain the best score possible. The NAEP has become a thermometer that measures the temperature of the room, with no incentive by educators to move a candle closer to provide the illusion of warmth. As a result, the NAEP becomes an appropriate measure to determine if students are learning more as a result of the policies implemented by the No Child Left Behind reauthorization of ESEA, or if states have cheated in their ceremonial compliance with the mandates set for by the federal government in order to maintain their funding and legitimacy.

2.6 - Other Research Comparing State Assessments and the NAEP

In this dissertation I will compare student performance at the state level on the state assessments and the NAEP to measure the growth of each over a multiple year period. The discrepancy in growth rates on the two assessments will determine a distortion index at the state level that will be explained using census data from Integrated Public Use Micro-data Series, IPUMS. My goal is to explain why states cheat.

Other research has been conducted using the student scores at the state level on state assessments and the NAEP assessment. This research method, using state assessment scores and the NAEP, generally falls into two categories; to determine the rigor of state standards based on comparing student performance on state assessment versus the NAEP, and to compare levels of proficiency on each test for subgroups to examine the achievement gap that exists within most school districts.

The first category, using the amount of students proficient or higher on state assessments aims to measure the rigor present in the current state standards. In McLaughlin, Bandeira de Mello, Blankenship, Chaney, Esra, Hikawa, Rojas, William, and Wolman (2008) the research team from the National Center for Educational Statistics compared the NAEP and state assessments from the 2002-2003 school year to examine levels of proficiency on each test to determine if those results were correlated across schools, and if the discrepancies in scores were consistent between subgroups of students.

This single year study focused on using the NAEP as the standard and evaluating each state, all of whom have different assessments, against one constant to determine the quality of each state assessment as well as demonstrate the range that existed within the country.

In Bandeira de Mello (2011) the researcher determined a NAEP scale equivalent score that compares the percentage of who scored proficient or higher on the NAEP assessment with the percentage of students that scored proficient or higher on their individual state assessment. A higher NAEP scale equivalent score indicates a better match between the two tests, and in the researcher's opinion an indication of more rigorous standards. This study, published by NAEP, supports the concept the NAEP is a more rigorous assessment because students do not do as well on the test.

In Linn, Graue, and Sanders (1990) the researchers compare the number of students proficient in each state on their state assessment with the number of students proficient on the NAEP assessment. The goal of this research is to determine how "demanding" each state's standards are for students. This article was published twelve years prior to NCLB and the authors are only able to obtain data from forty states, illustrating a strength of NCLB that requires states to publish their data. Similar types of studies were done with Bandeira de Mello, Blankenship, McLaughlin (2009), Schneider (2009), and Stoneberg (2007) to determine the rigor of state standards and has been used by Dr. Willard Daggett and others a rationale for the need to move to a Common Core Curriculum to ensure the same measure of rigor for all states, districts, and schools.

The second type of research in the literature with state assessment scores and NAEP scores involve the examination of the achievement gap between subgroups of students based on either their race or gender. In Lee and Reeves (2012) the researchers compare NAEP trend data from 1990 to 2009 to determine the impact of NCLB of the achievement gap. Using data prior to NCLB as a baseline the researchers tried to determine the growth of various subgroups, along with the gaps that existed between them based on each state's instructional capacity, ability to track data, fidelity of implementation, and rigor of their standards. The researchers examined number of students reported proficient by their racial identity on the state assessment and tried to compare those numbers to the number of students proficient on the NAEP assessment. They reported the same magnitude of discrepancy present on each test to illustrate which measure was more effective to identify the differences between groups.

In Ho and Reardon (2011) the researchers examine the achievement gap by designing a V statistic to better explain the differences in performance categories by students of different races on state assessments, Advanced Placement tests, and English proficiency tests when compared to the NAEP. The goal is similar to the Lee and Reeves (2012) study to determine how each assessment identifies differences in achievement by different students. Ho and Reardon incorporate the use of the V statistic to allow them to drill down and better identify the differences between groups of students.

Similar types of studies were done with Reardon, Greenberg, Kalogrides, Shores, Valentino, (2012) and Condon, Greenberg, Stephan, Williams, Gerdeman, van der Ploeg (2012) to determine the impact of NCLB on the achievement gap that exists within school districts by looking at how students of different races perform on their local state assessment versus how they perform on the NAEP.

This dissertation will add to the research in that it will develop a cheating index to quantify the discrepancy between the high stakes state assessment and the low stakes NAEP assessment. That cheating index will quantify the discrepancy in the rates of growth between the two assessments. Furthermore it will seek to explain why cheating occurs using census data to more fully explain the external factors associated with higher rates of cheating.

2.7 - Conclusion

Cheating occurs in education. The cheating that has occurred at the state level, as a result of the No Child Left Behind legislation, is a rational response in a high stakes culture. Under the pressure to produce results, in this case proficient students, states have followed Campbell's law which says, "the more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it was intended to monitor" (Nichols & Berliner, 2007). The corruption within the state is supported by two theories in organizational sociology, institutionalism and resource dependency, that has resulted in goal displacement by the states. The history of No Child Left Behind, traced back to Lyndon Johnson's War on Poverty, illustrates how a policy born out of good intentions can be corrupted under the pressure to produce results.

The baseline assessment for student achievement in this country is the National Assessment of Educational Progress, NAEP, because of its stratified, random sample and strategy for questioning students. The sampling and questioning techniques enable schools and students to be measured, free of the fear of retribution or loss of resources. Initially developed in 1969 to provide regional data, and revised in 1990 to provide state level data for fourth and eighth grade reading and math scores, the NAEP is a low stakes measure that allows policymakers to determine the progress of their state in a way that is less threatening to each states resources and legitimacy. States that receive title funds are required to administer the NAEP, but the test's structure and format do not allow the results to be traced back to an individual school or student making it low stakes at the local level.

Some states have been able to show remarkable progress on the rates of proficiency of their own state assessment, but very little or no progress on their NAEP assessment, those states have distorted their progress. Distortion, as defined in this dissertation, is the discrepancy over time of a state's performance on the state assessment versus the NAEP. States that show big gains on their state assessment, but little or no gain of their NAEP assessment have probably aligned their curriculum, manipulated their testing practices, and set the cut scores low enough that most students could attain proficiency. These responses to the pressure to maintain legitimacy are rational and explained by Campbell's law as well as organizational theory on resource dependency and institutionalism. In this dissertation, a distortion index will be created to measure the discrepancy over time between each states reported proficiency on the state assessment and their NAEP score.

The distortion index will allow me to quantify the level of distortion that has occurred at the state level. Once quantified, I will access census data using the Integrated

Public Use Micro-data Series, IPUMS, in an effort to suggest possible predictors to better understand why this distortion has occurred. This dissertation will provide a twist to most of the current research that has focus on comparing the two assessments to evaluate the rigor of individual state's standards or the achievement gap between groups of students. This dissertation will attempt to suggest predictors to the distortion that occurs and could potentially add to the discussion on the role of assessments in our schools what other factors influence the results we measure from our students.

Chapter 3

Methods

3.1 - Goals of this Dissertation

The goal of this dissertation is to examine the extent to which states distort their progress and suggest possible predictors to help explain why these distortions occur. To do this, I will evaluate the change of each state's reported growth on their own state assessment to measure levels of proficiency in reading and math when compared to the their growth on the NAEP assessment over the same period of time. States that reported similar levels of growth disparity from the NAEP assessment will then be grouped together and examined using Integrated Public Use Micro-data Series, IPUMS, to determine what demographic characteristics correlate with the different levels of disparity between the high stakes state assessment and the low stakes NAEP assessments.

The levels of distortion will be determined by the disparity between the reported progress on state assessments and reported progress on the NAEP assessment. States that show high levels of growth on their own state assessment, but very little or no growth on the NAEP, will be classified as the states that distort the most. The states that show a similar growth pattern on their state assessment as they do on the NAEP will be classified as distorting less.

The NAEP assessment provides state level data for analysis in fourth and eighth grades, as a result those grades will also be the primary focus on the state assessments when the comparison is completed.

3.2 - Data Sources

For this dissertation there will be three primary sources of data, the National Assessment of Educational Progress (NAEP) data explorer, state department of education websites, and census data.

The NAEP data explorer allows the researcher to customize the test data they intend to use. For my study I will look at reading test scores for fourth grade. The NAEP offers data for twelfth grade, but it is only national trend data and in this study I will compare states, which is not available for twelfth grade. The variable and description is listed in Table 3.1.

Table 3.1

Source	Code	Description
NAEP	er_n	State's Average Score on the Elementary Reading Test

The second source of data is the State Department of Education for each state in the study. This data source will be used to obtain the percentage of students the state has reported as proficient in reading in an elementary grade, usually third or fourth grade. Each state's website is different and some have a data explorer similar to the NAEP, but most do not. Most of this data was obtained by looking at the State Report Card for each year. On the State Report Card it typically reports the number of students proficient in reading and math by grade level. If the website did not have a data explored, or publish a state report card, then I contacted the state directly by email and asked for the scores. All the reported levels of proficiency in reading on the state assessment came from their individual state department of education and are reported as proficient or higher scores.

Most states did not offer this data disaggregated by subgroup so the numbers reported are for all students.

Table 3.2

Source	Code	Description
State Test	er_s	Elementary Students Proficient on the State Reading Assessment

The final source of data for this study is the United States Census. For individual and household level variable I was able to use the Integrated Public Use Micro-data Series, IPUMS, as a data explorer to identify variables and pull proxies for social, economic, and family characteristics tied to success in school. To aggregate this to the state level, the data was pulled by state code and the mean collapsed for each state to obtain a state level average for each variable listed below.

<u>Parent's Level of Education</u>

In IPUMS the variable has eleven different codes, as listed below:

00 – No schooling

01 – Nursery school to grade 4

02 – Grade 5, 6, 7, or 8

03 – Grade 9

04 – Grade 10

05 – Grade 11

06 – Grade 12

07 – 1 year of college

08 – 2 years of college

09 – 3 years of college

10 – 4 years of college

11 – 5+ years of college

To simplify this variable I used STATA to combine some of the categories and collapse the mean by the state. To recode and include everyone in the dataset I made the following changes.

IPUMS (00) became (01) – to represent those people with no schooling

IPUMS (01) became (02) – to represent those people with only elementary school

IPUMS (02) became (03) – to represent those people with only middle school

IPUMS (03, 04, 05, 06) became (04) to represent those people with high school

IPUMS (07, 08, 09, 10) became (05) to represent those people with college

IPUMS (11) became (06) to represent those people with graduate school

Once recoded I collapsed the mean by state for each of the years collected; 2003, 2005, 2007, and 2009. The STATA code is noted below.

```
gen school=1
replace school=2 if (educ==1)
replace school=3 if (educ==2)
replace school=4 if (educ==3 | educ==4| educ==5 |educ==6)
replace school=5 if (educ==7 | educ==8 | educ==9 |educ==10)
replace school=6 if (educ==11)
collapse (mean) school, by (statefip)
```

Father's Location in the Home

In IPUMS the variable is called "poploc" to indicate the father's location within the home. To interpret this data the researcher needs to know that any number greater than (00) indicates a father lives in the household. The number assigned by the census for this variable is person number listed on the census form that is returned. If the father happens to be listed first, he is number one, if he is listed second; he is number two, and so on. In my data the range of data went from one to fifteen. As a result, I recoded all numbers to be (01) if the father was present and (00 if he was not present. I then collapsed the mean by the state id number by the four years of data I collected; 2003, 2005, 2007, and 2009. The STATA code I used is listed below.

gen dad=0

replace dad =1 if (poploc==1 | poploc ==2 | poploc==3 | poploc ==4 | poploc==5 | poploc ==6 | poploc==7 | poploc ==8 | poploc==9 | poploc ==10 | poploc==11 | poploc ==12 | poploc==13 | poploc ==14 |poploc==15)

collapse (mean) dad, by (statefip)

<u>Poverty</u>

In IPUMS the three digit poverty variable expresses the percentage of the poverty threshold the family meets based on the previous year's income when compared to the poverty income determined by the Social Security Administration. For example, (050) would indicate the respondent has reported and income that is 50% of the poverty threshold income, while someone that reported (200) would be living at 200% of the income representing poverty. The higher the number indicates less poverty. In addition, this is a household variable so each member of the reporting family has the same number. To aggregate this to state level I simply collapsed the mean of each respondent to the state code for each of the years represented; 2003, 2005, 2007, and 2009. The STATA code is represented below.

