5-9-2014

The Impact of Georgia's Accountability System on School Performance and Subgroup Populations

Ashley Custard

PERMISSION TO BORROW

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this dissertation is:

Ashley Custard
391 J Ralph McGill Blvd. NE
Atlanta, GA 30312

The director of this dissertation is:

Dr. Mary Beth Walker
Department of Economics
Andrew Young School of Policy Studies
Georgia State University
P.O. Box 3992
Atlanta, GA 30302-3992

Users of this dissertation not regularly enrolled as students at Georgia State University are required to attest acceptance of the preceding stipulations by signing below. Libraries borrowing this dissertation for the use of their patrons are required to see that each user records here the information requested.

Name of User	Address	Date	Type of use (Examination only or copying)

THE IMPACT OF GEORGIA'S ACCOUNTABILITY SYSTEM ON SCHOOL PERFORMANCE AND SUBGROUP POPULATIONS

BY
ASHLEY LYNN CUSTARD

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2014

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair: Dr. Mary Beth Walker

Committee: Dr. Rachana Bhatt
Dr. Christine Roch
Dr. David Sjoquist

Electronic Version Approved:
Mary Beth Walker, Dean
Andrew Young School of Policy Studies
Georgia State University
May 2014

Acknowledgments

I am grateful to my advisor, Dr. Mary Beth Walker, for her continued guidance, patience and support during this process. Her belief in this project was invaluable to my completion.

I would like to thank Dr. Rachana Bhatt, Dr. Christine Roch, and Dr. David Sjoquist for serving as members of my committee. I recognize the time and effort required to do so and am grateful for your contributions.

I would also like to thank all of the faculty and staff within the Andrew Young School and Georgia State as a whole who have made my experience unforgettable.

To my mother, Brenda Mitchell, I would not be the person I am today without your love, sacrifice, and support. I am thankful for all that you have done for me, and am honored to be your daughter. I thank my father, Marcellus Custard Jr., for being a part of this journey. Your words of encouragement were instrumental.

To my grandparents, Elmar and Lloyd Wakefield, thank you for your unwavering love and support. You encouraged me to "Shine Bright," and I hope to make you proud. I thank my grandparents, Rose and Marcellus Custard Sr., for ensuring that I knew I was loved no matter the miles between us.

To my Aunties, Uncles, and Cousins: Thank you for teaching me to Love Big, Laugh Loud, and Dance Hard - In that order.

I am grateful to a countless list of friends who have helped me through this process. In particular I'd like to thank Brook Jones, Jennifer Olin, Valery Parham, Shequetta Perry, Meya Tidwell, and Andrea White. To be surrounded by such caring, intelligent, and hilarious women is truly a blessing.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

THE IMPACT OF GEORGIA'S ACCOUNTABILITY SYSTEM ON SCHOOL PERFORMANCE AND SUBGROUP POPULATIONS

By

ASHLEY LYNN CUSTARD

MAY 2014

Committee Chair: Dr. Mary Beth Walker

Major Department: Economics

This dissertation examines the impact of Georgia's accountability system on both school and student performance. We focus on two components within Georgia's accountability system – the sanctioning of failing schools and binding subgroup requirements. Schools within Georgia become subject to sanctions upon two consecutive years of failing to show Adequate Yearly Progress (AYP). The subgroup binding requirements, introduced by the No Child Left Behind (NCLB) Act, hold schools independently responsible for the performance of given subgroups contingent upon enrollment.

The first question of this dissertation examines the factors that influence a school's ability to meet assessment standards. We examine the relative importance of school characteristics, as they relate to accountability components, in determining AYP in practice. A binary response model is used as AYP is determined on a pass/fail basis. More specifically, we apply a correlated random effects probit model with a Chamberlain-Mundlak adjustment. The second question of this dissertation examines the impact of binding requirements on subgroup performance, where subgroup performance is defined as the percentage of students scoring at or above proficiency.

We employ a regression discontinuity design that compares the performance of bounded and unbounded subgroups to determine the treatment effect. Each question of this dissertation is addressed through evaluating both mean and distributional effects.

We find that imposing sanctions on failing schools has a positive impact on future performance. However, increasing the number of binding requirements has a negative impact on a school's probability of passage. This result suggests that heterogeneous schools, or schools with several large subgroup populations, are negatively impacted by the requirement. While we find that accountability components have a statistically significant impact on probability of AYP passage, factors related to school resources and quality appear to have a greater influence.

The mechanism for the negative impact of binding requirements remains unidentified as we also find that binding requirements have a slight positive impact on individual subgroup performance. The magnitude of this impact is dependent upon the subgroup examined, school type, and position of the subgroup within the Meets/Exceeds distribution. Overall, our results suggest the need for re-examination of the binding requirements as a method of targeting disadvantaged populations.

Chapter I

INTRODUCTION

Since the passage of No Child Left Behind (NCLB) in 2002, determining the impact of accountability systems has become increasingly important. Previous legislation focused on providing greater resources, but the emphasis has sense shifted to “standards, testing, and accountability” (Hanushek & Raymond 2005). One of the primary goals of NCLB, and any other accountability system, is to increase the performance of historically disadvantaged groups such as minorities and those of low socio-economic status. The persistence of the achievement gap, an issue of both class and race, highlights the need for policies targeting these populations. NCLB required the implementation of statewide accountability systems that applied uniform standards to all students and held schools independently accountable for the achievement of given subgroups. The objective of this dissertation is to investigate the success of NCLB components in Georgia, and determine whether holding schools responsible for subgroup populations (dependent on student enrollment) has been effective in increasing academic achievement within these groups. The results found reflect the influence of accountability structures on performance, as measured by subgroup achievement and ability to make Adequate Yearly Progress (AYP).

Under NCLB all states were required to create statewide accountability programs promoting improvement in student achievement. Those states with accountability systems in place prior to NCLB’s passage were forced to alter these systems to comply with federal requirements. Federal law mandated annual testing of all students in grades 3-8 and the testing of all high school students at least once during their high school academic career. The results of these tests are made available to the public and used to determine Adequate Yearly Progress, a measure evaluating overall school performance. A major component of AYP examines the percentage of students scoring at or above proficiency on state-administered standardized exams. Schools not meeting these criteria are labeled as “Needs Improvement” and become subject to sanctions. Available sanctions include school choice, the provision of mandatory tutoring, dismissal of faculty and/or administration, as well as government takeover. The premise of accountability systems being that threat of sanctions will alter school behavior.

Georgia’s response to the implementation of NCLB was the passage of the Single Statewide Accountability System (SSAS) in 2005. Scoring standards were applied to all students collectively as well as certain targeted populations or subgroups. The original conditions of NCLB mandated that Georgia show improvement in student achievement for every student category, including subgroups, until reaching 100 percent proficiency in the 2013-2014 school year[1]. Georgia’s designated subgroups included Asian/Pacific Islander, Black, Hispanic, American Indian/Alaskan, white, multi-racial, students with disabilities (SWD) [2], students with limited English proficiency (LEP), and the economically disadvantaged (ED).

The first question of this dissertation focuses on factors that influence AYP status, including but not limited to the components of the accountability system itself. While the ultimate determination of AYP uses three well defined measures, accountability structures may indirectly influence performance. In essence we are examining the relative importance of school

[1] In 2013, Georgia replaced the Adequate Yearly Progress objectives of NCLB with its own College and Career Readiness Performance Index (CCRPI). The CCRPI also includes measures of achievement for a set of subgroups – the economically disadvantaged, students with disabilities, and those with limited English proficiency.

[2] States may use modified achievement standards or alternate assessments to track the progress of students with disabilities. For those students who are required participate in Georgia’s testing procedures but have significant cognitive disabilities the Georgia Alternate Assessment (GAA) is administered.

characteristics, as they translate to accountability components, in determining AYP in practice. A binary response model is used to measure the effect of these factors since AYP is determined on a pass/fail basis. Data elements come from the Georgia Department of Education (GADOE) as well as the National Center for Education Statistics (NCES). Both static and dynamic models are estimated as the impact of underperformance and subsequent sanctions may carry over into future years.

The individual performance of a subgroup is used in a school's AYP determination if and only if student enrollment within that particular subgroup reaches a given threshold. The number of enrolled students must equal forty or total ten percent of students enrolled in AYP grades in order for a school to be held independently accountable for a particular subgroup. Therefore, all schools within the state will not be held accountable for every group. It is this aspect of SSAS that the second question of this dissertation will exploit in exploring the impact of accountability on subgroup achievement. A sharp regression discontinuity design is used to compare the achievement of those subgroups with enrollment just below the cutoff to those with enrollment just exceeding the cutoff. Subgroups and consequently schools with enrollment levels exceeding the threshold are categorized as 'treated' because they face greater pressure to improve achievement within binding subgroups. The data used contain detailed AYP reports for each school in Georgia dating from 2004-2011. Additional information on student body demographics and teacher characteristics was obtained from the GADOE and NCES.

When investigating the impact of accountability systems on student performance several key questions arise: Are these systems effective in improving student achievement? If effective, what factors contribute to success? Are there differential effects by race/ethnicity or location in the achievement distribution? The accountability literature consists of studies undertaken in an attempt to answer these questions. One branch of literature focuses on comparisons of accountability systems across states (Carnoy & Loeb 2002; Hanushek & Raymond 2003, 2005). These studies construct indices measuring the relative strength of each state's system, and compare student outcomes in 'weak' systems to those in 'strong' systems. These national studies serve to exploit variation between states, as all schools within a state face simultaneous implementation. Therefore, within state differences are not explored. This dissertation contributes to the existing literature by using variation in the *application* of accountability within the state of Georgia as opposed to variation in the time or method of implementation across states.

Studies focusing on individual state systems may examine state-specific components and therefore produce results that are highly specific and not generalizable. This dissertation investigates components within Georgia's system that are found nationwide –sanction threat and subgroup accountability. While the exact nature of each may vary from state to state, NCLB mandated that both be incorporated into all state systems.

The application of the regression discontinuity within this study is also distinctive. Though many authors have used a regression discontinuity design to evaluate the impact of issuing 'grades' to schools based on performance (Figlio & Rouse 2006; Rockoff & Turner 2010; Chiang 2009; Chakrabarti 2013) or to examine the impact of performance standards on achievement (Hemelt 2011), few have used the design to examine differences in populations for which schools are held accountable.

Again, this dissertation addresses two main questions: What is the impact of given accountability structures, the application of sanctions and the number of subgroup binding requirements, on probability of AYP passage? How do binding requirements impact subgroup performance? Each question is tackled by examining Georgia's accountability system, analyzing evidence found within the literature, presenting the data available, describing methodologies used, evaluating results, and reporting the conclusions drawn. Consistent with the literature (Carnoy & Loeb 2002; Hanushek & Raymond 2005; West & Peterson 2006; Figlio & Rouse 2006; Hastings & Weinstein 2008; Ahn & Vigdor 2013) we find that imposing sanctions has a positive and statistically significant impact on a school's probability of passing AYP. However, our findings also show that increasing the number of binding subgroup requirements has a negative impact on the probability of AYP passage. The mechanism for this negative relationship remains unidentified as we also find that binding requirements have a positive influence on the performance of subgroup populations.

The remainder of this dissertation proceeds as follows: Chapter II describes the theoretical framework for accountability systems, more specifically it speaks to the theories that support the use of accountability systems in improving student achievement; Chapter III examines the impact of accountability components on probability of AYP passage; Chapter IV analyzes the impact of binding requirements on subgroup performance; and Chapter V presents the conclusions of this study.

Chapter II

THEORETICAL FRAMEWORK

The education production function (EPF) maps the relationship between educational inputs and student achievement. The objective is to maximize student performance given a set of constraints. While there is no clear consensus on the factors relevant to determining student performance (Hanushek & Raymond 2005), the simplest form of the EPF is as follows:

$$Y_s = f\,(B_s, P_s, SC_s), \tag{1}$$

where Y_s represents the selected student outcome, B_s are student and family background characteristics, P_s are characteristics of peers within the school, and SC_s are school inputs. Socioeconomic status and parents' education are often used as measures of student background characteristics (Hanushek 2008). School demographics and/or school-level measures of achievement can serve as proxies for peer effects (Hanushek 2008). For example, measures of racial composition can be used to test the influence of minority concentration on achievement. School-level measures of income, such as the percentage of students qualifying for Free or Reduced Lunch, and school-level measures of performance, such passage rates on standardized tests, can also serve as proxies for peer effects. School inputs or resources are usually defined as teacher experience, pupil/teacher ratio, expenditures, etc.

The education production function establishes a relationship between student characteristics, school characteristics, and educational achievement. The introduction of an accountability system creates pressure for increased academic performance. Schools must adjust inputs within the production function to maximize output under an altered set of constraints. These adjustments can take several forms. Due to limitations in the ability of schools to significantly alter their student populations, modifications in student characteristics or family background as a means to improve achievement are limited. Literature suggests that the main mechanisms of increasing academic achievement occur through changes to school policy. For instance teachers may focus on marginal students, or students whose scores place them slightly below the passage point (Gillborn & Youdell 2000; Springer 2007; Hamilton et al. 2007; Rouse, Hannaway, Goldhaber, & Figlio 2013). Schools may reallocate resources to increase focus on tested subjects or test-specific skills (Jacob 2002; Reback 2008), increase the resources made available to teachers (Hemelt 2011), or extend the time devoted to instruction (Hamilton et al. 2007; Rouse, Hannaway, Goldhaber, & Figlio 2013; Hemelt 2011). While the purpose of this dissertation is not to identify the individual school policies through which improvement occurs, it is important acknowledge these mechanisms as it helps to define accountability systems within the structure of the education production function.

As noted above, accountability systems assume that schools have the capacity to respond to incentives and improve student outcomes through manipulation of the education production function. The threat of sanctions and the disclosure of quality ratings are meant to incentivize schools to increase student performance. However it is important to note that the degree to which institutions respond to accountability threats is dependent upon several factors. These factors include threat credibility, dissemination of information, institution autonomy, and availability of resources. Threat credibility refers to the perceived likelihood of sanctions. If states do not consistently impose sanctions, this negates the perceived threat and provides no incentive for improvement. NCLB and consequently SSAS mandate corrective action on schools who do not meet statewide standards. The sanctions are hierarchical in nature, with the most severe resulting in school restructuring. During the six year period from 2006 to 2011, an average of 117 schools per year faced restructuring in Georgia.

The dissemination of information refers to the degree to which student achievement results are made publicly available. In making achievement information available, the issue of asymmetric information between schools and their surrounding communities is reduced. Consumers are allowed to judge the quality of the good being provided, placing external pressure on schools to increase student achievement. Under NCLB low-performing schools are required to facilitate and fund the transfer of any student who wishes to do so. NCLB also required the publication of report cards documenting academic achievement. However, Georgia began publishing report cards in the year 2000, several years prior to the passage of NCLB.

Autonomy and the availability of resources dictate the extent to which schools are able to alter their curriculum and staff to adhere to accountability standards. For example, NCLB contains language which focuses on the improvement of teacher quality. The act mandates certification requirements for any teacher teaching a core subject such as math or reading. Each state was also required to implement programs to guarantee that all teachers were ‘highly qualified’ by the end of the 2005-2006 school year. In the state of Georgia, a ‘highly qualified’ teacher is defined as having a bachelor’s degree, full certification, and verified mastery of teaching skills and subject knowledge. The ability to hire and retain these teachers is influenced by the resources available, which can itself be influenced by previous performance.

The research presented here uses the education production function framework to examine two questions concerning the impact of accountability systems. The first studies the influence of accountability components as defined through school characteristics on AYP status. The second explores the impact of binding subgroup requirements on a different outcome, student performance as measured through achievement rates. While the two outcomes examined differ, the mechanisms for improvement remain unchanged. Under accountability systems schools are charged with improving student performance through the manipulation of educational inputs while facing given accountability structures and limited availability of school resources.

Chapter III

ADEQUATE YEARLY PROGRESS AND ACCOUNTABILITY COMPONENTS

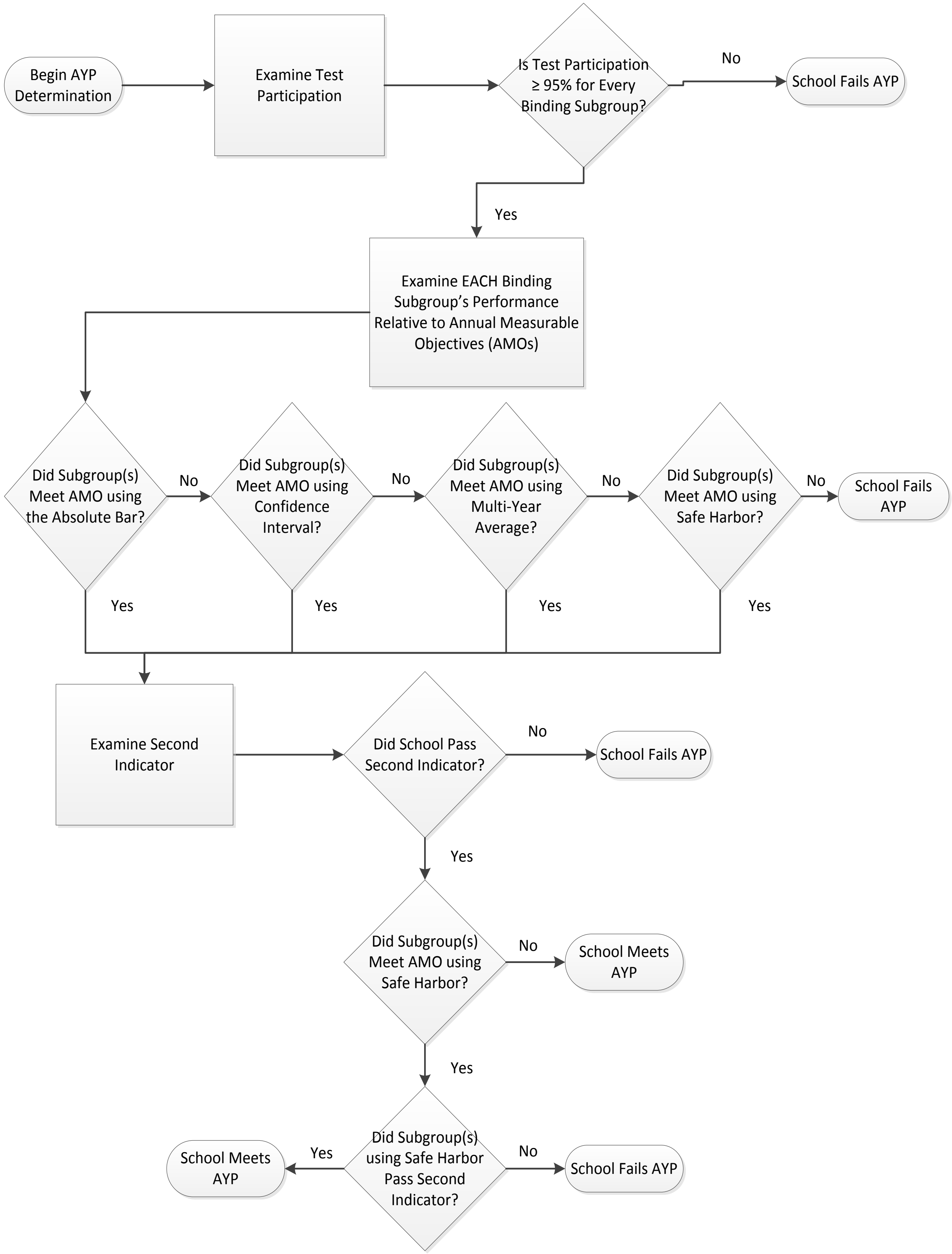
History of Accountability in Georgia

Prior to the passage of NCLB in 2002, the state of Georgia was in the process of developing its own accountability system entitled the “A+ Education Reform Act of 2000.” Like many other systems the focus of “A+” centered on improving teacher quality, decreasing the number of dropouts, and increasing student test scores. However, it also contained measures to decrease school violence, increase community involvement, and increase the level of integration between educational agencies within the state. This last goal was achieved through the creation of the Education Coordinating Council. The act also created the Office of Education Accountability (OEA) – later renamed the Governor’s Office of Student Achievement - whose primary responsibilities focused on developing and implementing Georgia’s accountability system. The original plan dictated that schools receive a grade of A-F based on the standards created by the OEA. However, prior to the enactment of A+, NCLB was passed and Georgia’s system was revamped to adhere to federal guidelines.

The new system, SSAS, held many of the same goals as its predecessor; however it went about achieving these goals in a different manner. In SSAS, AYP status was defined as a binary indicator of performance as opposed to scaled letter grades. Schools either pass or fail their AYP determination. Those who meet AYP have satisfied the following criteria: 1.) 95 % testing participation for every subgroup – including the 'all student' subgroup; 2.) Achievement in English/language arts (ELA) and mathematics in accordance with Annual Measurable Objectives; 3.) Passage of a Second Indicator. The first criterion, 95% participation in all subgroups, is meant to minimize the exclusion of underperforming students as a way to ensure success. Annual Measurable Objectives (AMOs) are benchmarks stating the minimum percentage of students scoring at or above proficiency required for a school to qualify as having made adequate progress. For instance, the AMO in math for high school students in the 2010-2011 school year was 81.2%, meaning that in order to pass at least 81.2% of students tested must score at or above proficiency[3]. It is important to note that in Georgia, if schools fail to meet the absolute bar set by the AMO, there are three other alternatives: confidence intervals, multi-year averaging, and a safe harbor method. The safe harbor method uses performance from the previous year as a benchmark and requires a decrease of ten percent in the number of students not meeting proficiency. The third criterion focuses on a second indicator of overall performance that is not generally applied to subgroup populations but to the school as a whole. Only those subgroup populations meeting AMO through the safe harbor method are also held accountable for the second indicator. Elementary and middle schools are allowed to use a variety of indicators such as attendance rates or the percent of students exceeding the standard in a given subject area as their second indicator. For high schools the second indicator is the graduation rate. The use of graduation rate as opposed to attendance or achievement scores may place high schools at a disadvantage in terms of ability to pass AYP[4]. Figure 1 is a visual representation of

[3] A listing of Annual Measurable Objectives (AMOs) by year and subject is available in Appendix A
[4] High School AYP Second Indicator Graduation Rate Standards are available in Appendix B

Figure 1. Adequate Yearly Progress Determination



the AYP determination process. After all calculations are performed an accountability profile is constructed for each school containing its AYP report, a performance index recognizing those schools making the greatest gains in achievement, and performance highlights listing each school's best performance indicators.

Once constructed accountability profiles are used to determine rewards received or sanctions administered. Those schools demonstrating the greatest gains and those with the highest absolute scores are eligible for rewards under SSAS. The law states that rewards can be both monetary and non-monetary. Banners of recognition, visits from the governor, banquets, and increased autonomy are the non-monetary awards offered. And while the possibility of financial rewards for above average performance is written into Georgia law, very few have actually been given. Title I schools who have made AYP for four consecutive years are the only school type that has received a financial reward for progress made. All other schools exhibiting excellence in performance have received banners of recognition or some other form on non-monetary reward.

In the 2002-2003 school year, the first year in which Georgia calculated Adequate Yearly Progress, 63.7% of Georgia schools met AYP while the state as a whole failed. The first year of the panel used for this analysis is 2004, and the percentage of schools achieving AYP increased to 79.76% in this year. As can be seen in Table1, the percentage of schools passing AYP peaked in the 2009-2010 school year with approximately 86% of schools meeting the standard. In 2011, approximately 27% of schools in Georgia failed to make AYP and faced possible sanction. This was the largest percentage of failures since the inception of NCLB in 2003. The sanction structure of SSAS was taken directly from NCLB. Whether or not a school faces sanctions is dependent not only on current year's performance, but on the previous year's status as well.

Table 1. Adequate Yearly Progress (AYP) Percentages by Year, 2004-2011

Year	# of Schools	Failed AYP	Met AYP
2004	1,932	20.24%	79.76%
2005	1,966	17.75%	82.25%
2006	2,011	20.14%	79.86%
2007	2,021	17.42%	82.58%
2008	2,052	19.40%	80.60%
2009	2,104	13.64%	86.36%
2010	2,114	21.85%	78.15%
2011	2,179	26.66%	73.34%

NOTE: Compiled using Georgia Department of Education, Needs Improvement Reports, 2004-2011

"Needs Improvement Status" contains the following five categories:

1. Distinguished – Has met standards for three consecutive years
2. Adequate Progress – Has met standards for two consecutive years
3. Adequate Progress, Did Not Meet – Met standards for the two previous years but failed to meet in the most recent.
4. Needs Improvement, Made AYP – Met standards for the most current year but was previously classified as Needs Improvement.
5. Needs Improvement – Has failed to meet AYP for two consecutive years.

During the first year of "Needs Improvement" schools must facilitate the transfer of students who wish to attend a better performing school and develop a school improvement plan. The next level of sanctions requires schools to provide supplemental services to low-achieving students. Each subsequent stage shows an increase in government intervention culminating in yearly evaluations by both the school system and Georgia Department of Education. In order to move out of the Needs Improvement category, schools must satisfy AYP for two consecutive years. Table 2 displays the number of schools present within each Needs Improvement category by year.

