Essays in Economics of Education

by

Jieun Choi

A Dissertation Submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Economics

Written under the direction of

Jennifer Hunt

And approved by

New Brunswick, New Jersey

October 2018

ABSTRACT OF THE DISSERTATION

ESSAYS IN ECONOMICS OF EDUCATION

By JIEUN CHOI

Dissertation Director:

Professor Jennifer Hunt

This dissertation studies the impact of school accountability on student outcomes - test scores and peer group composition. By examining both federal and state accountability systems, I seek to expand our understanding of the role of school accountability. Under the federal accountability, the No Child Left Behind policy, public schools are mandated to be held accountable for student achievement by race and ethnicity. The threat of sanctions may incentivize schools to reallocate school resources towards specific racial and ethnic groups to improve their achievement. In Chapter 2, I examine the policy to see whether schools improve achievement of black students at the expense of whites' achievement when the black subgroups miss annual proficiency goals, but the white groups meet them. Analyzing two nationwide data files, the Early Childhood Longitudinal Study and the Barnard/Columbia NCLB Database, I do not find strong evidence for this. Instead, I find that there is a large improvement in black student math scores and white reading scores when both blacks and whites fail. I find that the improvement comes from highly proficient students.

In Chapter 3, I seek to understand whether test-based school accountability decreases the racial diversity of a student population in schools to understand how school ratings affects a racial composition of children' peer groups. Examining North Carolina school accountability, I find that in low-performing schools during the post-

policy period, the share of black students statistically significantly falls relative to the control schools over time while the share of white students statistically significantly rises relative to the control group of schools. My findings imply that the racial diversity increases since white student enrollment increases in schools where the majority of a student population is black.

Acknowledgements

I cannot complete my doctoral program and this dissertation without my advisor's support. Professor Jennifer Hunt has motivated me to pursue my research interests. She has always suggested me invaluable guidance on whatever I brought to her office and helped me turn the ideas into interesting papers. Professor Hunt has also given me emotional support. She has deeply understood me when I had struggle with my doctoral program. She has always trusted me and never given up on me. This has encouraged me to accomplish what I wanted for a long time - being a researcher. I feel extremely lucky because I have met her as a mentor in my life at Rutgers. I will always be grateful to her.

I would also like to thank Professor Amanda Agan, Professor Ira Gang, and Professor Hilary Sigman. Professor Agan helped me polish my paper all the time, giving me very useful guidance and support until I finished my dissertation defense and this program. I cannot thank her enough. Professor Gang and Professor Sigman have taken a lot of time to discuss papers with me and given me invaluable comments and intuition. I am also thankful for Linda. I could not finish all things on time without her. I also thank Paula, Janet, Debbie, Donna, and Matt.

I would additionally like to thank Vicky, and all friends at Rutgers. I thank all participants at Rutgers Grad Workshops, annual conferences of AEA, AEFP, WEAI, MEA and EEA, and Barcelona GSE Labor Economics Summer School for their comments and discussion. I am grateful to the National Center for Education Statistics and the North Carolina Education Research Data Center for providing invaluable confidential data, and Ryan Womack for helping me do all the data process.

Finally, I want to thank my parents and sisters. Their dedication makes me what I am now.

Dedication

To my mom and dad, sisters, and my grand ma

Contents

List of Tables

List of Tables Continued

List of Figures

Chapter 1

Introduction

To improve public school systems and increase student achievement, many states implemented their own statewide test-based school accountability during the 1990s. Each state had own their rules and framework to make schools accountable for student academic proficiency. To increase school performance, some states introduced reward systems while other states introduced sanctions. As concerns about the nationwide level of student academic proficiency were growing, the Federal government implemented a nationwide accountability policy, the No Child Left Behind Act of 2001 (NCLB), to hold all public schools accountable for their students' test scores.

Under the NCLB framework, each state set their own proficiency thresholds for reading and math exams and other targets. The shares of students achieving the proficiency goal in schools was the policy measure. Starting from their baseline proficiency levels, each state was required to increase the percent of proficient students gradually with the goal of reaching 100% by 2013-2014. Then, public schools were forced to meet the targets in each year, making Adequate Yearly Progress (AYP). If the shares of proficient students fell below the target and schools could not make AYP, the federal government designated schools as "failing schools." Then, the government forced the schools to increase student achievement by placing them under a NCLB sanction regime where sanctions became increasingly harsh, based upon the number of consecutive failures.

The policy also has subgroup provisions to achieve proficiency for subgroups of disadvantaged students and narrow student test score disparities. In the subgroup framework, schools were also held accountable for student performance by subgroups such as racial and ethnic groups, groups of economically disadvantaged students, limited English proficient students, and disabled students. Although the NCLB act was replaced by the Every Student Succeeds Act (ESSA) in 2015, the federal government continues putting efforts into raising student academic achievement through accountability policy reform.

The impact of school accountability policies seems mixed. Numerous studies show that school accountability raised academic achievement of students who attend underperforming schools and improved school environments (Ahn and Vigdor, 2014; Chiang, 2009; Deming et al, 2015; Reback et al, 2014; Rouse et al, 2013; and Springer, 2008). However, other studies indicate that the policies result in unwanted consequences such as teaching to the test, academic gains of low achievers at the expense of high achieving students, and removal of low performing students from a testing pool (Chakrabarti, 2013; Figlio, 2006; Hanushek and Raymond, 2005; Jacob, 2005; Krieg, 2011; Neal and Schanzenbach, 2010; Reback, 2008).

The purpose of this dissertation is to address two important questions and contribute to broadening our understanding of the impact of school accountability. First, I examine whether schools improve black student test scores at the expense of white students' achievement when the black subgroups failed to meet annual proficiency goals, but the white groups met them. Second, I examine whether school accountability decreases the racial diversity of a student population in schools and causes more racial segregation between schools. My dissertation makes several contributions. I use low-stakes exam scores that are non-incentivized outcomes of the policy to see how the NCLB policy affects genuine knowledge of students by race. Additionally, I am the first to study the racial and ethnic subgroup-specific incentive of schools using nationwide data sets. Krieg (2011) finds evidence that the test scores of students in passing subgroups decline relative to students in failing groups when their schools failed using statewide data from Washington state. By separating school failures into racial subgroup-level failures, I extend this research to study how the subgroup failures affect racial test score disparities in a nationally representative data set. Lastly, I study the impact of accountability on an outcome other than student test scores, the racial composition of students' peer groups, expanding a literature which tends to limit the scope of analysis to test scores.

I use nationally representative data from the Early Childhood Longitudinal Study to analyze student test scores under the NCLB subgroup provision. I do not find that white students are academically adversely affected by black group failure in schools where whites passed. Also, I do not find that the black group failure statistically significantly increases reading and math skills of black students in a preferred specification, relative to those in the control schools. My results differ from Krieg (2011), which finds that schools are incentivized by accountability to pay less attention to passing subgroups which results in the adverse effect on student test scores. My results can be reconciled to the literature by viewing the ECLS tests as measuring true academic achievement and statewide exams as reflecting an important degree of the teaching-to-the test. Together, I conclude that the NCLB subgroup rules did not increase genuine academic knowledge of disadvantaged students, nor narrow the existing racial achievement disparities in black failing schools, at least at the early stage of NCLB implementation.