collapse (mean) poverty , by (statefip)

Race – Black

In IPUMS the RACBLK variable allows the respondent to indicate if they identify themselves as black. The respondent can mark as many races as they wish to identify themselves as for the census report. The census codes the responses as (01) for no, and (02) for yes. For this variable I collapsed the mean by state code for each of the years represented; 2003, 2005, 2007, and 2009. The STATA code is represented below.

collapse (mean) racblk , by (statefip)

Birthplace

In IPUMS the (BPL) Birthplace variable allows the respondent to indicate the state or country in which they were born. The state codes are the same as the state census codes from 1-99. All numbers above 100 represent another country outside the United States. For this variable I recoded all the respondents as (1) if they had a birthplace code of less than 100, to indicate they were born in the USA, and (0) if they had a birthplace code of greater than 100. For this variable I collapsed the mean by state code for each of the years represented; 2003, 2005, 2007, and 2009. The STATA code is represented below.

gen bp=1

replace bp=0 if (bpl<100)

collapse (mean) bpl, by (statefip)

Manufacturing

In IPUMS the (IND) Industry Code assigns a four-digit number based on the reported occupation of the responder. The codes are then categorized into the seventeen categories listed below:

1. Not Applicable (0000)
2. Agriculture, Forestry, Fishing and Hunting (0170 – 0290)
3. Mining (0370 – 0490)
4. Construction (0770)
5. Manufacturing (1070 – 3990)
6. Retail Trade (4070 – 5790)
7. Transportation and Warehousing (6070 – 6390)
8. Utilities (0570 – 0690)
9. Information and Communications (6470 – 6780)
10. Finance, Insurance, Real Estate, and Rental and Leasing (6870 – 7190)
11. Professional, Scientific, Management, Administrative, and Waste Management Services (7270 – 7790)
12. Education, Health, and Social Services (7860 – 8470)
13. Arts, Entertainment, Recreation, Accommodations, and Food Services (8560 – 8690)
14. Other Services (Except Public Administration) (8770 – 9290)
15. Public Administration (9370 – 9590)
16. Armed Forces (9670 – 9870)
17. Unemployed (9920)

To combine categories in order to classify each occupation as either "manufacturing" or "not manufacturing" I recoded all the job codes as (1) and then changed every code less than 999 or greater than 4000 to (0). This created two classifications, manufacturing (1) and not manufacturing (0). I then collapsed the mean by state code for the years 2003, 2005, 2007, and 2009. The STATA code is represented below.

```
gen manuf=1
replace manuf=0 if (ind<999 | ind>4000)
collapse (mean) manuf, by (statefip)
```

Population

To obtain state level variable, population, I utilized the data explorer on the census.gov website. I obtained the population for each year in the study; 2003, 2005, 2007 and 2009.

Per Pupil Expenditure

To find the (PPE) Per Pupil Expenditure I researched the website http://datacenter.kidscount.org/data/acrossstates/Rankings.aspx?ind=5199 to obtain the data for the years 2003, 2005, 2007, and 2009. The Kids Count website uses census data in a search engine to allow access to state level data.

Listed in table 3.3 is a summary of the IPUMS and Census variables collected and used in this study, as well as the code I used in STATA.

Table 3.3

Source	**STATA Code**	**Description**
IPUM	dad	State's Level of Fathers Residing in the Home
IPUM	school	State's Level of Parental Educational Attainment
IPUM	poverty	State's Level of Citizens in Poverty
IPUM	manuf	State's Level of Manufacturing
IPUM	black	State's Level of African-Americans
IPUM	bp	Birth place

Census	st_pop	Population of the State
Census	ppe	Per Pupil Expenditure for Education

3.3 - Data Collection

For this dissertation there will be three primary sources of data, the National Assessment of Educational Progress (NAEP) data explorer, state department of education websites, and census data. To collect this data it involved using the resources available for each entity.

To obtain the NAEP data I used the NAEP data explorer. In August of 2011 I was fortunate enough to attend a seminar in Washington DC, sponsored by the National Center for Educational Statistics (NCES), over how to use the NAEP databases. The training included a history of the NAEP assessment as well as the potential and limitations of how the data could be used. At this training, I learned the NAEP data explorer would be the best resource to obtain state level measures on the NAEP assessment. The NAEP is a unique measure because of its stratified, random sampling technique that does not provide student or school level data. The smallest reliable measure is at the state level, so it quickly became ideal for this study.

The goal of this study is to compare NAEP data with state assessment levels of proficiency. However, the NAEP scores are not reported in categories relative to proficiency, it is only reported by the average score in each subgroup. This will allow a reliable and consistent measure in which to compare state's levels of performance.

To obtain the each state's levels or reported proficiency on state reading and math assessments required the most work. In the absence of a state assessment data explorer, with each state using their own individual assessment, it required a seminar on each state's testing plan to identify the test they used and how they reported their results. To try to obtain consistent data I tried to obtain fourth and eighth grade levels of proficiency in reading for the years 2003, 2005, 2007, and 2009. However, this was not always possible. Some states did not test those grades prior to 2006, the year they were required to do so as part of NCLB. If they were testing grades three or seven I continued that trend until 2009.

Some states were missing state assessment data. Minnesota and New Hampshire did not have elementary or middle school reading levels for 2003 or 2005. New Mexico and Ohio did not have middle school reading data for 2003. Rhode Island did not have elementary or middle school reading data for 2005. Utah, Vermont, and West Virginia did not have elementary or middle school reading data for 2003. To account for the missing data I matched it to the closest reported year the state did have data so that it reflected no progress. The purpose of this study is to show the misrepresentation of progress at the state level, and filling in the missing data with the next data point established no gain. See Appendix 1.

Each state publishes some version of a State Report Card. On the Report Card they will report the proficiency level in reading and math for selected grades. Finding these report cards for the years 2003, 2005, 2007, and 2009 was the method I used to obtain this data. Some report cards reported multiple years of scores, but most states did not. To obtain the data for all 50 states, with four years of data, required locating approximately 200 state report cards.

To obtain the census level data I used two sources, the Integrated Public Use Micro-data Series, IPUMS, data explorer and the census website. The individual and household level variables are accessible through the IPUMS data explorer. Using that tool I was able to obtain the six variables outlined above for the years; 2003, 2005, 2007, and 2009.

To obtain the state level variables for the population and financial information I used the census.gov data explorer. Similar to IPUMS it allowed the researcher to build reports and charts customized to your study. I chose the population data and the financial information because I am interested in looking at not only the expenditures per pupil but also the expenditures per citizen in states that are identified as cheating the most.

3.4 - Analysis Strategy

Standardization of Data

The goal of this study is to suggest predictors to why states distort their educational progress. To do this requires the standardization of variables to allow for comparison. The data measuring the state's reported proficiency level on their measure used for No Child Left Behind, their state assessment score in reading for elementary and middle school, is measured in percentage of the testing sample scoring proficient or higher. The low-stakes NAEP score for the corresponding grade level and year is measured in raw score out of a possible 500 points. The state population is measure in the number of registered citizens. The per-pupil expenditure is measured in dollars spent per pupil on education. The manufacturing level of each state was calculated based on the number census respondents reporting a manufacturing occupation. The poverty variable is a combination of all the respondents' relationship to the poverty level established by the Social Security Administration. The educational attainment of the parents is the combination of a six-level matrix developed to simplify an eleven-level matric and aggregated to the state level. The race variable, father's location, and birthplace are bivariate variables aggregated to the state level for four different years. The number of different variables and measures required each to be standardized in order to compare and interpret the data.

To standardize the data the first task was to determine the data set. For this study, three different ways to standardize the data were discussed. The first was a vertical standardization by year. Each variable would be nested within the year across all states, and the standard error would be accounted for within the year. The second was a grand standardization by comparing each data point against the entire dataset for all states in all years, thereby losing any of the benefits of a panel dataset. The third way, which was the final one used in the analysis, was a horizontal standardization, thereby containing the standard error within the state by using each state's data for all four years to create a Z-score by calculating the mean and standard deviation of the four data points (2003, 2005, 2007, and 2009) to standardize each data point used in the study. For example:

State	2003	2005	2007	2009
Alabama – Raw Score	51	68	71	75
Alabama – Z Score	-1.44	.165	.449	.828

Mean of the Alabama raw scores is: 66.25

Standard Deviation of the Alabama raw scores is: 10.56

Z Score = (Raw Score – Mean)/Standard Deviation

The reason this standardization method was selected was to ensure that each state contained its own standard error, and it allowed for the benefits of a panel dataset, allowing each state to nest within itself, and the z-transformation approach is based on procedures outlined by Koretz (2008). Each variable in this study; state assessment proficiency levels, NAEP scores, per citizen expenditures on education, state population, parents level of education, poverty status, father's location in the home, African-American status, and white status were all standardized in this way to allow for comparison and interpretation.

<u>Growth Models</u>

The basis of this study is comparing the growth on the state assessment versus the growth on the NAEP assessment. To determine the growth I looked at the change in scores from 2003 to 2005, from 2005 to 2007, and from 2007 to 2009 to give each state three measures of growth over the seven years of the study. To calculate the growth I used the following equation:

$$\text{Growth} = \frac{t_{2-}\, t_1}{\max - t_1}$$

However, two significant adjustments had to be made in order for the model to accurately reflect the growth on the state assessment and NAEP scores. First, the number used for the max. On the state assessment the theoretical maximum possible for each state to achieve in any given year was 100%, and some states approached this number. The theoretical maximum score possible on the NAEP assessment is a score of 500, and no state scored above 250. As a result it was determined to develop a growth measure based on the practical growth within a year based on the maximum score on each assessment within a given year. This modifies the above equation in the following way:

$$\text{Growth} = \frac{t_{2-}\, t_1}{t_{1Max} - t_1}$$

The second adjustment occurred by adjusting the "practical max" score to a "standardized practical max" score by determining where the practical max existed within the standard distribution of a set of scores within a given year. When combined with the existing data points it creates the following Practical Normed Growth equation:

$$\text{Practical Normed Growth (PNG)} = \frac{t_{2-}\, t_1}{t_{1Max} - t_1}$$

t_2 = time 2

t_1 = time 1

t_{1Max} = Maximum Score within a given year and then standardized based on the mean and standard deviation of all the scores across all states within that given year.

<u>Distortion Index Calculation</u>

To calculate the distortion index I subtracted the practical normed growth on the NAEP assessment from the practical normed growth on state assessment, represented in the equation below. The purpose of calculating it in this way allows the distortion metric to remain positive when distortion has occurred and it is anticipated that most states will show less growth on the NAEP than they did their own state assessment. In addition, this method of calculation allows for the distortion to reflect the discrepancy in growth measures on the two tests, the foundation of this study. The equation is represented below:

$$\text{Distortion Index} = PNG_{state} - PNG_{NAEP}$$

Fixed Effects Time Variant Panel Regression Model

To analyze this data, based on how it has been standardized, I will use a panel regression model that allows the standard error to be nested within the state, since it is the unit of analysis. In STATA, this approach does the same thing as creating a dummy code to calculate the error for each data point, as it mean centers the data. The commands are listed below based on how the variables are identified within the program.

Elementary Reading Stata Commands

xi: xtreg ch_png_hz_er per_cit_exp_hz st_pop_hz, i(statenum) fe

xi: xtreg ch_png_hz_er per_cit_exp_hz st_pop_hz manuf_hz school_hz, i(statenum) fe

xi: xtreg ch_png_hz_er per_cit_exp_hz st_pop_hz manuf_hz school_hz poverty_hz dad_hz black_hz white_hz, i(statenum) fe

Chapter 4

Results of the Research

In this chapter I will outline the results of the two research questions posed in this dissertation.

1. Do states misrepresent their progress on their own state assessments?
2. If states do distort their progress, are their predictors to suggest why this distortion occurs?

4.1 Research Question #1 – Do States Misrepresent their Progress?

1. Do states misrepresenting their progress on their own state assessments?

In order to determine if states have misrepresented their progress I calculated the Practical Normed Growth (PNG) for each state's individual state assessment and the NAEP for three time periods; 2003 to 2005, 2005 to 2007, and 2007 to 2009. The formula is listed below and explained in Chapter 3 of this dissertation.