Table 2. Percentage of Schools in each Needs Improvement Status by Year, 2004-2011

School Year	# of Schools	Needs Improvement	Needs Improvement - Made AYP	Adequate - Did NOT Make AYP	Adequate	Distinguished
2004	1,932	14.60%	6.73%	5.02%	60.66%	12.99%
2005	1,966	9.82%	7.73%	7.63%	18.92%	55.90%
2006	2,011	10.04%	4.92%	9.70%	12.33%	63.00%
2007	2,021	10.74%	4.90%	6.04%	11.88%	66.45%
2008	2,052	9.26%	4.78%	9.89%	12.23%	63.84%
2009	2,104	8.46%	4.23%	5.04%	15.54%	66.73%
2010	2,114	9.37%	2.93%	11.83%	11.45%	64.43%
2011	2,179	13.22%	3.17%	13.03%	10.56%	60.03%

NOTE: Table compiled using Georgia Department of Education, Needs Improvement Reports, 2004-2011

Figures 2 and 3 document the passage rate for the Criterion-Referenced Competency Tests (CRCT)[5] and Georgia High School Graduation Tests (GHSGT) in math and English/language arts from 2003 through 2011. Students in third through eighth grades are tested in English/language arts and mathematics using CRCTs. High school students, 11th graders specifically, are tested in the same subject areas using the GHSGTs. For schools serving grades not covered by the CRCT or GHSGT, AYP determinations are made using other forms of assessments as the U.S. Department of Education requires that all schools be subject to AYP evaluations. For example, End of Course Tests (EOCTs) are used for ninth grade centers, and the Kindergarten Assessment Program is used for Kindergarten-only schools.

At the inception of NCLB, approximately 78% of students taking the CRCT in English/ language arts met the standard, while only 71% of students met the standard in mathematics. As can be seen from Figure 3, the percentage of high school students scoring at or above proficiency

[5] Criterion-reference tests are designed to measure how well student achievement adheres to a specific curriculum, which in Georgia included Quality Core Curriculum (QCC) and Georgia Performance Standards (GPS). Norm-referenced tests measure more general subjects/skills that are taught throughout the country.

Figure 2. Meets/Exceeds Rate by Subject & Year, CRCT – All Students, 2003-2011

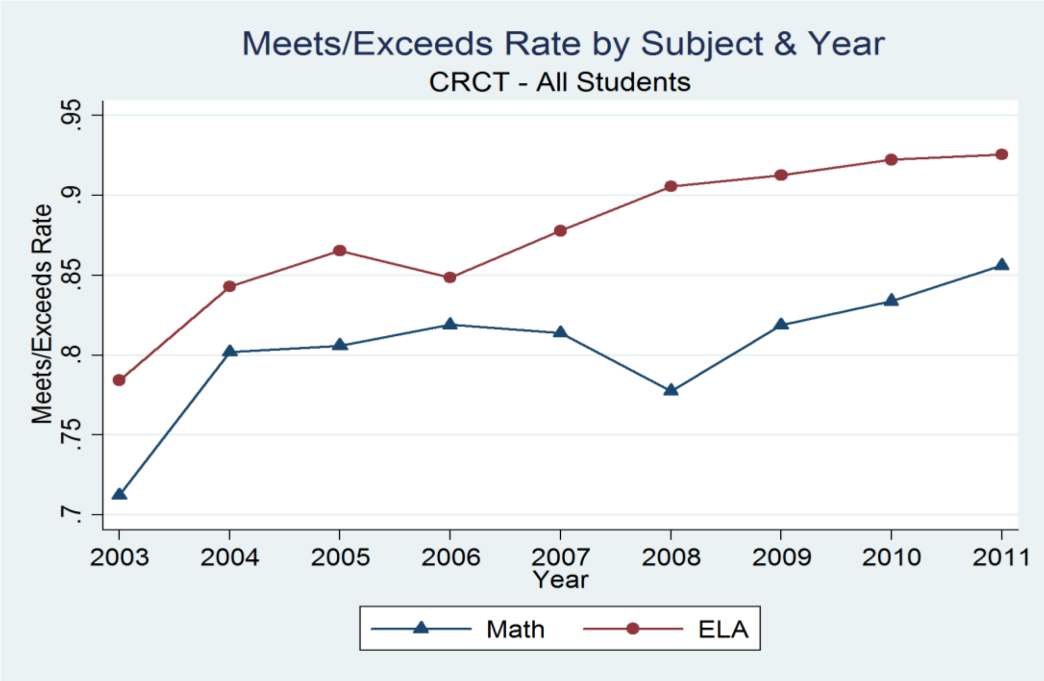
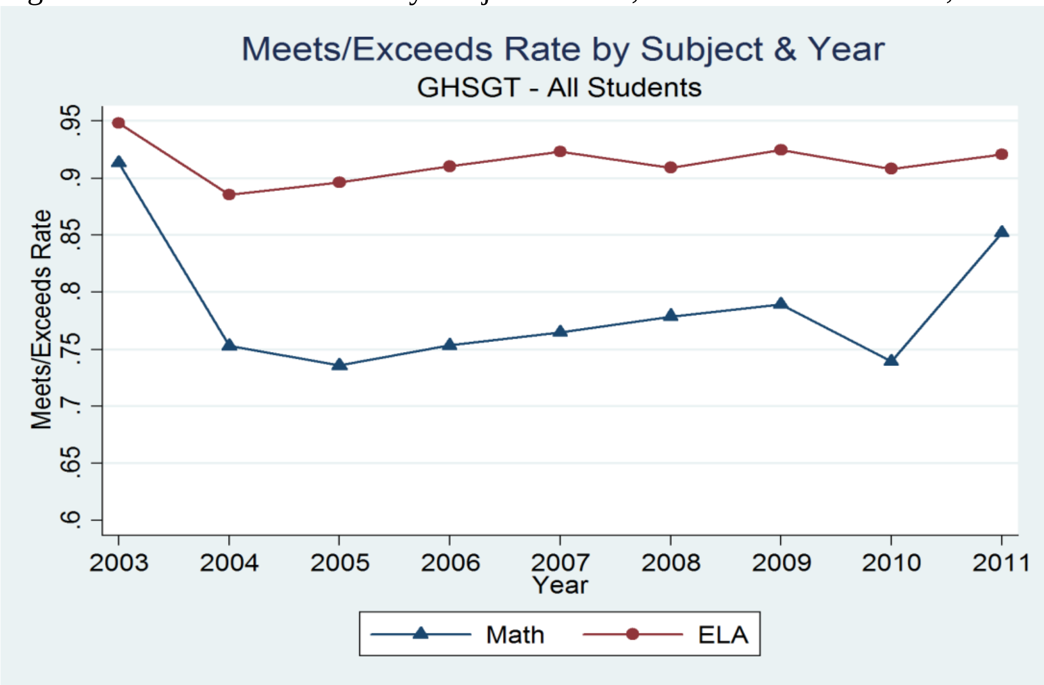


Figure 3. Meets/Exceeds Rate by Subject & Year, GHSGT – All Students, 2003-2011



in ELA hovered around 90% for the duration of the period. The passage rate for the mathematics portion of the GHSGT began at around 90% in 2003, but decreased to approximately 74% in 2004 after the introduction of the “Enhanced-GHSGT.” Both figures highlight that a larger percentage of students meet the standard in ELA than in mathematics. The difference in percentage passing is more pronounced for students taking the GHSGT than students taking the CRCT.

Literature Review – Impact of Accountability Systems on Achievement

National Studies

As stated previously, NCLB required that each state create an accountability system that held all schools within the state to a uniform set of standards. While NCLB mandated system creation, state governments were given discretion in implementation. This discretion created variation in the standards applied across states. The presence of pre-existing systems combined with variation in standards makes a national study of the impact of NCLB difficult. One approach has looked at trends in the National Assessment of Educational Progress (NAEP) and sought to determine if the implementation of NCLB had a noticeable impact. Ladd (2007) performs such a test, looking at trends in math and reading for both 4^{th} and 8^{th} graders from 1992 to 2007. She discovers the decline in eighth grade reading scores and growth in math scores follow the pre-NCLB trend. Therefore, neither eighth nor fourth grade scores provide clear evidence of a positive impact of NCLB. Cronin, Kingsbury, McCall, and Bowe (2005) also investigate the initial impact of NCLB on a national scale. Using individual level data from 200 districts in 23 states, the authors compare pre and post- NCLB scores to determine the act’s impact on performance levels and growth. Findings show that growth in student scores decreased for every ethnic group, grades with mandatory testing experienced greater gains than those without, and minority students fared worse than whites in both growth and levels of attainment.

Other studies have chosen to focus on the impact of accountability as a whole as opposed to simply focusing on NCLB. In 2002, Carnoy and Loeb conducted a study investigating the impact of the relative strength of an accountability system on student outcomes. The authors developed an index of accountability strength using the presence of state achievement tests, rewards and/or sanctions, and high school exit exams. The scale ranged from zero to five. Those states with no tests or state standards received zeros, and those with statewide tests, well-developed reward/sanction systems, and mandatory high school exit exams received fives. Georgia's pre-NCLB system of 1999-2000 warranted an index rating of two. While statewide testing and high school exit exams were present, the state lacked a system of rewards or sanctions. Using math scores at the eighth grade level, the authors determine that those states with 'stronger' systems saw greater improvements in student performance. This study not only showed that holding schools responsible increases performance, but that holding students accountable, in this case through high school exit exams, has a positive impact on performance.

Using NAEP data Hanushek and Raymond (2005) investigate the impact of state accountability systems on student achievement. The authors divide states into two categories, "Consequential" states that add direct consequences to underperformance, and "Report Card" states that simply publish results. The study focused on the performance of Black, Hispanic, and white eighth graders, and found that accountability programs had a positive impact. Those states which implemented consequential systems early showed the greatest improvement. Overall the authors find a significant gain in performance with the introduction of accountability.

State-specific Studies

Another branch of the literature focuses on state or district-specific systems. Jacob (2002) studies the impact of high-stakes testing in Chicago public schools. Implemented in the 1996-1997 school year, the system held students, teachers, and schools accountable. Students faced the elimination of 'social promotion.' A passing grade on state administered tests was now required for promotion of students in third, sixth, and eighth grades. Teachers and administrators were subject to dismissal or reassignment, and schools were subject to sanctions including restructuring if students consistently failed to meet state standards. As seen in most studies of this type the positive impact of accountability appears to be larger for math than for reading. One possible explanation being that the skills needed to increase math scores can be taught in the classroom, while the skill set for reading is more greatly influenced by outside factors. This study also examined whether improvement in scores occurred due to an increase in general or test-specific skills. If improvement was due to an increase in general skill, gains made should carry over to other state administered tests. Results showed no improvement in any other state administered test, and therefore implied that students increased their test-specific skills. In fact, Jacob found that accountability had a slightly negative impact on scores of alternate tests. Previous achievement had a noticeable impact on gains in tested subjects such as math and science, but little to no impact on gains in non-tested subjects. Suggesting that accountability systems cause shifts in resources towards tested subjects.

Rockoff and Turner (2010) study New York City's accountability system. In 2007 New York City implemented a program which assigned schools letter grades ranging from A-F based on student performance (30%), student progress (55%), and school environment (15%). The calculations made are separate from New York's NCLB determinations. Each school's performance is evaluated relative to similar schools within the city. Schools of the same type and schools with similar student populations are considered comparable to one another. Well-performing schools receive financial rewards such as increases in per pupil spending and principal bonuses. Low-performing schools are subject to sanctions including student transfer, closure, and other corrective action. Letter grades are released in September, and students are tested the following January leaving the schools several months to respond to their rating. The authors use the discontinuous nature of grade assignment to identify the impact of accountability on performance. Schools are assigned letter grades based on their relative position in the distribution of comparable schools, but are later able to obtain additional credit. The authors note that at least 161 schools received additional credit which moved them to a higher grade. The results show schools receiving a grade of D or F in math and those receiving an F in English showed significant improvement. The short-run impact of receiving an F was a reduction in the achievement gap between C and F schools. There was an 18% reduction in math, and a 20% reduction in English. Schools receiving D's saw a 16% reduction in the A-D school achievement gap in math.

Mechanisms for Improvement

Rouse, Hannaway, Goldhaber, and Figlio (2013) not only determine the impact of accountability on student achievement but investigate the mechanisms through which these improvements take place. Florida's system, like many others, assigns letter grades based on performance measures. Those schools receiving an F in the previous year saw a long-lasting increase in both math and reading scores, implying a positive impact of sanctions. The authors determine the proportion of gains made that can be attributed to changes in school policy. In response to accountability, schools increased the focus on low achieving students, increased resources made available to teachers, extended time devoted to instruction (summer school, tutoring, etc.), and decreased principal control. The authors also estimate the impact of school failure and the subsequent implementation of corrective school policy. They control for school-level variables such as enrollment, expenditure per student, percentage receiving Free or Reduced Lunch etc. Implementing corrective school policy accounted for over 15 percent of reading gains and 38 percent of gains in math.

Hamilton et al. (2007) find evidence suggesting that schools within the state of Georgia altered school policy in response to accountability pressure. Instructional time appeared to be reallocated from non-tested towards tested subjects in the time period immediately following the passage of the act. This effect was seen most prominently in elementary schools, as elementary teachers have more freedom in their schedule and can thus reallocate time more easily. The study also found that schools began to give additional attention to those students closest to the passage benchmark.

Response to Failure

How do schools respond to previous bad ratings? Do low-performing schools respond to low-ratings through improvements in achievement in later years? Several studies have found that failed or threatened schools show persistent improvements in achievement (Chiang 2009; Chakrabarti 2013). Hemelt (2011) finds negative impacts of failure in the short-run, but improved performance of failed subgroups in the long run. And similar to the Hamilton et al. (2007) results, accountability threats resulted in increased spending on instruction and teacher development. Figlio and Rouse (2006) find the scores of students in low-performing schools showed a larger increase than those of students in schools with high performance. Gains were seen in both low and high-stakes tests. However, gains in low-stakes reading exams were explained by differences in student characteristics, and low-stakes math gains were only present in grades that were also subject to high-stakes testing. The authors also note that voucher threats were not responsible for the improvements seen in math scores; instead they attribute the growth to other accountability pressures such as performance bonuses or the stigma of receiving a low grade. While voucher threat was dismissed by Figlio and Rouse (2006), others have introduced school choice as a primary mechanism for improvement (West & Peterson 2006; Hastings & Weinstein 2008). In a study investigating Florida schools West and Peterson (2006) find students at low-performing schools performed better than their counterparts attending schools not subject to accountability threats. The authors found that Blacks, those eligible for Free or Reduced Lunch, and those with low initial test scores were most affected. The authors also note that targeted sanctions are more effective than general systems. While these results highlight the threat of school choice as a mechanism for improvement, others have found that without credibility (i.e. availability of school choice options or increased competition) the threat of school choice is not sufficient (Hastings & Weinstein 2008; Chakrabarti 2013).

The evidence concerning the impact of NCLB and other accountability systems on student performance is mixed. While there is no clear evidence suggesting large overall gains in student achievement, certain components within the system appear to yield positive results. The strength of accountability systems, or their ability to impose sanctions, is positively correlated with student performance. In particular, students at under-performing schools subject to sanctions showed the greatest gains. However, it is important to note that not all sanctions yield the same results. Simply publishing results does not provide adequate incentive as the greatest gains were seen in schools subject to 'Consequential' sanctions. Ahn and Vigdor (2013) examine the impact of sanction severity and find similar results. The authors conclude that the strongest positive results are seen with the most severe sanctions, i.e. school restructuring, and that gains in sanctioned schools were concentrated among the lowest-performing students.

Data & Methodology – Adequate Yearly Progress

Data – Adequate Yearly Progress

This study uses school-level Needs Improvement reports and Certified Personnel Information for school years 2004 to 2011 obtained from the Georgia Department of Education (GADOE), as well as Common Core Data covering the same time period from the National Center for Education Statistics (NCES). The GADOE data documents the AYP and Needs Improvement Status for every Georgia school. AYP status is a coded binary indicator with a value equal to one if a school met all requirements and zero otherwise. Needs Improvement Status is a categorical variable documenting the level of sanction/intervention present within each school. Schools are categorized as Distinguished, Adequate, Adequate – Did Not Meet AYP, Needs Improvement – Made AYP, and Needs Improvement. Once a school has been designated as "Needs Improvement" they become subject to hierarchical sanctions, which begin with school choice and end with school restructuring.

This study uses a sample of 2,326 schools from the Needs Improvement Status reports provided by the GADOE. Schools were included in the sample based on the availability of covariates, and account for 94.5% of all schools within Georgia during this period. The sample is an unbalanced panel with approximately seventy-seven percent of schools present in all eight years. This leaves twenty-three percent of schools with observed time periods less than eight. Covariates were constructed using enrollment data from NCES, as well as AYP reports and Certified Personnel Information reports from GADOE. The number of binding subgroups, those populations subject to additional requirements, in both mathematics and ELA is a proxy for the amount of AYP pressure each school faces. As the number of subgroups for which a school is held accountable increases, so do the opportunities for failure. Using AYP reports, a binding subgroup count is obtained by summing the number of groups held accountable in both mathematics and English/language arts by year and school. Every subgroup measured under NCLB, including the All Student group, is included in the count. Data from NCES is used to create school-level demographic and teacher covariates. More specifically, enrollment counts are used to determine the percentage of students qualifying for Free or Reduced Lunch and the student to teacher ratio. Certified Personnel information is used to categorize teachers by level of experience and calculate the percentage of teachers having less than one year's experience. School type is also taken from NCES data as the impact of accountability pressure may vary by grades served[6]. Lastly, a binary variable indicating whether or not a school was subject to sanctions in the previous year is created. Within a given year all schools categorized as "Needs Improvement" or "Needs Improvement – Made AYP" are subject to sanctions. Therefore, the lagged value for Needs Improvement status is an indicator of whether or not a school faced sanctions in the previous year. As this is a binary variable, the value "1" indicates sanctions were imposed.

Table 3 displays the descriptive statistics for the variables generated. Approximately 80% of schools across all years met Adequate Yearly Progress, implying a failure rate of twenty percent. Sixteen percent of schools faced sanction pressure due to repeated failures in previous years. Table 4 compares sample averages by school type.

[6] NCES defines institution level based on both lowest and highest grade served. Primary schools are defined as having a lowest grade range of Prekindergarten through 3rd grade; highest grade up to 8th. Middle schools have a lowest grade served between 4th and 7th; highest grade between 4th and 9th. The lowest grade served in high schools ranges between 7th and 12th; highest grade of 12th. All other possibilities are categorized as "Other"

Table 3. Descriptive Statistics for Adequate Yearly Progress Measures

Variables	Mean	St. Dev	Min	Max	Count
AYP Status	0.8031	0.3977	0	1	16379
# of Binding Subgroups	8.0471	2.3126	2	18	16379
Sanctions	0.1654	0.3715	0	1	16379
% FRL	0.5682	0.2468	1	1	16379
Pupil / Teacher Ratio	15.1656	2.3709	1	88	16379
% of Teachers with Experience less than 1 year	0.0534	0.0599	0	1	16379

Table 4. Sample Means of Adequate Yearly Progress Measures by School Type

Variable	All	Primary	Middle	High	Other
AYP Status	0.8031	0.9193	0.7064	0.5311	0.4750
# of Binding Subgroups	8.0471	7.8539	9.4685	6.8856	8.3375
Sanctions	0.1654	0.0611	0.3242	0.3186	0.4250
% FRL	0.5682	0.5914	0.5709	0.4772	0.6383
Pupil / Teacher Ratio	15.1656	14.8622	15.0653	16.4768	13.5188
% of Teachers with Experience less than 1 year	0.0534	0.0495	0.0597	0.0584	0.0666
Number of Observations	16379	9888	3569	2762	160

When comparing AYP passage rates and sanctions pressure the need to control for school type becomes apparent. For instance, 91.9% of primary schools pass AYP while only 53.1% of high schools met the requirements. Since primary schools are more likely to pass AYP they are also less likely to face sanction pressure due to previous years' failures. Approximately 31.8% of high schools within the sample faced sanction pressure as compared to 6.1% of primary schools.

The average number of binding subgroup requirements equaled approximately eight for all schools, and did not vary significantly across school type. The lack of variation in binding group requirements suggests that there are other mechanisms working to create the large differences between passage rates across school types. One possibility is the lack of student preparation becomes more pronounced as a student ages and reaches later grades. Another is the use of graduation rate as a secondary indicator for high schools, while elementary and middle schools are allowed to use measures such as attendance or scores on standardized tests in non-high stakes subjects. Chakrabarti (2013) finds that within threatened schools attendance rates improve, with little to no evidence of improvements for graduation rates.

Methodology – Correlated Random Effects Probit

The purpose of this chapter is to determine the impact of NCLB components on passage rates while controlling for other student demographics and school resources. As mentioned previously, three measures are used in AYP determinations: test participation, academic performance, and a second indicator specific to school type. It is important to note that while the formal calculation of AYP status is based solely on these three factors, allowances within each make the process more complex. For instance, when measuring academic performance each school is allowed to meet AMOs through absolute measurement, confidence intervals, multi-year averaging, or the safe harbor method. The final determination combines all three measures to form a single binary pass/fail indicator. Therefore, the structure of the dependent variable requires the use of a binary response model.

Accountability components, or structures present within an accountability system, can indirectly affect a school's ability to meet AYP. The two main components of interest for this dissertation are the number of binding subgroups and the implementation of sanctions. The number of binding subgroups refers to the number of groups for which a school is held independently responsible. Having a large number of binding subgroups may increase the probability of failure due to the resulting increase in the number of criteria required to meet AYP. Imposing sanctions can create pressure that subsequently impacts academic performance and secondary indicators such as attendance and graduation rates. One way to address these issues is to use regression analysis to compare schools facing varying accountability components while controlling for other school demographics and teacher characteristics. The subsequent estimates yield measures of the relative importance of school characteristics in determining AYP. The estimates highlight factors that contribute to determining AYP in practice but that are unaccounted for in the formal AYP calculation. The formal rules of determination as described do not account for the variation in requirements faced by schools due to dissimilar demographics and past performance.

We start with a simple static specification that models AYP status as a function of the factors mentioned above. We employ the Chamberlain-Mundlak random effects probit model using maximum likelihood estimation. The first specification focuses on the number of binding subgroups, while the second adds an additional measure of accountability pressure. More specifically, this second measure indicates whether a school faced sanctions in the previous year. To address the possibility of state dependence, the previous specifications are also explored using a dynamic framework.

When using panel data two common approaches to estimating binary response models are fixed effects and random effects. One benefit of the fixed effects model is that it assumes no relationship between unobserved heterogeneity and model covariates. However, this benefit also causes restrictions in the covariates and data available for use. Time invariant covariates are eliminated from the model, and observations with no variation in outcome are eliminated from the sample. Restricting the estimation sample and excluding critical time invariant variables such as school type is not desired. As shown in Table 4, the passage rates for primary, middle, and high schools differ greatly. Therefore, the use of the random effects model is explored. The random effects model assumes complete independence between unobserved heterogeneity and covariates, thus allowing for the estimation of time constant variables. However, the complete independence assumption is not applicable in many cases, creating the need for adjustments. The Chamberlain-Mundlak model provides such an adjustment by specifying the precise relationship between unobserved heterogeneity and covariates.

We begin with a general latent variable model,

$$y_{it}^* = \theta_{it} R_{it} + \boldsymbol{x}_{it}\boldsymbol{\beta} + \lambda z_i + c_i + e_{it} \tag{2}$$

$$y_{it} = \begin{cases} 1 \; if \; y_{it}^* > 0 \\ 0 \; if \; y_{it}^* \leq 0 \end{cases}$$

The model stipulates that there is an unobserved continuous variable, y_{it}^*, that determines AYP status. In the case of accountability systems, the latent variable can be thought of as the sum of production with regards to student performance. We do not observe y_{it}^*, only the end result of passage or failure. The latent variable is modeled as a function of the number of binding subgroup requirements (R_{it}), a vector of time-varying covariates ($\boldsymbol{x}_{it}$), the time constant variable of school type (z_i), unobserved heterogeneity (c_i), and an error term (e_{it}). The outcome actually observed is denoted by y_{it}.