To analyze the impact of pre-NCLB state school accountability policies on racial compositions of schools, I use administrative data on grades 3 through 5 from North Carolina. In low-performing schools during the post-policy period, I find that the black student enrollment share falls relative to the control schools, but the white student share rises statistically significantly relative to the control schools. Parents of black students might recognize a relatively good quality school at the cheap cost under the policy and move to another school. Also, schools might invest in school quality to attract additional white students after the policy began. My findings imply that school accountability reduces racial segregation since the majority of a student population in low-performing schools is black.

Chapter 2

The Role of Sanctions Threat under the No Child Left Behind on Racial Test Score Disparity

2.1 Introduction

The No Child Left Behind Act of 2001 (NCLB) was introduced to raise the nationwide level of student academic proficiency. Each state set annual proficiency rate thresholds in statewide math and reading exams. If a school's share of students achieving proficiency fell below the threshold, the school was designated as a failing school to make Adequate Yearly Progress (AYP). Consecutive school failures placed the school under a NCLB sanction regime where sanctions became increasingly harsh.

In addition to this, the federal government set subgroup provisions to achieve proficiency for subgroups of disadvantaged students and narrow student test score disparities. In the framework, students were categorized into racial and ethnic groups such as Whites, Blacks and Hispanics and into groups of economically disadvantaged students, limited English proficient students, and disabled students. If the size of groups was larger than a state-chosen minimum, schools were required to have each subgroup pass the annual threshold to make AYP.

Numerous studies document that accountability induces schools to increase student achievement through pedagogical improvement and increased school expenditures (e.g. Figlio and Rouse, 2006; Reback et al; 2014; Rouse et al, 2013; Chiang, 2009). However, there is concern that the NCLB policy causes unintended consequences. Evidence shows that schools target students whose test scores are just below the policy threshold at the expense of low performing or high achieving students' achievement (Chakrabarti, 2014; Jacob, 2005; Krieg, 2008; Neal and Schanzenbach, 2010; Reback, 2008). Due to the subgroup rules, one would expect schools to have devoted resources to certain racial subgroups who missed the AYP proficiency target in order to raise their scores in statewide exams. It is plausible that the pressure to raise scores of particular racial groups could incentivize schools and teachers to pay less attention to other racial groups. If so, the failure of the racial subgroups could adversely affect achievement of others. This could happen if it is hard for schools to raise achievement of all racial groups given limited school resources and time.

Schools could assign experienced teachers to classes with a high proportion of black students, and teachers could increase classroom interactions - asking questions and helping do homework or solve problem sets together - with black students relative to white students. By shifting resources from the passing group toward the failing group, schools could increase the black student test scores but possibly decrease white student test scores. The shifting of resources could entail an increase in general academic knowledge or merely an increase in test-specific knowledge on the part of black students. Alternatively, schools could identify inefficiencies and eliminate them to improve teaching practice for all students.

In this paper, I examine whether proficiency failure by the black student subgroup brings an increase in the black student test scores and whether this comes at the cost of white students who passed the proficiency threshold. My paper makes several contributions. First, I use low-stakes exam scores that are non-incentivized outcomes of the policy to see how the policy affects genuine knowledge of students by race. When studying school performance under the NCLB, many researchers examine statewide exam scores. However, the test scores are directly incentivized by the policy, and it is hard to understand the policy impact on genuine knowledge of students. Second, this paper is a first nationwide analysis studying the racial and ethnic subgroup-specific incentive of schools. There are earlier nationwide studies examining achievement gaps under school accountability (e.g. Hanushek and Raymond, 2005; Reback et al, 2014). Hanushek and Raymond (2005) conduct a nationwide analysis using the National Assessment of Educational Progress (NAEP), another non-incentivized measure, and find that school accountability does not help narrow the score gap between blacks and whites. In the meanwhile, Krieg (2011) examines whether passing subgroups are adversely affected by failures of other subgroups. Using statewide data from Washington state, he finds evidence that students in passing subgroups fall their test scores relative to students in failing groups when their schools failed. However, these papers do not disaggregate school failures and failing subgroup performance by race and ethnicity.[1]

For the nationwide analysis, I employ two nationwide data files of the individual-level Early Childhood Longitudinal Study (ECLS) and the school-level Barnard/Columbia No Child Left Behind Database. The ECLS from the National Center for Education Statistics (NCES) provides low-stakes assessment scores of math and reading that are comparable across states. These test scores are non-incentivized measures, which were not counted for the purpose of the NCLB proficiency and reflected a low degree of the teaching-to-the test compared to statewide assessments used by many earlier studies. The Barnard/Columbia data provides first two NCLB records for 2002-2003 and 2003-2004, including AYP status and proficiency rates. I link the NCLB of 2002-2003 file to the ECLS data. This covers students who were in third grade in Spring 2002 and those in fifth grade in Spring 2004. Using the combined data, I estimate the policy effect at an early stage of NCLB implementation using a difference-in-differences framework; I compare student test scores before and after the NCLB implementation in schools in which blacks failed only to those in schools in which both blacks and whites passed, separately by two racial groups, blacks and whites.

I find no strong evidence that black students have academic gains in their math and reading exam scores when their racial group failed to meet the AYP proficiency goals: in schools where blacks failed to pass the math cutoff but whites passed the target, black student math scores decrease by 0.115 standard deviation significantly at the 10% level, but the students increase their scores 0.027 standard deviation, but it is not statistically significant, once I control for the school district-level differences within states. The size of the coefficient is modest in the preferred specification

[1]In California, the policy worsens school performance by burdening public schools serving diverse student populations because such schools have more subgroups than other schools (Sims, 2013)

with district dummies, when compared to results of prior studies showing that the NCLB policy and state accountability policies increase about 0.1 standard deviation in math scores (e.g. Chiang, 2009; Krieg, 2011; Reback et al, 2014). Also, the black proficiency failure did not have statistically significantly adverse effect on white math scores within schools that whites passed the target.

In addition, there is an insignificant negative effect on the black reading scores when only blacks missed the reading proficiency target while there is no statistically significantly adverse impact on white reading scores in the treatment schools. In terms of the black-white test score gap, the gap in the treatment schools remains stable, relative to the existing gap in the control schools. Third, I find that there is a large improvement in black student math scores and white student reading scores when both blacks and whites failed. However, I find that the big improvement comes from highly proficient students, who are not regarded as a typical treatment group under NCLB.

There are several plausible explanations. It is possible that schools find it difficult to target resources at particular racial groups, and they may be willing to increase school resources for all students if both racial groups failed. Another possibility is that schools responded to incentivized measures of NCLB by teaching to the test, resulting in no effect on genuine math and reading ability as measured by the ECLS assessments.

Together, my results imply that the NCLB subgroup rules did not increase academic knowledge of disadvantaged students, nor narrow racial achievement disparities, at least at the early stage of NCLB implementation.

2.2 Overview of the No Child Left Behind Act

As a part of Title 1 of the Elementary and Secondary Education Act of 1965, federal funds are granted to high poverty public schools to help all students meet state academic standards. The NCLB policy affected about 47,000 Title 1 schools serving 14.9 million students - about 30% of K-12 public school enrollment until it has recently been reformed (Department of Education).[2]

The goal of NCLB was to guarantee that 100% of students at Title 1 schools achieve a certain

[2]"https://www2.ed.gov/rschstat/eval/disadv/title1-factsheet.pdf," Visited on November 5, 2017

proficiency level in statewide exams by the 2013-2014 academic year.[3] Starting from their baseline proficiency levels, each state was required to increase the percent of proficient students gradually with the goal of reaching 100% by 2013-2014. For example, Florida public schools started with a goal of 31% proficient in reading in 2003, and this goal increased each year to have all students achieve proficiency by 2013-2014.[4]

The federal policy also required states to achieve proficiency for subgroups of disadvantaged students. Each state created its definition of the subgroups and categorized students into racial and ethnic groups such as Whites, Blacks and Hispanics and into a group of economically disadvantaged students, limited English proficient students, and disabled students. Each racial group and a group of economically disadvantaged students face the same annual proficiency threshold while the other subgroups have adjusted threshold. The states also determined the minimum size of the subgroups required to hold schools accountable for the subgroups. For instance, some states set the minimum size to be 40. This forced any subgroup of more than 40 students in the tested grades to meet the proficiency thresholds for both reading and math exams and other NCLB objectives.