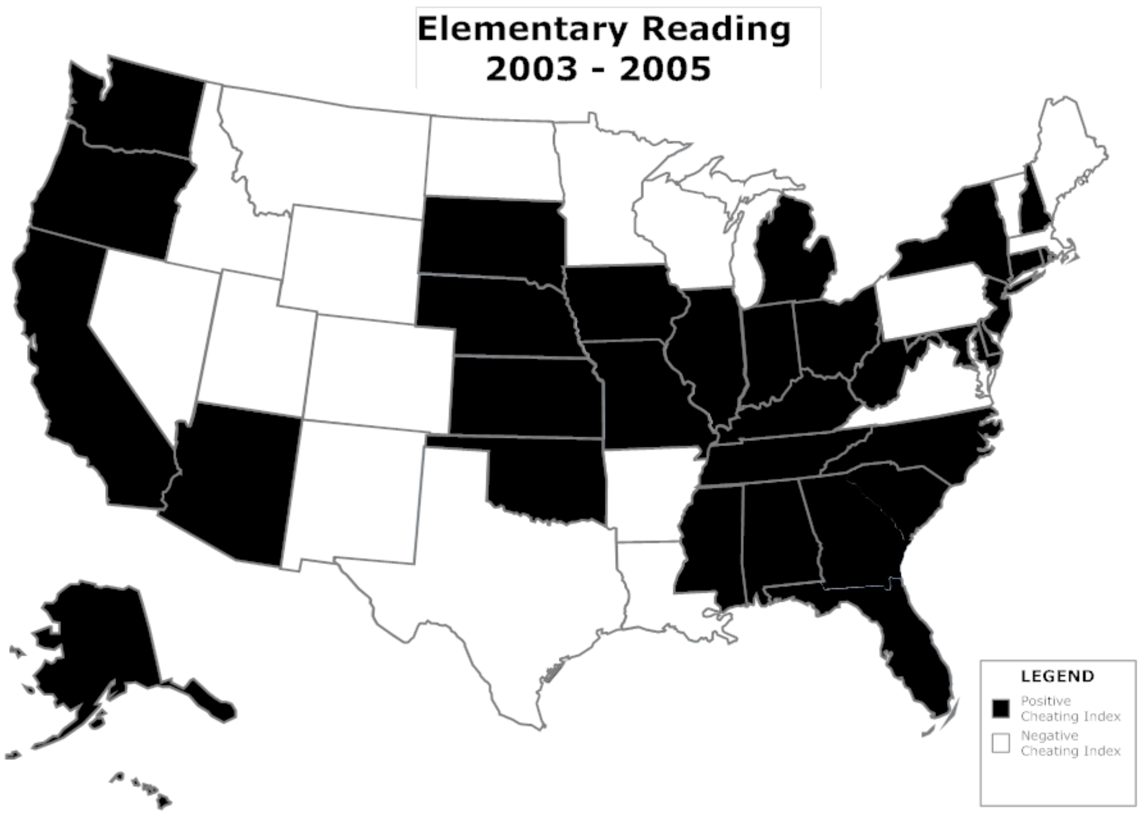
$$\text{Practical Normed Growth (PNG)} = \frac{t_{2-}\, t_1}{t_{1Max} - t_1}$$

Once the PNG had been calculated for each test within each time period, I then calculated a cheating index by subtracting the PNG of the NAEP assessment from the PNG of the state assessment. The formula is listed below and is also explained in Chapter 3 of this dissertation.

$$\text{Distortion Index} = PNG_{state} - PNG_{NAEP}$$

A positive distortion index indicates the state is reporting more growth on their state assessment than they achieved on the NAEP assessment. A negative distortion index indicates the NAEP assessment grew faster than the state assessment and the state underrepresented their progress to their constituents.   This process was followed for both the elementary reading assessment, which is represented in Table 4.1, 4.2, and 4.3.

<u>Elementary Reading from 2003 – 2005</u>

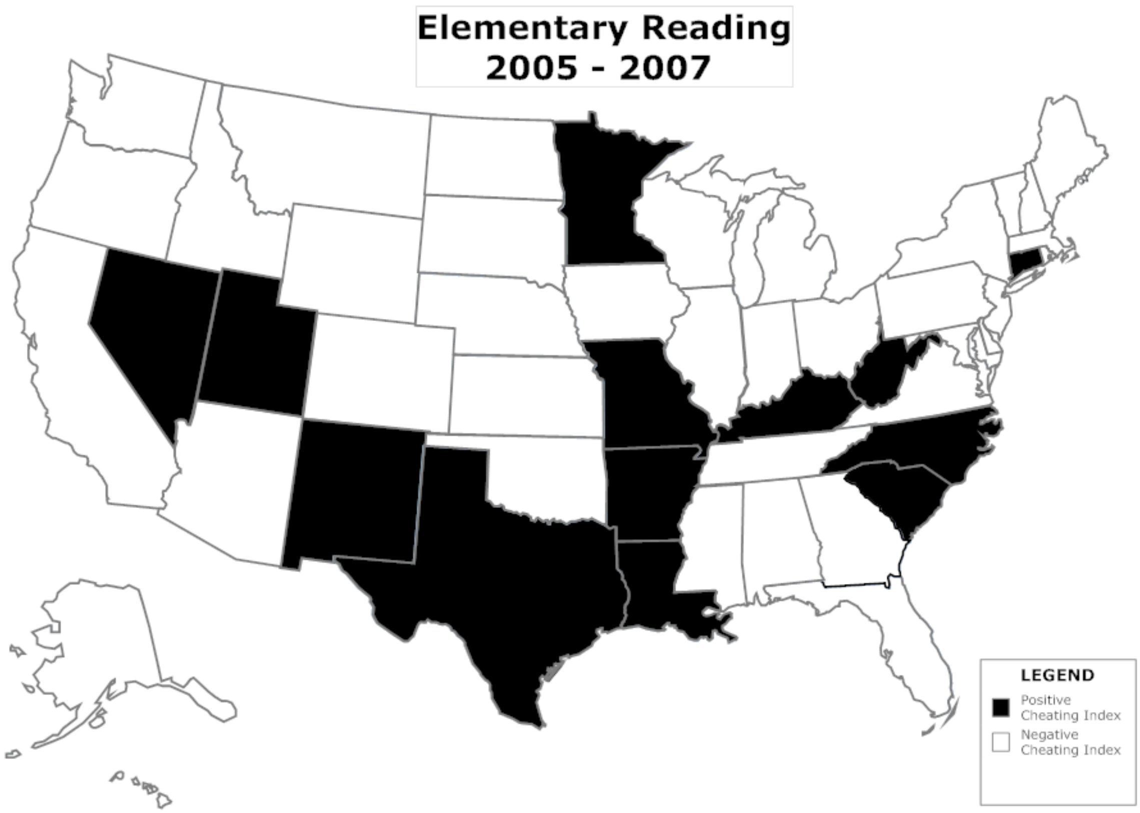


The map above and Table 4.1 provide a summary of the state assessment and NAEP Practical Normed Growth (PNG) calculations from 2003 to 2005 on the elementary reading assessment. Thirty-two out of fifty states have a positive distortion index, indicating they showed more growth on their own state assessment than they did on the NAEP. The four states with the highest distortion index on these tests, for this time period are West Virginia (18.1243), North Carolina (7.6970), Connecticut (1.7369), and Iowa (1.6484). West Virginia reported 81% proficient on their state reading assessment in 2003, and 81% again in 2005. However, their NAEP score dropped from 219 to 215. North Carolina reported a state assessment reading proficiency of 81% in 2003, and an increase to 82% in 2005, but their NAEP score dropped from 221to 217. Connecticut reported 68% proficient or higher on their state reading assessment in both 2003 and 2005, but their NAEP score dropped from 228 to 226. Iowa reported 70% of its 4th graders proficient on the Iowa Test of Educational Development (ITED) in 2003, and 79% proficient in 2005, but their scores on the NAEP assessment fell from 223 to 221. The results of all 50 states, including the 18 states with a negative distortion index are listed in Table 4.1.

Table 4.1 - Elementary Reading 2003 - 2005

	State	NAEP	DI		State	NAEP	DI
Alabama	0.6088	0.0512	0.5576	Montana	-0.1488	0.4037	-0.5526
Alaska	0.5995	-0.1559	0.7554	Nebraska	0.3120	0.2617	0.0503
Arizona	0.4788	-0.8635	1.3423	Nevada	-0.0912	0.0400	-0.1312
Arkansas	-0.7788	0.6981	-1.4769	New Hampshire	0.0000	-0.1814	0.1814
California	0.2960	0.1766	0.1194	New Jersey	0.3442	-0.2333	0.5775
Colorado	-18.6861	-0.0023	-18.6839	New Mexico	0.0000	0.3726	-0.3726
Connecticut	0.0000	-1.7369	1.7369	New York	0.4147	0.1915	0.2232
Delaware	0.9002	0.7656	0.1346	North Carolina	0.1049	-7.5921	7.6970
Florida	0.4010	0.1554	0.2456	North Dakota	0.2522	0.4910	-0.2388
Georgia	0.3506	0.1242	0.2263	Ohio	0.4671	0.1451	0.3219
Hawaii	0.3322	0.2278	0.1044	Oklahoma	0.7141	0.0573	0.6569
Idaho	-1.2743	0.5687	-1.8431	Oregon	0.3973	-0.4794	0.8766
Illinois	0.3420	0.0441	0.2978	Pennsylvania	0.2649	0.4401	-0.1751
Indiana	0.7182	-0.6730	1.3912	Rhode Island	0.2654	-0.0079	0.2733
Iowa	0.5761	-1.0722	1.6484	South Carolina	0.4244	-1.0330	1.4574
Kansas	0.4173	0.0554	0.3619	South Dakota	0.2758	0.1051	0.1707
Kentucky	0.3875	0.1189	0.2686	Tennessee	0.4331	0.3755	0.0577
Louisiana	-0.4110	0.8210	-1.2320	Texas	-0.4060	0.5898	-0.9958
Maine	0.1425	0.3801	-0.2376	Utah	0.0000	0.7071	-0.7071
Maryland	0.4067	0.1436	0.2632	Vermont	0.0000	0.2312	-0.2312
Massachusetts	-3.4462	0.3635	-3.8097	Virginia	-5.9648	0.4915	-6.4563
Michigan	0.5932	-0.3582	0.9514	Washington	0.7915	0.6376	0.1539
Minnesota	0.0000	0.7652	-0.7652	West Virginia	0.0000	-18.1243	18.1243
Mississippi	0.0000	-0.1808	0.1808	Wisconsin	0.0000	0.1213	-0.1213
Missouri	0.0567	-0.5602	0.6169	Wyoming	0.0823	0.3406	-0.2583

Elementary Reading from 2005 – 2007

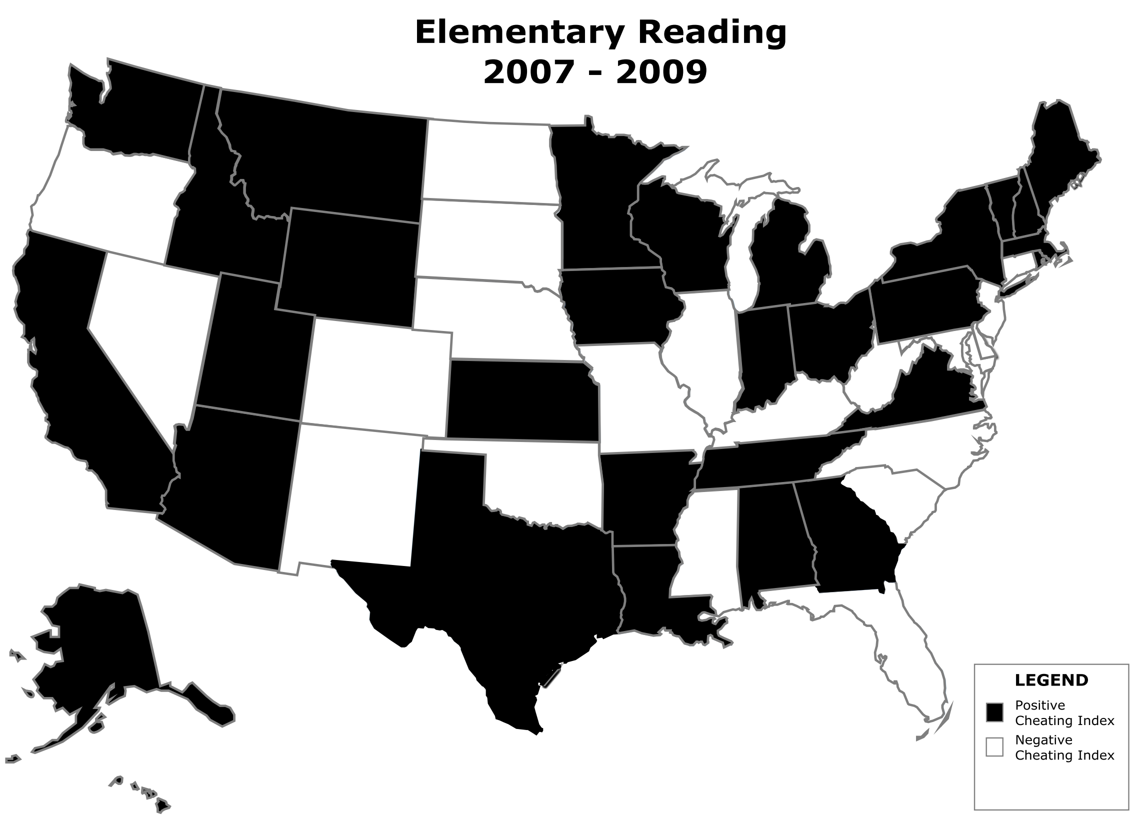


The map above and Table 4.2 provide summary of the state assessment and NAEP Practical Normed Growth (PNG) calculations from 2005 to 2007 on the elementary reading assessment. Thirteen out of fifty states have a positive distortion index, indicating they showed more growth on their own state assessment than they did on the NAEP. The four states with the highest distortion index on these tests, for this time period are; Louisiana (2.1092), Missouri (.6169), South Carolina (.5368), and Connecticut (.4651). Louisiana's reading proficiency increased from 52% to 58% on their state reading assessment of fourth graders, but their NAEP score dropped from 209 to 207. Missouri's state assessment scores jumped from 35% to 46%, but their NAEP score was stagnant at 221. South Carolina increased their reading scores from 79% to 82%, but the NAEP scores only increased from 213 to 214. Connecticut increased their state assessment scores from 68% to 71%, but their NAEP scores rose at a much slower rate increasing from 226 to 227.