The number of binding subgroups represents the number of additional requirements faced by each school. This requirement was added to NCLB to ensure that all students were considered when calculating accountability performance. An unintended consequence of this calculation method is the disproportional amount of accountability pressure placed on heterogeneous schools - the more diverse the student population, the greater the number of subgroups for which a school can be held accountable. We hypothesize that as the number of binding subgroups increases, the more difficult it becomes for a school to meet AYP. This implies a negative relationship between the dependent variable and the number of subgroup requirements. Time-varying covariates include the percentage of students qualifying for Free or Reduced Lunch within the school as a whole, the student to teacher ratio, and the percentage of teachers with less than one year's experience[7]. The percentage of students qualifying for Free or Reduced Lunch is a school-level measure of income, and the student to teacher ratio is used as a measure of school resources. The percentage of teachers with experience less than one year is yet another proxy for school resources. While the literature shows that teacher effectiveness improves with experience, Hanushek et al. (2005) suggests that experience is relevant in the first year only. Others (Grissmer et al. 2000; Gordon, Kane, and Staiger 2006) find similar results, with gains in experience concentrated in the first and second year. The format of Certified Personnel Information allows us to identify those teachers within a school with less than one year's experience. The next available experience grouping contains teachers with one to ten years of experience, which does not allow for the identification of a teacher's early years. As the level of resources available is a key determinant of the ability of schools to meet state-mandated criteria, it is important to account for these resources. Schools lacking sufficient resources may lack the capacity to respond to the incentives created under accountability.

The second specification adds a measure indicating if a school faced sanctions in the previous year which is represented by D_{it} .

$$y_{it}^{*} = \phi_{it} D_{it} + \theta_{it} R_{it} + \boldsymbol{x}_{it} \boldsymbol{\beta} + \lambda z_i + e_{it} \qquad (3)$$

As mentioned earlier, two consecutive years of failing AYP subjects a school to sanctions. The intensity of these sanctions increases with each subsequent failure. This variable is intended to measure the efficacy of sanctions imposed, with a positive relationship suggesting that sanctions lead to improvement in performance as defined by probability of passing AYP. However, it is

[7] All models also contain controls for year effects.

also possible that the enforcement of sanctions on an already failing school could have a negative impact on school performance.

Employing the Chamberlain-Mundlak model relaxes the assumption of complete independence between unobserved heterogeneity and covariates. It assumes a specific a relationship between the individual effects and covariates. To begin, we define $\boldsymbol{w}_{it} = \{\boldsymbol{AC}_{it}, \boldsymbol{x}_{it}\}$ which represents a vector of accountability components ($\mathbf{AC}_{it}$) and a vector of time-varying covariates ($\mathbf{x}_{it}$). The structure of the accountability vector is dependent upon specification. In the first specification it contains only the number of binding subgroups, but in the second it also contains the sanction status of the previous year. Due to the unbalanced nature of the panel, we must also define a selection indicator for each school across periods, $\{s_{i1}, \dots, s_{iT}\}$ (Wooldridge 2010). This indicator is equal to one if the observation can be used within estimation, and zero otherwise. Therefore the number of periods available for each observation is defined as $T_i = \sum_{r=1}^{T} s_{ir}$ (Wooldridge 2010). Equation (4) depicts the relationship between the individual effect and observable covariates as stated by the model.

$$c_i = \psi + \overline{\boldsymbol{w}}_i \xi + a_i \tag{4}$$

The individual effect is modeled as a linear function of time averages for all time-varying covariates ($\overline{\boldsymbol{w}}_i$) and a random error term (a_i). The time-average for covariates is defined as $\overline{w}_i = T_i^{-1} \sum_{r=1}^{T} s_{ir}\, x_{ir}$ (Wooldridge 2010). Since this is an unbalanced panel, this vector also includes time averages for year dummies as not all observations are present across all years. If the relationship in Equation 4 is correctly specified, time-averages account for unobserved heterogeneity and consistent estimates are obtained. The final static specifications then become

$$y_{it} = \theta_{it} R_{it} + \boldsymbol{x}_{it}\boldsymbol{\beta} + \lambda z_i + \psi + \overline{\boldsymbol{w}}_i \xi + a_i + e_{it} \tag{5}$$

$$y_{it} = \phi_{it} D_{it} + \theta_{it} R_{it} + \boldsymbol{x}_{it}\boldsymbol{\beta} + \lambda z_i + \psi + \overline{\boldsymbol{w}}_i \xi + a_i + e_{it} \tag{6}$$

The question of whether or not passage in the current year is influenced by passage in previous year introduces the issue of state dependence. AYP status may be an accumulation of previous years' performances as failing in a previous year could negatively or positively impact current year's performance. For instance, the negative stigma that comes with failing AYP for a given year could make it more difficult for schools to attract quality teachers or encourage higher performing students to seek education elsewhere. However, it is also possible that the consequences of school failure serve to correct ineffective policies/structures and promote growth in performance. The issue of state dependence is addressed by including the first lag of the dependent variable into both specifications. It is important to note that the first lag is a different measure from the sanction variable. The sanction variable measures whether or not a school was subject to sanctions in the previous year, and sanction status is dependent upon two consecutive years of performance. The first lag is an indicator of the previous year's performance only. There are two categories within the Needs Improvement status that highlight these differences. Any school classified in the "Needs Improvement – Made AYP" category made AYP in the current year but is still subject to sanctions as it takes two consecutive years of passing AYP to exit Needs Improvement status. Also, schools in their first year of failure are classified as "Adequate – Did Not Meet" and are not subject to sanctions due to the reliance on two consecutive years of performance. Therefore whether or not a school is subject to sanctions in the previous year and the first lag of the dependent variable bring different information to the model.

In order to obtain consistent estimates using a dynamic model, the initial conditions problem must be addressed. If left ignored, the model assumes the start of the student performance process begins with the first observation. And while there are new entrants into the panel during the period investigated it is unreasonable to assume that all other schools start the achievement process with the initial observation. Therefore, we need to account for the possibility of endogeneity within initial conditions. The method used by Wooldridge (2005) suggests including the initial condition (y_{i0}) in the definition of the individual effect

$$c_i = \psi + \gamma_i y_{i0} + \bar{\boldsymbol{w}}_i \xi + a_i \qquad (7)$$

The unbalanced nature of the panel requires the allowance of heterogeneous initial conditions as each school did not enter the panel during the same time period. The final specifications for the dynamic model are thus,

$$y_{it} = \theta_{it} R_{it} + \rho y_{it-1} + \boldsymbol{x}_{it}\boldsymbol{\beta} + \lambda z_i + \psi + \gamma_i y_{i0} + \bar{\boldsymbol{w}}_i \xi + a_i + e_{it} \qquad (8)$$

$$y_{it} = \phi_{it} D_{it} + \theta_{it} R_{it} + \rho y_{it-1} + \boldsymbol{x}_{it}\boldsymbol{\beta} + \lambda z_i + \psi + \gamma_i y_{i0} + \bar{\boldsymbol{w}}_i \xi + a_i + e_{it} \qquad (9)$$

All models are estimated using maximum likelihood. Average marginal effects are calculated and displayed in the results section of this chapter.

Results – Adequate Yearly Progress

Correlated Random Effects – Average Partial Effects

This section presents the results found when examining the relationship between accountability structures and passage rates. We estimate both the static and dynamic correlated random effects (CRE) probit models as outlined previously[8]. Note that coefficients obtained from a probit model are limited in their interpretation. The direction of the relationship between covariate and dependent variable is determined by the sign, but the magnitude of the coefficient is in terms of the standard deviation of the latent variable. The expected change in the conditional mean of the outcome given changes in covariates is not directly estimated. Therefore,

[8] Fixed effects estimates were performed for comparison and are available upon request.

other than direction, the coefficients do not lend themselves to meaningful interpretation. It is for this reason that marginal effects are estimated for every binary response model presented. Marginal effects are generally estimated in two ways – the marginal effect at the mean or the mean marginal effect. The marginal effect at the mean gives the marginal effect for a particular covariate when all other covariates are held at their mean value, $\bar{\boldsymbol{x}}$. Average marginal effects or average partial effects (APEs) yield the population averaged marginal effect. The marginal effect is calculated for each case, and then these effects are averaged over the entire population. APEs encompass all covariate values and lend themselves to more general interpretations. However, in order for APEs to be consistent, the distribution of the covariates must be representative of the population. For this study, we choose to estimate the APE as it incorporates the population as a whole as opposed to focusing on a single and possibly non-representative covariate value.

The first model specifies AYP passage as a static function of binding subgroup requirements, proxies for student income and school resources, and school type. Table 5 shows a negative and statistically significant relationship between binding subgroup requirements and passage rates. As the number of binding subgroups increases, the probability of AYP passage decreases. An increase in subgroup requirements increases the number of accountability standards that must be met, increasing the opportunities for failure. Therefore, heterogeneous schools may have a more difficult time passing AYP as they face more binding requirements and consequently more accountability criteria. The subgroup requirements were put in place to ensure improvement for all students, but the structure of the system may actually increase the likelihood of failure. A single unit increase in the number of binding subgroups decreases the probability of a school meeting AYP by approximately 5.6 percentage points. While the coefficient does not appear extremely large in magnitude, it is important to remember that an increase in a specific subgroup population could add up to two additional binding requirements as math and ELA are counted separately. Therefore, growth in one subgroup population could lead to a decrease in the probability of passing AYP that is greater than 5.6 percentage points. The results found here are aligned with literature (Sims 2013; Kane & Staiger 2003) that suggests the imposition of additional subgroup requirements increases the probability of school failure.

All covariates controlling for student income and school resources have the expected sign, but not all are significant. Both proxies for school resources are negative and significant, but the impact of the percentage of students receiving Free or Reduced Lunch is negative and insignificant. An increase in the percentage of teachers with less than one year of experience creates a large decrease in probability of passing AYP. The negative nature of this relationship is consistent with the literature, but the magnitude of the effect appears to be relatively large. One possible explanation focuses on measures of school quality. The level of resources or the ability of a school to attract experienced teachers can be considered a function of school quality. Those schools with a greater percentage of teachers with less than one year's experience may be classified as low quality or under-performing. Therefore, this percentage not only acts as a proxy for school resources, but may encompass other factors associated with underperforming or low quality schools leading to an amplified effect.

The second specification introduces a binary covariate indicating whether a school faced sanctions in the previous year. Similar to the results found in the literature, sanctions have a positive impact on performance. Schools facing sanctions in the previous year were 5.3 percentage points more likely to meet AYP. The size of the effect implies that incentives generated by the sanction system are positive and moderate in magnitude. Being subject to sanctions in a previous year can improve student performance through the creation of greater incentives or the implementation of structural changes. When examining the moderate magnitude of the sanctions effect, it is important to note the tiered nature of Georgia's system and the implication of these tiers on sanction severity. Although two consecutive years of AYP failure places a school in the Needs Improvement status, it takes three to four consecutive failures for a school to face corrective sanctions. Prior to corrective action sanctions are limited to school choice and mandatory tutor offerings.

In the second specification, both the magnitude and sign of the binding subgroup requirement remain unchanged. In fact, the sign and significance of all covariates except the for Free or Reduced Lunch variable remain the same. The effect of the percentage of students receiving Free or Reduced Lunch remains negative but both the magnitude and significance differs from the first specification. The first specification depicted an insignificant negative APE ranging from 6.8 to 8.2 percentage points, while the second specification shows a significant but still negative APE of 11.9 percentage points.

The dynamic model is estimated to allow for state dependence. High levels of state dependence suggest the probability of passing AYP is persistent over time, and mobility from state to state is limited. As can be seen in Table 5 the coefficient of the lagged dependent variable in the first specification is positive and significant. Passing in a previous year increases the probability of passing in current year by 8.6 percentage points. However, the magnitude of the coefficient implies that the current AYP state is not highly dependent upon the previous year's performance. The initial condition in Specification (1) is also positive, significant, and relatively small in magnitude. These findings, in combination with positive impact of sanctions, support the idea of mobility between states. Between the years of 2005 and 2011, only five to ten percent of schools failed to reach standards for three consecutive years and faced school choice. When examining school restructuring the percentage range decreases further, with merely four to six percent of schools facing this most extreme sanction.

Facing sanctions in the previous year increases the probability of passing AYP by 7 percentage points in the dynamic model, compared to the 5.3 percentage point increase found in in the static model. Both effects are moderate in magnitude and in line with the positive impact of sanctions. The effect of the previous year's performance remains positive in the dynamic model. Schools who met the standard in the previous year were 9.2 percentage points more likely to meet the standard. The initial condition coefficient remains positive but becomes insignificant and is greatly reduced in magnitude. When moving to the dynamic model, the impact of binding subgroup requirements remains moderate and negative in sign; the effect of sanctions remains positive. It is important to note that the magnitude of the change in the sanctions coefficient was significant as the coefficient increased from 5.3 percentage points to 7 percentage points, an increase of approximately 32%. These results suggest that while the impact of sanctions on passage rates is influenced by state dependence, the impact of subgroup requirements is not.

The magnitude and significance of all other covariates remained relatively unchanged. However, it is important to note that rho, a measure of individual heterogeneity, decreased greatly when moving from the static to dynamic model. The decline in rho suggests that accounting the previous year's state accounts for a significant portion of individual heterogeneity. As a robustness check, estimates using pooled maximum likelihood methods are available in Appendix D [9].

[9] When estimating the correlated random effects probit model with pooled maximum likelihood methods, clustering on the individual school ID is necessary.

Table 5. Correlated Random Effects Probit Model, 2004-2011

	Static				Dynamic			
	(1)		(2)		(1)		(2)	
Variables	Coeff.	APE	Coeff.	APE	Coeff.	APE	Coeff.	APE
Lagged Value of AYP Status					0.424***	0.086***	0.452***	0.092***
					(0.043)	(0.010)	(0.045)	(0.011)
Initial Condition for AYP Status					0.267***	0.050***	0.017	0.003
					(0.047)	(0.009)	(0.042)	(0.007)
Sanctions			0.338***	0.053***			0.454***	0.070***
			(0.041)	(0.006)			(0.044)	(0.006)
# of Binding Subgroups	-0.314***	-0.056***	-0.309***	-0.053***	-0.291***	-0.053***	-0.287***	-0.051***
	(0.018)	(0.003)	(0.018)	(0.003)	(0.018)	(0.003)	(0.017)	(0.003)
Rounded % of FRL Students	-0.461	-0.082	-0.690*	-0.119*	-0.373	-0.068	-0.672**	-0.119**
	(0.367)	(0.065)	(0.355)	(0.061)	(0.312)	(0.057)	(0.341)	(0.061)
Pupil/Teacher Ratio	-0.044***	-0.008**	-0.042***	-0.007***	-0.045***	-0.008***	-0.044**	-0.008**
	(0.017)	(0.003)	(0.016)	(0.003)	(0.016)	(0.003)	(0.017)	(0.003)
% of Teachers w. Experience less than 1 year	-1.511***	-0.268***	-1.598***	-0.276***	-1.423***	-0.258***	-1.532***	-0.272***
	(0.303)	(0.054)	(0.332)	(0.058)	(0.323)	(0.058)	(0.315)	(0.055)
Middle School	1.017***	0.299***	0.857***	0.227***	0.869***	0.248***	0.752***	0.195***
	(0.069)	(0.019)	(0.055)	(0.015)	(0.053)	(0.015)	(0.051)	(0.014)
Primary	2.144***	0.495***	1.382***	0.319***	1.637***	0.386***	1.184***	0.273***
	(0.069)	(0.016)	(0.060)	(0.016)	(0.066)	(0.016)	(0.059)	(0.016)
Constant	2.073***		1.541***		1.316***		1.106***	
	(0.284)		(0.296)		(0.295)		(0.230)	
rho	0.251		0.102		0.148		0.026	
	(0.018)		(0.016)		(0.017)		(0.048)	
N	16379	16379	16379	16379	16379	16379	16379	16379

Note: All specifications include time dummies; Clustered standard errors are in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Mundlak-Chamberlain equation includes time averages of all time-varying variables including time dummies. Dependent variable is a binary indicator equal to one if a school meets AYP and zero otherwise.

Correlated Random Effects – Distribution of the Covariates

Next, we estimate the marginal effect at specified values of two covariates: the percentage of students qualifying for Free or Reduced Lunch and the percentage of teachers with less than one year's experience. More specifically, we examine the marginal effect at the 25th, 50th, 75th, and 90th percentiles of these covariates to determine if marginal effects differ dependent upon location within each covariate distribution. In the static model of Specification (2), schools with a high proportion of students qualifying for Free or Reduced Lunch experience greater negative marginal effects from the subgroup requirements. Moving from the 25th to the 90th percentile in the distribution of percentage of students qualifying for Free or Reduced Lunch causes the marginal effect to increase by 12 percentage points or 25 percent. Examining the distribution of inexperienced teachers yields similar results. The marginal effect increases approximately 6 percentage points or 12 percent from the 25th to the 90th percentile, highlighting that those schools with the most inexperienced teachers are the most negatively affected. However, results also indicate that applying sanctions is most effective for those schools with a large proportion of economically disadvantaged students and a relatively high percentage of inexperienced teachers. These results suggest that sanction interventions were most helpful in schools with low values for school income and school resource proxies. When transitioning to the dynamic model, there is a uniform decrease in the magnitude of the marginal effect of subgroup requirements. The approximate 30% increase in the marginal effect of sanctions seen in both covariate distributions, reinforces the importance of accounting for state dependence within the model.

Table 6. Average Partial Effects of Accountability Measures at Specified Values of Covariates

	Static		Dynamic	
Variables	Binding Requirements	Sanctions	Binding Requirements	Sanctions
% of FRL Students				
25th Percentile	-0.049***	0.048***	-0.047***	0.064***
	(0.004)	(0.006)	(0.004)	(0.006)
50th Percentile	-0.053***	0.052***	-0.051***	0.070***
	(0.003)	(0.006)	(0.003)	(0.006)
75th Percentile	-0.058***	0.057***	-0.055***	0.075***
	(0.004)	(0.007)	(0.004)	(0.008)
90th Percentile	-0.061***	0.060***	-0.058***	0.079***
	(0.005)	(0.009)	(0.005)	(0.010)
% of Teachers with Exp < 1				
25th Percentile	-0.051***	0.050***	-0.049***	0.067***
	(0.003)	(0.006)	(0.003)	(0.006)
50th Percentile	-0.053***	0.052***	-0.051***	0.069***
	(0.003)	(0.006)	(0.003)	(0.006)
75th Percentile	-0.055***	0.054***	-0.052***	0.072***
	(0.003)	(0.006)	(0.003)	(0.007)
90th Percentile	-0.057***	0.056***	-0.054***	0.074***
	(0.003)	(0.007)	(0.003)	(0.007)
N	16379	16379	16379	16379

Note: All specifications include time dummies; Clustered standard errors are in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Mundlak-Chamberlain equation includes time averages of all time-varying variables including time dummies. Dependent variable is a binary indicator equal to one if a school meets AYP and zero otherwise.

Correlated Random Effects – Stratified by School Type

AYP results published yearly by the GADOE suggest that high schools are less likely to pass AYP than other school types. In 2011, 72% of all schools met the requirements. However, when broken down by school type primary schools had a passage rate of 83%, approximately 70% of middle schools made AYP, and only 41% of high schools met the standard. In the results presented in Table 5 high schools are the omitted category, and as evident from the estimates both middle and primary schools were more likely to pass AYP. When compared to high schools, middle schools were 19.5 to 29.9 percentage points more likely to pass and primary schools were 27.3 to 49.5 percentage points more likely meet AYP.

We examine potential differences and test the robustness of previous findings through analyzing subsets of the data stratified by school type. Stratifying by school type allows for distinct slopes on all estimated variables, while previous estimates allowed for differing intercepts only. When examining schools across types, patterns similar to the full sample emerge. Sanctions have a relatively small positive and significant impact on AYP status, increases in binding subgroups have a negative impact, and there is evidence to suggest past performance has an effect on current passage rates. The magnitudes of the effects vary by school type, with primary schools showing the least response to accountability measures. However the differences in magnitude observed across all covariates are not large enough to generate the dissimilarities in passage rates seen by school type. Therefore, the mechanisms examined do not appear to generate the differences in passage rates between elementary, middle, and high schools. Dissimilarities by school type in the nature of the AYP second indicator as well as flexibility of curriculum delivery and content may account for the differences observed in passage rates. As stated previously high schools are subject to a different secondary indicator than primary and middle schools. While middle and primary schools are allowed to use attendance and performance on student assessments as a secondary indicator, high schools must use graduation rates. Passage of the secondary indicator is mandatory for meeting AYP, therefore difficulty in maintaining the mandated graduation rate may put high schools at a disadvantage. The structured nature of high school may also contribute to the increased probability of failure. Teachers in both primary and middle schools have more freedom to adjust curriculum and content to focus on test-specific skills and subjects.

Table 7. Average Partial Effect by School Type, Chamberlain-Mundlak Model, 2004-2011

Variables:	Static (1) Average Partial Effect	Static (2) Average Partial Effect	Dynamic (1) Average Partial Effect	Dynamic (2) Average Partial Effect
High Schools				
Lagged Value of AYP Status			0.107***	0.119***
			(0.029)	(0.027)
Initial Condition for AYP Status			0.053*	0.016
			(0.031)	(0.023)
Sanctions		0.085***		0.117***
		(0.019)		(0.017)
# of Binding Subgroups	-0.054***	-0.047***	-0.052***	-0.045***
	(0.008)	(0.008)	(0.008)	(0.008)
% of FRL Students	-0.511**	-0.560**	-0.479**	-0.549***
	(0.242)	(0.233)	(0.201)	(0.189)
Pupil/Teacher Ratio (Nearest Integer)	0.006	0.007	0.005	0.006
	(0.009)	(0.009)	(0.008)	(0.007)
% of Teachers w. Experience < 1 year	-0.027	0.015	-0.059	-0.006
	(0.248)	(0.218)	(0.234)	(0.242)
N	2762	2762	2762	2762
Middle Schools				
Lagged Value of AYP Status			0.065***	0.083***
			(0.020)	(0.019)
Initial Condition for AYP Status			0.091***	0.020
			(0.023)	(0.020)
Sanctions		0.136***		0.152***
		(0.014)		(0.013)
# of Binding Subgroups	-0.092***	-0.086***	-0.089***	-0.083***
	(0.009)	(0.007)	(0.009)	(0.008)
% of FRL Students	-0.242	-0.337**	-0.238	-0.342**
	(0.197)	(0.169)	(0.168)	(0.162)
Pupil/Teacher Ratio (Nearest Integer)	-0.029***	-0.025***	-0.029***	-0.024***
	(0.005)	(0.005)	(0.005)	(0.005)
% of Teachers w. Experience < 1 year	-0.415***	-0.415***	-0.434***	-0.432***
	(0.136)	(0.132)	(0.129)	(0.136)
N	3569	3569	3569	3569

Primary Schools				
Lagged Value of AYP Status			0.030***	0.038***
			(0.011)	(0.014)
Initial Condition for AYP Status			0.027***	-0.008
			(0.009)	(0.008)
Sanctions		0.048***		0.054***
		(0.005)		(0.006)
# of Binding Subgroups	-0.033***	-0.034***	-0.033***	-0.034***
	(0.003)	(0.003)	(0.003)	(0.003)
% of FRL Students	0.086*	0.080	0.087	0.083
	(0.050)	(0.054)	(0.056)	(0.053)
Pupil/Teacher Ratio (Nearest Integer)	-0.004**	-0.004*	-0.004*	-0.004*
	(0.002)	(0.002)	(0.002)	(0.002)
% of Teachers w. Experience < 1 year	-0.146***	-0.161***	-0.141***	-0.156***
	(0.050)	(0.047)	(0.054)	(0.048)
N	9888	9888	9888	9888

Note: All specifications include time dummies; Clustered standard errors are in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Mundlak-Chamberlain equation includes time averages of all time-varying variables including time dummies. Dependent variable is a binary indicator equal to one if a school meets AYP and zero otherwise

Chapter IV

THE IMPACT OF ACCOUNTABILITY ON SUBGROUP PERFORMANCE

Motivation for Research

The level of Annual Measureable Objectives (AMOs) set under NCLB highlights the need for separate requirements for subgroup populations. During no point from 2003-2011 was one-hundred percent proficiency required for every student category. This was not to happen until the 2013-2014 school year. Once the system required 100% proficiency for all students, the need for separate subgroup requirements would have become unnecessary. However, the 2011 AMOs allow 24.3% of all students to fall below proficiency in elementary and middle school math; 20% of all elementary and middle school students to fall below proficiency in English and language arts; approximately 24% of all high school students to score below the math standard; and 11.2% all of students to score below proficiency in high school English and language arts[10]. These figures highlight a pattern present in all years of NCLB, the percentage of students allowed to fail the CRCT Math examination is higher than any other category. Therefore the largest margin for underperformance or lack of attention occurs within elementary and middle school math. For schools in which disadvantaged subgroups constitute a small proportion of the population, allowing a failure rate of 24.3% excuses these schools from concentrating on subgroup performance - given that a sufficient amount of the total population scores at proficiency.

[10] 2011 AMOs: CRCT Math – 75.7%; CRCT ELA – 80%; GHSGT Math – 76%; GHSGT ELA – 90.8%

The negative results of the binding requirements found in the previous chapter motivate the examination of the impact of these requirements within subgroup themselves as opposed to the school as a whole. In other words, the second question of this dissertation examines the impact of accountability systems on the performance of particular subgroups within the student population. Findings from Chapter III suggest that an increase in the number of requirements has a negative impact on passage rates - a finding which holds across school types. These results raise the question of the mechanism through which an increase in the number of subgroup requirements generates an increase in probability of school failure. For instance, is the increase in the probability of failure simply due to an increase in the number of accountability criteria that must be met? Or does a lack of improvement in the performance of binding subgroups produce increased probability of school failure? The intent of this chapter is not only to examine the impact of binding requirements on subgroup performance, but also to determine if the effects vary dependent upon a subgroup's position in the performance distribution.