Additionally, the NCLB policy mandated states to establish several objectives other than proficiency targets. Schools must meet targets for test participation rates, attendance rates, and graduation rates as well as proficiency rates in order to make AYP each year. For instance, schools in states had to show at least 95% of test participation rate in state assessments overall and for each subgroup. If schools met all NCLB targets other than a goal for test participation rate, they were regarded as schools that did not make AYP in that year.

Sanctions were applied to schools that failed to make AYP in consecutive years. A profile of potential sanctions was possible such as implementation of a school choice option to harsher actions like replacement of school staff. For instance, schools that missed AYP goals once did not receive any sanctions. However, if they did not make AYP in either math or reading subject in two consecutive years, they were identified as the schools in need of improvement. This placed

[3]At the very beginning stage, states were allowed to choose testing grades. By the 2006-2007 school year, all 3 - 8 grades were required to test.

[4]There are two adjustments allowed to the NCLB thresholds that can save a school from failing AYP even if they fail to make the appropriate threshold: a confidence interval adjustment and a safe harbor. The confidence interval adjustment lowered the effective proficiency target for some groups - it was meant to provide leniency due to small numbers of students tested in some subgroups. If a subgroup still does not meet the proficiency threshold, but there is at least a 10% reduction in the percentage of students who were not proficient in that group compared to last year, the school would still be considered as meeting AYP through the “safe harbor” rule.

them on a first stage of sanctions, in which school districts must offer students a transfer choice to another public school within the same districts. Then, sanctions became increasingly harsher. If schools missed AYP in six consecutive years, reconstruction plans were applied, which could include states' takeover to turn them into charter schools, replacement of school administrators, or closure of the public schools.[5]

Due to consecutive failures, many schools were placed at risk of closure. For example, there were around 3,600 Title 1 schools under reconstruction plans in 2007-2008 (The Center on Education Policy, 2008). Before academic year 2013-2014 when all public schools were expected have every student meet each state's proficiency goal, the necessity of granting NCLB waivers to states appeared in order to save the schools from the closure. In 2011, the federal government introduced a formal application process for the waivers, and most states were granted the waivers for two years. States could develop a renewal plan by taking flexibility in certain NCLB provisions into account. Once they submitted it to the U.S. Department of Education and were approved for their proposal, they could additionally extend the NCLB waivers. While granting states renewals, the federal government could take the time to reform the accountability policy. Finally, the NCLB act was replaced by the Every Student Succeeds Act (ESSA) in 2015.[6]

2.3 Data

For a nationwide analysis, this study uses the confidential-use version of the ECLS and the Barnard/Columbia NCLB Database.[7] The ECLS collected information on a nationally representative cohort of children who attended kindergarten in 1998-1999. This survey followed the same cohort until they were in their middle schools in 2006-2007. The survey was administered in seven rounds, providing data in Fall 1998, in Fall and Spring 1999, in Spring 2000 when most children were first graders, and in Spring 2002, 2004 and 2007 when most students were third, fifth and eighth graders respectively.

The NCLB data collected school report cards across states in 2002-2003 and 2003-2004. This

[5] Ahn and Vigor (2014) summarize sanctions profile in details.

[6] "https://obamawhitehouse.archives.gov/issues/education/k-12/reforming-no-child-left-behind" and U.S. Department of Education, Every Student Succeeds Act (ESSA) "https://www.ed.gov/ESSA" Visited on September 12, 2018

[7] Reback, Rockoff, Schwartz and Davidson compiled this database based upon school report cards from each state Department of Education and made it publicly available online at http://www.gsb.columbia.edu/nclb.

data covers about 80,000 public schools that reported their AYP status in 2003. It contains numbers of tested students, the test participation rates, and math and reading proficiency rates by subgroups as well as overall proficiency rates for each school. However, not all states required schools to report subgroup-level proficiency rates at the early stage of the NCLB policy. Thus, many of 80,000 schools in the data have missing values for subgroup proficiency rates. I take this data incompleteness into consideration in my analysis.

To construct a sample, I link the ECLS data to the NCLB data through the NCES school identifier. Each public school was uniquely assigned its identifier by the NCES. While the NCLB data includes only public schools, the ECLS also includes students in private schools and home learners. Thus, I drop the non-public schools from my sample. 9,250 out of 11,820 students from the ECLS data representing 1,960 public schools were successfully matched to the NCLB data.[8] From the linked data, I drop 20 children whose race variables are missing and then drop 2,620 children whose schools reported themselves as non-Title1 schools. I keep those in schools that did not report their status of Title 1 and those students in magnet public schools, leaving 6,610 students in 1,400 schools. For simplicity, I keep only whites and blacks, and this leaves 4,200 students. For analysis, 2,840 observations are employed due to missing control variables.

2.3.1 Dependent Variable

The ECLS assessed students' math and reading knowledge and skills in each survey round. They include Item Response Theory (IRT) scaled scores and t-scores that standardized IRT scores to a mean of 50 and variance of 10. I standardize the t-scores to z-scores with a mean of zero and variance of one. The main dependent variable is the difference in the student's z-score of Spring 2004 and Spring 2002 for math and reading assessment respectively.

2.3.2 Independent Variable

The main independent variable of interest is an indicator for whether a student is in a school where black students failed to pass proficiency cutoffs but whites passed. These will be compared to students in schools where both blacks and whites passed. To make this comparison, I use two

[8] By the confidential policy of the NCES, all observation numbers are rounded to nearest 10.

steps. First, I create two indicators, one for whether blacks passed proficiency targets or not and one for whether whites passed. Next, I separate the sample of schools into four cases using the two indicators.

Let subscript b and w represent blacks and whites respectively. Let subscript j and s represent a subgroup and a school respectively. Employing subgroup proficiency rates for each j and state proficiency cutoffs for each subject respectively, I create a subgroup failure indicator whether j met AYP proficiency targets for math and reading respectively in the following way. For each $j \in \{b, w\}$, I assign value one if j's proficiency rate is above the cutoff, and otherwise zero. However, j could pass the target through a confidence interval provision and a safe harbor provision even if the proficiency rate is below the cutoff.[9] So, I reassign one if j's proficiency rate is below the cutoff but s met AYP. This indirectly takes the provisions into consideration.

Then, denoting C_s equal to one if both w and b met, define treatment variables for each subject in the following way.

$$T_{1,s} = \begin{cases} 1 & \text{if} \quad w \text{ met and } b \text{ did not meet} \\ 0 & \quad\ \text{otherwise} \end{cases}$$

$$T_{2,s} = \begin{cases} 1 & \text{if} \quad \text{both } w \text{ and } b \text{ did not meet} \\ 0 & \quad\ \text{otherwise} \end{cases}$$

$$T_{3,s} = \begin{cases} 1 & \text{if} \quad w \text{ did not meet and } b \text{ met} \\ 0 & \quad\ \text{otherwise} \end{cases}$$

The main variable of interest is $T_{1,s}$ for math and reading scores respectively.