Table 4.2 - Elementary Reading 2005 - 2007

	State	NAEP	DI		State	NAEP	DI
Alabama	0.0519	0.6981	-0.6462	Montana	0.6478	0.7929	-0.1452
Alaska	0.2993	0.9484	-0.6491	Nebraska	0.3779	0.7061	-0.3282
Arizona	0.2756	0.6369	-0.3613	Nevada	0.7524	0.6653	0.0871
Arkansas	0.3892	-0.0292	0.4184	New Hampshire	0.0000	0.6905	-0.6905
California	0.2102	0.4901	-0.2799	New Jersey	-0.1312	0.7833	-0.9145
Colorado	-0.0527	0.0357	-0.0884	New Mexico	0.9510	0.7970	0.1540
Connecticut	0.8209	0.3558	0.4651	New York	-0.2362	0.4938	-0.7299
Delaware	-5.4093	-1.3132	-4.0961	North Carolina	0.3514	0.1729	0.1785
Florida	0.3347	0.5162	-0.1815	North Dakota	0.3372	0.4610	-0.1238
Georgia	-1.8893	0.7583	-2.6477	Ohio	0.3506	0.7971	-0.4465
Hawaii	0.5527	0.8761	-0.3234	Oklahoma	-0.0892	0.6664	-0.7556
Idaho	-0.5603	0.5659	-1.1262	Oregon	-3.6251	-0.8537	-2.7713
Illinois	0.6236	0.7216	-0.0980	Pennsylvania	0.5046	0.6910	-0.1864
Indiana	-1.2743	0.6206	-1.8949	Rhode Island	0.0000	0.3379	-0.3379
Iowa	0.1510	0.8587	-0.7077	South Carolina	0.7373	0.2005	0.5368
Kansas	0.3581	0.7548	-0.3967	South Dakota	0.1904	0.9353	-0.7448
Kentucky	0.4218	0.3833	0.0385	Tennessee	0.0000	0.4040	-0.4040
Louisiana	0.2913	-1.8180	2.1092	Texas	0.7219	0.3188	0.4030
Maine	0.5818	0.8250	-0.2433	Utah	0.7015	-0.0586	0.7601
Maryland	0.4571	0.5876	-0.1305	Vermont	-1.8133	0.5358	-2.3491
Massachusetts	0.6459	0.6952	-0.0493	Virginia	-0.6117	0.5171	-1.1289
Michigan	0.1822	0.9273	-0.7451	Washington	-0.8759	0.3776	-1.2535
Minnesota	0.0000	-0.3853	0.3853	West Virginia	0.1907	0.0781	0.1126
Mississippi	0.2926	0.4884	-0.1958	Wisconsin	-2.3503	0.8992	-3.2495
Missouri	0.0567	-0.5602	0.6169	Wyoming	0.8369	0.8821	-0.0452

Elementary Reading from 2007 – 2009



The map above and Table 4.3 provide a summary of the state assessment and NAEP Practical Normed Growth (PNG) calculations from 2007 to 2009 on the elementary reading assessment. Thirty-one out of fifty states have a positive distortion index, indicating they showed more growth on their own state assessment than they did on the NAEP. The four states with the highest distortion index on these tests, for this time period are; Alaska (17.6027), Wisconsin (13.1808), Michigan (12.9353), and Maine (8.9547). In Alaska the scores dropped on both the state assessment and the NAEP. On their state reading assessment it fell from 79% to 78%, for a negative PNG of -.4272, but the NAEP had a much larger negative PNG of -18.0299 as it's scores fell from 214 to 211. In Wisconsin the state assessment proficiency levels stayed the same from 2007 to 2009, holding steady at 81%. However, the Wisconsin NAEP assessment dropped from

223 to 220. The same thing happened in Michigan with scores of 84% both years on the state assessment, but the NAEP score did not drop as much as Wisconsin, only falling two points as it dropped from 220 to 218. Finally, in Maine the state assessment scores increased from 67% to 71%, but the NAEP score dropped from 226 to 224.

On the elementary reading assessment, over the three growth periods examined, 32 states had a positive distortion index from 2003 to 2005, 13 states had a positive distortion index from 2005 to 2007, and 31 states had a positive distortion index from 2007 to 2009. In 150 possible opportunities to cheat, 50 states in three time periods, states cheated 76 times, or 51% of the time.

Table 4.3 - Elementary Reading 2007 - 2009

	State	NAEP	Cheating		State	NAEP	Cheating
Alabama	0.1641	-0.0309	0.1951	Montana	0.3678	-3.6483	4.0161
Alaska	-0.4272	-18.0299	17.6027	Nebraska	0.6076	-0.5975	1.2051
Arizona	0.3805	0.3457	0.0348	Nevada	-0.3377	0.1760	-0.5137
Arkansas	0.7964	-0.5697	1.3661	New Hampshire	0.9510	0.1799	0.7711
California	0.7984	0.5930	0.2055	New Jersey	-1.9718	-0.6161	-1.3557
Colorado	0.0501	0.9504	-0.9003	New Mexico	-19.4237	-3.2332	-16.1905
Connecticut	-1.5280	0.6732	-2.2012	New York	0.8597	0.5692	0.2905
Delaware	0.0000	0.3272	-0.3272	North Carolina	-2.8895	0.3561	-3.2456
Florida	0.5031	0.5607	-0.0575	North Dakota	-3.0523	-0.2022	-2.8501
Georgia	0.7473	-0.7322	1.4796	Ohio	0.2699	-1.4227	1.6926
Hawaii	0.1235	-5.1970	5.3206	Oklahoma	-2.2115	0.1517	-2.3632
Idaho	0.7182	-2.0096	2.7278	Oregon	0.3563	0.7645	-0.4082
Illinois	-0.2762	-0.1978	-0.0784	Pennsylvania	0.4366	-1.6697	2.1063
Indiana	0.5603	0.4481	0.1123	Rhode Island	0.9031	0.8668	0.0363
Iowa	0.1779	-5.1654	5.3433	South Carolina	-6.5480	0.8342	-7.3822
Kansas	0.4185	-0.5395	0.9579	South Dakota	-2.5873	-17.8009	15.2136
Kentucky	0.3648	0.7874	-0.4225	Tennessee	0.7640	0.4386	0.3255
Louisiana	0.8905	0.0291	0.8614	Texas	0.0000	-0.4012	0.4012
Maine	0.3974	-8.5573	8.9547	Utah	0.0000	-2.2992	2.2992
Maryland	0.2105	0.3798	-0.1693	Vermont	0.6445	0.4128	0.2317
Massachusetts	-0.7296	-1.0216	0.2919	Virginia	0.2277	-0.4931	0.7208
Michigan	0.0000	-12.9353	12.9353	Washington	-0.4669	-3.1570	2.6901
Minnesota	0.9510	-1.4200	2.3710	West Virginia	-2.2379	-0.1422	-2.0957
Mississippi	-2.5405	0.7537	-3.2942	Wisconsin	0.0000	-13.1808	13.1808
Missouri	0.1772	0.8933	-0.7161	Wyoming	-2.1994	-9.7308	7.5314

Table 4.4 - Number of States with Positive Cheating Index

	2003-05	2005-07	2007-09
Elementary Reading	32	13	31

The first research question of this dissertation was to determine if states distort their progress educational progress by misrepresenting their progress on their own state assessments? This research suggests they do cheat more than half of the time. On the elementary reading assessment comparison the states cheated 76 out of 150 times, or 51%. These findings lead to the second research question, are there any predictors that help to explain why this cheating occurs.

4.2 Research Question #2 – Do Any Predictors Explain Why States Distort their Progress?

Fixed Effects Time Variant Panel Regression Model

To analyze this data, based on how it has been standardized, I used a fixed effects time variant panel regression model that allows the standard error to be nested within the state, since it is the unit of analysis. In STATA, this approach does the same thing as creating a dummy code to calculate the error for each data point, as it mean centers the data for each of the eight predictors used in the models.

In order to produce a stepwise approach I created three models for each analysis. The first model uses predictors contributed by the home, the second model combines the home predictors with community predictors, and the third model combines home and community predictors with the state level predictors. In the first model of fixed effects time variant regression, the African-American status, father's location in the home, parents level of education, and born in the USA were compared to the Practical Normed Growth (PNG) Cheating Index created by subtracting the NAEP's PNG from the state's PNG.

In the second model the first four predictors, the African-American status, father's location in the home, parents level of education, and born in the USA remained and the state's level of manufacturing along with poverty were added. In the third model, all the predictors were included; per pupil expenditure, state population, manufacturing, parent's education, poverty status, father's location in the home, African-American status, and born in the USA were included to determine if any or all were statistically significant. The purpose of structuring the process in this way allows model one to examine the contribution of the home factors to the distortion by using the African-American status, father's location in the home, parents level of education, and born in the USA as predictors. Model two combines the factors contributed by the home with factors within the community that impact the educational progress of most students, poverty and occupational status, as measured by the manufacturing level.

The third model combines the home factors of model one of the African-American status, father's location in the home, parents level of education, and born in the USA, the community factors of model two with poverty and manufacturing level, with the state factors of population and funding, as measured with the per pupil expenditures. At the conclusion of running all models a Hausman test was conducted in STATA and it supported the fixed effects model over the random effects.

Table 4.5 illustrates the variables as well as their mean, max, min, and standard deviation.

Table 4.5 - Description of Variables

Model	Variable	STATA Code	Mean	Min	Max	SD	Predicted Impact
	State Assessment Growth	png_hz_er_s	-0.3755	-19.4237	0.951	2.504	
	NAEP Assessment Growth	png_hz_er_n	-0.6929	-18.1243	0.9504	3.2938	
	Discrepancy Index	ch_png_hz_er	0.3174	-18.6839	18.1243	3.9891	
Model One	African-American Status	black_hz	0	-1.4882	1.4977	0.8682	Increased African-American Status could increase discrepancy due to racial bias of tests.
	Parent's Level of Education	school_hz	0	-1.4974	1.4193	0.8682	Lower the level of Parent's Education may increase discrepancy.
	Born in the USA	bp_hz	0	-1.4748	1.4858	0.8682	Higher state population of immigrants may result in a higher discrepancy
	Father's Location in the Home	dad_hz	0	-1.3663	1.4882	0.8682	Fewer fathers at home could lead to a higher test discrepancy
Model Two	State's Level of Manufacturing	manuf_hz	0	-1.4999	1.4994	0.8682	Increased manufacturing could lead to a higher discrepancy
	Poverty Status	poverty_hz	0	-1.4581	1.4983	0.8682	Increased poverty may increase test discrepancy.
Model Three	Per Pupil Expenditure	ppe_hz	0	-1.3305	1.4251	0.8682	States with less funding may have a higher discrepancy
	State Population	st_pop_hz	0	-1.4822	1.4997	0.8682	Smaller states could have a higher discrepancy due to fewer resources and smaller testing pool.