The remainder of this chapter describes the subgroup requirements in Georgia, provides a literature review of the differential impact of accountability systems on varying student populations, introduces the data and methodology used to investigate the questions proposed, and lastly presents the results found.

History of Subgroup Requirements in Georgia

Whether or not a particular subgroup enters into a school's accountability profile is dependent upon the size of the overall student population as well as the size of the subgroup itself. As such not every school is held accountable for every subgroup. Again, the designated subgroups for the state of Georgia are as follows: All Student subgroup, Asian/Pacific Islander, American Indian/Alaskan, Blacks, Hispanics, whites, multi-racial, students with disabilities (SWD), limited English proficiency (LEP), and the economically disadvantaged (ED)[11].

Table 8 details the percentage of schools subject to a particular subgroup requirement in both mathematics and English/language arts[12]. It is important to note that the populations of Tables 8, 9, and 10 are limited to those schools with at least one subgroup requirement other than the "All Student" group. Due to data restrictions they are also limited to those schools serving grades in which the CRCT and GHSGT are administered[13]. While the percentage of schools individually responsible for the Asian subgroup remains relatively small throughout the course of this study, the vast majority of schools were responsible for the economically disadvantaged. The percentage peaked at 93% during both the 2010 and 2011 school years. Approximately 70% of schools each year were accountable for the Black subgroup, while the percentage of schools responsible for the white subgroup decreased from 75% in 2004 to 70% in 2011. Hispanics were the only subgroup to see significant growth in binding requirements during the period, and students with disabilities were the only group to see a dramatic decline. The source of these trends is uncertain as they could reflect changes in the population or classification shifts. For example, during the period studied Georgia's Hispanic population grew which most likely accounts for the increase in the number of schools held accountable for this particular subgroup. However, population shift may not be the most suitable explanation for the decrease seen within the SWD subgroup. Difficulty in meeting performance standards may have led to fewer students being identified and tested within this group, ultimately causing the decrease in the percentage of schools held accountable for SWD.

[11] The American Indian/Alaskan and multi-racial subgroups are excluded from all analysis in this chapter due to the limited instances of enrollment reaching the binding threshold.
[12] Note that it is possible for schools to be held accountable for a subgroup in one subject area but not the other.
[13] Detailed subgroup assessment performance data is only available for the CRCT and GHSGT.

Table 8. Percentage of Schools facing Binding Requirements in both Math and ELA, 2004-2011

Year	# of Schools	Asian/ Pacific Islander	Blacks	Hispanics	Whites	SWD	LEP	ED
2004	2002	5.29%	72.18%	14.49%	75.87%	46.80%	6.59%	88.71%
2005	2029	3.30%	71.07%	12.67%	73.93%	40.17%	5.27%	88.37%
2006	2058	3.69%	70.94%	13.99%	72.98%	39.65%	6.37%	89.36%
2007	2095	3.91%	71.17%	15.99%	71.89%	37.09%	6.21%	90.69%
2008	2133	4.13%	71.26%	16.88%	71.21%	32.96%	6.70%	91.42%
2009	2160	4.58%	71.62%	17.87%	70.32%	28.15%	7.64%	91.71%
2010	2214	4.52%	71.59%	21.59%	71.14%	27.37%	9.35%	93.41%
2011	2239	5.09%	70.03%	22.20%	70.12%	27.24%	9.74%	92.99%

NOTE: Compiled using Georgia Department of Education, Adequate Yearly Progress Reports, 2004-2011; Multi-Racial subgroup excluded from table above as less than one percent of schools are bound by this requirement.

While Table 8 gives detail concerning the number of schools bound by subgroup requirements, Table 9 shows how these schools and subgroups performed. More specifically, it shows the percentage meeting Annual Measurable Objectives in mathematics and English/ language arts for those schools in which the subgroup requirement is binding. All subgroups defined by race or ethnicity, except Blacks, maintain a passage rate of over 90 percent throughout the panel. The passage rates for Blacks fell to 82% in both the 2010 and 2011 school years due to a large decrease in the number of Blacks meeting the math requirement. The three subgroups not defined by race or ethnicity have lower passage rates, the lowest occurring within the students with disabilities subgroup.

Of those schools meeting standards in the “All Student” group, a significant percentage fail to meet AYP due to inadequate achievement in other subgroup populations. Table 10 gives a summary of the impact of subgroup accountability on making AYP for a school as a whole. In 2004, approximately 14% of schools who met both mathematics and ELA requirements for all students collectively did not make AYP due to the failure of a particular subgroup. For these schools, the performance of at least one subgroup lagged behind the performance of the student

Table 9. Percentage of Subgroups Meeting Annual Measurable Objectives (AMOs), 2004-2011

Year	Asian/Pacific Islander	Blacks	Hispanics	Whites	SWD	LEP	ED
2004	100.00%	97.09%	96.55%	100.07%	77.80%	90.15%	95.15%
2005	100.00%	91.33%	93.77%	100.00%	77.68%	78.50%	93.30%
2006	100.00%	90.06%	90.28%	99.93%	78.77%	69.47%	90.05%
2007	100.00%	91.42%	98.21%	99.80%	78.25%	84.62%	92.63%
2008	100.00%	87.30%	98.06%	99.47%	81.91%	90.21%	89.28%
2009	100.00%	90.95%	99.22%	99.67%	86.84%	95.73%	92.63%
2010	98.00%	82.27%	96.85%	99.05%	77.24%	92.23%	85.58%
2011	95.61%	82.91%	96.77%	99.11%	75.08%	81.19%	85.88%

NOTE: Compiled using Georgia Department of Education, Adequate Yearly Progress Reports, 2004-2011; Multi-Racial subgroup excluded from table above as less than one percent of schools are bound by this requirement.

Table 10. Failure of AYP due to Additional Subgroup Requirements, 2004-2011

Year	# of Schools	# of Schools with Additional Subgroup Requirements	# of Schools Passing "All Subgroup"	#% Failing Additional Subgroup Requirement
2004	2002	1978	1953	13.93%
2005	2029	2004	1948	13.86%
2006	2058	2033	1917	13.72%
2007	2095	2063	1970	11.52%
2008	2133	2102	1950	11.59%
2009	2160	2121	2011	7.56%
2010	2214	2173	1936	11.00%
2011	2239	2193	1969	13.76%

NOTE: Compiled using Georgia Department of Education, Adequate Yearly Progress Reports, 2004-2011

population as a whole. The percentage of schools failing solely due subgroup performance reached its lowest point in 2009 at 7.56%, but climbed to 13.76% by 2011.

All previous tables and figures have focused on the performance of binding subgroups in relation to the school as a whole. However, in order to perform an analysis on the impact of the requirements within each subgroup you must include the entire population – those held accountable and those not held accountable. The following figures show the percentage of all students meeting or exceeding the standard within a given year by subgroup, subject and test.

Blacks, Hispanics, and the economically disadvantaged have similar trend patterns in English/language arts. While Blacks and the economically disadvantaged also move in similar patterns in mathematics, the Hispanic subgroup outperforms both. During the time period studied the statewide curriculum in Georgia transitioned from Quality Core Curriculum (QCC) to Georgia Performance Standards (GPS). The staging of the transition from QCC to GPS differed by grade and subject area[14]. In 2006 students taking the CRCT in ELA were assessed using GPS, while dependent upon grade level the mathematics CRCT transitioned to GPS from 2006 through 2008. The 2008 school year also saw a transition away from QCC towards GPS for students taking the GHSGT in ELA. The mathematics portion of the GHSGT switched to GPS in the 2011 school year. Transition years can be identified in the graphs below by the uniform decreases in the percentage of students meeting standard across all subgroups.

The stipulations of NCLB required that Georgia measure progress as changes in aggregate levels of performance and not as improvements in individual student performance. Therefore, although SSAS promotes adequate yearly progress, it is a status-based model and not a growth approach (Figlio & Ladd 2008). The distinction between the two models is important because each creates different incentives. Status-based models provide a uniform target for all groups and force schools to focus on low-achieving students that may have otherwise been ignored - a primary goal of NCLB and SSAS. However, one disadvantage of status-based models is that they do not take into account students' original levels of achievement or students' ending positions. Whether or not a student meets the proposed standard may be highly dependent on where they began. It is because of this that status-based systems may favor schools serving more

[14] A description of the transition from QCC to GPS is available in Appendix C.

Figure 4. Meets/Exceeds Rate by Subject, Subgroup, and Year – CRCT ELA

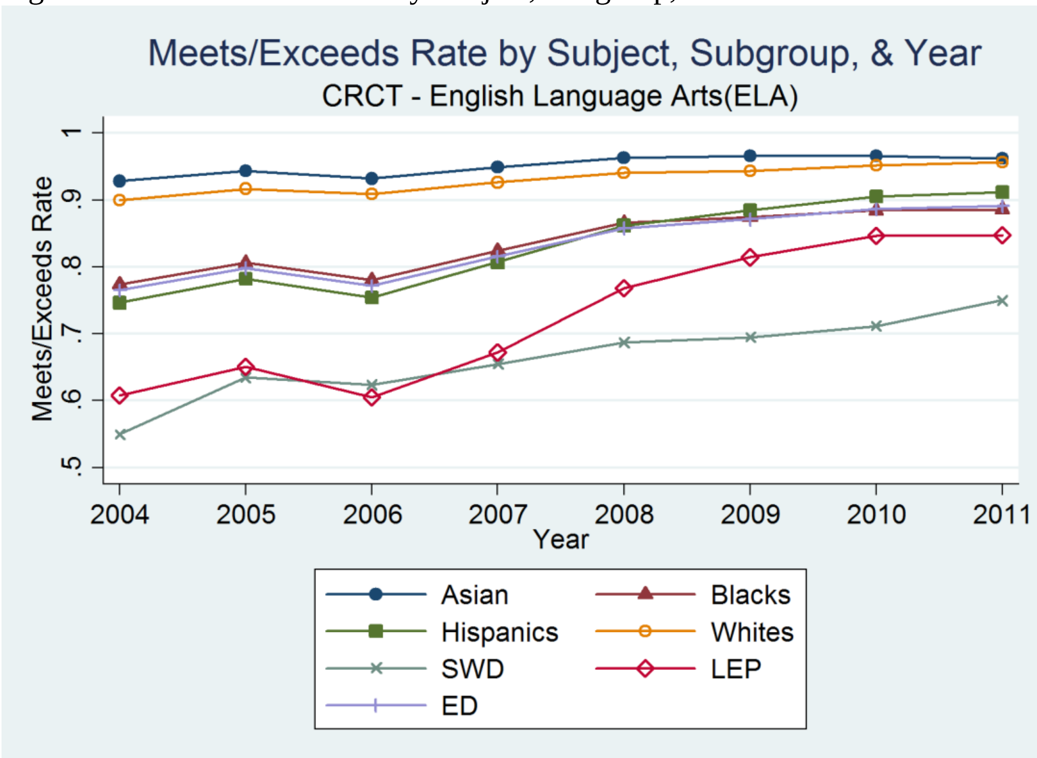


Figure 5. Meets/Exceeds Rate by Subject, Subgroup, and Year – CRCT Mathematics

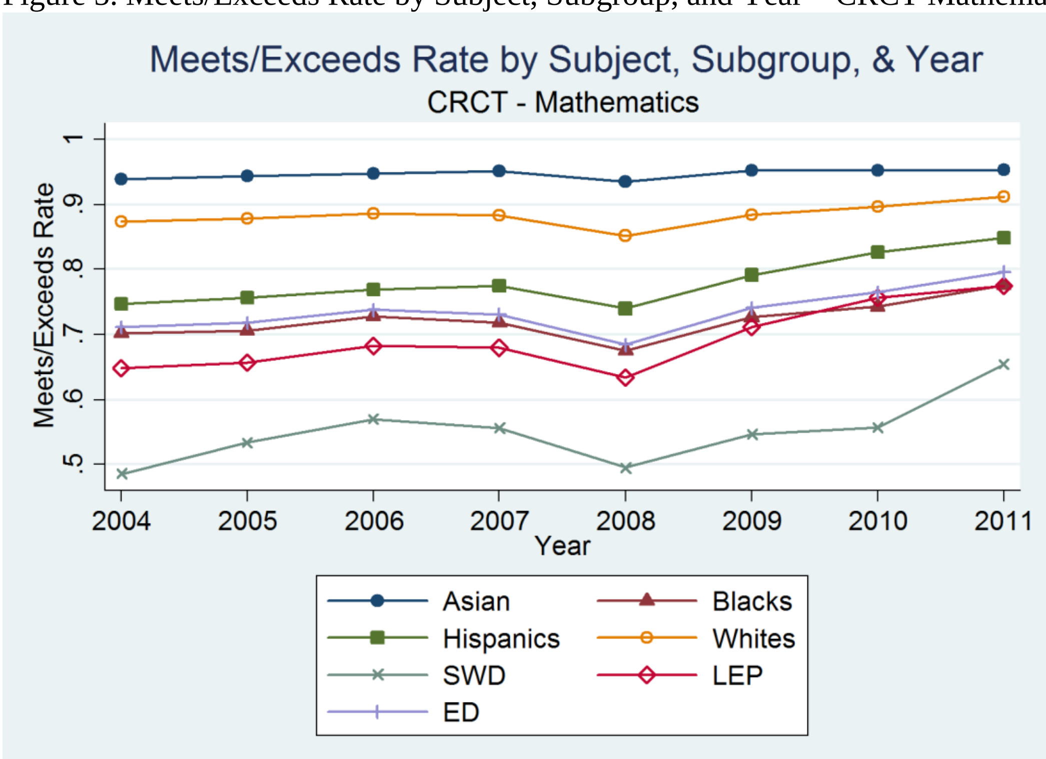


Figure 6. Meets/Exceeds Rate by Subject, Subgroup, and Year – GHSGT ELA

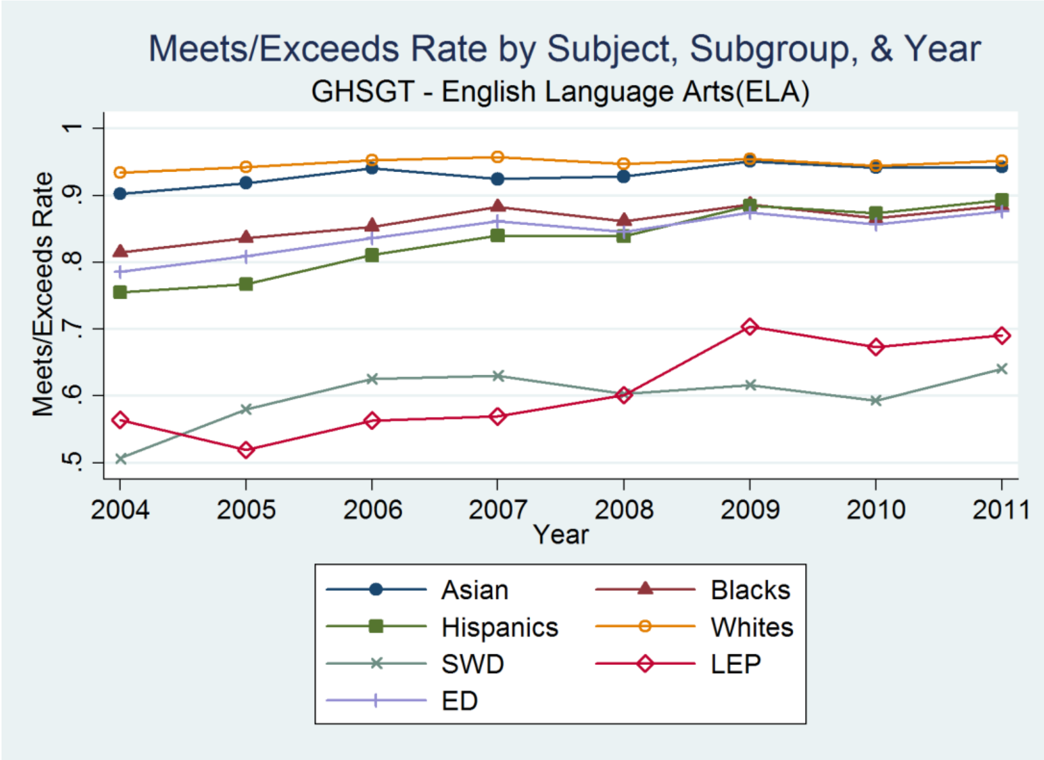
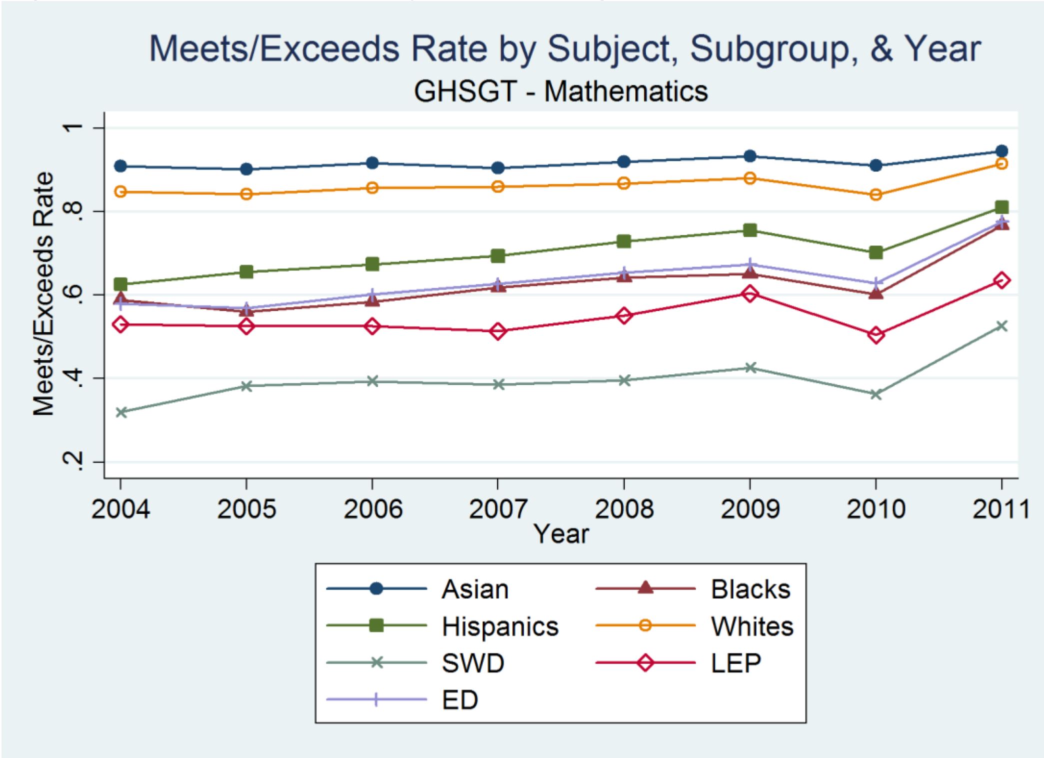


Figure 7. Meets/Exceeds Rate by Subject, Subgroup, and Year – GHSGT Mathematics



advantaged students, as there is a high correlation between socio-economic status and student achievement (Ladd 2007). More advantaged students begin with higher scores and therefore are able to reach and surpass thresholds more easily than their lower socio-economic counterparts. The status-based nature of SSAS also creates incentives as to which students receive additional attention. Those students scoring just below proficiency receive additional attention, and the number of students falling into this category is dependent upon the proficiency threshold. Given the nature of SSAS and the incentives generated, the following literature review investigates the possibility that accountability systems create differential impacts dependent upon student characteristics.

Literature Review – Differential Impact of Accountability Systems

A large number of studies indicate that accountability has a small positive effect on achievement. But there is also evidence to suggest that the impact of accountability does not affect all students uniformly. Differences can be seen across ethnicity groups and achievement levels. The notion of 'educational triage' (Gillborn & Youdell 2000) suggests rationing school resources to promote achievement for a targeted group of students. Schools and teachers may choose to target only those students whose scores are used in determining overall status, to focus on students at a particular point in the achievement distribution, or increase the placement of students into categories unaffected by accountability measures.

Race/Ethnicity and Socioeconomic Status

Two categories included in determining subgroup requirements are ethnic/racial categories and socioeconomic status. Many studies have investigated the differential impact of accountability on disadvantaged populations, but the evidence is mixed. Carnoy and Loeb (2002) find that Blacks and Hispanics are most affected when looking at NAEP results. West and Peterson (2006) find similar results while investigating the Florida system. For those schools previously receiving a failing grade, Blacks and the economically disadvantaged saw a significant improvement in scores while other groups did not. Receiving a grade of ‘D’ appeared to have a broader effect as whites, Blacks, and all income groups saw significant improvements. Hemelt (2011) and Figlio et al. (2009) also find positive impacts of accountability for minorities and the economically disadvantaged. Lauen and Gaddis (2012) found positive results for both groups in mathematics, while the results for reading were less consistent. However, others (Kane & Staiger 2003; Cronin et. al., 2005; Hanushek & Raymond, 2005) have found accountability to have a neutral to negative effect on the achievement of minority populations. Cronin et al. (2005) documented that both Blacks and Hispanics experienced lower rates of growth than their white counterparts under NCLB. Hanushek and Raymond (2005) find similar negative results for Blacks and Hispanics despite a gain in scores for all students as a whole. Not only did Blacks experience lower growth rates than their white counterparts, they were also the group least impacted by the system. Therefore, although accountability increased overall student performance, it contributed to growth in the Black-white achievement gap. Clotfelter, Ladd, and Vigor (2009) find mixed results with respect to the influence of accountability measures on minority populations. The achievement of Blacks at the bottom of the performance distribution increased at the expense of those at the top.

Distributional Effects

Focusing on students based on their position within the distribution of achievement scores is another common option of educational triage. Reback (2008) used individual student level data from Texas to determine the impact of accountability on the distribution of student achievement. He hypothesized that a student will receive more attention in a particular subject if that student's score is on the passing margin, if the school as a whole is underperforming in that subject relative to others, or if a significant portion of the student's class is on the passing margin. Findings showed that students on the margin experienced greater than expected achievement gains in both math and reading. Neal and Schanzenbach (2010) also find evidence that those students closest to the threshold experience the largest gain in scores. Studying Chicago public schools, the authors find no evidence that those at the bottom of the distribution experienced score gains, and mixed evidence supporting gains for high-achievers. Using data from Wisconsin, Chakrabarti (2013) also finds that students at the margin benefit the most from accountability pressure. However, improvements made were not obtained at the expense of students at the lower end of the distribution. Given the evidence in the literature concerning the performance of students on the margin it is important to note that while distributional differences in gains can be explained using "educational triage" theory, differences in student effort is yet another plausible explanation. Those students on the margin have greater incentive to put forth greater effort, while those at both the top and bottom of the distribution have little to no incentive to alter their effort level (Betts & Costrell 2001).

Ballou & Springer (2009), using student-level test score data from seven states, also investigate the distributional effects of NCLB. The authors compare test scores across years as grades enter the accountability system. Years in which test scores within a grade do not contribute to NCLB calculations are classified as low-stake years, and those in which test scores are used in calculations are labeled high-stakes years. Findings show that accountability increased achievement in high-stakes years, and those at the low end of the distribution benefited most at the expense of high-achievers. Reback (2008) suggests that while students at the margin benefit most from accountability measures, students at the bottom of the distribution can benefit as well. If a higher rating is attainable through a small increase in performance in a given subject area, greater amounts of resources are devoted to this subject and low-achieving students may experience larger than expected gains.

North Carolina implemented a growth based accountability system prior to NCLB, and with the act's passage added status-based components in the 2003-2004 school year. Ladd and Lauen (2010) use individual student level data to determine how the two reform approaches affect students at various points in the performance distribution. For those students below proficiency both systems generate gains in achievement. However, for those students located at the top of the distribution, status-based reform caused a decrease in gains. More specifically, failing to make AYP or pressure under the status-based system generates gains in math scores for the lower end of the distribution but slows growth for those at the top. In reading, test scores at the lower end of the distribution increased, but large losses were incurred at the upper end of the distribution. Therefore, under the status-based system, low-achieving students improve their test scores at the expense of high-achievers.

In all, the results are mixed when examining the impact of accountability systems on subgroup populations. While a branch of the literature show gains in scores (Carnoy & Loeb 2002; West & Peterson 2006), others note the relative magnitude of these gains maintain the achievement gap between minority populations and their white counterparts (Cronin et al. 2005; Hanushek & Raymond 2005). The results concerning the distributional impact are also mixed. Several studies find those at the margin, or close to the passing value, show the greatest improvement (Reback 2008; Neal and Schanzenbach 2010). While others claim low-achievers are the beneficiaries at the expense of high-achieving students (Ballou & Springer 2009; Ladd and Lauen 2010).