2.3.3 Control Variables

I use the following variables from the ECLS as control variables. First, the ECLS computed parental socioeconomic status (SES) scores using parental education, occupation and household

[9] Please see section 2 for the provisions.

income information collected in each survey round. The SES score is an average score of the normalized parental information components, which ranges between -2.48 and 2.54. I use the difference in the SES scores in Spring 2004 and in Spring 2002 to control for differences in family backgrounds across students over time. Second, I employ school-level variables from the ECLS to control for differences in school characteristics. The ECLS provides the percentage of students who are eligible for free lunch at schools. I calculate the difference in the variables in between Spring 2004 and Spring 2002. I also use the percentage of racial minority students at schools. The ECLS provides a continuous variable of the percentage of minority in Spring 2002 by calculating nonwhite students in schools while they offer a categorical variable of the percentage of minority in Spring 2004. Thus, I turn the continuous variable of spring 2002 into a categorical variable. I calculate the difference in the two categorical variables to use it as a control variable.

2.3.4 Descriptive Statistics

I present descriptive statistics by groups of schools. Table 2.1 describes individual level variables of students in each group. I normalize the scores to a mean of zero and variance of one over a full sample of the raw data. The raw data contains students attending private schools too. After keeping only students in Title 1 public schools, the means of the test scores over the final sample become lower than zero. In the schools where blacks fail but whites meet, the means of students' scores are low relative to those in the schools where both whites and blacks fail. In addition to this, parental SES scores are lower than those in the control group. In the final sample, there is the high proportion of black students in the treatment schools where blacks fail but whites pass in 2004, around 40%, compared to the proportion of the control group of schools. On average, the proportion of black students is around 18%.

Table 2.2 depicts descriptive statistics of school characteristics. The treatment schools are more likely to have nonwhite students, compared to the control group of schools. In 2004, about 40% of schools exhibit the high proportion of nonwhites, more than 50% in a student population. The percentage minority students in 2002 are around 42% on average while the percentage minority in 2002 is 66% in the group of interest, which is higher than the average. Title 1 public schools exhibit the high percentage of students who are eligible for free lunch. On average, it is around

44% in 2004, and even the treatment group of schools show around 56% of students eligible for free lunch. This implies that the treatment schools where only blacks fail are more likely to have students from relatively low-income households.

Table 2.3 describes the proportion of each category of school in the NCLB data. In the sample of schools for analysis, about 86% of them fall into the group of schools where both blacks and whites meet proficiency targets for both subjects. Among the sample, the schools where blacks fail but whites pass represent about 10%.

2.4 Empirical Strategy

My goal is to measure whether the proficiency failure of black student group improves black student scores and has a negative effect on white student scores. To address this question, I need to exploit disaggregated racial subgroup-level failures, rather than simply comparing failing schools to non-failing schools as most prior studies do. To distinguish the black group failure from school-level performance, I use the constructed treatment variables that I describe in a previous section. By comparing changes in student scores of schools where only blacks failed to those of schools where both blacks and whites passed the proficiency thresholds, I employ the difference-in-differences framework. I estimate Equation (1) separately for black students and white students.

$$\Delta Y_i = \alpha + \gamma T_{1,s} + \delta T_{2,s} + \theta T_{3,s} + \beta \Delta X_i + \psi \Delta Z_s + \mu_{state} + \Delta \varepsilon_i \tag{2.1}$$

The dependent variable is the score difference between Spring 2004 and Spring 2002, ΔX_i is the changes in parental SES scores between 2002 and 2004, ΔZ_s is the changes in school-level co-variates between 2002 and 2004, and μ_{state} is a state-fixed effects to control for a time trend in each state. Differencing removes unobserved fixed individual characteristics such as ability, and removes unobserved fixed school characteristics like school quality or neighborhood quality of surrounding areas for students staying at the same school. To control for time-varying effects across school districts, I additionally include district dummies to the baseline model in Equation (1).[10] With the school district-fixed effects, I expect to control for changes in schooling policies other

[10]For instance, each school district could introduce its own school policies that can also affect student test scores. Or, school district finance could vary over time.

than NCLB, schooling cultures, or investment in public school performance at district-level.

The main coefficient of interest is γ, which is the estimate of the effect of the subgroup proficiency failure, conditioned on individual-level and school-level characteristics and the time trend in states and districts. If γ is statistically significantly positive for the black student sample, I interpret that schools focus on increasing the test scores of black students when blacks failed, compared to the scores of students in schools where both blacks and whites passed. If γ is statistically significantly negative for the white, non-Hispanic sample, I interpret that there is the adverse effect of the black group failure on the non-targeted group within schools.

There are possible concerns about this estimation. First, one may concern that the main treatment variable of interest could not be exogenous if schools respond to school failures prior to NCLB. Analyzing the first year of NCLB implementation is one way to address the concern. Each state set the annual proficiency cutoff that was unknown to school administrators and teachers before NCLB implementation. They also did not expect whether they could barely pass the threshold or easily pass. In addition to this, there is transitory shock in student populations as the literature indicates (e.g. Kane and Staiger, 2002). Thus, I argue that the status of treatment is unexpected at the beginning of the implementation.

Second, it is also plausible that schools could target a particular group of students within racial groups. Rather than targeting all minority students, schools could focus on those whose test scores are close to the proficiency cutoff, which could help them increase the proportion of students who pass the target in the very short term. Also, it may not be plausible to conjecture that white students who are close to the cutoff are adversely affected. To avoid a sanction in the following year, the treatment schools may be likely to raise the test scores of whites who at least close to the cutoff even if whites passed. On the other hand, high performing whites could possibly receive less school resources and teachers' attention than before, given the limited time and school resources. This idea aligns with many earlier studies (e.g. Neal and Schanzenbach, 2010; Reback, 2008).

To address this, I must distinguish whose test scores are close to the proficiency threshold using individual-level statewide test scores, which I do not have. Alternatively, I use the ECLS test scores of Spring 2002 and calculate tertiles of the score distribution.[11] Then, I create two

[11] I calculate the tertiles on a Title 1 sample after dropping students in private schools. The Title 1 sample includes all races.

dummy variables L_i and M_i indicating for whether a student is in a low achieving or in a middle scoring group. Since high performing students would easily pass the proficiency threshold and low performing students would find it hard to meet the cutoff, it is highly likely that the middle level group includes those who relatively close to the cutoff. Using the proficiency dummies, I extend the baseline model to a triple-differences model in the following way.

$$\begin{aligned}\Delta Y_i = \alpha + \gamma_1 T_{1,s} + \gamma_2 T_{1,s}\mathrm{M}_i + \gamma_3 T_{1,s}\mathrm{L}_i + \delta_1 T_{2,s} + \delta_2 T_{2,s}\mathrm{M}_i + \delta_3 T_{2,s}\mathrm{L}_i \\ + \theta_1 T_{3,s} + \theta_2 T_{3,s}\mathrm{M}_i + \theta_3 T_{3,s}\mathrm{L}_i + \eta_1 \mathrm{M}_i + \eta_2 \mathrm{L}_i \\ + \beta \Delta X_i + \psi \Delta Z_s + \mu_{state} + \Delta \varepsilon_i\end{aligned} \tag{2.2}$$

In this specification, I test if γ_2 is statistically significantly positive for the treatment group of black students. For high achieving white students, if γ_1 is statistically significantly negative, I interpret this as evidence of reallocation of school resources, which driving the adverse effect on the non-targeted group.