In Appendix D there are scatter plots comparing each of the predictors above to the distortion index to illustrate the relationship each state has with each preditor.

Elementary Reading Cheating Regression Models

Model One is designed to measure the impact of the home on test distortion between the state and NAEP assessments. By combining the predictors contributed solely by the home; your race, whether or not your father lives with you, how educated your parents are, and if you are an immigrant, the goal of this model is to examine those characteristics to suggest possible impact on test distortion. Out of the four predictors in this model, only one is statistically significant. In Model One the predictor that is statistically significant is African-American status. The coefficient of 1.7303 indicates that as the African-American status of a state increases so does the level of cheating on the elementary reading state assessment. This finding can be explained by the research that black students score lower on standardized assessments than their white counterparts (Fryer & Levitt, 2006) resulting in the state's need to distort their progress in order to maintain their legitimacy and resources. As the African-American status of a state increases, the state will also have an increase in African-American students in their schools. This increase in students that do not perform as well on standardized assessments, coupled with the "one outcome fits all" mentality of No Child Left Behind, results in the state's need to produce the same outcomes as other states with students that perform at a higher level on standardized assessments.

In Model Two the home factors are combined with the community factors of poverty and manufacturing level to examine the impact of the community has on the home factors of Model One. Poverty is often viewed as the biggest contributor to student outcomes, as targeted in the LBJ War on Poverty with title funds. According to Rury and Saatcioglu (2011), "the percent of adults holding manufacturing jobs is a measure of occupational status, and can be interpreted as reflecting the blue collar character of the *total* suburban or central city population in a given metro area." A state's level of manufacturing can also be used as a proxy to measure the occupational status of a state. The states with higher manufacturing, and by proxy higher occupational status, typically have more income, but less education. Both predictors added in Model Two examine the impact a community can have on the predictors you get from home.

In Model Two only one predictor was statistically significant, African-American status, with a coefficient of 1.8224 indicating a more positive association with state level distortion than Model One of 1.7303. The standard error did increase slightly from .7522 to .7688. This finding indicates the community factors of poverty and occupational status were not significant enough to overcome the home factors, specifically race.

In Model Three the predictors from home are combined with the predictors from the community and are added to those from the state by factoring in state population and the per pupil expenditures spent on education. In this model, two predictors were statistically significant, race and state population. African-American status increased its coefficient from 1.8224 to 2.0060, and lowered its p-value to .01. The second statistically significant predictor in Model Three is the state's population. The negative coefficient (-1.5967) indicates that cheating goes up as the state's population decreases. This finding could be explained by the limited resources of smaller states and the pressure to produce the same results as larger states. The "one result fits all" in the form of proficiency or higher on state assessments is the signature of No Child Left Behind. The institutional pressure the states feel to produce this result could manifest itself in the need to cheat to produce the same result of larger states. Those states with limited resources are required to meet the same standard as the larger states, and with fewer advantages. In addition to potentially fewer resources, there is less room for error within the testing pool of students. In a large state, with a larger testing pool to draw for students, more students must pass, but more can fail and the state can still succeed. In a smaller state, fewer students take the test, and as a result each one is more valuable to the overall success of the state. This increased pressure created by a smaller testing pool could also result in the smaller states being force to cheat in order to maintain resources and legitimacy by incorporating institutional testing practices to produce the best result possible for their state.

Table 4.6 - Elementary Reading Distortion Regression Models

State Level Predictors	Model One			Model Two			Model Three		
	Coeff.	Std. Err.		Coeff.	Std. Err.		Coeff.	Std. Err.	
African-American Status	1.7303	0.7522	**	1.8224	0.7688	**	2.0060	0.8061	***
Father's Location in the Home	1.3366	1.0483		1.1820	1.1215		1.3355	1.1681	
Parent's Level of Education	0.5182	1.2574		0.5916	1.3301		0.6535	1.3561	
Born in the USA	-0.1754	0.6288		-0.1101	0.6432		-0.0626	0.6553	
Poverty Status				0.1262	0.6501		0.2092	0.6482	
State's Level of Manufacturing				0.4446	0.7795		0.4060	0.8194	
Per Pupil Expenditure							1.2051	1.3411	
State Population							-1.5967	.9225	*
Constant	0.3510	0.3547		0.3978	0.3647		0.4935	0.4244	
R-Squared	.0360			.0369			.0519		
Change in R-Squared				.0009			.0150		
F	2.07		*	1.45			1.49		

*** Significant at 0.010.

** Significant at 0.050.

* Significant at 0.100.

4.3 - Discussion of Results

When examining the distortion that occurs on the elementary reading assessments it is interesting to see different predictors statistically significant for the different assessments. To explain the cheating on the elementary reading assessments the predictor of African-American status was statistically significant in all models indicating a strong finding that maintained as additional predictors were added to explain the distortion that occurs on high stakes assessments. The African-American status illustrates conceptually a bias in both the standardized assessment and practices used to administer it to students. The achievement gap between African-American students and others is well researched. The state's need to circumvent that bias by manufacturing favorable results when more African-Americans are present represent bureaucratic policies aimed towards compensating for inefficiencies within a system.

The second statistically significant predictor, state population, suggests that smaller states distort their progress more than larger states. This finding could be explained in several different ways. It could be a function of the testing pool with smaller states being unable to absorb the range of student performance within the mean and as a result facilitating a system that distorts education progress more than larger states. It could be a result of limited resources within smaller states to provide to education. In most states over half of their state budget is spent on education, and smaller states would have less resources available to fund their educational systems.

Chapter 5

Conclusion

In this chapter I will summarize the findings of this dissertation, address potential concerns with the data, method, and section of predictors, as well as suggest potential directions for future research.

5.1 Summary of Findings

This dissertation addresses two research questions:

1. Do states misrepresent their progress on their own state assessments?
2. If states do distort their progress, are their predictors to suggest why this distortion occurs?

The first research question requires that distortion be defined. For the purposes of this dissertation I calculated the growth from 2003 to 2005, 2005 to 2007, and 2007 to 2009 on each state's individual state assessment and the NAEP. To calculate the growth I used a modified growth equation that subtracts the two scores and divides that by the maximum score on that test in that year, from the first score. This calculation produces a Practical Normed Growth (PNG) for the state assessment as well as the NAEP. To determine the distortion index I subtract the NAEP's PNG from the state assessment's PNG. A positive distortion index indicates the state assessment's growth was greater than the NAEP's growth and the state cheated. A negative distortion index indicates the NAEP's PNG was greater than the state assessment's PNG and the state did not misrepresent their progress. This analysis was done on the elementary reading assessment. This assessment includes three growth periods to compare, creates three observations of the 50 states, or 150 data points possible for distortion.

The first research question, do states distort their progress? The answer is yes. On the elementary reading assessments the states had a positive cheating index 76 times out of a possible 150, or 51%. The observed distortions came from three basic models. First, the state assessment scores went up, but the NAEP scores went down or stayed the same. Second, the state assessment scores stayed the same, but the NAEP scores went down. Third, the state assessment scores went down, but the NAEP scores went down more. In each of these possible scenarios the states have misrepresented the education progress of their state to their stakeholders. In the first scenario, if the scores on the state assessments go up while the NAEP assessment scores go down indicates a narrowed curriculum and an overemphasis on the state assessment. In the second scenario, if the scores on the state assessment stayed the same while the NAEP scores dropped indicates a less effective focus on the state assessment, at the expense of the NAEP. The third scenario, both the state assessment and NAEP scores fall, with the NAEP scores falling faster indicates a state struggling to do anything well.

The second research question, are there predictors to suggest why this cheating occurs? The answer is yes. On the elementary reading assessment comparison, two predictors were statistically significant in the final model, state population and African-American status. The coefficients of each of these predictors indicate that states with lower populations and higher numbers of African-Americans distort their progress more on the elementary reading assessments.

The combination of the calculation to determine cheating, and the regression analysis to explain distortions, support the resource dependency and institutional theory outlined in the review of literature. States in an effort to ensure resources have ceremonially adopted state policies, curriculum, and testing practices to ensure those resources continue. The accelerated growth on the state assessment demonstrates each state's commitment to the alignment of their stated public goal of reading proficiency in both middle and elementary schools. The lack of growth on the NAEP assessment indicates those policies were adopted ceremonially and overall educational program has remained unchanged or regressed. States have behaved like organization theory suggests, with the primary goal being survival. The alignment to, and growth of, state assessment proficiency rates in both middle and elementary schools, combined with a stagnant or declining growth on the NAEP assessment, provide a powerful example of each state's willingness to do what is necessary to maintain their legitimacy and survive.

5.2 Theory of Why States Distort their Results

As a result of the implementation of NCLB, there are at least two plausible theories to explain why states distort their results. The first involves two agencies, the states and federal governments, working together to ceremonially adopt changes and manufacture progress to gain public trust to ensure survival. The second involves the same two agencies competing against one another, and the state cheating in order to maintain resources and legitimacy.

In 2002, when NCLB was passed, the standard for success by 2014 allowed no room for error. The intent of the reform was to eliminate the achievement gap between subgroups of students and restore the faith in public education damaged by the 1983 Nation at Risk report. In the first theory, the state and federal governments have the same goal, restore faith in public schools. The federal government sets forth an ambitious goal of complete proficiency and the states adopt timelines, curriculums, assessments, and policies ceremonially to meet the goal. The federal government is a co-conspirator who does not deliver any real accountability, only false promises of tough action with little or no follow through. They allow the states to create and deliver less challenging assessments to enable the proficiency rates to grow quickly and restore public trust in public schools. The state distorts their results, as demonstrated by the rapid growth of state assessment proficiency rates when compared to NAEP assessment scores, but the federal government looks the other way as the legitimacy of public schools is restored by the results they are able to produce on the state tests.

The second theory is that NCLB was a conservative attack to end public schooling in this country, and states have distorted their progress to survive. One of the basic components of NCLB is the concept of choice. As part of the reform if a student attends a school that does not make AYP, they are allowed to choose another school that did make AYP, with the home school paying the cost. The theory of competition between schools for students resulting in a higher quality educational program is based market concept of competition creating quality and efficiency. It is plausible to argue the reason NCLB received unilateral support at its inception, from both the right and left, is both sides viewed this as an opportunity to legitimize their belief in the role of public schools in America, with the conservative right seizing on the opportunity to take the first steps towards a voucher system. If enough public schools were not able to make AYP, a logical next step would be to allow those students to choose a charter or private school option to meet their educational needs, in effect ending public education. This high stakes attack on survival resulted in states cutting corners and doing whatever was necessary to produce the results required by the reform to ensure the maintenance of resources and legitimacy. In this theory the federal government is used to attack the legitimacy of public schools and push the country to a privatized system of education built on competition and results.

5.3 Addressing Concerns

Data

To obtain the state assessment data I had to go to each state's individual website and mine the data. The lack of centralized location to obtain this state level data created some problems as some states were missing state assessment data. Minnesota and New Hampshire did not have elementary or middle school reading levels for 2003 or 2005. New Mexico and Ohio did not have middle school reading data for 2003. Rhode Island did not have elementary or middle school reading data for 2005. Utah, Vermont, and West Virginia did not have elementary or middle school reading data for 2003. To account for the missing data I matched it to the closest reported year the state did have data so that it reflected no progress. The purpose of this study is to show the misrepresentation of progress at the state level, and filling in the missing data with the next data point established no gain.

In an ideal study each state would have all of the data for each year, but not all states published their data for the years of this study. The NAEP data was mined from the NAEP Data Explorer in less than an hour. The state assessment data required becoming familiar with each state's assessment, website, testing protocols, and reporting format. This process took hundreds of hour and still was unable to produce a complete data set. However, when examining growth patterns it is my hope that filling in the missing data to show no growth does not diminish from the value or worth of the study.