Data & Methodology – Binding Subgroup Requirements

Data – Binding Subgroup Requirements

The data used are detailed school-level Adequate Yearly Progress Reports ranging from the 2004 school year through the 2011 school year. The reports give the percentage meeting, exceeding, and failing the standard for every subgroup and assessment (CRCT and GHSGT exams) within a school for a given year. The data most critical to this chapter are the various counts of Full Academic Year (FAY) students, represented in the AYP reports through the following variables: Students in AYP grade levels, First Time Test Takers, enrollment during testing window, and test participants. Calculation of the binding enrollment cutoff and a school's position relative to this cutoff are performed given this information. Whether or not a school is held independently responsible for a particular subgroup is dependent upon on these calculations. Certified Personnel Information (CPI) and Common Core data are mined to create covariates for model robustness checks. The covariates included are estimated at the school-level. Estimates are performed by subgroup, but this does not create an issue as school and subgroup levels are equivalent. The only deviation occurs when a school administers both the CRCT and GHSGT, which occurs in less than two percent of the population. Covariates used include the percentage of all students qualifying for Free or Reduced Lunch, the pupil to teacher ratio, the percentage of teachers with less than one year's experience, and school type. It is important to note that these covariates apply to the student population as a whole, while the subgroup measures apply to students within AYP tested grades only.

Methodology - Regression Discontinuity & Quantile Estimation

While it is possible to evaluate students at the individual level to determine the impact of an accountability system, the structure of a status-based system creates incentives for the entire school. Jacob (2002) finds that response to accountability occurs at the school-level as opposed to the student level. Prior school achievement had a greater impact on outcomes than prior student scores. Combining this result with the presence of indivisibilities in production within schools, the need for school-level analysis becomes apparent.

Following the methodology originally proposed by Thistlethwaite and Campbell (1960), and further defined by various authors (Hahn et al. 2001; Lee 2008; van der Klaauw 2002 and 2008) this chapter uses a regression discontinuity design (RDD) to identify the impact of subgroup requirements on achievement. By focusing on schools with subgroup populations close to the requirement threshold the impact of this accountability policy can be determined (van der Klaauw 2008). Evaluation begins with the acknowledgment of two potential outcomes $Y_i(0)$ and $Y_i(1)$, where $Y_i(0)$ represents the outcome for untreated schools/subgroups and $Y_i(1)$ represents the outcome of those treated. The outcome of interest in this study is the percentage of students within a subgroup who meet or exceed the proficiency standard in a given school year. The causal effect is identified as $Y_i(1) – Y_i(0)$. However since we cannot observe both $Y_i(0)$ and $Y_i(1)$ simultaneously, the observed outcome is defined as follows:

$$Y_i=(1-w_i)*Y_i(0) + w_i*Y_i(0) \, , \qquad (10)$$

where w_i is an indicator equal to one if treatment is received and equal to zero otherwise. The keystone of RDD is that assignment to treatment is determined by the value of a predictor also known as a forcing or running variable, X_i. The forcing variable in this study is the number of Full Academic Year students tested within a subgroup population[15]. In order to determine if a subgroup will be included separately in AYP determinations the Georgia Department of Education (DOE) evaluates the following equation:

*max (.10*Students in AYP Grade Levels, 40),* with a cap set at 75 students. (11)

Threshold values are dependent upon school size. More specifically, they are dependent upon the number of Full Academic Year students present within tested grades or AYP grade levels. This relationship is outlined below in Table 11.

Table 11. Binding Requirement Thresholds

Students in AYP Grade Levels	Threshold Value
FAY Students ≤ 400	40
400 > FAY Students > 750	41 - 75
FAY Students ≥ 750	75

NOTE: Generated using the Binding Requirement Rule as described in Equation (11)

A school becomes accountable for a subgroup if one of the two following conditions holds:

1.) Subgroup population consists of 10% of all students in AYP grade levels and is a minimum of 40 students

2.) Subgroup population >= 75

Since classification as treatment or control is completely determined through the forcing variable, sharp RDD is most applicable in this case. All schools with values of the forcing variable exceeding one of the stipulated cutoffs are located within the treatment group, and all schools with values below the cutoff point are classified as untreated.

[15] As the ability to exclude students is limited the number of students tested is highly correlated with the subgroup population as a whole. It is also important to note that Georgia makes distinctions between all students and those enrolled for a Full Academic Year (FAY). Only FAY students are used in making AYP determinations.

It is important to emphasize that for the purpose of this study a subgroup is considered to be ‘treated’ if their performance directly impacts the overall AYP determination. Therefore, ‘treatment’ consists of a school being held independently accountable for the performance of a given subgroup population. This is of course in addition to being held responsible for the performance of the school as a whole. The following thought experiment explains the nature of the ‘treatment’ group: Consider the additional responsibility imposed due to binding requirements as enrollment in a separate program promoting and requiring higher achievement for each bounded subgroup. Treatment is then defined as enrollment into a program we label SPARE[16]. Within SPARE, every school is held accountable for the performance of each of their bounded subgroups.

Again, this analysis uses school-level as opposed to student-level data, but it is worth nothing the methodology through which a student is assigned to a particular subgroup. Each student within a school is simultaneously classified within a minimum of two groupings, the “All Student” group and one racial/ethnic group. Whether or not the student also qualifies for entrance into the three remaining subgroups (limited English proficiency, students with disabilities, or the economically disadvantaged) is dependent upon specific student characteristics. Therefore, students can be simultaneously classified within multiple subgroups. How does this classification process impact the study as designed? More specifically, what are the implications of a student within an unbounded group being simultaneously classified within a bounded group? The data available do not allow for the identification of individual cases simultaneous classification. Moreover, treatment as defined within this study consists of a school being held independently responsible for a particular subgroup. We examine how formal binding requirements impact the performance of a subset of students, and in essence determine the

[16] Subgroup Population Achievement REquirements (SPARE)

impact of targeting particular categories of students within a school. Consequently, treatment is defined at an aggregate level. Therefore, if results indicate that unbounded groups outperform their bounded counterparts and we are unable to examine individual outcomes, simultaneous classification becomes a viable explanation. A student may receive additional attention due to membership in a bounded subgroup that can translate into increased performance of an unbounded group.

To validate the use of a regression discontinuity design, a simple graph of the conditional mean outcome must show a discontinuity around the threshold (Wooldridge & Imbens 2007). If no evidence of a discontinuity is found, the likelihood of finding a treatment effect with more complex methods is unlikely. The figure also aids in selection of functional form.

The figures below show the mean Meets/Exceeds rate conditioned on the value of the centered forcing variable[17]. Again, the running variable is defined as the number of Full Academic Year (FAY) students tested within a particular subject and subgroup. If this count exceeds the threshold value, as determined by the GADOE rule defined in Equation (11), the subgroup requirement is binding. We subtract the appropriate threshold value, X_c, from the number of Full Academic Year students tested within a given subgroup and subject, X_i, to calculate the centered forcing variable, $(X_i\text{-}X_c)$. The now centered variable is more easily interpreted as all values greater than or equal to zero indicate a bounded subgroup, while all values less than zero indicate an unbounded subgroup.

Visual inspection shows a discontinuity in the conditional mean at the threshold, with those subject to the subgroup requirement performing better than their unbounded counterparts. The graphs do not eliminate a linear relationship between the outcome and forcing variable, but

[17] Figures 8 and 9 are for all subgroup observations combined; Graphs for individual subgroups can be found in Appendix E.

Figure 8. Average Meets/Exceeds Rate in Mathematics for Bounded and Unbounded Subgroups

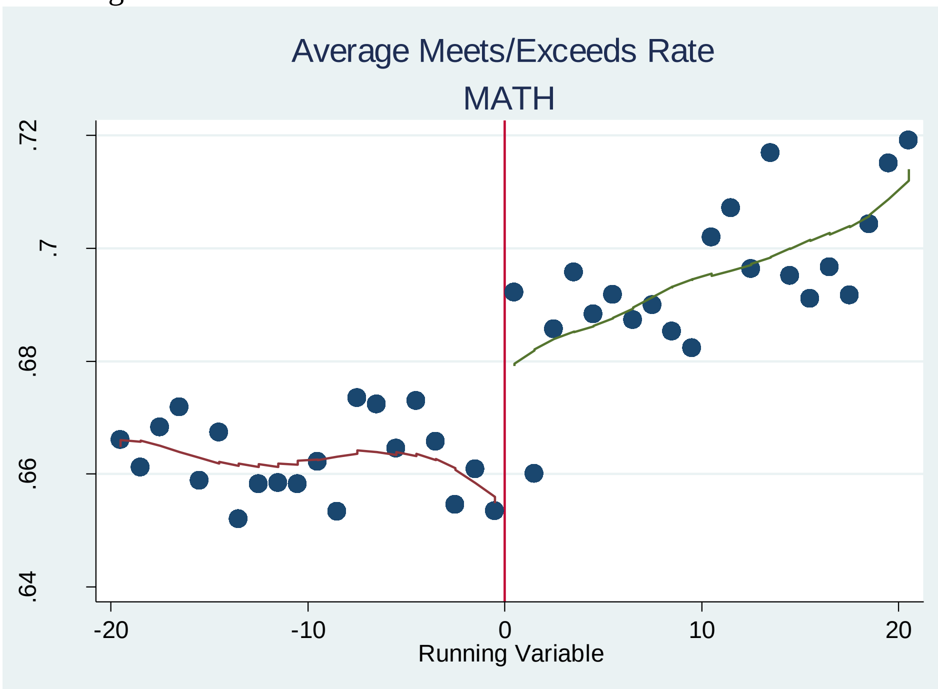
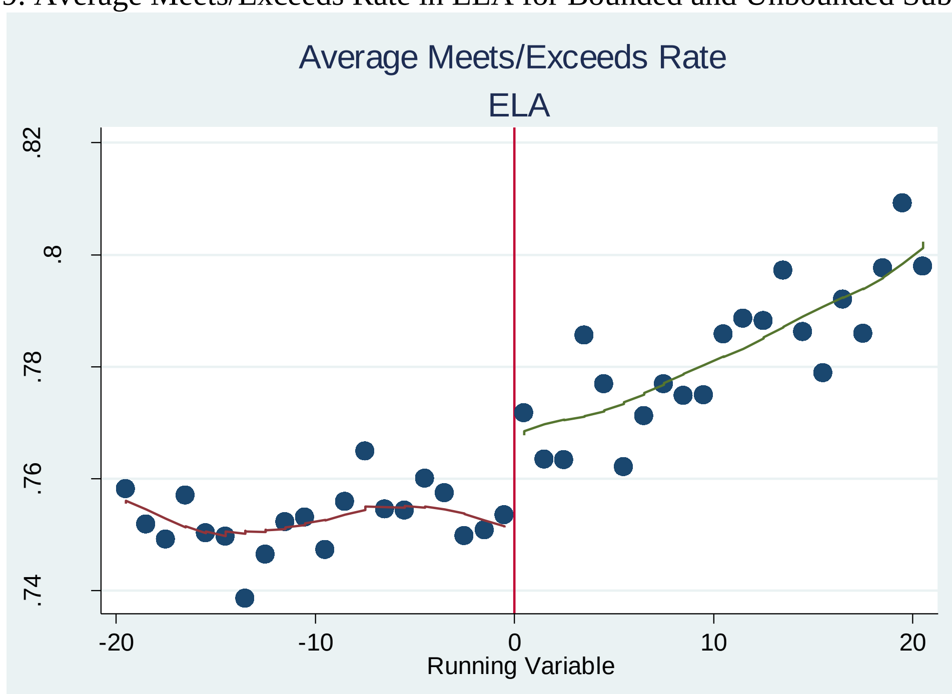


Figure 9. Average Meets/Exceeds Rate in ELA for Bounded and Unbounded Subgroups

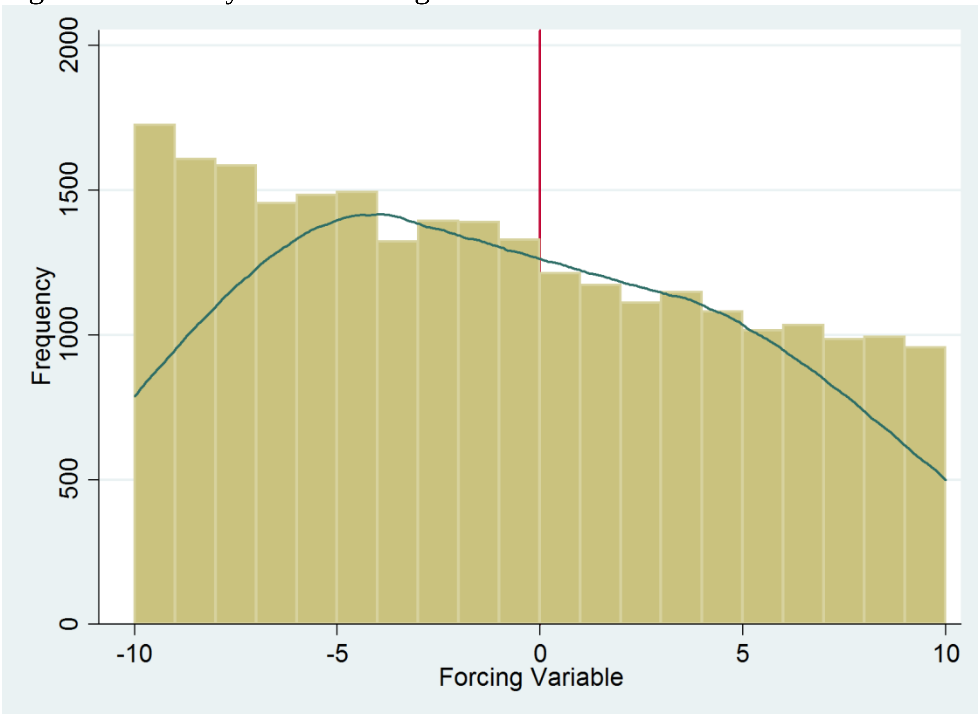


suggest the need for more tests to determine functional form.

Theory allows for the forcing variable to be correlated with outcome, as long as the correlation itself has no discontinuities. For example, as previous literature has implied the poverty concentration and therefore the Free or Reduced Lunch subgroup enrollment may be correlated with student outcomes. However, this relationship must be continuous in order for RDD to produce valid results. Another way of stating the previous assumption is to assume that schools are unable to precisely manipulate the forcing variable, creating randomized variation in the treatment near the cutoff point (Lee & Lemieux 2010). The assumption is pivotal because if it holds then any discontinuity in the distribution of the outcome (Y_i) conditioned on the forcing variable (X_i) at the cutoff point can be interpreted as the causal effect of the treatment (Wooldridge & Imbens 2007). The running variable, the number of Full Academic Year (FAY) students, is calculated based on student enrollment. In order to qualify as FAY students must be continuously enrolled in the same school from the Fall Full-Time Equivalent (FTE) count through the end of the state testing window[18]. The continuous enrollment requirement implies that the Fall FTE gives the maximum number of students present within each subgroup. Therefore, manipulation of the forcing variable would require removing or eliminating students from the FAY category. Testing for possible forcing variable manipulation is an important step in any regression discontinuity design. Wooldridge and Imbens (2007) suggest a testing methodology that involves visual inspection of the density of the forcing variable around the cutoff point. The suggested graph divides the centered forcing variable into bins surrounding the cutoff point with the bin width defined such that no bin lies on both sides of the cutoff. For the purposes of this study the bin width is equal to one. The bin mid-point is then plotted against

[18] Fall FTE Count occurs on the first Tuesday in October; Testing Window for the GHSGT ends in March; Testing Window for CRCT ends in April or May

Figure 10. Density of the Forcing Variable



the number of observations falling into each bin. The resulting graph, depicted in Figure 10, shows a lack of discontinuity at the cutoff point and therefore RD is deemed an appropriate design[19]. The average causal effect of treatment at the discontinuity point can be written as:

$$\tau_{SRD} = E\,[Y_i(1) - Y_i(0) \mid X_i = c], \text{ where c is the cutoff} \tag{12}$$

The simplest form of the model is as follows:

$$Y_i = \beta_0 + \beta_1 w_i + \beta_2(X_i - X_c) + \beta_3(X_i - X_c)*w_i + \mathbf{Z}_i + e_i \tag{13}$$

In the specification outlined in Equation (13), the variable of interest is β_1, as it predicts the treatment effect. Again, the running variable, X_i, is centered for ease of interpretation. An

[19] The McCrary density test, developed in McCrary (2008), uses local linear estimation to look for the presence of a discontinuity in the density of the forcing variable. The forcing variable in this study is discrete and therefore use of the McCrary density test may not be appropriate. However, in the interest of exploration the McCrary test is performed and the results indicate no discontinuity in the forcing variable at the cutoff value. Results are available upon request.

interaction term between the centered running variable and the treatment variable is added to allow the slope of the treatment variable to vary on either side of the cutoff point. Additional covariates are represented by the vector Z_i.

The discrete nature of the running variable does not allow for the comparison of observations in "arbitrarily" small windows surrounding the cutoff point, which is the methodology used in non-parametric RD estimation. Instead, estimation of the treatment effect is performed using a flexible polynomial as data not near the cutoff are included (Card et al. 2004; Oreopoulos 2006; Lee & Card 2008; Guiteras 2008; van der Klaauw 2002). Determining the correct functional form of the relationship between the assignment variable and the outcome variables is crucial to obtaining unbiased estimates. Visual inspection of Figures 8 and 9, along with goodness of fit statistics are used to determine functional form. Since visual inspection suggests a nonlinear relationship, goodness of fit statistics are used to test polynomials up to the fourth degree. Two such statistics are implemented. The first, presented by Lee and Card (2008), is used in the case of a discrete running variable with heteroskedastic variance. The second is the Akaike Information Criterion (AIC) measure. The results of all testing do not show a distinct advantage in the usage of higher order functional forms. Therefore, in the interest of robustness all models are estimated using both linear and quadratic forms[20]. Lee and Card (2008) also suggest clustering standard errors at values of the discrete running variable to account for group structure and obtain consistent estimates[21].

The inclusion of covariates within a regression discontinuity design can help to decrease sample biases arising from the use of values of the forcing variable not close to the cutoff. The

[20] Robustness checks are conducted using varying bandwidths and functional forms. Estimates using a cubic form at bandwidths of 10, 20, and 30 students were also performed and are available upon request.
[21] Each observation of the outcome is weighted by the number of tested students. We use the 'aweight' command in STATA to take into account the number of students generating the Meets/Exceeds percentage.

covariates included here are those that can be found in a traditional education production function. As the model is specified at the school-level, all covariates are measured at the school-level as well. The first covariate is the share of students qualifying for Free or Reduced Lunch and measures socio-economic composition. Measures of educational resources, such as teacher experience and the pupil-to-teacher ratio are also used. Prior to inclusion, the distributions of all covariates were tested for discontinuities near the cutoff value. These tests are done to ensure that the covariates used are appropriate for inclusion in the RD design. Testing was performed using two methods. First, we perform a "placebo test" in which we re-estimate the RD design using each covariate as the dependent variable (Wooldridge & Imbens 2007). Significance of the treatment variable would indicate a discontinuity in the covariate at the cutoff point, and therefore make the use of that covariate inappropriate. For all the covariates tested, the treatment variable remained insignificant and therefore no discontinuities were found. We also perform a visual inspection by constructing graphs showing the average value of each covariate by the centered forcing variable. The graph divides the forcing variable into bins surrounding the cutoff point, with bin width equal to one. We then plot the average value of the covariate within each bin against the bin mid-points (Wooldridge & Imbens 2007). All densities were found to be continuous in the forcing variable[22]. Based on the results as outlined above, all covariates tested are appropriate for inclusion in the RD design. We also include indicator dummies for school type to account for the possibility of differential effects for primary, middle, and high schools.

Another issue of concern is whether or not the structure of the assessments administered changed during the time frame examined. Research shows that several changes were made to the Georgia High School Graduation Test (GHSGT). The test was enhanced to meet NCLB and SSAS standards. Changes were also made to Criterion-Referenced Competency Test (CRCT),

[22] Graphical representations of the densities of covariates are available upon request.

and proficiency standards were increased to adhere to Georgia Performance Standards. Differences in test structure across years can create shifts in the average difficulty of the test and limit comparisons across time. To account for this possibility, we include year effects in all models.

The last issue to be addressed is the bounded nature of the dependent variable. The Meets/Exceeds rate is the percentage of students who meet the standard and is therefore bounded by zero and one. Ordinary Least Squares (OLS) regression can be problematic as the predicted values are not guaranteed to be within the bounded interval. One course of action suggests using the log-odds transformation, log[y/(1-y)], to map the dependent variable. The transformed dependent variable can then be used in OLS regression. However, this method is only viable if the dependent variable lies strictly between zero and one. As our data contains observations at the boundary values, this method is not applicable. Also, using the log-odds transformation yields estimation coefficients that are often difficult to interpret. Instead, we use the fractional logistic method developed by Papke and Wooldridge (1996). The authors develop a generalized linear model that uses quasi-maximum likelihood estimation to obtain results. The first advantage of this method is the inclusion of boundary values, and the second centers around coefficient interpretation. When partial effects from the fractional logistic regression model are evaluated at sample averages, the results become comparable to the coefficients obtained using linear regression. Therefore, all models are estimated using both OLS and the quasi-maximum likelihood estimation method as outlined by Papke and Wooldridge (1996). The results from each method are then compared.

The regression discontinuity model as described yields the mean effect of binding subgroup requirements on performance. However, the average effect can mask results that become apparent when examining various points in the distribution of passage rates. For example, an insignificant mean effect may be the result of gains made at the lower end of the distribution being negated by losses suffered towards the top. For this reason, we employ a quantile regression discontinuity analysis to pursue the possibility of heterogeneous treatment effects. It is important to note that we are examining passage rates across schools with respect to the subgroup(s) examined. Therefore the term “distribution” refers to the distribution of Meets/Exceeds rates across schools for the subgroup(s) in question.

When estimating quantile treatment effects in a regression discontinuity design, it is possible to employ the instrumental variable quantile regression model. The link between the instrumental variables estimator and the fuzzy regression discontinuity (FRD) design was established by Hahn, Todd, and van der Klaauw (2001) as well as Angrist and Lavy (1999). More specifically, the estimator used in the FRD is algebraically equivalent to the Wald estimator used in the instrumental variables setting. It is important to note that while the two estimators are numerically equivalent, both the motivation and assumptions used to justify their usage differ (Hahn, Todd, and van der Klaauw 2001). For instance, the instrumental variables design requires the instrument selected to be unrelated to the regression error and omitted variables. This assumption is not necessary within the FRD framework. Estimates obtained within FRD will be consistent as long as the continuity assumption, the assumption that in the absence of treatment those on either side of the cutoff would have similar outcomes, is satisfied (Hahn, Todd, and van der Klaauw 2001). The linkage between the FRD and instrumental variables estimators allows the use of quantile instrumental variables models when estimating heterogeneous effects across outcomes. Note that the subgroup requirement rule as defined by GADOE dictates a sharp design. However, as shown by Hahn, Todd, and van der Klaauw (2001) the sharp design can be defined as a special case of the FRD in which the discontinuity in the probability of treatment at the cutoff is equal to one. Using this reasoning we investigate the use of the instrumental variables quantile method for regression discontinuity designs proposed by Abadie, Angrist, and Imbens (2002). To begin, the authors introduce the conditional linear quantile regression model as outlined by Koenker and Bassett (1978)

$$\left(\hat{\beta}^{\tau}, \hat{\delta}^{\tau}\right) = \arg min_{\beta,\delta} \sum \rho_{\tau} \left(Y_i - \boldsymbol{x}_i \boldsymbol{\beta} - D_{\boldsymbol{i}} \delta\right) \quad (14)$$

for any quantile index $\tau \epsilon (0,1)$, where $\boldsymbol{x}_i$ is a vector of covariates and D_i is the treatment indicator. Estimates are obtained through the minimization of the check or loss function, $\rho_{\tau}(u) = u(\tau - \mathbf{1}\{u < 0\})$. The model assumes selection on observables and estimates conditional effects, meaning the estimates obtained are conditional on the values of model covariates. The parameter of interest is the quantile treatment effect, estimated by $\hat{\delta}^{\tau}$. Abadie, Angrist, and Imbens (2002) apply the Koenker and Bassett model to the instrumental variables framework through the introduction of a weighting component meant to identify "compliers" or observations who participate in treatment if and only if they are selected. However, as is the case in this study, the Abadie, Angrist, and Imbens instrumental variables formulation is reduced to the Koenker and Bassett conditional quantile estimation presented in Equation (14) if the binary treatment variable exactly equals the binary instrument.

Results – Binding Subgroup Requirements

Regression Discontinuity – Mean Effects

If we assume that an increase in the number of subgroup requirements has a negative impact on a school's probability of making AYP, examining the achievement of subgroup populations can help to identify the mechanism for this relationship. For example, do the requirements have a negative impact because they do not translate into increased subgroup performance? Or is the relationship between additional requirements and increased probability of failure a function of an increase in the number of criteria required for passage?