Third, one may concern that proficient students in treatment schools could move to passing schools after the release of NCLB report cards in 2003. To further study whether results could be driven by student sorting behaviors, some studies use the percentage of proficient students as a dependent variable (e.g. Chakrabarti, 2013, 2014). If I can identify the time when a student moved to another school, I could replicate the approach. Unfortunately, in the ECLS data, students moved between two survey rounds of 2001-2002 and of 2003-2004. They moved to another school either after the NCLB report card released in 2003 or before the academic year of 2002-2003 started, and I cannot identify when they moved. Furthermore, a subset of movers was followed while all stayers were tracked. Thus, to study whether student sorting drives the findings of this study, I estimate the baseline model by dropping movers. I report findings of the stayers in section 2.5.1.

I cluster all standard errors at the school-level. All reported regressions are unweighted. I report findings of both math and reading exam scores in the following section.

2.5 Results

Table 2.4 describes the baseline results of the math scores for black students. In column (1), I only include the treatment variable. In schools where only black students failed, black student math scores rise by 0.008 standard deviation, but not statistically significantly. I add the changes in parental SES scores to column (1) and present the findings in column (2). The estimate of black math failure in treatment schools does not change. When I add school characteristics in column (3), the black student math scores fall by 0.0004 standard deviation in treatment schools, but not statistically significantly, and the size of the estimate is close to zero. Column (4) shows the results with the state-fixed effects. In schools where the black groups failed but the white groups met the AYP math proficiency target, the black student scores fall by 0.115 standard deviation, significantly at the significance level of 10%. There may be unobserved or observed characteristics such as neighborhood quality or school budget varying across school districts within a state. To control for the district-level differences, I add the district-fixed effects. Then, the estimate becomes 0.028, but is statistically insignificant. Thus, there is no strong evidence that black students academically benefit from their racial group‘s failure.

Interestingly, the black student scores increase significantly when white students also failed. Looking at estimates in a second row, relative to black students in the control schools, black students statistically significantly increase their scores by 0.279 standard deviation with the state-fixed effects. Even controlling for the school district-level differences, their scores rise 0.352 standard deviation at the significance level of 1%. The size of the coefficients is large when compared to results of prior studies showing that the NCLB policy and state accountability policies increase about 0.1 standard deviation in math scores (e.g. Chiang, 2009; Krieg, 2011; Reback et al, 2014). In a third row, in schools that only whites failed, the black student math scores significantly fall by 0.471 standard deviation in the preferred specification with the district-fixed effects, compared to the control schools. It may be plausible to conjecture that black students are adversely affected by the white group failure, relative to those in the control schools. However, in this case, the insufficient number of schools in this group requires very careful interpretation.

Table 2.5 presents the estimates of the black group’s reading failure on the changes in black reading scores. With the state-fixed effects, black students decrease their score when blacks failed but whites met. Relative to those in the control group, their scores fall by 0.102 standard deviation, significant at the 10% level. Again, even within a same state, each school district could have different neighborhoods, schooling practice, or academic environment. Taking this into account, I add the district-fixed effects, and find that the estimate becomes more negative though it is not statistically significant. In schools where both groups failed, the black student reading scores fall, but it is not statistically significantly. In schools where only whites failed, the black student reading scores significantly fall in any specification.

Table 2.6 presents the effect of the black group's math failure on changes in white math test scores. In columns (1), the results show that white student math scores in the treatment schools statistically significantly fall by about 0.09 standard deviation, relative to whites in schools where both blacks and whites passed. However, once I add covariates with the school-fixed effects, the scores statistically insignificantly fall by 0.067 standard deviation in the treatment schools, relative to those in the control schools. A lower bound of the confidence interval is -0.159, and I could not reject that white students are not adversely affected by the black group failure. Even I control for the district-level differences, the size of the coefficient becomes small. The white student math scores fall by 0.002 standard deviation, but the lower bound of the coefficient that I can reject is -0.224, implying that there is no evidence of an adverse effect. In schools where both groups failed, none of the estimate is statistically significant in any specification. In schools where only whites failed, the white students decrease their math scores by 0.059, but it is not statistically significantly with the state-fixed effects. The estimate becomes statistically significantly negative and large, -0.379, once I add the district-fixed effects.

Table 2.7 depicts the impact of the black group's reading failure on changes in white reading scores. With the state-fixed effects, the coefficient of interest is -0.012 in column (4), which means that the white reading scores fall by 0.012 standard deviation relative to those in schools where both whites and blacks met the reading proficiency threshold. However, the estimate is not statistically significant. With the district-fixed effects, the coefficient of interest is 0.005 that is very close to zero and not statistically significant. This suggests that there is no strong negative effect on the white student achievement. Interestingly, white students in schools where both groups failed show improvement. In column (5), their scores significantly rise by 0.232 standard deviation at the significance level of 1%, compared to whites in schools where both groups passed. In addition to this, whites in schools where only whites failed significantly increase the reading score by 0.117 standard deviation at the significance level of 1%. For this group of schools, the coefficient is statistically significant in any specification and the size is large.

To summarize the baseline results, in the treatment schools, black student math scores significantly fall with the state-fixed effects while the scores insignificantly rise in the preferred specification with the school district-fixed effects. Relative to the control group, black reading scores in the treatment schools statistically insignificantly fall. In the preferred specification with the school district-fixed effects, the reading scores rise, but the magnitude of the coefficient of interest is small. For the white students, their math scores fall in both the state-fixed and the district-fixed effects, and the magnitude of it is small in the preferred specification. In the meanwhile, white reading scores statistically significantly fall with the state-fixed effects, and the scores rises with the district-fixed effects. The rise in the preferred specification is small and statistically insignificant.

Next, I examine whether students whose test scores are close to the proficiency cutoff raise their score while high achieving students do not. To avoid sanction in the following year, the treatment schools may be likely to raise the student test scores just below and above the cutoff by shifting school resources from high performing students, given the limited resources. Within a same treatment school, student scores may differently change based upon students' proficiency level. Also, white students who are at the proficiency threshold could increase their scores even if their racial group passed. To take this idea into account, I estimate the triple-differences model in Equation (2) by a group of blacks and whites respectively and present the results in Table 2.8 and Table 2.9.

In Table 2.8, black students in the treatment group do not show positive effect on their math scores when only blacks failed. In schools where both groups failed, black students at the middle level do not show improvement relative to high performing students. Relative to the scores of high performing students, low proficient black students' test scores statistically significantly fall by 0.590 standard deviation with the district-fixed effects. The improvement presented in Table 2.4 comes from the high performing black students as they show a significant increase in the scores of

0.793 standard deviation with the district-fixed effects. I also test whether the overall effect is positive on the middle and low proficient student test scores when both blacks and whites failed.[12] The black students at the middle level statistically significantly statistically insignificantly increase their math scores by 0.18 standard deviation with the district-fixed effects. The scores of low performing students increase by 0.299 standard deviation with the district-fixed effects at the significance level of 1%.