<u>Method</u>

The most complex task of this study is the standardization of the data to allow for interpretation. When comparing proficiency percentages against test scores against eight different measures of predictors, the task of making sense of it all is a tall one. To do this it was determined that every data point should be standardized to allow for comparison and easier interpretation. To standardize any data point you must first determine the data set to calculate the mean and standard deviation. With a mean and standard deviation the calculation is simple, but the theory behind creating the dataset is complex. Three methods were discussed. First, standardize each number by year. For each data point compare it against how all fifty states did within a given year. This would allow the standard error to be nested within a given year. Second, standardize each number by state. This would create a dataset for each variable by state, and standardize it based on how the state performed on each of the four observations. This would allow the standard error to be nested within the state's characteristics. Third, standardize each number by state and year. This creates the largest dataset for each variable by allowing the standardization to occur over all fifty states in a four year span. The second option was the one selected for this study, standardizing the data by state, with the rationale the state was the unit of study, not the year. This study looks at the state's need to cheat to maintain resources and legitimacy, and the best way to do this would be to allow all the standard error to be nested within the state, allowing the full benefits of the panel data created.

Selection of Predictors

The second research question of this dissertation, are there predictors to explain why states cheat, is what makes this study unique. The selection of the eight predictors; per pupil expenditure on education, state population, state's level of manufacturing, parents level of education, poverty status, father's location in the home, African-American status, and immigrant status of a state were educated guesses at what variables could serve as predictors for why a state would need to cheat on their state assessments. There may be other predictors that could better explain the cheating outlined in this dissertation and those could be the basis of future research. The variables selected explain only 5% of the distortion on the elementary reading assessments as indicated by the r-squared values of the final models. Ideally this model would explain more of the distortion that occurs on state level testing progress when compared to the NAEP, and the selection of additional predictors could increase this value.

5.4 Future Direction of Research

The foundation of this research creates several different possibilities for future research. Those areas include examining the cut scores used by state to determine proficiency, the rigor of the state standards used to create the assessments, teacher licensure requirements, and additional predictors to explain the cheating that is outlined in this dissertation.

Cut scores

The data obtained for this study regarding the state assessment was categorized by those students who scored proficient or higher on the elementary or middle school state assessment for reading. Each state determines its standard for proficiency. For example in Kansas a student must score a 68% in reading to be classified as proficient, but only needs a 50% to be classified as proficient in math. Another possible extension of this research could be to calculate the cheating index based on the proficiency data contained in this dissertation combined with the cut score needed to be classified as proficient. With each state determining the cut score needed to be classified as proficient, and Campbell's Law telling us "the more any quantitative social indicator is used for social decision making, the more subject it will be to corruption pressures and the more apt it will be to distort and corrupt the social processes it was intended to monitor," it is likely states may be using the arbitrary classification of proficiency as a way to cheat. The combination of the percentage of students classified as proficient, coupled with the cut score needed to meet that classification, could serve as a powerful indicator of the cheating that occurs at the state level.

State Standard Rigor

Each state determines its own standards for reading and math. The state assessments administered by the state are built on those state standards. The rigor of those standards determines the difficulty of the state assessment. Multiple researcher studies (Bandeira de Mello, Blankenship, McLaughlin (2009), Schneider (2009), Stoneberg (2007)) have been done to examine the rigor of the standards used to develop state assessments. An extension of this dissertation could be to couple the rigor of state standards to the cheating that occurs at the state level. Do states with more rigorous standards cheat more or less? Do states that cheat more have more rigorous standards cheat more or less? Do states that have less rigorous standards cheat more or less? These are some possible topics that could explored with state level cheating tied to standard rigor.

Teacher Licensure

Under No Child Left Behind each state determines the criteria for a teacher to be considered "highly qualified." Some states require a college degree in education; some require a standardized exam in their content area, others in professional practice. Some states have alternative licensure programs, others do not. All states have licensing requirements that could be used as predictors for teacher quality. The quality of teachers a state has in its education system could be linked to the state's need to cheat. Another possible research study could examine a state's cheating index in combination with the requirements to teach in that state. Do states with lower teacher quality cheat more or less? Does the state have to compensate for an inadequate teaching force by cheating? These are topics that could be explored in future studies based on this research.

Additional Variables as Predictors

In this study the two different assessments, elementary and middle school reading, had different predictors for cheating. This finding suggests a possible topic for future research based on the emphasis of assessment and the curriculum at the different levels of education, and how this change in emphasis could lead to cheating. The complexity of the cheating outlined in this dissertation could not be captured within the eight variables used within this study. Other possible predictors could be the addition of proxies aimed at measuring the impact of the home structure on the state's need to compensate by cheating. Examples include the mother's location in the home, number of siblings, hours the parents work per week, and other factors demonstrated to have an impact on how a student typically performs in school.

As a policy study, this dissertation only looked at per citizen expenditure as the funding predictor for cheating. Additional variables could be explored to better tell the story of cheating that occurs in states. Possible variables could the amount of money spent on special education, title funds received by the state per student, or other targeted funds aimed at improving the educational opportunity of those students needing additional support outside the general education spectrum.

In addition, future studies could look at state and NAEP math cheating indexes to determine if similar predictors apply to the cheating that occurs on those assessments. A frequent topic in education circles is the quality and abundance of math and science teachers. Many states typically score lower on their NAEP and state math assessments, do these lower scores indicate more or less cheating? Has the perception of lower teacher quality in math and science forced states to cheat more? These are possible topics for future studies based on the foundation established in this dissertation.

Bibliography

Amrein-Beardsley, Audrey. 2009. "The Unintended, Pernicious Consequences of "Staying the Course" on the United States' No Child Left Behind Policy." *International Journal of Education Policy and Leadership*, Vol 4, No 6.

Auge', K., Simpson, K. (2012, May 20). As stakes rise, so do allegations of cheating on standardized tests. The Denver Post.

Bandeira de Mello, V. 2011. *Mapping State Proficiency Standards onto NAEP Scales: Variation and Change in State Standards for Reading and Mathematics 2005-2009* (NCES 2011-458). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: Government Printing Office.

Bandeira de Mello, V., Blankenship, C., McLaughlin, D. 2009. *Mapping State Proficiency Standards onto NAEP Scales: Variation and Change in State Standards for Reading and Mathematics 2005-2007* (NCES 2010-456). National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC: Government Printing Office.

Berliner, D.C. and Nichols, S.L. 2007. Collateral Damage: How High-Stakes Testing Corrupts American Schools. Cambridge: Harvard Education Press.

Booher-Jennings, Jennifer. 2005. "Below the Bubble: 'Educational Triage' and the TexasAccountability System." *American Educational Research Journal* 42:231-68.

Carey, Kevin. 2006. "Hot Air: How States Inflate Their Educational Progress Under NCLB." Education Sector Online.

Condon, Chris., Greenberg, A., Stephan, J., Williams, R., Gerdeman, R., van der Ploeg, A. 2012. "Performance in Science on the Minnesota Comprehensive Assessments – Series II for Students in Grades 5 and 8." Regional Educational Laboratory Midwest, Naperville, IL.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Iomorphism and Collective Rationality." *American Sociological Review* 48: 147-60.

Dolezalek, Holly. 2009. Standardized Testing in Schools. ABDO Publishing Company.

Fryer, Roland G., and Steven D. Levitt, 2006. "The Black-White Test Score Gap Through Third Grade," American Law and Economics Review, Oxford University Press, vol. 8(2), pages 249-281.

Gabriel, T. (2010, June 10). Under Pressure, Teachers Tamper with Tests. The New York Times.

Hall, Daria. 2005. "Getting Honest About Grad Rates: How States Play the Numbers and Students Lose." Education Trust. Washington, DC.

Haney, Walt., George Madaus, and Anne Wheelock. 2003. "DOE Report Inflates MCAS Pass Rates." National Board of Education Testing and Public Policy.

Harris, Donna. 2012. "Urban Schools, Accountability, and Equity: Insights Regarding NCLB and Reform." Education and Urban Society vol. 44 no. 2 **203-210.**

Ho, Andrew, D. and Reardon, S. 2011. "Estimating Achievement Gaps From Test Scores Reported in Ordinal 'Proficiency' Categories." Journal of Educational and Behavioral Statistics. pp. 1-29.

Jacob, Brian A. and Steven D. Levitt. "Rotten Apples: An Investigation Of The Prevalence And Predictors Of Teacher Cheating," Quarterly Journal of Economics, 2003, v118(3,Aug), 843-878.

Koretz, Daniel. 2008. Measuring Up: What Educational Testing Really Tells Us. Cambridge, MA: Harvard University Press.

Lee, Jaekyung, and Todd Reeves. 2012. Revisiting the Impact of NCLB High-Stakes School Accountability, Capacity, and Resources: State NAEP 1990–2009 Reading and Math Achievement Gaps and Trends. *Educational Evaluation and Policy Analysis.* 34: 209-23.

Linn, R. L., Graue, M. E. and Sanders, N. M. 1990. Comparing State and District Test Results to National Norms: The Validity of Claims That "Everyone Is Above Average". Educational Measurement: Issues and Practice, 9: 5–14.

Loveless, Tom. 2007. The Peculiar Politics of No Child Left Behind: Standards-Based Reform and the Poverty Gap: Lessons for No Child Left Behind. Washington, D.C.: Brookings Institution.

McLaughlin, D.H., Bandeira de Mello, V., Blankenship, C., Chaney, K., Esra, P., Hikawa, H., Rojas, D., William, P., and Wolman, M. 2008. Comparison Between NAEP and State Mathematics Assessment Results: 2003. National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC.

Mellon, E. (2011, July 13). HISD finds signs of cheating on state exams. The Houston Chronicle.

Meyer, John and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83:340-63.

Pfeffer, J. and Gerald R. Salancik. 1978. *The External Control of Organizations*. New York, NY: Harper and Row. Excerpts: 23-38, 39-61, 92-112.

Reardon, Sean F., Greenberg, E., Kalogrides, D., Shores, K., Valentino, R. 2012. "Trends in Academic Achievement Gaps in the Era of No Child Left Behind." Society for Research on Educational Effectiveness. Sheridan Road, Evanston, IL.

Rothstein, Richard. (2009, January 23). "Getting Accountability Right." Education Weekly.

Rury, John L., and Argun Saatcioglu. "Suburban Advantage: Opportunity Hoarding and Secondary Attainment in the Postwar Metropolitan North." *American Journal of Education* 117.3 (2011): 307-342.

Schneider, M. 2009. "Math in American High Schools: The Delusion of Rigor." *Education Outlook* No. 10.

Stoneberg, B. 2007. "Using NAEP to Confirm State Test Results in the No Child Left Behind Act." *Practical Assessment, Research & Evaluation* Volume 12: Number 5.

Suchman, Mark. 1995. "Managing Legitimacy: Strategic and Institutional Approaches." *The Academy of Management Review* 20: 571-610.

Tolbert, Pamela S., and Lynne G. Zucker. 1983. "Institutional Sources of Change in the Formal Structure of Organizations: The Diffusion of Civil Service Reform, 1880-935." *Administrative Science Quarterly* 28:22-39.

Toppo, G. Amos, D., Gillum, J. and Upton, J. (2011, March 6). When test scores seem too good to believe. USA Today.