The first stage of analysis consists of using parametric methods to reconstruct the result found in Figures 8 and 9. As stated previously these figures are graphic representations of the overall mean effect of the binding requirements across subgroups. The regression discontinuity design as outlined in the previous section is implemented on subgroup performance data. Bandwidth selection becomes important when parametric estimation is used as observations not close to the cutoff are included. When including observations far from the threshold, the similarity in characteristics between the treated and non-treated can come into question. It is for this reason, that all models are estimated at two bandwidths. The first includes observations +/- 10 students around the threshold, while the second widens the range and includes observations +/- 20 students from the cutoff point.

To get a baseline estimate for the effect of binding subgroup requirements on student performance, we first examine all subgroups collectively. When using OLS, the results indicate a small positive impact of requirements. Those subgroups facing binding requirements outperformed their counterparts by 1.4 to 3.4 percentage points in mathematics across bandwidths and functional forms[23]. The estimated impact in ELA is slightly smaller with a total positive effect of 1.1 to 2.6 percentage points. The slightly greater effect for math than for ELA is in line with existing literature, and the magnitude of the effect is similar to that seen in Figures 8 and 9. Taken together, this evidence suggests a positive but small impact of subgroup requirements on student performance. While the evidence presented above is encouraging, it is not sufficient. As outlined in the previous section, the fractional nature of the dependent variable

[23] The estimates obtained using OLS can be found in Appendix F

requires the use of alternate methods of estimation. The results found using the fractional logit model are similar to the OLS results in both sign and magnitude. Across specifications, those subject to the binding requirements in mathematics saw a 1.3 to 3.4 percentage point increase in the Meets/Exceeds rate, while the ELA estimates are smaller and range from 1.2 to 2.6 percentage points. While those schools facing subgroup requirements outperformed their counterparts, the binding requirements were unable to substantially increase the performance of bounded subgroup populations.

The results presented above examine all subgroups collectively, but the effect of binding requirements may differ within subgroups and school type[24]. As evident from Chapter III, differential impacts of accountability components exist by school type. Keeping with this result we estimate the impact of binding subgroup requirements for elementary, middle, and high schools separately. Dividing the population by type not only adheres to previous findings but strengthens the validity of the regression discontinuity design as the control and treatment groups become more comparable. Tables 12 through 18 display the results by school type for the following five categories: all subgroups collectively, Blacks, Hispanics, students with disabilities, and the economically disadvantaged[25]. Positive and significant effects were found across all bandwidths, functional forms, and estimation methods when examining all subgroups collectively within elementary or primary schools. Results range from 1 to 2.5 percentage points for both mathematics and ELA. The subgroup driving this positive impact appears to be students with disabilities, as they also experience positive results across all specifications and bandwidths. The effects for mathematics and ELA are of relatively equal magnitude for this subgroup.

[24] Analysis for individual subgroups combined across school type is available upon request

[25] Results presented are limited to the quadratic specification of the subgroups stated above; Linear results and those for excluded subgroups are available upon request. Asian, white, and LEP subgroups were not included due to lack of statistically significant findings.

Table 12. Average Partial Effect of Binding Subgroup Requirements: All subgroups

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mathematics								
Treatment	0.0201***	0.0161***	0.0207**	0.0143*	0.0200**	0.0137	0.0338***	0.0243**
	(0.0057)	(0.0051)	(0.0087)	(0.0076)	(0.0092)	(0.0084)	(0.0109)	(0.0096)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	27,248	27,105	13,255	13,180	27,248	27,105	13,255	13,180
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Linear	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic	Quadratic
ELA								
Treatment	0.0132***	0.0120***	0.0158***	0.0133***	0.0141***	0.0114**	0.0257***	0.0226***
	(0.0047)	(0.0038)	(0.0056)	(0.0046)	(0.0052)	(0.0046)	(0.0074)	(0.0044)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	27,215	27,072	13,237	13,162	27,215	27,072	13,237	13,162
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Linear	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience.

Table 13. Average Partial Effects of Binding Subgroup Treatment, Primary Schools, All Subgroups

All Subgroups				
Variables	(1)	(2)	(3)	(4)
Mathematics				
Treatment	0.0205***	0.0175***	0.0236***	0.0191***
	(0.00609)	(0.00522)	(0.00732)	(0.00545)
Include Year Effects?	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes
N	17,835	17,811	8,725	8,715
Bandwidth	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic
ELA				
Treatment	0.0191***	0.0158***	0.0245***	0.0193***
	(0.00557)	(0.00457)	(0.00710)	(0.00551)
Include Year Effects?	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes
N	17,810	17,786	8,716	8,706
Bandwidth	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience.

Table 14. Average Partial Effects of Binding Subgroup Treatment, Primary Schools, Individual Subgroups

	Blacks				Hispanics			
Variables	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mathematics								
Treatment	-0.0123	-0.0154	-0.0199	-0.0214*	0.00942	0.0105	-0.00151	0.00412
	(0.0127)	(0.0102)	(0.0172)	(0.0129)	(0.00750)	(0.00754)	(0.00629)	(0.00772)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	2,875	2,869	1,425	1,420	2,134	2,133	961	960
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
ELA								
Treatment	-0.0166*	-0.0176*	-0.0139	-0.0115	-0.00586	-0.00513	-0.0171	-0.0112
	(0.00957)	(0.00896)	(0.0120)	(0.00916)	(0.00822)	(0.00710)	(0.0108)	(0.0104)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	2,874	2,868	1,426	1,421	2,132	2,131	957	956
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience.

Table 14 (cont'd). Average Partial Effects of Binding Subgroup Treatment, Primary Schools, Individual Subgroups

	Students with Disabilities (SWD)				Economically Disadvantaged (ED)			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mathematics.								
Treatment	0.0427***	0.0315***	0.0583***	0.0443***	-0.0112	-0.0108	0.0201	0.0239
	(0.00593)	(0.00806)	(0.00590)	(0.0105)	(0.0151)	(0.0149)	(0.0179)	(0.0175)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	7,409	7,406	3,856	3,856	1,580	1,579	793	793
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
ELA								
Treatment	0.0413***	0.0292***	0.0559***	0.0405***	-0.00995	-0.00896	0.00412	0.00887
	(0.00650)	(0.00578)	(0.00638)	(0.00748)	(0.0124)	(0.0120)	(0.0152)	(0.0144)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	7,400	7,397	3,853	3,853	1,579	1,578	797	797
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience.

However, it is important to note that while bounded SWD groups outperform their counterparts the effect is not large in magnitude.

Over the course of the panel for those schools failing AYP due to inadequate subgroup performance, fifty percent show failure within the SWD subgroup. Consequently, for this subgroup, and this subgroup only, continuous failure resulted in a systematic and artificial increase in the reported achievement for bounded subgroups. Beginning in the 2004-2005 school year, schools missing AYP solely due to failure in the SWD category were granted a federal adjustment. The adjustment added a proxy percentage to the Meets/Exceeds rate based on the percentage of SWD students within the state. For example, during the 2005-2006 school year this proxy percentage added 16 percentage points to every qualifying school's Meets/Exceeds rate for students with disabilities. The analysis presented here is performed using the unadjusted rate.

High schools have the lowest passage rate for AYP. When examining the high school population as a whole, positive effects for both mathematics and ELA were found in all specifications. The effect in math was found to be larger than that for ELA. Looking at all subgroups collectively, the estimates range from two to seven percentage points. When examining each subgroup individually, Hispanics saw uniform positive effects across specifications. Those subject to the subgroup requirements had an average Meets/Exceeds rate that was significantly higher than those not bounded by the requirements. Hispanics were the only individual subgroup to see consistently positive effects across bandwidths and specifications. Therefore, within high schools, Hispanic students appear to be the group most affected by accountability requirements. There is also some evidence suggesting a positive effect for the economically disadvantaged. Three specifications saw positive impact for both math and ELA. It is important to remember that since this category is not based on racial or ethnic

Table 15. Average Partial Effects of Binding Subgroup Requirements, High Schools, All Subgroups

All Subgroups				
Variables	(1)	(2)	(3)	(4)
Mathematics				
Treatment	0.0423***	0.0444***	0.0654***	0.0633***
	(0.00979)	(0.00898)	(0.0113)	(0.0111)
Include Year Effects?	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes
N	4,526	4,517	2,007	2,004
Bandwidth	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic
ELA				
Treatment	0.0214**	0.0278**	0.0332**	0.0358**
	(0.0106)	(0.0110)	(0.0158)	(0.0161)
Include Year Effects?	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes
N	4,521	4,512	2,011	2,008
Bandwidth	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience.

Table 16. Average Partial Effects of Binding Subgroup Requirements, High Schools, Individual Subgroups

	Blacks				Hispanics			
Variables	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mathematics								
Treatment	0.00733	0.00989	-0.0438	-0.00679	0.0761**	0.0651***	0.117***	0.112***
	(0.0336)	(0.0233)	(0.0297)	(0.0228)	(0.0302)	(0.0245)	(0.0382)	(0.0302)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	940	936	461	459	437	437	191	191
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
ELA								
Treatment	-0.0008	0.00794	-0.0237**	0.00181	0.0661***	0.0564***	0.102***	0.102***
	(0.0117)	(0.00927)	(0.00951)	(0.00813)	(0.0192)	(0.0167)	(0.0229)	(0.0209)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	934	930	461	459	440	440	190	190
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year’s experience.

Table 16 (cont'd). Average Partial Effects of Binding Subgroup Requirements, High Schools, Individual Subgroups

	Students with Disabilities (SWD)				Economically Disadvantaged (ED)			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mathematics								
Treatment	0.0470	0.0368	0.0962***	0.0311	0.0325*	0.0467***	0.0149	0.0471***
	(0.0344)	(0.0249)	(0.0316)	(0.0286)	(0.0174)	(0.0128)	(0.0249)	(0.0162)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	1,438	1,437	516	516	910	908	461	460
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
ELA								
Treatment	-0.0176	0	0.0165	-0.00925	0.0231**	0.0340***	0.0195*	0.0343***
	(0.0214)	(0.0148)	(0.0251)	(0.0193)	(0.0109)	(0.0105)	(0.0102)	(0.00980)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	1,435	1,434	519	519	911	909	463	462
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience.

Table 17. Average Partial Effects of Binding Subgroup Requirements, Middle Schools, All Subgroups

All Subgroups				
Variables	(1)	(2)	(3)	(4)
Mathematics				
Treatment	-0.0160	-0.00910	0.0277	0.0183
	(0.0289)	(0.0293)	(0.0327)	(0.0374)
Include Year Effects?	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes
N	4,373	4,345	2,201	2,190
Bandwidth	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic
ELA				
Treatment	-0.0139	-0.00351	0.0293**	0.0292
	(0.0171)	(0.0202)	(0.0140)	(0.0201)
Include Year Effects?	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes
N	4,371	4,343	2,189	2,178
Bandwidth	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience.

Table 18. Average Partial Effects of Binding Subgroup Requirements, Middle Schools, Individual Subgroups

	Blacks				Hispanics			
Variables	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Mathematics								
Treatment	-0.112**	-0.0669	-0.0917*	0.0264	0.0312	0.0116	-0.0109	-0.0365**
	(0.0463)	(0.0413)	(0.0501)	(0.0351)	(0.0231)	(0.0196)	(0.0227)	(0.0180)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	369	365	179	179	730	725	341	338
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
ELA								
Treatment	-0.0791**	-0.0488	-0.0622*	0.0271	0.00942	-0.000406	0.00694	-0.00894
	(0.0337)	(0.0350)	(0.0319)	(0.0267)	(0.0128)	(0.0112)	(0.0203)	(0.0148)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	370	366	179	179	729	724	336	333
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience

Table 18 (cont'd). Average Partial Effects of Binding Subgroup Requirements, Middle Schools, Individual Subgroups

	Students with Disabilities (SWD)				Economically Disadvantaged (ED)			
Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mathematics								
Treatment	-0.0410	-0.0271	0.0108	-0.00922	0.0270	0.0190	0.0256	0.0129
	(0.0326)	(0.0260)	(0.0368)	(0.0337)	(0.0591)	(0.0447)	(0.0819)	(0.0562)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	2,112	2,106	1,168	1,165	145	139	67	65
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
ELA								
Treatment	-0.0253	-0.0100	0.0127	0.000606	0.0302	0.0207	0.0480	0.0302
	(0.0186)	(0.0152)	(0.0181)	(0.0182)	(0.0533)	(0.0359)	(0.0686)	(0.0473)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	2,109	2,103	1,163	1,160	143	137	67	65
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Fractional logit model with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience.

makeup, it encompasses a broader range of students. And as is evident from Figures 4-7, the trend lines for performance of Blacks and Hispanics often closely follow that of the economically disadvantaged. Therefore, improvement within the economically disadvantaged could indirectly result in small improvements within certain racial categories and vice versa.

Middle schools saw no overall positive impacts for subgroup requirements. In fact, estimates for the Black subgroup were negative and significant in several specifications for both math and ELA. These results suggest that the impact of subgroup requirements for Blacks was at best neutral and at worst detrimental to ELA achievement in middle schools. This neutral to negative result is in line with a section of the literature that suggests Blacks are negatively impacted by accountability systems (Kane & Staiger 2003; Cronin et. al., 2005; Hanushek & Raymond, 2005; Clotfelter, Ladd, and Vigor 2009). However, we must also note that the widest variation in student population, and therefore cutoff values, occurs within the middle school sample. In all other categories, separation by school type decreases the variation in student population and cutoff values, thereby making populations more comparable.

Regression Discontinuity – Quantile Estimation

The results above indicate a small but positive average effect of binding requirements on subgroup performance. However, it is possible that the effect differs by quantile or specific percentiles within the Meets/Exceeds rate distribution. To be clear, when describing quantiles in reference to subgroup performance we are referring to the distribution of the Meets/Exceeds rate (MER) for a given subgroup across schools. For example, a Black subgroup in the 90^{th} percentile has a Meets/Exceeds rate that is higher than 90 percent of Black subgroups across all schools. We again focus on narrow windows around the cutoff point, estimate both linear and quadratic specifications, and split the sample by school type to make the treatment and control groups more comparable. The mean treatment effect in mathematics for elementary schools with respect to all subgroups was positive and small in magnitude. Quantile estimation yields similar results. When examining all subgroups collectively, positive effects are found in both ELA and mathematics. Estimates also indicate that these effects are largest and most significant at lower percentiles. These results suggest that under-performing subgroups, or those towards the lower end of the MER distribution, benefit most from the binding subgroup requirements. As depicted in Figures 4-7, Blacks, Hispanics, SWD, and the economically disadvantaged are under-performing relative to all others. Therefore, we focus on these four categories when performing individual subgroup analyses.

Among Blacks in elementary school, the binding subgroup requirement actually has a statistically significant negative impact in ELA for the bottom portion of the distribution. Across specifications, estimates at the 25th quantile show an approximately two percentage points advantage for unbounded subgroups over bounded groups. This difference is underestimated by the mean effect. There is no consistent evidence supporting a negative or positive effect on the economically disadvantaged or Hispanics in elementary schools. However, similar to the mean effects discussed in the previous subsection, students with disabilities show positive and significant effects in both mathematics and ELA. The largest effects are seen in the upper tails. Estimates range from two to six percentage points for mathematics, and one to six percentage points for ELA. Both of which are larger than the mean effect. Therefore high-performing schools with respect to the SWD subgroup generate the greatest gains from the binding requirements.

The pattern when looking at all high school subgroups collectively is similar to the results found in the previous subsection. We see a positive effect in both mathematics and ELA, with larger effects in the lower end of the distribution. Therefore, we again focus on the impact for Hispanics, students with disabilities, and the economically disadvantaged. The Black subgroup is excluded from this discussion as the majority of estimates were negative but insignificant. Again, suggesting that binding requirements were at best neutral in improving achievement among Black students in high school. For Hispanic subgroups in the upper quantiles, binding requirements had positive and significant effects in mathematics. For those scoring in the 90th quantile, the requirements were associated with a five to ten percentage point increase in the Meets/Exceeds rate. The economically disadvantaged also saw large positive effects for mathematics in higher quantiles with no evidence of positive or significant effects in the lower end of the distribution. For instance, estimates at the 80th quantile show that bounded subgroups have a MER six to eight percentage points higher than unbounded subgroups. Hispanics also saw positive treatment effects within ELA scores. However, unlike mathematics the effects are not concentrated in the upper end of the distribution. Hispanic subgroups scoring in the 40th quantile see a treatment effect of four to eight percentage points, while those scoring in the 80th quantile display a treatment effect of four to six percentage points. In conclusion, when examining high schools it appears under-performing subgroup categories such as Hispanics, the economically disadvantaged, and SWD benefit from binding requirements. However, the benefits are somewhat concentrated in the upper quantiles of these subgroups, suggesting that high-achieving Hispanic, ED, or SWD high schools are more able to adjust student performance in response to the binding requirements than their low-achieving counterparts.

Middle schools were the only category to show no overall positive effects of binding requirements in mathematics or ELA. The most interesting result is the negative and significant effect found in mathematics for Blacks in the upper end of the distribution. The 75th quantile of

Table 19. Quantile Effects of Binding Subgroup Requirements, Primary Schools

Subgroup Categories	Mean Effect	10th	25th	50th	75th	90th
Mathematics						
All Subgroups	0.0191***	0.0181	0.0315**	0.0213	0.0106	-0.00550
	(0.00545)	(0.0228)	(0.0159)	(0.0166)	(0.0102)	(0.0134)
Blacks	-0.0214*	-0.0387	-0.0187	-0.0215	-0.0146	-0.0377**
	(0.0129)	(0.0246)	(0.0244)	(0.0227)	(0.0181)	(0.0169)
Hispanics	0.00412	0.0506	0.0162	-0.0369	0.00602	0.0138
	(0.00772)	(0.0500)	(0.0370)	(0.0340)	(0.0349)	(0.0292)
Students with Disabilities	0.0443***	0.0310	0.0298*	0.0413**	0.0558***	0.0462**
	(0.0105)	(0.0203)	(0.0181)	(0.0178)	(0.0192)	(0.0217)
Economically Disadvantaged	0.0239	0.0347	0.0295	0.00949	0.0102	0.0563*
	(0.0175)	(0.0598)	(0.0493)	(0.0403)	(0.0395)	(0.0326)
Bandwidth	10	10	10	10	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
ELA						
All Subgroups	0.0193***	0.0365***	0.0255**	0.0182*	0.0188**	0.0167**
	(0.00551)	(0.0136)	(0.0106)	(0.0104)	(0.00813)	(0.00676)
Blacks	-0.0115	-0.0362	-0.0117	0.00124	-0.0168*	-0.00603
	(0.00916)	(0.0302)	(0.0119)	(0.0145)	(0.0101)	(0.0116)
Hispanics	-0.0112	-0.0352	-0.0344	-0.00242	0.0206	0.0105
	(0.0104)	(0.0366)	(0.0271)	(0.0184)	(0.0198)	(0.0215)
Students with Disabilities	0.0405***	0.0156	0.0377**	0.0362**	0.0412***	0.0627***
	(0.00748)	(0.0168)	(0.0183)	(0.0161)	(0.0116)	(0.0149)
Economically Disadvantaged	0.00887	0.0229	0.0326	0.0102	-0.000649	0.0124
	(0.0144)	(0.0735)	(0.0364)	(0.0229)	(0.0171)	(0.0259)
Bandwidth	10	10	10	10	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: *** significant at 1%; ** significant at 5%; * significant at 10%; Time dummies and the following school-level covariates included in all models: the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience; Dependent variable: Percentage of students Meeting/Exceeding the standard.

Table 20. Quantile Effects of Binding Subgroup Requirements, High Schools

Subgroup Categories	Mean Effect	10th	25th	50th	75th	90th
Mathematics						
All Subgroups	0.0633***	0.117**	0.0806	0.0519**	0.0386	0.0217
	(0.0111)	(0.0572)	(0.0492)	(0.0257)	(0.0359)	(0.0267)
Blacks	-0.00679	0.00976	-0.00913	0.00524	-0.0383	-0.0374
	(0.0228)	(0.0523)	(0.0533)	(0.0445)	(0.0653)	(0.0578)
Hispanics	0.112***	0.177*	0.0882	0.0599	0.0928**	0.102*
	(0.0302)	(0.0966)	(0.0628)	(0.0590)	(0.0412)	(0.0573)
Students with Disabilities	0.0311	0.0644	0.0823	0.0554	-0.00332	0.00791
	(0.0286)	(0.0679)	(0.0517)	(0.0529)	(0.0631)	(0.0871)
Economically Disadvantaged	0.0471***	-0.0676	-0.00683	0.0529	0.102**	-0.00737
	(0.0162)	(0.0649)	(0.0485)	(0.0375)	(0.0415)	(0.0602)
Bandwidth	10	10	10	10	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
ELA						
All Subgroups	0.0358**	0.0238	0.0634	0.0375	0.00483	0.00543
	(0.0161)	(0.0504)	(0.0531)	(0.0295)	(0.0184)	(0.0123)
Blacks	0.00181	-0.0313	-0.00710	0.0103	-0.0341	0.00507
	(0.00813)	(0.0594)	(0.0254)	(0.0229)	(0.0259)	(0.0245)
Hispanics	0.102***	0.141	0.109*	0.0827***	0.103**	0.0241
	(0.0209)	(0.0972)	(0.0622)	(0.0282)	(0.0500)	(0.0519)
Students with Disabilities	-0.00925	-0.0119	0.00765	0.00225	0.00123	-0.0205
	(0.0193)	(0.0582)	(0.0565)	(0.0487)	(0.0500)	(0.0361)
Economically Disadvantaged	0.0343***	0.0545	0.0301	0.0289	0.0357	0.0272
	(0.00980)	(0.0335)	(0.0260)	(0.0243)	(0.0452)	(0.0333)
Bandwidth	10	10	10	10	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: *** significant at 1%; ** significant at 5%; * significant at 10%; Time dummies and the following school-level covariates included in all models: the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience; Dependent variable: Percentage of students Meeting/Exceeding the standard.

Table 21. Quantile Effects of Binding Subgroup Requirements, Middle Schools

Subgroup Categories	Mean Effect	10th	25th	50th	75th	90th
Mathematics						
All Subgroups	0.0183	-0.0293	-0.0288	0.0167	0.0332	0.0393
	(0.0374)	(0.0313)	(0.0280)	(0.0264)	(0.0424)	(0.0419)
Blacks	0.0264	-0.0218	0.0129	-0.0648	-0.0172	0.00551
	(0.0351)	(0.119)	(0.0967)	(0.0811)	(0.109)	(0.0702)
Hispanics	-0.0365**	0.0615	-0.0279	-0.0576	0.0120	0.0315
	(0.0180)	(0.0808)	(0.0526)	(0.0360)	(0.0351)	(0.0509)
Students with Disabilities	-0.00922	-0.0268	-0.00740	0.00405	0.00996	0.00247
	(0.0337)	(0.0369)	(0.0231)	(0.0258)	(0.0262)	(0.0383)
Economically Disadvantaged	0.0129	0.0689	0.0975	0.0169	-0.0611	-0.183
	(0.0562)	(0.156)	(0.152)	(0.142)	(0.161)	(0.143)
Bandwidth	10	10	10	10	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic
ELA						
All Subgroups	0.0292	-0.00193	0.000614	0.0237	0.0346*	0.0230
	(0.0201)	(0.0355)	(0.0362)	(0.0176)	(0.0197)	(0.0186)
Blacks	0.0271	0.0346	0.0591	0.0539	0.0292	0.0268
	(0.0267)	(0.218)	(0.0575)	(0.0480)	(0.0229)	(0.0291)
Hispanics	-0.00894	0.0209	0.00781	0.0121	-0.0240	-0.0295
	(0.0148)	(0.0538)	(0.0234)	(0.0362)	(0.0364)	(0.0477)
Students with Disabilities	0.000606	-0.00258	-0.0430	-0.0224	0.0136	0.0206
	(0.0182)	(0.0361)	(0.0364)	(0.0252)	(0.0272)	(0.0282)
Economically Disadvantaged	0.0302	-0.0654	0.0180	-0.0330	-0.0316	-0.0616
	(0.0473)	(0.105)	(0.0958)	(0.0704)	(0.0726)	(0.0732)
Bandwidth	10	10	10	10	10	10
Functional Form	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: *** significant at 1%; ** significant at 5%; * significant at 10%; Time dummies and the following school-level covariates included in all models: the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience; Dependent variable: Percentage of students Meeting/Exceeding the standard.

MER for bounded Black subgroups is four to eight percentage points lower than their unbounded counterparts. It is important to note that negative estimates for Blacks in mathematics were present in both primary and high schools, but middle schools were the only category in which the effects were both negative and significant. Again, when considering the effect of minority concentration on Black students the negative effects obtained are in accordance with literature. Overall, the results show a lack of positive response to binding requirements for middle schools.