In the meanwhile, there is no statistically significant negative effect on the score of high proficient white students when blacks failed and whites passed. In addition to this, white students at either the middle or the high proficient level do not show significant effect relative to the high performing students in the schools. For whites in schools where both failed, I cannot find evidence that high proficient students are adversely affected by school resource allocation and that the middle proficient students academically improve. The overall effect on students at the middle level is statistically significantly positive, 0.136 standard deviation in the preferred specification with the school-district fixed effects. The scores of students at the low proficient level statistically significantly rise by 0.188 standard deviation at the 1% level in the preferred specification. When only whites failed, high proficient white students show the significant negative effect. The size of the coefficient is large. The white student math scores fall by 0.692 in the preferred specification when they are at the high proficient level. Interestingly, students in either the middle level or the low proficient level significantly increase their math scores relative to the high performing students when only whites failed.

Table 2.9 shows that blacks in the middle level of proficient in the treatment group reduce the reading scores by 0.102 standard deviation with the state-fixed effects and by 0.041 standard deviation relative to high performing black students, but not statistically significantly. The low proficient black students in the treatment schools statistically insignificantly fall their reading scores. In the meanwhile, middle proficient black students in schools where both blacks and whites failed show no positive effect on the relative scores. With the state fixed effects, the black student reading scores statistically significantly fall by 0.418 standard deviation at the 5% significance level. Low performing black students statistically significantly decrease their reading scores by about

[12]To calculate the coefficient and its standard error, I use a stata command “lincom.”

0.77 standard deviation in both specifications, relative to high performing students. However, high proficient black students in the schools show large improvement of 0.447 standard deviation at the significance level of 1%. There is the statistically insignificant negative overall effect on the black students at the middle proficient level while there is the statistically significant positive overall effect on the low proficient black student reading scores. They raise their reading scores by 0.27 standard deviation in the preferred specification at the significance level of 5%.

For white students in schools where only blacks failed, there is no statistically significant positive effect on students at the middle proficient level and low performing students relative to high performing students. The high proficient whites in the schools decrease their scores with the state-fixed effects and increase the reading scores in the preferred specification, but none of the effects are statistically significant. In the schools where both failed, the scores of whites at the middle proficient level statistically insignificantly fall by 0.383 in the preferred specification, and low proficient whites significantly fall their reading scores by 0.594 standard deviation relative to high proficient white students at the significance level of 10% in the preferred specification. The size of the coefficient is large. The high proficient white students in the schools increase their reading scores by 0.620 standard deviation with the district-fixed effects at the significance level of 5%. The overall effect on whites at both middle and low proficient levels is statistically significantly positive. Relative to high proficient white students, the middle proficient and low proficient whites raise their scores by 0.142 standard deviation and by 0.352 standard deviation respectively in both specifications.

To summarize the results of the extended model, in the treatment schools, neither black students nor white students show statistical changes in either math or reading scores when they were at the middle level of proficient level. Also, low performing students of both racial groups also do not show statistically significant score changes, relative to high performing students. Together, I do not find that students in the treatment schools differently exhibit their score changes based upon their proficiency level.

2.5.1 Sensitivity Checks

In this section, I further test whether the results were driven by the following possibility. First, I test if trends in initial characteristics play a role in estimating the coefficient of interest. It is possible that students have initial characteristics starting at a different level, but my baseline model exploits only differences in variables. So, I add level variables of 2002 to the baseline equation to control for the trends. Results are presented in Table 2.10 and Table 2.11. I find this makes no difference.

Next, I estimate the baseline model with a sample of stayers as I discuss in the end of section 2.4. This yields similar results to the baseline findings. In Table 2.12 and Table 2.13, I find no strong evidence that the black student test scores increase while the white scores fall when only blacks failed. Although the black math scores decrease by 0.111 standard deviation with the state-fixed effects and 0.027 standard deviation with the district-fixed effects, none of them are statistically significant. White students also show a decrease in their math scores, but it is not significant in the treatment schools.

As similar to Table 2.5, in Table 2.13, the black reading scores fall by 0.056 standard deviation in the preferred specification with the district-fixed effects, which is not significantly when only blacks failed. A difference is that the reading scores decrease for the black stayers when both blacks and whites failed, but the decrease is not statistically significant. The coefficients of interest for white students are positive but are not statistically significant. Compared to the estimate presented in Table 2.7, the size of the coefficient for the white reading scores becomes large, 0.030 with the state-fixed effects and 0.043 with the district-fixed effects. In schools where both blacks and whites failed, the white student math scores significantly increase by 0217 standard deviation with the district-fixed effects only. The size of this coefficient is also large and similar to the coefficient in Table 2.7.

Overall, the results suggest that there is no strong evidence that the black students have academic gains when their racial group failed to meet the AYP proficiency goals. Although black students in schools where both groups failed show a large improvement in their math scores relative to those in the control schools, the score gains come from the highly proficient black students. While there is no effect on the black reading scores when both blacks and whites failed, the black students at the high proficient level improve their scores while the scores of the low performing students fall once I include the proficiency dummies. In addition to this, white students are not academically adversely affected by their schools when whites met but blacks missed the goals. When whites also failed to meet the reading threshold, there is a huge improvement of the white reading scores, but this is due to highly proficient white students.

2.6 Discussion and Conclusion

As researchers indicate that accountability policies increase student academic proficiency, it seems that the NCLB policy achieves its primary purpose. The policy framework includes subgroup provisions to narrow the racial achievement gaps, and it is important to study how the NCLB subgroup policy alters the racial achievement disparities in order to evaluate the policy, but few studies examine this. This paper studies whether the pressure of schools on the black group failure improve the black student achievement but adversely affect white student test scores.

To examine the impact of the black subgroup failure on their group performance and white group performance, I separate school failures into racial subgroup-level failures. In the treatment schools that black students failed to pass the proficiency target, but the white group passed, white students are not academically adversely affected by the black group failure, compared to those in schools that both racial groups passed. Also, I do not find that there is an increase in the reading and math skills of black students, relative to those in the control schools.

My results can be reconciled to the literature by viewing the ECLS tests as measuring true academic achievement and statewide exams as reflecting an important degree of the teaching-to-the test. Schools incentivized by the NCLB sanction pressure could teach students exam skills and topics only relevant to the statewide exams. This may be one explanation why changes in student test scores appearing in statewide exams do not appear in the ECLS tests, the low-stakes exams. This is consistent to some prior studies as they show that there is no positive policy effect on other low-stakes exam test scores. This would be one reason why I do not find evidence presented by Krieg (2011) indicating finds that test scores of students in passing subgroups in failing schools fall relative to those in non-failing schools.

To delve in the impact of the NCLB subgroup policy, further analysis could be carried out. So far, I do not consider other racial groups of students' NCLB performance. If I consider the NCLB performance of Hispanics and Asians and their effects on black and white student test scores, it would be interesting. It would be also interesting to see the effects of reading scores on the math scores. From this, we can learn how school failures in a certain subject affect the other subject test scores. I leave the ideas for future research.

From my results, I conclude that the NCLB subgroup rules did not contribute to inducing the treatment schools to increase genuine knowledge of disadvantaged students whose subgroups failed to meet the NCLB goal at least at the early stage of implementation.