APPENDIX A

STATA Printouts of Cheating Regression Models

```
. xi: xtreg ch_png_hz_er  school_hz  dad_hz black_hz bp_hz , i(statenum) fe

Fixed-effects (within) regression               Number of obs      =       150
Group variable: statenum                        Number of groups   =        50

R-sq:  within  = 0.0794                         Obs per group: min =         3
       between = 0.0028                                        avg =       3.0
       overall = 0.0360                                        max =         3

                                                F(4,96)            =      2.07
corr(u_i, Xb)  = -0.1524                        Prob > F           =    0.0908

------------------------------------------------------------------------------
ch_png_hz_er |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   school_hz |   .5182599   1.257447     0.41   0.681    -1.977752    3.014272
      dad_hz |   1.336655   1.048386     1.27   0.205    -.7443737    3.417684
    black_hz |   1.730326    .752274     2.30   0.024      .237074    3.223578
       bp_hz |  -.1754249   .6288745    -0.28   0.781    -1.423731    1.072881
       _cons |   .3510334   .3547292     0.99   0.325    -.3530985    1.055165
-------------+----------------------------------------------------------------
     sigma_u |  2.2642732
     sigma_e |  4.0174014
         rho |  .24108085   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(49, 96) =     0.92              Prob > F = 0.6272

. xi: xtreg ch_png_hz_er  manuf_hz school_hz poverty_hz dad_hz black_hz bp_hz ,
i(statenum) fe

Fixed-effects (within) regression               Number of obs      =       150
Group variable: statenum                        Number of groups   =        50

R-sq:  within  = 0.0845                         Obs per group: min =         3
       between = 0.0032                                        avg =       3.0
       overall = 0.0369                                        max =         3

                                                F(6,94)            =      1.45
corr(u_i, Xb)  = -0.1685                        Prob > F           =    0.2057

------------------------------------------------------------------------------
ch_png_hz_er |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    manuf_hz |   .4446202   .7795305     0.57   0.570    -1.103156    1.992396
   school_hz |    .591623   1.330175     0.44   0.658    -2.049471    3.232717
  poverty_hz |   .1262143   .6501807     0.19   0.846    -1.164735    1.417163
      dad_hz |   1.182043   1.121597     1.05   0.295    -1.044914       3.409
    black_hz |   1.822449   .7688497     2.37   0.020     .2958799    3.349018
       bp_hz |  -.1101685   .6432659    -0.17   0.864    -1.387388    1.167051
       _cons |   .3978872   .3647916     1.09   0.278     -.326415    1.122189
-------------+----------------------------------------------------------------
     sigma_u |  2.2801599
     sigma_e |  4.0486809
         rho |  .24080136   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(49, 94) =     0.91              Prob > F = 0.6407
```

```
. xi: xtreg ch_png_hz_er  st_pop_hz manuf_hz school_hz poverty_hz dad_hz black_hz bp_hz
ppe_hz , i(statenum) fe

Fixed-effects (within) regression               Number of obs      =       150
Group variable: statenum                        Number of groups   =        50

R-sq:  within  = 0.1144                         Obs per group: min =         3
       between = 0.0015                                        avg =       3.0
       overall = 0.0519                                        max =         3

                                                F(8,92)            =      1.49
corr(u_i, Xb)  = -0.1939                        Prob > F           =    0.1733

------------------------------------------------------------------------------
ch_png_hz_er |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   st_pop_hz |  -1.596762   .9225054    -1.73   0.087    -3.428937    .2354135
    manuf_hz |    .406041   .8194533     0.50   0.621    -1.221464    2.033546
   school_hz |   .6535592   1.356107     0.48   0.631    -2.039786    3.346904
  poverty_hz |   .2092734   .6482187     0.32   0.748    -1.078145    1.496692
      dad_hz |    1.33556    1.16816     1.14   0.256    -.9845074    3.655628
    black_hz |   2.006099   .8061815     2.49   0.015     .4049531    3.607245
       bp_hz |   -.062695   .6553324    -0.10   0.924    -1.364242    1.238852
      ppe_hz |   1.205118   1.341107     0.90   0.371    -1.458437    3.868672
       _cons |   .4935275   .4244248     1.16   0.248    -.3494168    1.336472
-------------+----------------------------------------------------------------
     sigma_u |  2.3138886
     sigma_e |  4.0250176
         rho |  .24839334   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(49, 92) =     0.90             Prob > F = 0.6604

xi: xtreg ch_png_hz_er  st_pop_hz manuf_hz school_hz poverty_hz dad_hz black_hz bp_hz
ppe_hz i.year, i(statenum) fe
i.year            _Iyear_2003-2009    (naturally coded; _Iyear_2003 omitted)
note: _Iyear_2009 omitted because of collinearity

Fixed-effects (within) regression               Number of obs      =       150
Group variable: statenum                        Number of groups   =        50

R-sq:  within  = 0.1681                         Obs per group: min =         3
       between = 0.0054                                        avg =       3.0
       overall = 0.1044                                        max =         3

                                                F(10,90)           =      1.82
corr(u_i, Xb)  = -0.1237                        Prob > F           =    0.0684

------------------------------------------------------------------------------
ch_png_hz_er |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   st_pop_hz |  -2.037217   .9222669    -2.21   0.030    -3.869461   -.2049727
    manuf_hz |   1.116114   .8554784     1.30   0.195    -.5834431    2.815671
   school_hz |   .5730169   1.366165     0.42   0.676    -2.141109    3.287142
  poverty_hz |   .0994065   .6683017     0.15   0.882    -1.228291    1.427104
      dad_hz |   1.034643   1.200483     0.86   0.391    -1.350326    3.419612
    black_hz |   1.659551   .8171379     2.03   0.045     .0361642    3.282938
       bp_hz |  -.3682929   .6571598    -0.56   0.577    -1.673855    .9372697
      ppe_hz |  -1.293108   2.363061    -0.55   0.586    -5.987741    3.401526
 _Iyear_2005 |  -6.578347   5.177365    -1.27   0.207    -16.86409    3.707392
 _Iyear_2007 |  -4.363207   2.282684    -1.91   0.059    -8.898156    .1717428
 _Iyear_2009 |          0  (omitted)
       _cons |   5.297553   3.417122     1.55   0.125    -1.491156    12.08626
```

```
-------------+----------------------------------------------------------------
     sigma_u |  2.2356262
     sigma_e |   3.944149
         rho |  .24316156   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(49, 90) =     0.86             Prob > F = 0.7097
```

APPENDIX B

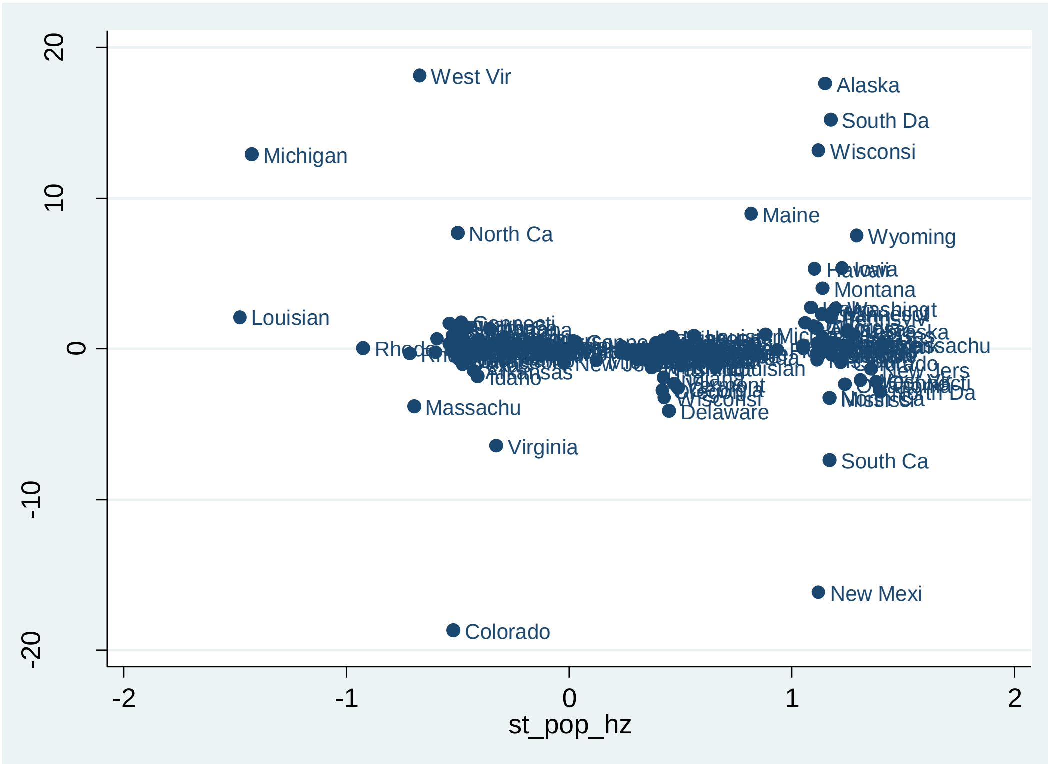
	Middle School Reading							
	State Assessment Scores				NAEP Assessement Scores			
State	2003	2005	2007	2009	2003	2005	2007	2009
Alabama	51	68	71	75	253	252	252	255
Alaska	67	79	85	81	256	259	259	259
Arizona	56	63	63	69	255	255	255	258
Arkansas	52	58	64	70	258	258	258	258
California	65	72	73	79	251	250	251	253
Colorado	88	64	63	64	268	265	266	266
Connecticut	76	76	76	77	267	264	267	272
Delaware	69	78	82	81	265	266	265	265
Florida	49	44	49	54	257	256	260	264
Georgia	81	83	89	96	258	257	259	260
Hawaii	38	37	59	68	251	249	251	255
Idaho	34	82	86	92	264	264	265	265
Illinois	65	74	83	83	266	264	263	265
Indiana	65	67	67	68	265	261	264	266
Iowa	69	71	73	74	268	267	267	265
Kansas	70	76	78	84	266	267	267	267
Kentucky	57	61	66	65	266	264	262	267
Louisiana	52	62	56	62	253	253	253	253
Maine	45	44	65	71	268	270	270	268
Maryland	59	66	67	81	262	261	265	267
Massachusetts	66	66	69	70	273	274	273	274
Michigan	61	72	77	83	264	261	260	262
Minnesota	63	63	63	66	268	268	268	270
Mississippi	61	56	80	49	255	251	250	251
Missouri	32	32	42	50	267	265	263	267
Montana	70	64	79	81	270	269	271	270
Nebraska	76	85	89	95	266	267	267	267
Nevada	49	50	56	60	252	253	252	254
New Hampshire	66	66	66	77	271	270	270	271
New Jersey	73	72	73	81	268	269	270	273
New Mexico	51	51	55	62	252	251	251	254
New York	45	48	57	69	265	265	264	264
North Carolina	85	87	87	66	262	258	259	260
North Dakota	78	79	79	74	270	270	268	269
Ohio	78	78	80	72	267	267	268	269
Oklahoma	79	81	79	67	262	260	260	259
Oregon	61	63	66	69	264	263	266	265
Pennsylvania	63	64	75	82	264	267	268	271
Rhode Island	40	59	59	64	261	261	258	260
South Carolina	66	74	71	67	258	257	257	257
South Dakota	78	79	78	74	270	269	270	270
Tennessee	84	86	92	92	258	259	259	261
Texas	83	82	89	93	259	258	261	260
Utah	75	75	79	81	264	262	262	266
Vermont	65	65	68	74	271	269	273	272
Virginia	53	53	49	45	268	268	267	266
Washington	47	69	68	59	264	265	265	267
West Virginia	80	80	80	61	260	255	255	255
Wisconsin	79	84	84	83	266	266	264	266
Wyoming	39	39	75	64	267	268	266	268