Chapter V

CONCLUSIONS

The structure of Georgia's accountability system translates certain school characteristics into accountability components. School demographics are used to determine the number of binding subgroups, and underperformance leads to corrective measures. Using an unbalanced panel covering the school years 2004-2011, we investigate the impact of these characteristics on satisfying Adequate Yearly Progress[26]. The application of sanctions appears to have a positive impact on school performance. Schools facing sanctions in the previous year were more likely to meet standards. However, the magnitude of the impact was relatively small. Increasing the number of binding requirements has a negative impact on probability of passage. Similar to results found in literature, our findings imply heterogeneous schools or those facing a greater number of binding subgroup requirements fare worse than their homogenous counterparts. This effect remains when accounting for the previous year's performance.

It is important to note the possibility of spillover effects on unbounded populations when examining the impact of binding requirements. The impact of these effects can be positive or negative. When students of a bounded group are present in a class also containing unbounded students, instructors may reallocate time and resources towards the bounded group at the expense of the rest of the class. However, it is also possible that the presence of a bounded subgroup

[26] As a robustness check models are also estimated using balanced subset only. The results obtained are similar to those found using the unbalanced panel and are available upon request.

increases the quality of instruction and resources available to the class as a whole; thus creating a positive impact of binding requirements for students in both bounded and unbounded subgroups. Determining the significance and direction of spillover effects is a topic for future research, as it requires the comparison of unbounded subgroups across schools and varying levels of binding requirements.

While each proxy for school resources has the expected sign, having inexperienced teachers yields a result that is larger than expected. One possible explanation being the percentage of inexperienced teachers serves a proxy not only for school resources, but for overall school quality. State dependence plays a factor in determining passage rate as those meeting the standard in the previous year are more likely to pass in the current. However, similar to the accountability measures examined, the magnitude of the effect is relatively small. When examining all schools collectively, high schools are least likely to meet AYP, but the mechanisms for this remain unclear.

Accountability components have statistically significant but relatively small impacts on whether or not a school meets AYP. However, factors relating to the school resources, quality, and type appear to have a greater influence. Despite this, the negative impact of binding subgroup requirements on passage rates calls into question if these measures can be effective in increasing student performance. While AYP status is a global measure of a schools well-being, it is a binary and by definition does not allow for measuring improvements in student achievement. Gains in student performance and AYP failure can occur simultaneously if the gains made are insufficient to meet the Annual Measureable Objective. For this reason, a regression discontinuity design was implemented to determine if binding requirements had a positive impact on subgroup performance. The outcome of interest measures the percentage of students within each subgroup who meet or exceed the standard. The initial estimates, when examining all subgroups collectively, indicate a slight positive impact of binding requirements on both mathematics and ELA performance. Next, we explored the possibility of heterogeneous effects by both subgroup and school type. Distinguishing by subgroup allows for the possibility of varying effects by race, ethnicity, disability, and socio-economic status. Stratifying the sample by school type increases the comparability of the subsamples examined. Positive effects on student achievement were found but concentrated within different subgroups dependent upon school type. In elementary schools, students with disabilities were the only group to show positive and statistically significant gains from binding requirements. Within high schools, improvements in student performance were concentrated within Hispanics and the economically disadvantaged. For middle schools, no positive effects were found and the performance of Blacks appeared to suffer from the implementation of subgroup requirements.

While the mean effect was determined to be small but positive in nature, differences in both magnitude and sign are possible when examining different points in the performance distribution. For this reason, we employed quantile estimation. When examining all subgroups collectively, those in the lower end of the distribution benefited most from binding requirements. During the time period studied Blacks, Hispanics, students with disabilities, and the economically disadvantaged were considered lower-performing relative to all others. Therefore, these groups appear to generate the positive results found when examining all subgroups collectively. Heterogeneous effects may exist within individual groups, and it is for this reason that quantile estimates were obtained for each group separately. Similar to the mean effects discussed, within elementary schools students with disabilities bounded by requirements outperform their counterparts in both mathematics and ELA. Gains are seen throughout the entire distribution and not confined to high or low achievers. However, the greatest gains were seen towards the upper end of the distribution. When examining high schools, gains are concentrated in the upper ends of the mathematics distribution for both Hispanics and the economically disadvantaged. Therefore, higher performing schools, with respect to these particular subgroups, showed the greatest improvements in student performance. While historically underperforming subgroups appear to benefit from binding requirements, it is the higher achieving schools with respect to these subgroups that benefit most. The ELA results for Hispanic subgroups are similar as those high schools in the middle to upper ends of the distribution benefit most.

In conclusion, our results indicate that imposing sanctions on failing schools has a positive impact on the probability these schools will meet AYP in the following year. Therefore, sanctions are an effective mechanism for improved school performance. Our findings mirror those found in previous literature which highlights the importance of sanction severity and credibility. During the 2006-2011 time period, an average of 117 schools per year faced possible restructuring – the most severe sanction available under Georgia's SSAS. The consequential structure of Georgia's accountability system, which mandates the sanction of failing schools, served to increase performance and decrease the probability of repeated failures.

NCLB implemented the use of subgroup classifications to ensure the performance of under-represented groups was not overlooked in the calculation of AYP. Chapter III of this dissertation shows that as the count of binding requirements increases within a school, the probability of AYP passage decreases. Therefore, these findings imply that heterogeneous schools have a lower probability of passage than their more homogenous counterparts. However, the mechanism for the decrease in probability of passage remains unidentified. The results of Chapter IV of this dissertation indicate that bounded groups outperformed unbounded groups within the SWD, Hispanic, and ED subgroups. However, the magnitudes of these effects were relatively small and largest in comparatively higher performing schools. The persistence of the achievement gap mandates the need for targeting subgroup populations; however the results as presented suggest the need to re-evaluate binding requirements as the tool used to accomplish this goal. The requirements do not yield substantial growth in all targeted populations, and may prove detrimental to heterogeneous schools in relation to AYP passage. Any method implemented must acknowledge the obstacles faced by schools whose student body contains a large percentage of disadvantaged students.

While there may be unintended negative consequences of binding subgroup requirements on the probability of passing AYP, there is also evidence to suggest that bounded subgroups marginally outperform their unbounded counterparts. Therefore, the mechanism for detrimental impact of binding subgroup requirements on passing AYP is undetermined. Increases in the number of passage requirements could create downward pressure on probability of passing or generate improvements in student performance for bounded groups that are insufficient relative to the proficiency threshold. However, it is important to note that this study investigates student performance as a whole and does not address student performance in relation to the AMO set by the state. The impact of subgroup requirements with respect to a schools ability to meet AMOs is a subject for future research.

Beginning in the 2012-2013 school year, Georgia implemented the College and Career Ready Performance Index (CCRPI) which eliminated subgroup binding requirements as defined by NCLB. The new system uses a combination of indicators measuring overall school performance, individual student growth, and the relative performance of low-achieving students. The index also contains specific measures regarding students with disabilities, those with limited

English proficiency, and the economically disadvantaged. However, examining the performance of these subgroups is not a central component of the index, and is simply used as a method for earning additional points. It has yet to be determined if Georgia's new system, with the complete elimination of specific subgroup requirements, will result in improved scores for disadvantaged populations. Comparing the outcomes of these groups under NCLB and CCRPI is a question for future research.

Appendix A: Annual Measurable Objectives

Georgia's Annual Measureable Objectives (AMOs)

English/Language Arts	
CRCT Grades 3-8	
School Year	**% of Students Meeting or Exceeding Standard**
2002-2003	60
2003-2004	60
2004-2005	66.7
2005-2006	66.7
2006-2007	66.7
2007-2008	73.3
2008-2009	73.3
2009-2010	73.3
2010-2011	80
2011-2012	86.7
2012-2013	93.3
2013-2014	100

Mathematics	
CRCT Grades 3-8	
School Year	**% of Students Meeting or Exceeding Standard**
2002-2003	50
2003-2004	50
2004-2005	58.3
2005-2006	58.3
2006-2007	58.3
2007-2008	59.5
2008-2009	59.5
2009-2010	67.6
2010-2011	75.7
2011-2012	83.8
2012-2013	91.9
2013-2014	100

English/Language Arts	
GHSGT	
School Year	**% of Students Meeting or Exceeding Standard**
2002-2003	88
2003-2004	81.6
2004-2005	81.6
2005-2006	84.7
2006-2007	84.7
2007-2008	87.7
2008-2009	87.7
2009-2010	87.7
2010-2011	90.8
2011-2012	93.9
2012-2013	96.9
2013-2014	100

Mathematics	
GHSGT	
School Year	**% of Students Meeting or Exceeding Standard**
2002-2003	81
2003-2004	62.3
2004-2005	62.3
2005-2006	68.6
2006-2007	68.6
2007-2008	74.9
2008-2009	74.9
2009-2010	74.9
2010-2011	76
2011-2012	84
2012-2013	92
2013-2014	100

Following the March 2005 administration of the Enhanced GHSGT, AMOs reset for High Schools.
Following the Spring 2008 administration of the mathematics CRCT, AMOs for grades 3-8 reset.
Following the Spring 2011 administration of the mathematics GHSGT, AMOs reset for High Schools.

Source: State of Georgia Consolidated State Application Accountability Workbook, U.S. Department of Education

Appendix B: Graduation Rate Requirements for Second Indicator of AYP

High School Second Indicator
Graduation Rate

School Year	**High School AYP Second Indicator Graduation Standard**
2006-2007	65 % or greater; OR 1) Apply multi-year average to achieve 65% 2) Increase by 10% from the preceding year from a minimum threshold of 50%
2007-2008	70% or greater; OR 1) Apply multi-year average to achieve 70% 2) Increase by 10% from the preceding year from a minimum threshold of 50%
2008-2009	75% or greater; OR 1) Apply multi-year average to achieve 75% 2) Increase by 10% from the preceding year from a minimum threshold of 55%
2009-2010	80% or greater; OR 1) Apply multi-year average to achieve 80% 2) Increase by 10% from the preceding year from a minimum threshold of 60%
2010-2011	85% or greater; OR 1) Apply multi-year average to achieve 85% 2) Increase by 10% from the preceding year from a minimum threshold of 60%
2011-2012	90% or greater; OR 1) Apply multi-year average to achieve 90% 2) Increase by 10% from the preceding year from a minimum threshold of 70%
2012-2013	95% or greater; OR 1) Apply multi-year average to achieve 95% 2) Increase by 10% from the preceding year from a minimum threshold of 70%
2013-2014	100% or greater; OR 1) Apply multi-year average to achieve 100% 2) Increase by 10% from the preceding year from a minimum threshold of 80%

NOTE: From 2002-2006 the Graduation Standard was set at 60%
Source: State of Georgia Consolidated State Application Accountability Workbook, U.S. Department of Education

Appendix C: Transition from Quality Core Curriculum (QCC) to Georgia Performance Standards (GPS).

Grade	School Year GPS Used in Testing for English / Language Arts	School Year GPS Used in Testing for Mathematics
3	05-06	07-08
4	05-06	07-08
5	05-06	07-08
6	05-06	05-06
7	05-06	06-07
8	05-06	07-08
11	07-08	10-11

Training in the new GPS curriculum was given in the school year directly preceding implementation for Grades 3-8.

Training for the GPS High School Math Curriculum was administered in the 2007-2008 school year. The new curriculum was then implemented for all ninth graders entering in the 2008-2009 school year culminating in GPS standards being used on the Math portion of the GHSGT administered to 11th graders in the 2010-2011 school year.

Training for the GPS High School ELA Curriculum was administered in the 2004-2005 school year. The new curriculum was implemented for all ninth graders entering in the 2005-2006 school year culminating in GPS standards being used on the ELA portion of the GHSGT administered to 11th graders in 2007-2008 school year.

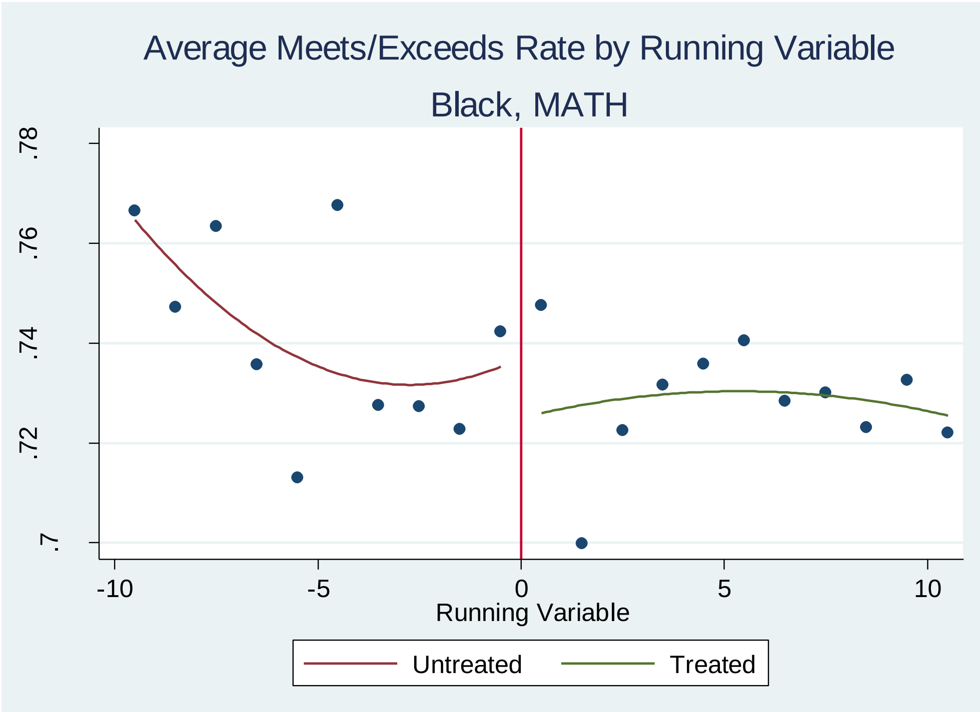
Source: State of Georgia, Consolidated State Application Accountability Workbook, U.S. Department of Education

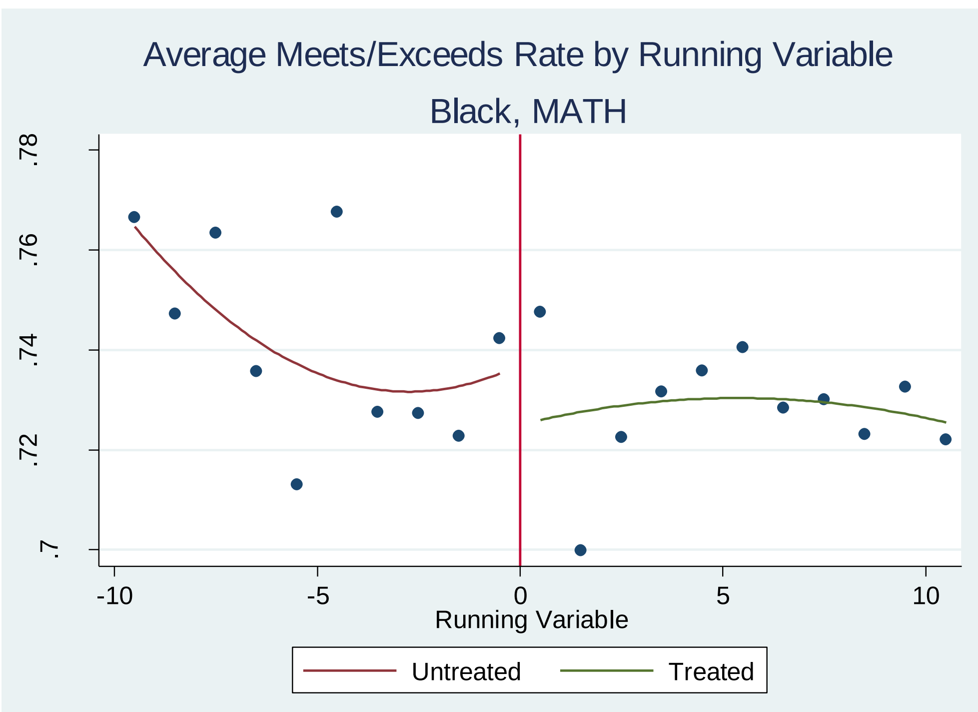
Appendix D: Correlated Random Effects Probit Model, Pooled Maximum Likelihood Estimation

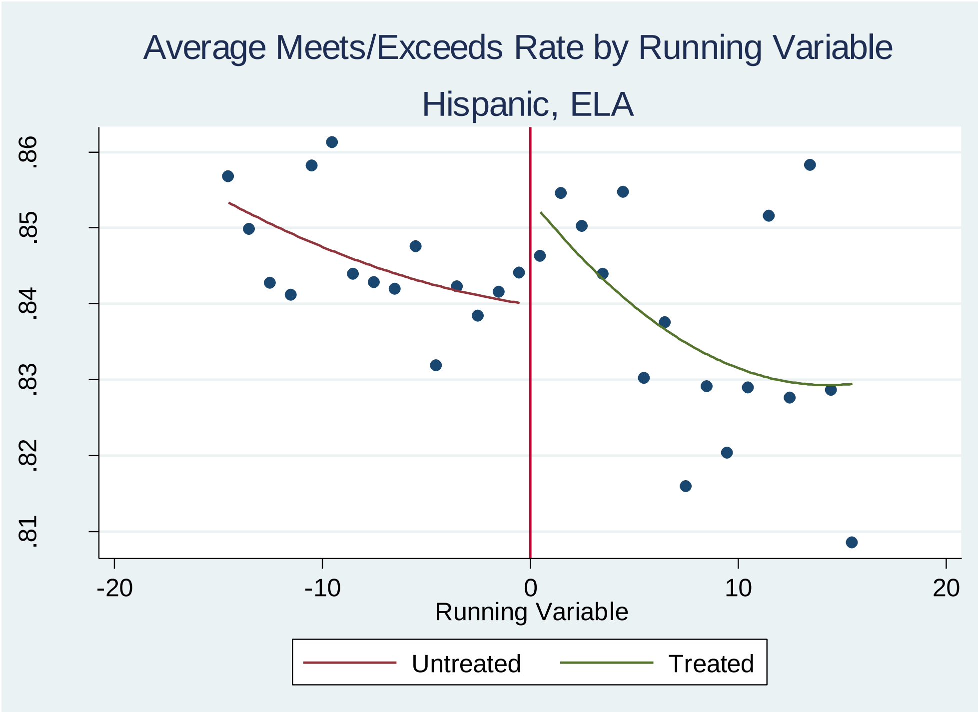
	Static				Dynamic			
	(1)		(2)		(1)		(2)	
Variables	Coeff.	APE	Coeff.	APE	Coeff.	APE	Coeff.	APE
Lagged Value of AYP Status					0.620***	0.140***	0.483***	0.101***
					(0.036)	(0.009)	(0.039)	(0.009)
Initial Condition for AYP Status					0.180***	0.036***	0.009	0.002
					(0.041)	(0.008)	(0.039)	(0.007)
Sanctions			0.329***	0.054***			0.461***	0.071***
			(0.042)	(0.006)			(0.042)	(0.006)
# of Binding Subgroups	-0.269***	-0.054***	-0.290***	-0.053***	-0.261***	-0.050***	-0.281***	-0.051***
	(0.016)	(0.003)	(0.017)	(0.003)	(0.015)	(0.003)	(0.016)	(0.003)
% of FRL Students	-0.352	-0.071	-0.642*	-0.118*	-0.285	-0.055	-0.659**	-0.119**
	(0.308)	(0.062)	(0.340)	(0.062)	(0.297)	(0.057)	(0.330)	(0.059)
Pupil/Teacher Ratio	-0.038***	-0.008***	-0.040***	-0.007***	-0.043***	-0.008***	-0.043***	-0.008***
	(0.013)	(0.003)	(0.014)	(0.003)	(0.014)	(0.003)	(0.014)	(0.003)
% of Teachers w. Experience less than 1 year	-1.319***	-0.264***	-1.526***	-0.280***	-1.297***	-0.249***	-1.511***	-0.272***
	(0.281)	(0.056)	(0.311)	(0.057)	(0.295)	(0.056)	(0.317)	(0.057)
Middle School	0.898***	0.280***	0.815***	0.224***	0.766***	0.224***	0.737***	0.193***
	(0.058)	(0.017)	(0.049)	(0.013)	(0.050)	(0.015)	(0.045)	(0.012)
Other	0.289*	0.093*	0.235	0.072	0.150	0.048	0.181	0.053
	(0.174)	(0.056)	(0.156)	(0.046)	(0.153)	(0.048)	(0.141)	(0.040)
Primary	1.912***	0.483***	1.310***	0.316***	1.461***	0.359***	1.161***	0.271***
	(0.055)	(0.013)	(0.053)	(0.014)	(0.059)	(0.015)	(0.054)	(0.014)
Constant	1.726***		1.409***		1.036***		1.057***	
	(0.270)		(0.241)		(0.255)		(0.236)	
N	16379	16379	16379	16379	16379	16379	16379	16379

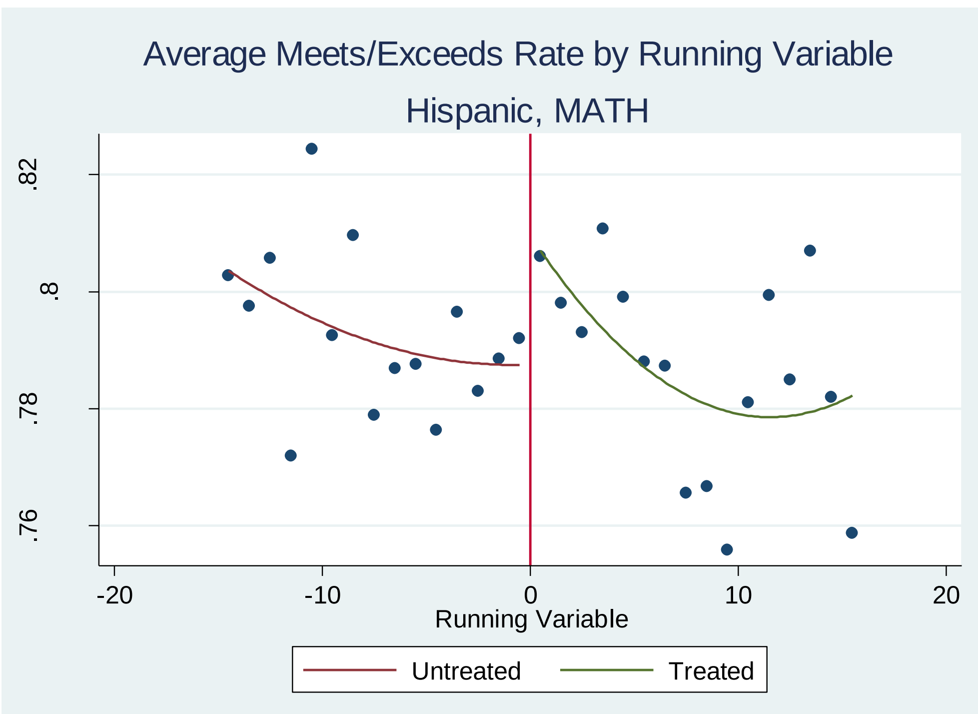
Note: All specifications include time dummies; Clustered standard errors are in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Mundlak-Chamberlain equation includes time averages of all time-varying variables including time dummies. Dependent variable is a binary indicator equal to one if a school meets AYP and zero otherwise

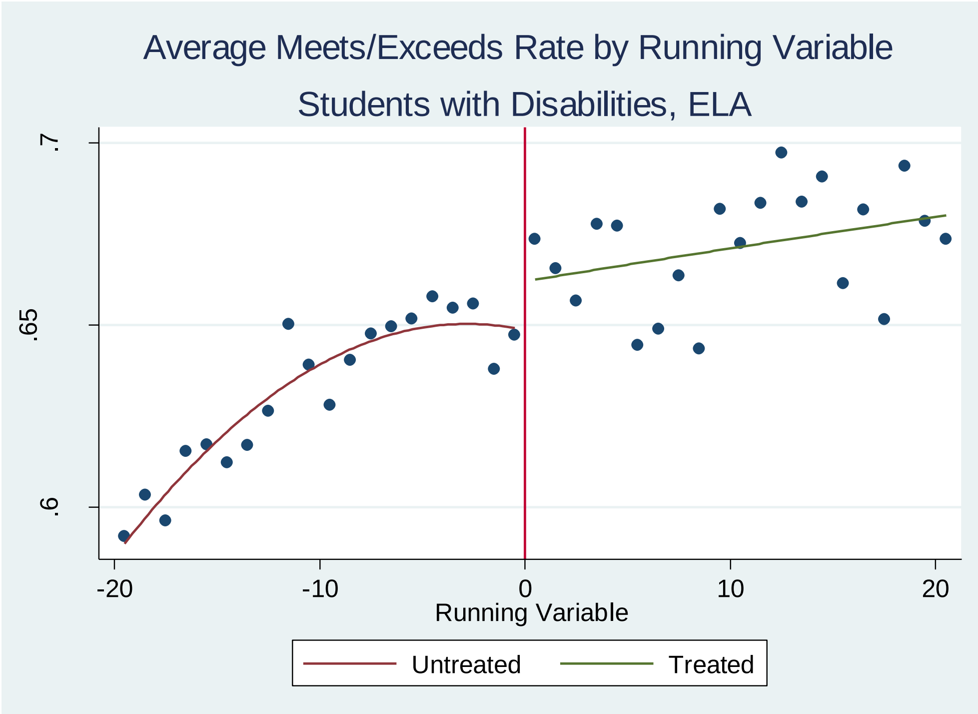
Appendix E: Average Meets/Exceeds Rate by Subject and Subgroup