Tables

Table 2.1: **Summary Statistics of Student Characteristics**

	Pooled Sample	Blacks and Whites passed	Blacks failed, Whites passed	Both failed	Blacks passed, Whites failed
Math	-0.06	-0.04	-0.31	-0.02	-0.65
	[2840]	[2560]	[200]	[70]	[20]
Reading	-0.04	-0.03	-0.15	-0.06	-0.46
	[2840]	[2550]	[200]	[70]	[20]
Math in 2002	-0.03	-0.01	-0.22	-0.10	-0.74
	[2840]	[2560]	[200]	[70]	[20]
Reading in 2002	-0.03	-0.01	-0.17	-0.11	-0.44
	[2820]	[2540]	[200]	[70]	[20]
SES in 2004	-0.14	-0.13	-0.20	-0.24	-0.62
	[2840]	[2560]	[200]	[70]	[20]
SES in 2002	-0.14	-0.13	-0.21	-0.20	-0.56
	[2840]	[2560]	[200]	[70]	[20]
Race (%)					
White	81.53	83.43	58	75.76	94.44
Black	18.47	16.57	42	24.24	5.56
Stayer	0.83	0.84	0.73	0.54	0.94

Note: 1. The number of observations are rounded. 2. Numbers in brackets are numbers of observations. 3. A stayer variable is a dummy variable, and statistics are means.

Table 2.2: **Summary Statistics of School Characteristics**

	Pooled Sample	Blacks and Whites passed	Blacks failed, Whites passed	Both failed	Blacks passed, Whites failed
% of minority in 2004					
less than 10%	25.76	27.97	6.90	22.22	50.00
10% to less than 25%	13.06	15.45	3.45	-	-
25% to less than 50%	19.68	20.67	15.52	11.11	-
50% to less than 75%	13.77	12.53	20.69	22.22	25.00
75% or more	27.19	23.38	53.45	44.44	25.00
	[560]	[480]	[60]	[20]	[10]
% of minority in 2002	42.20	38.92	65.82	55.96	30.28
	[560]	[480]	[60]	[20]	[10]
% of students eligible for	44.44	42.52	55.90	59.07	42.04
free lunch in 2004	[560]	[480]	[60]	[20]	[10]
% of students eligible for	41.65	39.27	58.85	48.60	46.21
free lunch in 2002	[560]	[480]	[60]	[20]	[10]

Note: 1. The number of observations are rounded. 2. Numbers in brackets are numbers of observations.

Table 2.3: **The Numbers of Schools in Each Group**

Groups of Schools by Math and Reading Subjects	Mean	N
Math Proficiency Failure		
Schools that both blacks and whites met math proficiency targets	0.857	480
Schools that Blacks missed math proficiency targets but whites met	0.104	60
Schools that both blacks and whites missed math proficiency targets	0.032	20
Schools that Blacks met math proficiency targets but whites missed	0.007	10
Reading Proficiency Failure		
Schools that both blacks and whites met reading proficiency targets	0.856	480
Schools that Blacks missed reading proficiency targets but whites met	0.108	60
Schools that both blacks and whites missed reading proficiency targets	0.025	20
Schools that Blacks met reading proficiency targets but whites missed	0.011	10

Note: 1. Variables are school-level. 2. The number of observations are rounded.

Table 2.4: **The Effect of the Subgroup Policy on Changes in Black Math Scores**

Δ math	(1)	(2)	(3)	(4)	(5)
Blacks failed, Whites passed	0.008	0.008	-0.0004	-0.115^{*}	0.028
	(0.057)	(0.057)	(0.057)	(0.060)	(0.086)
Blacks and Whites failed	0.550^{***}	0.553^{***}	0.553^{***}	0.279^{***}	0.352^{***}
	(0.082)	(0.083)	(0.082)	(0.094)	(0.128)
Blacks passed, Whites failed	0.052^{**}	0.053^{**}	0.039	-0.255	-0.471^{***}
	(0.021)	(0.021)	(0.041)	(0.171)	(0.065)
Δ SES		-0.119^{**}	-0.119^{**}	-0.157^{***}	-0.137^{***}
		(0.046)	(0.047)	(0.047)	(0.053)
Δ% of free lunch			-0.001^{*}	-0.002^{**}	-0.001
			(0.001)	(0.001)	(0.001)
Δ % of minority			0.004	0.013	0.003
			(0.034)	(0.033)	(0.065)
State-fixed effects				Y	Y
District -fixed effects					Y
N	530	530	530	530	530

Note: 1. Δ% of minority is the difference in categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses. 3. The regressions are unweighted.

Table 2.5: **The Effect of the Subgroup Policy on Changes in Black Reading Scores**

Δ reading	(1)	(2)	(3)	(4)	(5)
Blacks failed, Whites passed	0.011	0.008	-0.003	-0.102^{*}	-0.114
	(0.056)	(0.055)	(0.056)	(0.062)	(0.097)
Blacks and Whites failed	0.077	0.085	0.080	-0.049	-0.020
	(0.088)	(0.087)	(0.088)	(0.111)	(0.145)
Blacks passed, Whites failed	-0.666^{***}	-0.681^{***}	-0.715^{***}	-0.851^{***}	-1.263^{***}
	(0.125)	(0.136)	(0.138)	(0.188)	(0.065)
Δ SES		-0.092^{*}	-0.088	-0.093^{*}	-0.069
		(0.053)	(0.053)	(0.054)	(0.065)
Δ% of free lunch			-0.002	-0.002^{*}	-0.002
			(0.001)	(0.001)	(0.001)
Δ % of minority			0.046	0.019	0.082
			(0.048)	(0.046)	(0.067)
State-fixed effects				Y	Y
District -fixed effects					Y
N	520	520	520	520	520

Note: 1. Δ% of minority is the difference in categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses. 3. The regressions are unweighted.

Table 2.6: **The Effect of the Subgroup Policy on Changes in White Math Scores**

Δ math	(1)	(2)	(3)	(4)	(5)
Blacks failed, Whites passed	-0.091^{*}	-0.092^{*}	-0.088^{*}	-0.067	-0.002
	(0.051)	(0.051)	(0.052)	(0.047)	(0.113)
Blacks and Whites failed	-0.033	-0.036	-0.044	-0.050	0.006
	(0.156)	(0.157)	(0.158)	(0.145)	(0.111)
Blacks passed, Whites failed	0.034	0.032	0.029	-0.059	-0.379^{***}
	(0.023)	(0.023)	(0.023)	(0.066)	(0.098)
Δ SES		-0.045	-0.045	-0.058	-0.089^{*}
		(0.041)	(0.041)	(0.041)	(0.047)
Δ% of free lunch			0.001	0.001	0.003^{*}
			(0.001)	(0.001)	(0.002)
Δ % of minority			-0.023	-0.039^{**}	-0.056
			(0.017)	(0.017)	(0.036)
State-fixed effects				Y	Y
District -fixed effects					Y
N	2320	2320	2320	2320	2320

Note: 1. Δ% of minority is the difference in categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses. 3. The regressions are unweighted.

Table 2.7: **The Effect of the Subgroup Policy on Changes in White Reading Scores**

Δ reading	(1)	(2)	(3)	(4)	(5)
Blacks failed, Whites passed	-0.020	-0.020	-0.014	-0.012	0.005
	(0.038)	(0.038)	(0.039)	(0.042)	(0.057)
Blacks and Whites failed	0.229*	0.229*	0.229*	0.168	0.232***
	(0.134)	(0.134)	(0.136)	(0.118)	(0.063)
Blacks passed, Whites failed	0.283***	0.283***	0.281***	0.296**	0.117***
	(0.053)	(0.053)	(0.055)	(0.115)	(0.025)
Δ SES		-0.022	-0.021	-0.038	-0.014
		(0.046)	(0.046)	(0.048)	(0.054)
Δ% of free lunch			-0.001	-0.0010	0.0010
			(0.001)	(0.001)	(0.002)
Δ % of minority			-0.024	-0.035	-0.013
			(0.018)	(0.022)	(0.058)
State-fixed effects				Y	Y
District -fixed effects					Y
N	2300	2300	2300	2300	2300

Note: 1. Δ% of minority is the difference in categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses. 3. The regressions are unweighted.