Missing Scores

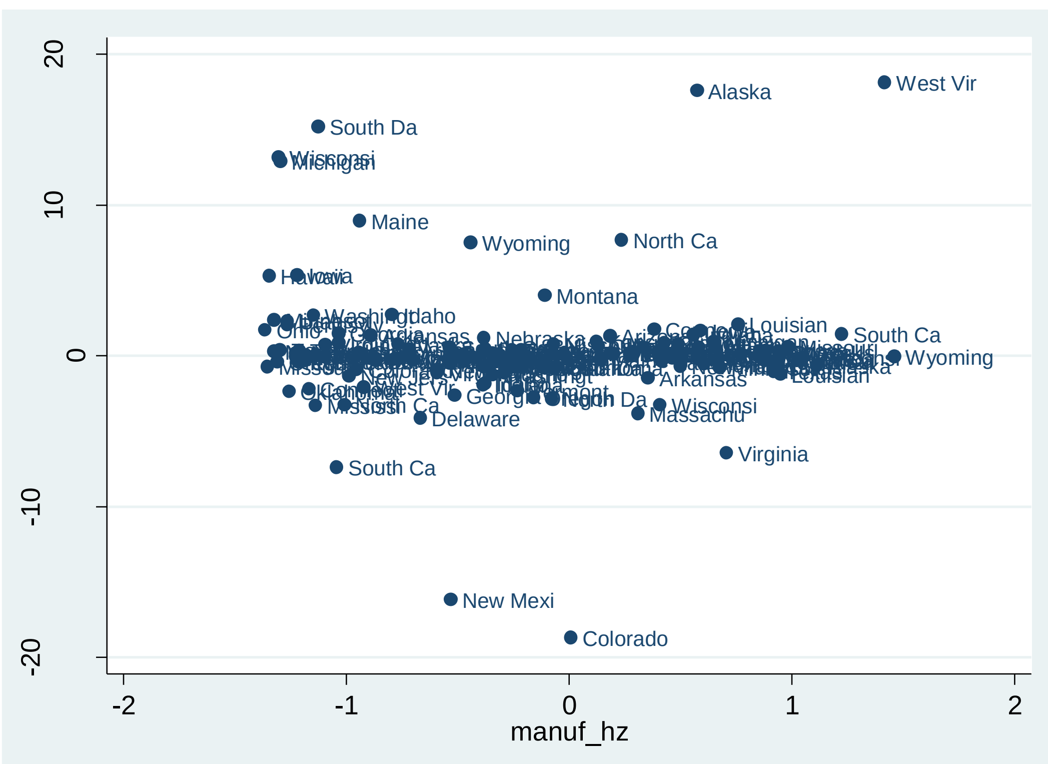
APPENDIX C

	Elementary School Reading							
	State Assessment Scores				NAEP Assessement Scores			
State	2003	2005	2007	2009	2003	2005	2007	2009
Alabama	53	83	84	87	207	208	216	216
Alaska	73	78	79	78	212	211	214	211
Arizona	57	67	70	73	209	207	210	210
Arkansas	61	52	60	70	214	217	217	216
California	71	75	77	83	206	207	209	210
Colorado	87	69	68	69	224	224	224	226
Connecticut	68	68	71	70	228	226	227	229
Delaware	79	84	81	81	224	226	225	226
Florida	63	67	69	71	218	219	224	226
Georgia	90	92	85	93	214	214	219	218
Hawaii	42	51	61	62	208	210	213	211
Idaho	87	84	81	87	218	222	223	221
Illinois	63	68	74	73	216	216	219	219
Indiana	73	75	74	75	220	218	222	223
Iowa	70	79	80	81	223	221	225	221
Kansas	68	76	80	83	220	220	225	224
Kentucky	62	68	72	74	219	220	222	226
Louisiana	58	52	58	71	205	209	207	207
Maine	49	53	67	71	224	225	226	224
Maryland	75	81	85	86	219	220	225	226
Massachusetts	56	50	55	53	228	231	236	234
Michigan	75	83	84	84	219	218	220	218
Minnesota	71	71	71	74	223	225	225	223
Mississippi	88	88	95	52	205	204	208	211
Missouri	34	35	46	47	222	221	221	224
Montana	76	75	80	81	223	225	227	225
Nebraska	78	84	89	94	221	221	223	223
Nevada	43	42	51	50	207	207	211	211
New Hampshire	75	75	75	78	228	227	229	229
New Jersey	77	81	80	63	225	223	231	229
New Mexico	51	51	54	51	203	207	212	208
New York	64	70	68	77	222	223	224	224
North Carolina	81	82	85	69	221	217	218	219
North Dakota	85	87	89	77	222	225	226	226
Ohio	66	76	80	82	222	223	226	225
Oklahoma	63	91	90	63	214	214	217	217
Oregon	80	82	71	76	218	217	215	218
Pennsylvania	58	63	70	73	219	223	226	224
Rhode Island	61	63	63	68	216	216	219	223
South Carolina	76	79	82	75	215	213	214	216
South Dakota	85	87	88	77	222	222	223	222
Tennessee	84	87	87	90	212	214	216	217
Texas	81	79	84	84	215	219	220	219
Utah	76	76	77	77	219	221	221	219
Vermont	69	69	68	69	226	227	228	229
Virginia	63	56	51	54	223	226	227	227
Washington	66	79	76	73	221	223	224	221
West Virginia	81	81	83	64	219	215	215	215
Wisconsin	82	82	81	81	221	221	223	220
Wyoming	44	47	75	63	222	223	225	223

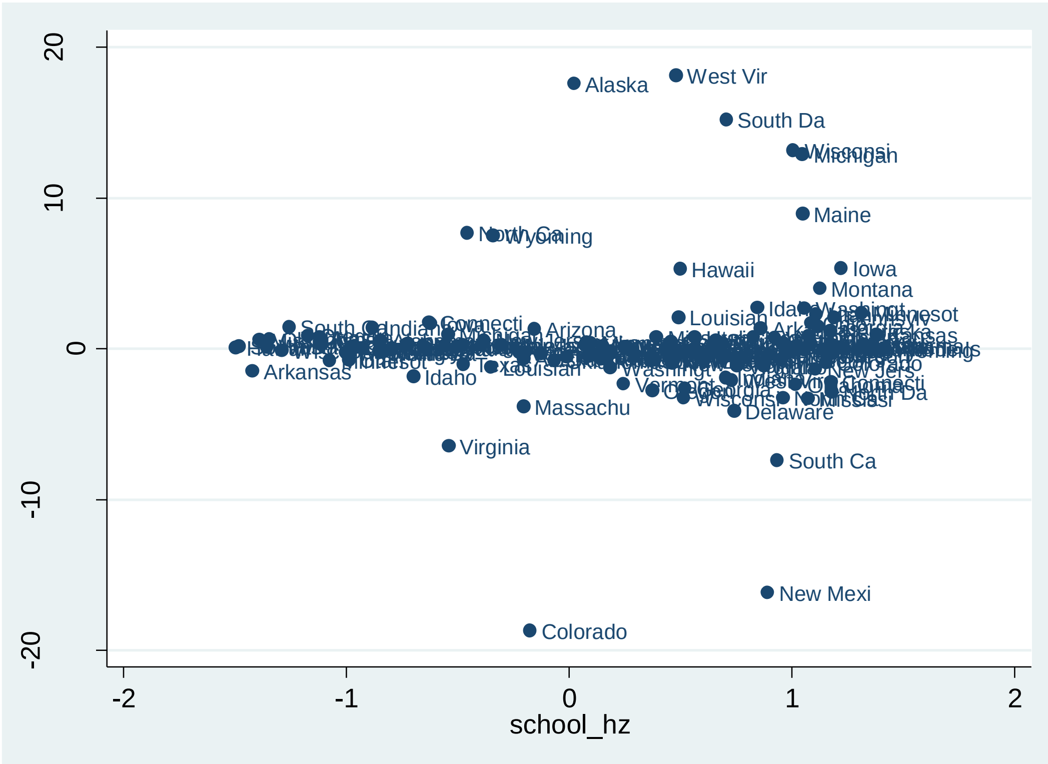
Appendix D

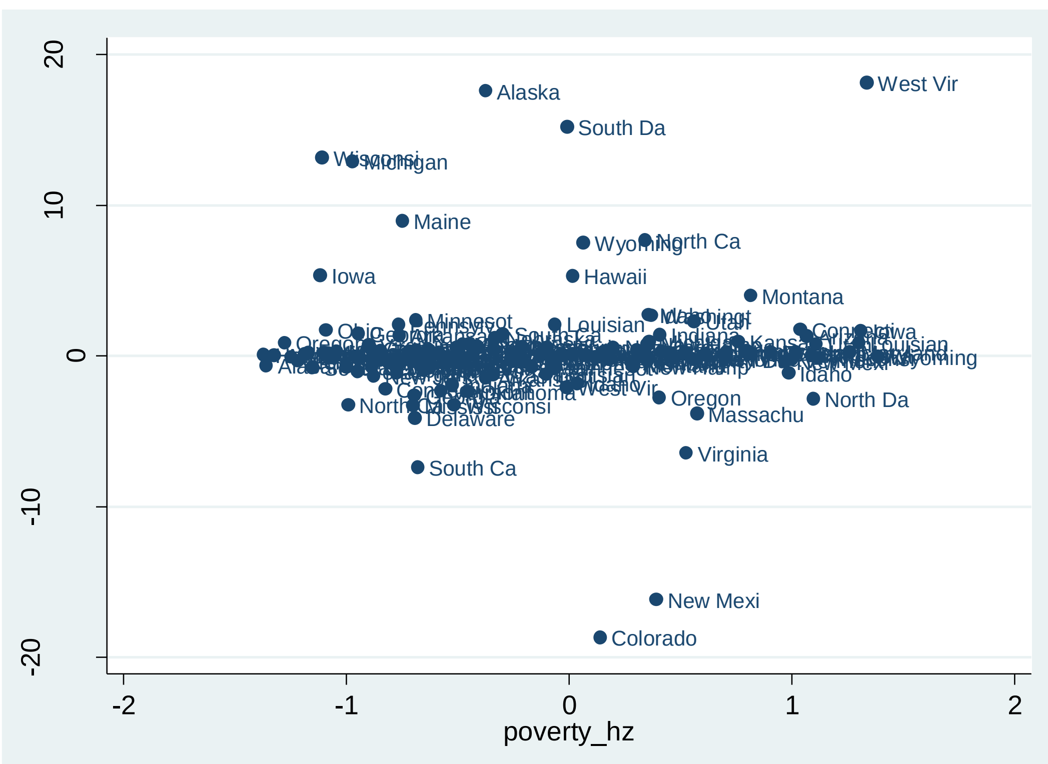


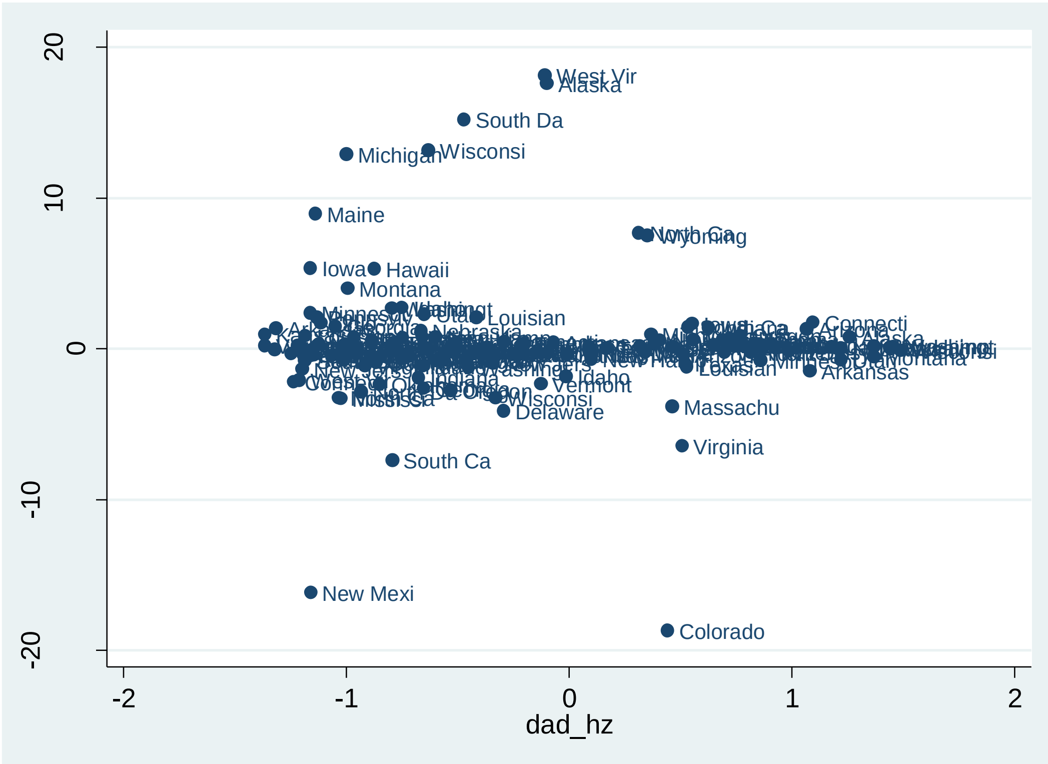
20
10
0
-10
-20
-2
-1
0
1
2
st_pop_hz
West Vir
Alaska
South Da
Wisconsi
Michigan
Maine
North Ca
Wyoming
Montana
Louisian
Massachu
Virginia
Delaware
South Ca
New Mexi
Colorado



20
10
0
-10
-20
-2
-1
0
1
2
manuf_hz
West Vir
Alaska
South Da
Maine
Wyoming
North Ca
Montana
Louisian
South Ca
Wyoming
Arkansas
Wisconsi
Massachu
Delaware
Virginia
South Ca
New Mexi
Colorado


20
10
0
-10
-20
-2
-1
0
1
2
school_hz
Alaska
West Vir
South Da
Michigan
Maine
Wyoming
Hawaii
Iowa
Montana
Louisian
Arizona
Arkansas
Idaho
Louisian
Massachu
Virginia
Delaware
South Ca
New Mexi
Colorado


20
10
0
-10
-20
-2
-1
0
1
2
poverty_hz
Alaska
West Vir
South Da
Michigan
Maine
North Ca
Iowa
Hawaii
Montana
Louisian
Idaho
Oregon
North Da
Massachu
Virginia
Delaware
South Ca
New Mexi
Colorado



20
10
0
-10
-20
-2
-1
0
1
2
dad_hz
West Vir
Alaska
South Da
Michigan
Wisconsi
Maine
Wyoming
Iowa
Hawaii
Montana
Louisian
Connecti
Nebraska
Idaho
Vermont
Wisconsi
Delaware
Arkansas
Massachu
Virginia
South Ca
New Mexi
Colorado