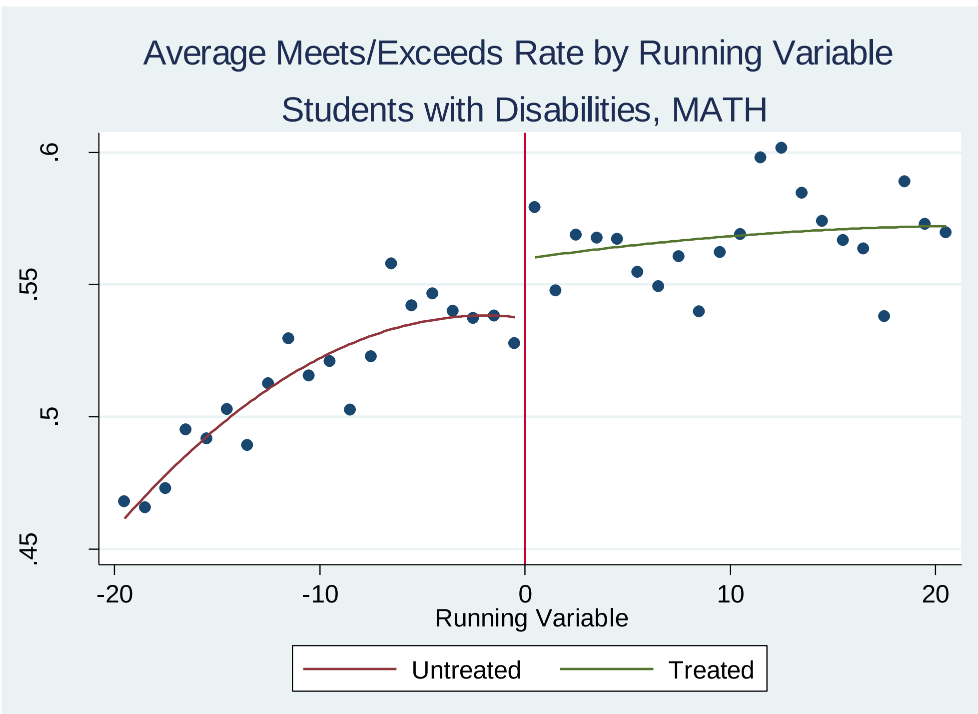


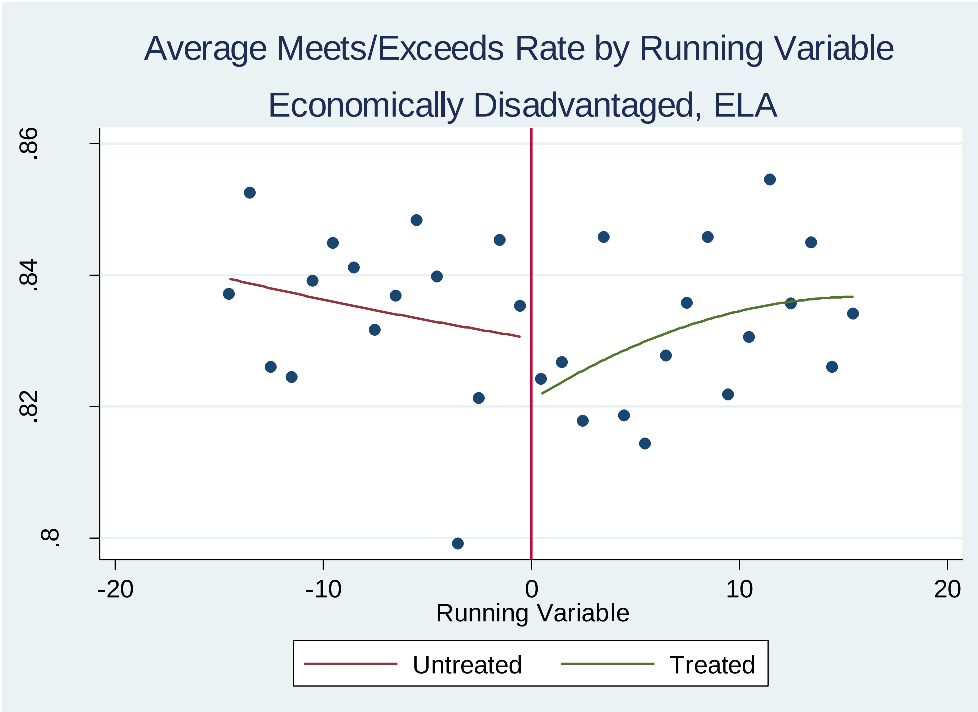


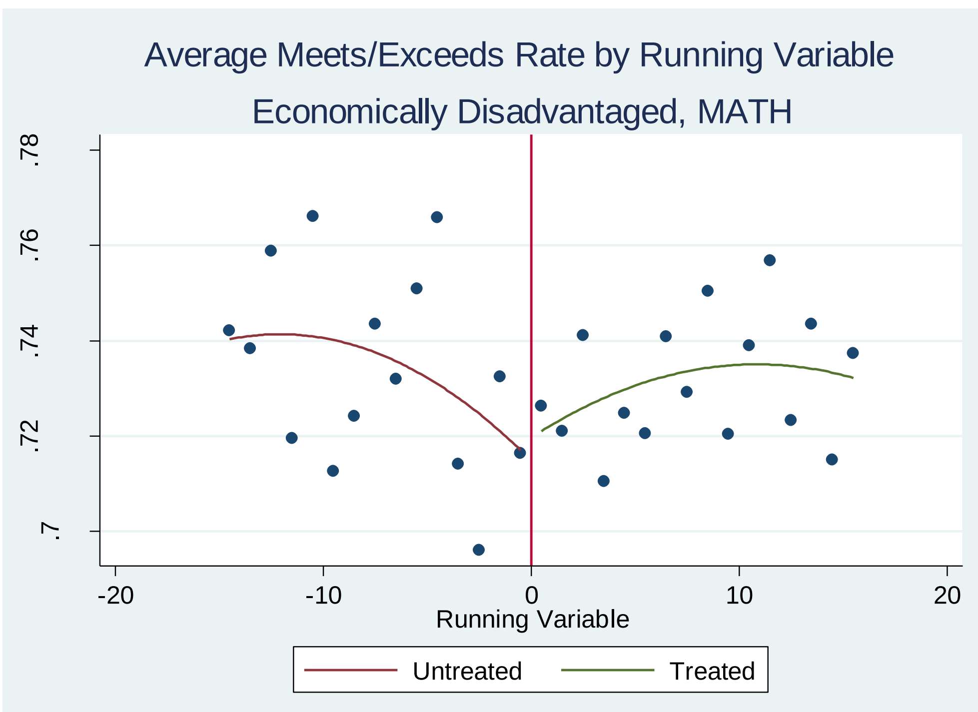

Average Meets/Exceeds Rate by Running Variable
Hispanic, ELA
.81
.82
.83
.84
.85
.86
-20
-10
0
10
20
Running Variable
Untreated
Treated


Average Meets/Exceeds Rate by Running Variable
Hispanic, MATH
.76
.78
.8
.82
-20
-10
0
10
20
Running Variable
Untreated
Treated


Average Meets/Exceeds Rate by Running Variable
Students with Disabilities, ELA
.6
.65
.7
-20
-10
0
10
20
Running Variable
Untreated
Treated


Average Meets/Exceeds Rate by Running Variable
Students with Disabilities, MATH
.45
.5
.55
.6
-20
-10
0
10
20
Running Variable
Untreated
Treated


Average Meets/Exceeds Rate by Running Variable
Economically Disadvantaged, ELA
.8
.82
.84
.86
-20
-10
0
10
20
Running Variable
Untreated
Treated


Average Meets/Exceeds Rate by Running Variable
Economically Disadvantaged, MATH
.7
.72
.74
.76
.78
-20
-10
0
10
20
Running Variable
Untreated
Treated

Appendix F: Treatment Effects of Binding Subgroup Requirements using Ordinary Least Squares (OLS), All Subgroups

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Mathematics								
Treatment	0.0206***	0.0166***	0.0211**	0.0151*	0.0203**	0.0142*	0.0343***	0.0248**
	(0.0055)	(0.0049)	(0.0085)	(0.0073)	(0.0092)	(0.0084)	(0.0109)	(0.0096)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	27248	27105	13255	13180	27248	27105	13255	13180
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Linear	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic	Quadratic
ELA								
Treatment	0.0140***	0.0126***	0.0162***	0.0142***	0.0143***	0.0118***	0.0261***	0.0228***
	(0.0045)	(0.0036)	(0.0054)	(0.0042)	(0.0052)	(0.0043)	(0.0075)	(0.0046)
Include Year Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Include School-Level Covariates?	No	Yes	No	Yes	No	Yes	No	Yes
N	27215	27072	13237	13162	27215	27072	13237	13162
Bandwidth	20	20	10	10	20	20	10	10
Functional Form	Linear	Linear	Linear	Linear	Quadratic	Quadratic	Quadratic	Quadratic

NOTE: Ordinary Least Squares Regression with clustered standard errors in parentheses; *** significant at 1%; ** significant at 5%; * significant at 10%; Dependent variable: Percentage of students Meeting/Exceeding the standard; School-level Covariates include the percentage of all students receiving Free or Reduced Lunch, the student/teacher ratio, and the percentage of teachers with less than one year's experience.

BIBLIOGRAPHY

Abadie, Alberto, Joshua Angrist, and Guido Imbens. 2002. “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings.” *Econometrica*, 70 (1): 91-117.

Ahn, Thomas, and Jacob L. Vigdor. 2013. "The Impact of No Child Left Behind's Accountability Sanctions on School Performance: Regression Discontinuity Evidence from North Carolina." Unpublished Manuscript.

Alexander, Nicola A. 2006. "Being on Track for No Child Left Behind - Examining the Organizational Capacity of Massachusetts Public Eighth-Grade Programs." *Educational Policy*, 20 (2): 399-428.

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, 114 (2): 533-575.

Bacolod, Marigee, John DiNardo, and Mireille Jacobson. 2012. "Beyond Incentives: Do Schools Use Accountability Rewards Productively?" *Journal of Business & Economic Statistics*, 30 (1): 149-163.

Ballou, Dale, and Matthew G. Springer. 2009. "Achievement Trade-Offs and No Child Left Behind." *The Urban Institute*. Unpublished Manuscript.

Betts, Julian R., and Robert M. Costrell. 2001. "Incentives and Equity Under Standards-Based Reform." In *Brookings Papers on Education Policy*, edited by Diane Ravitch, p. 9-74. Washington D.C.: The Brookings Institute.

Booher-Jennings, Jennifer. 2005. "Below the Bubble: 'Educational Triage' and the Texas Accountability System." *American Educational Research Journal*. 42 (2): 231-268.

Brewer, Dominic J., Guilbert C. Hentschke, Eric R. Eide, Ally Kuzin, and Michelle B. Nayfack. 2008. "The Role of Economics in Education Policy Research." In *Handbook of Research Education Finance and Policy*, edited by Helen F. Ladd and Edward B. Fiske, p. 23-41. New York: Routledge.

Card, David, Carlos Dobkin, and Nicole Maestas. 2004. "The Impact of Nearly Universal Insurance Coverage on Health Care Utilization and Health: Evidence from Medicare." *National Bureau of Economic Research*. Working Paper 10365.

Carnoy, Martin, and Susanna Loeb. 2002. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis." *Educational Evaluation and Policy Analysis*, 24 (4): 305-331.

Chakrabarti, Rajashri. 2013. "Accountability with Voucher Threats, Responses, and the Test Taking Population: Regression Discontinuity Evidence from Florida." *Education Finance & Policy*, 8 (2): 121-167.

Chamberlain, Gary. 1984. "Panel data." In *Handbook of Econometrics Volume 2*, edited by Zvi Grilches and Michael D. Intrillgator, p. 1247-1318. Amsterdam: Elsevier.

Chay, Kenneth Y., Patrick J. McEwan and Miguel Urquiola. 2005. "The Central Role of Noise in Evaluating Interventions That Use Test Scores to Rank Schools." *The American Economic Review*, 95 (4): 1237-1258.

Chiang, Hanley. 2009. "How Accountability Pressure on Failing Schools Affects Student Achievement." *Journal of Public Economics*, 93 (9-10): 1045-1057.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2004. "Do School Accountability Systems Make It More Difficult for Low-Performing Schools to Attract and Retain High-Quality Teachers?" *Journal of Policy Analysis & Management*, 23 (2): 251-271.

Clotfelter, Charles T., Helen F. Ladd, and Jacob L. Vigdor. 2009. "The Academic Achievement Gap in Grades 3 to 8." *The Review of Economics and Statistics*, 91 (2): 398-419.

Cronin, John, G. Gage Kingsbury, Martha S. McCall, and Branin Bowe. 2005. "The Impact of the No Child Left Behind Act on Student Achievement and Growth: 2005 Edition." *Northwest Evaluation Association*. Northwest Evaluation Association Technical Report.

Cullen, Julie Berry, and Randall Reback. 2006. "Tinkering Toward Accolades: School Gaming Under a Performance Accountability System." *National Bureau of Economic Research.* Working Paper 12286.

Figlio, David N., and Lawrence S. Getzler. 2002. "Accountability, Ability and Disability: Gaming the System " *National Bureau of Economic Research.* Working Paper 9307.

Figlio, David N., and Helen F. Ladd. 2008. "School Accountability and Student Achievement." In *Handbook of Research in Education Finance and Policy*, edited by Helen F. Ladd and Edward B. Fiske, p. 166-182. New York: Routledge.

Figlio, David N., and Maurice E. Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *The American Economic Review*, 94 (3): 591-604.

Figlio, David N., and Cecilia E. Rouse. 2006. "Do accountability and voucher threats improve low-performing schools?" *Journal of Public Economics*, 90 (1-2): 239-255.

Figlio, David N.. Cecilia E. Rouse, and Analia Schlosser. 2009. "Leaving No Child Behind: Two Paths of School Accountability." *NCLB: Emerging Findings Research Conference*. Washington, D.C.

Frandsen, Brigham R., Markus Frolich, and Blaise Melly. 2012. "Quantile Treatment Effects in the Regression Discontinuity Design." *Journal of Econometrics*, 168 (2): 382-395.

Frolich, Markus, and Blaise Melly. 2010. "Estimation of Quantile Treatment Effects with STATA." *STATA Journal*, 10 (3): 423-457.

Gillborn, David, and Deborah Youdell. 2000. *Rationing Education: Policy, Practice, Reform and Equity*. London, England: Open University Press.

Gordon, Robert, Thomas J. Kane, and Douglas Staiger. 2006. "Identifying Effective Teachers Using Performance on the Job." *Hamilton Project Discussion Paper 2006-01*. Washington, D.C.: The Brookings Institute.

Grissmer, David W., Ann Flanagan, Jennifer H. Kawata, and Stephanie Williamson. 2000. *Improving Student Achievement: What State NAEP Test Scores Tell Us*. Santa Monica, California : RAND.

Guiteras, Raymond. 2008. "Estimating Quantile Treatment Effects in a Regression Discontinuity Design." Mimeo. University of Maryland.

Hahn, Jinyong, Petra Todd, and Wilbert van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69 (1): 201-209.

Hamilton, Laura S., Brian M. Stetcher, Julie A. Marsh, Jennifer Sloan McCombs, Abby Robyn, Jennifer Russell, Scott Naftel, and Heather Barney. 2007. *Standards-based Accountability under No Child Left Behind: Responses of Superintendents, Principals, and Teachers in Three States*. Santa Monica, California: RAND.

Hanushek, Eric A. 2008. "Education Production Functions." In *The New Palgrave Dictionary of Economics*, edited by Steven N. Durlauf and Lawrence E. Blume. Basingstoke: Palgrave Macmillan.

Hanushek, Eric A. 2010. "Education Production Functions: Developed Countries." In *Economics of Education* edited by Dominic J. Brewer and Patrick J. McEwan, p. 132-136. Amsterdam: Elsevier.

Hanushek, Eric A., John F. Kain, Daniel M. O'Brien, and Steven G. Rivkin. 2005. "The Market for Teacher Quality." *National Bureau of Economic Research*. Working Paper 11154.

Hanushek, Eric A., John F. Kain, and Steven G. Rivkin. 2009. "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." *Journal of Labor Economics*, 27 (3): 349-383.

Hanushek, Eric A., and Margaret E. Raymond. 2003. "Improving Educational Quality: How to Best Evaluate Our Schools." In *Education in the 21st Century: Meeting the Challenges of a Changing World*, edited by Yolanda K. Kodrzycki, p. 193-224. Boston, Massachusetts: Federal Reserve Bank of Boston.

Hanushek, Eric A., and Margaret E. Raymond. 2004. "The Effect of School Accountability Systems on the Level and Distribution of Student Achievement." *Journal of the European Economic Association*, 2 (2-3): 406-415.

Hanushek, Eric A., and Margaret E. Raymond. 2005. "Does School Accountability Lead to Improved Student Performance?" *Journal of Policy Analysis & Management*, 24 (2): 297-327.

Hastings, Justine S., and Jeffrey M. Weinstein. 2008. "Information, School Choice, and Academic Achievement: Evidence from Two Experiments." *Quarterly Journal of Economics*, 123 (4):1373-1414.

Hemelt, Steven W. 2011. "Performance effects of failure to make Adequate Yearly Progress (AYP): Evidence from a regression discontinuity framework." *Economics of Education Review*, 30 (4): 702-723.

Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics*, 115 (4):1239-1285.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics*, 142 (2): 615-635.

Jacob, Brian A. 2002. "Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *National Bureau of Economic Research*. Working Paper 8968.

Jacob, Brian A., and Lars Lefgren. 2004. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *The Review of Economics and Statistics*, 86 (1): 226-244.

Kane, Thomas J., and Douglas O. Staiger. 2003. "Unintended Consequences of Racial Subgroup Rules." In *No Child Left Behind? The Politics and Practice of Accountability*, edited by Paul E. Peterson and Martin R. West, p. 152-176. Washington, D.C.: Brookings Institution Press.

Kodrzycki, Yolanda K. 2002. "Education in the 21st Century: Meeting the Challenges of a Changing World." *New England Economic Review*, Q4: 3-18. Boston, Massachusetts: Federal Reserve Bank of Boston.

Koenker, Roger, and Gilbert Bassett Jr. 1978. "Regression Quantiles." *Econometrica*, 46 (1): 33-50.

Krieg, John M. 2008. "Are Students Left Behind? The Distributional Effects of the No Child Left Behind Act." *Education Finance & Policy*, 3 (2): 250-281.

Krieg, John M. 2011. "Which students are left behind? The racial impacts of the No Child Left Behind Act." *Economics of Education Review*, 30 (4): 654-664.

Ladd, Helen F. 2007. "Holding Schools Accountable Revisited." *Association for Public Policy Analysis and Management*. 2007 Spencer Foundation Lecture in Education Policy and Management.

Ladd, Helen F., and Douglas L. Lauen. 2010. "Status versus Growth: The Distributional Effects of School Accountability Policies." *Journal of Policy Analysis and Management,* 29 (3): 426-450.

Lauen, Douglas Lee, and S. Michael Gaddis. 2012. "Shining a Light or Fumbling in the Dark? The Effects of NCLB's Subgroup-Specific Accountability on Student Achievement." *Educational Evaluation & Policy Analysis*, 34 (2): 185-208.

Lee, David S. 2008. "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics,* 142 (2): 675-697.

Lee, David S., and David Card. 2008. "Regression Discontinuity Inference with Specification Error." *Journal of Econometrics*, 142 (2): 655-674.

Lee, David S., and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48 (2): 281-355.

Lee, Jaekyung, and Kenneth K. Wong. 2004. "The Impact of Accountability on Racial and Socioeconomic Equity: Considering Both School Resources and Achievement Outcomes." *American Educational Research Journal*, 41 (4):797-832.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142 (2): 698-714.

Monk, David H. 1989. "The Education Production Function: Its Evolving Role in Policy Analysis." *Educational Evaluation & Policy Analysis*, 11 (1):31-45.

Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica*, 46 (1): 69-85.

Neal, Derek and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *The Review of Economics & Statistics*, 92 (2): 263-283.

Oreopoulos, Philip. 2006. "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter." *American Economic Review*, 96 (1): 152-175.

Papke, Leslie E., and Jeffrey M. Wooldridge. 1996. "Econometric Methods for Fractional Response Variables With an Application to 401 (K) Plan Participation Rates." *Journal of Applied Econometrics*, 11 (6): 619-632.

Reback, Randall. 2008. "Teaching to the rating: School accountability and the distribution of student achievement." *Journal of Public Economics*, 92 (5-6): 1394-1415.

Rockoff, Jonah, and Lesley J. Turner. 2010. "Short-Run Impacts of Accountability on School Quality." *American Economic Journal: Economic Policy*, 2 (4): 119-147.

Rouse, Cecilia Elena, Jane Hannaway, Dan Goldhaber, and David Figlio. 2013. "Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure." *American Economic Journal: Economic Policy*, 5 (2):251-281.

Sielke, Catherine C. 2000. "Georgia's A Plus Education Reform Act of 2000." In *School Finance at the Start of the Millennium: The State of the States and Provinces 2000*, edited by Marge Plecki, p. 51-99. Seattle, Washington: The University of Washington.

Sims, David P. 2013. "Can failure succeed? Using racial subgroup rules to analyze the effect of school accountability failure on student performance." *Economics of Education Review*, 32 (2013): 262-274.

Springer, Matthew G. 2007. "Accountability Incentives: Do Failing Schools Practice Educational Triage?" *Education Next*, 8(1): 74-79.

Thistlethwaite, Donald L., and Donald T. Campbell. 1960. "Regression-discontinuity analysis: An alternative to the ex post facto experiment." *Journal of Educational Psychology*, 51 (6): 309-317.

U.S. Department of Education. 2008. "State of Georgia Consolidated State Application Accountability Workbook" prepared by the Office of Elementary and Secondary Education. Washington, D.C.: United States Government Printing Office.

van der Klaauw, Wilbert H. 2002. "Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression–Discontinuity Approach." *International Economic Review*, 43 (4): 1249-1287.

van der Klaauw, Wilbert H. 2008. "Regression–Discontinuity Analysis: A Survey of Recent Developments in Economics." *LABOUR: Review of Labour Economics & Industrial Relations*, 22 (2): 219-245.

van der, Klaauw, Wilbert H. 2008. "Breaking the link between poverty and low student achievement: An evaluation of Title I." *Journal of Econometrics,* 142 (2): 731-756.

West, Martin R., and Paul E. Peterson. 2006. "The Efficacy of Choice Threats Within School Accountability Systems: Results from Legislatively Induced Experiments." *The Economic Journal*, 116 (510): C46-C62.

Wooldridge, Jeffrey M. 2005. "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity." *Journal of Applied Econometrics*, 20 (1): 39-54.

Wooldridge, Jeffrey M. 2010. "Correlated Random Effects with Unbalanced Panels." *Manuscript*. Michigan State University.

Wooldridge, Jeffrey M. 2013. "Correlated Random Effects Panel Data Models" [PDF document]. Retrieved from http://www.iza.org/conference_files/SUMS_2013/slides_1_linear_iza.pdf

Wooldridge, Jeffrey M., and Guido W. Imbens. 2007. "What's New in Econometrics Lecture Notes 3 - Regression Discontinuity Designs" [PDF document]. Retrieved from http://www.nber.org/WNE/lect_3_rd_fig.pdf

VITA

Ashley Lynn Custard was born on May 27, 1984 in Greenville, South Carolina, USA. She holds a Bachelor of Arts degree in Economics (Minor: Financial Economics) from Vanderbilt University located in Nashville, Tennessee. After completion of her Bachelor's degree she attended the American Economic Association's Summer Training and Minority Scholarship Program hosted by Duke University.

Ashley began the Doctoral Program in Economics at Georgia State University in 2006. She chose to study public finance with an interest in the economics of education. She has worked as a graduate research assistant for Dr. Inas Rashad and Dr. Mary Beth Walker. From 2006-2009 Ashley was the recipient of the Andrew Young School of Policy Studies Dean's Fellowship. During the summer of 2008, she worked with the Congressional Research Service as an Analyst Assistant. She also served as an economics tutor to undergraduate students (Fall 2008, Spring 2009). Ashley was awarded the Dan Sweat Dissertation Fellowship in 2010. In September of 2011 she accepted a position with the Georgia Student Finance Commission (GSFC) as a Data Liaison. She began serving as Education Policy Researcher within GSFC in November 2011, and as Senior Education Policy Researcher in September 2012.

Ashley received her Doctorate of Philosophy in Economics from Georgia State University in May of 2014. To date, she plans to continue her work as Senior Education Policy Researcher with the Georgia Student Finance Commission.