Table 2.8: **The Effect of the Subgroup Policy on Math Scores by Proficiency Level**

Δ math	Black (1)	Black (2)	White (1)	White (2)
Blacks failed, Whites passed				
Mid level of proficient	-0.108	-0.149	-0.029	-0.011
	(0.198)	(0.300)	(0.083)	(0.091)
Low level of proficient	-0.096	-0.091	-0.143	-0.092
	(0.201)	(0.325)	(0.091)	(0.106)
Blacks and Whites failed				
Mid level of proficient	0.053	-0.068	-0.086	-0.114
	(0.167)	(0.194)	(0.228)	(0.256)
Low level of proficient	-0.546^{***}	-0.590^{***}	0.014	-0.217
	(0.171)	(0.200)	(0.244)	(0.201)
Blacks passed, Whites failed				
Mid level of proficient	(empty)	(empty)	0.094^{***}	0.086^{***}
			(0.036)	(0.026)
Low level of proficient	(omitted)	(omitted)	0.443^{***}	0.495^{***}
			(0.043)	(0.080)
Blacks failed, Whites passed	-0.030	0.136	-0.045	0.020
	(0.173)	(0.273)	(0.061)	(0.118)
Blacks and Whites failed	0.656^{***}	0.793^{***}	-0.023	0.104
	(0.160)	(0.200)	(0.154)	(0.136)
Blacks passed, Whites failed	-0.274	-0.493^{***}	-0.375^{***}	-0.692^{***}
	(0.171)	(0.067)	(0.061)	(0.070)
Mid level of proficient	0.104	0.18	0.147^{***}	0.136^{***}
	(0.088)	(0.113)	(0.024)	(0.026)
Low level of proficient	0.249^{***}	0.299^{***}	0.211^{***}	0.188^{***}
	(0.081)	(0.104)	(0.030)	(0.033)
ΔSES	Y	Y	Y	Y
Δ% of free lunch	Y	Y	Y	Y
Δ% of minority	Y	Y	Y	Y
State-fixed effects	Y	Y	Y	Y
District -fixed effects		Y		Y
N	530	530	2320	2320

Note: 1.Δ% of minority is the difference in categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses. 3. The regressions are unweighted. 4. Due to the relatively small number of observations, there is no observation in high or middle performing groups, and students in the low proficient group are omitted for analysis.

Table 2.9: **The Effect of the Subgroup Policy on Reading Scores by Proficiency Level**

Δ reading	Black (1)	Black (2)	White (1)	White (2)
Blacks failed, Whites passed				
Mid	-0.102	-0.041	0.006	0.085
	(0.157)	(0.191)	(0.092)	(0.106)
Low	-0.182	-0.055	-0.075	-0.034
	(0.155)	(0.205)	(0.114)	(0.123)
Blacks and Whites failed				
Mid	-0.418^{**}	-0.257	-0.399	-0.383
	(0.177)	(0.187)	(0.350)	(0.317)
Low	-0.768^{***}	-0.779^{***}	-0.619^{**}	-0.594^{**}
	(0.175)	(0.193)	(0.263)	(0.233)
Blacks passed, Whites failed				
Mid	(empty)	(empty)	0.283	0.378
			(0.338)	(0.383)
Low	(omitted)	(omitted)	0.329	0.273
			(0.270)	(0.334)
Blacks failed, Whites passed	-0.031	-0.146	-0.010	0.037
	(0.127)	(0.175)	(0.055)	(0.069)
Blacks and Whites failed	0.466^{***}	0.447^{***}	0.525	0.620^{**}
	(0.108)	(0.140)	(0.333)	(0.245)
Blacks passed, Whites failed	-1.003^{***}	-1.266^{***}	0.008	-0.060
	(0.143)	(0.062)	(0.179)	(0.243)
Mid level	-0.006	-0.074	0.143^{***}	0.142^{***}
	(0.085)	(0.100)	(0.027)	(0.030)
Low level	0.329^{***}	0.270^{**}	0.357^{***}	0.352^{***}
	(0.093)	(0.109)	(0.033)	(0.038)
ΔSES	Y	Y	Y	Y
Δ% of free lunch	Y	Y	Y	Y
Δ% of minority	Y	Y	Y	Y
State-fixed effects	Y	Y	Y	Y
District -fixed effects		Y		Y
N	520	520	2300	2300

Note: 1. Δ% of minority is the difference in categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses. 3. The regressions are unweighted. 4. Due to the relatively small number of observations, there is no observation in high or middle performing groups, and students in the low proficient group are omitted for analysis.

Table 2.10: **The Effect of Initial Characteristics on Changes in Math Scores**

Δ math	Black (1)	Black (2)	White (1)	White (2)
Black failed, White passed	-0.119^{**}	0.001	-0.079^{*}	-0.030
	(0.057)	(0.085)	(0.047)	(0.102)
Black and White failed	0.307^{***}	0.355^{***}	-0.051	0.007
	(0.091)	(0.130)	(0.147)	(0.113)
Black passed, White failed	-0.286	-0.551^{***}	-0.046	-0.399^{***}
	(0.182)	(0.084)	(0.067)	(0.098)
Δ SES	-0.134^{***}	-0.094	-0.049	-0.077
	(0.051)	(0.058)	(0.041)	(0.047)
Δ % of free lunch	-0.003^{***}	-0.003^{*}	0.001	0.001
	(0.001)	(0.002)	(0.001)	(0.002)
Δ % of minority	0.031	-0.011	-0.032	-0.051
	(0.033)	(0.064)	(0.020)	(0.040)
Initial characteristics				
SES in 2002	0.033	0.070	0.025	0.027
	(0.034)	(0.045)	(0.016)	(0.017)
% of free lunch in 2002	-0.002	-0.002	-0.0003	-0.003^{*}
	(0.001)	(0.002)	(0.001)	(0.002)
% of minority in 2002	0.002^{*}	0.0003	0.001	0.001
	(0.001)	(0.002)	(0.001)	(0.002)
State-fixed effects	Y	Y	Y	Y
District-fixed effects		Y		Y
N	530	530	2320	2320

Note: 1. Δ% of minority is the difference between categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses.
3. The regressions are unweighted.

Table 2.11: **The Effect of Initial Characteristics on Changes in Reading Scores**

Δ reading	Black (1)	Black (2)	White (1)	White (2)
Black failed, White passed	-0.090	-0.079	-0.030	-0.006
	(0.065)	(0.099)	(0.045)	(0.066)
Black and White failed	-0.071	-0.038	0.155	0.233^{***}
	(0.129)	(0.168)	(0.135)	(0.077)
Black passed, White failed	-0.839^{***}	-1.283^{***}	0.313^{**}	0.107^{***}
	(0.185)	(0.087)	(0.123)	(0.032)
Δ SES	-0.074	-0.042	-0.042	-0.014
	(0.061)	(0.076)	(0.049)	(0.055)
Δ % of free lunch	-0.0002	0.001	-0.001	-0.0001
	(0.002)	(0.002)	(0.001)	(0.002)
Δ % of minority	-0.011	0.055	-0.022	-0.008
	(0.051)	(0.070)	(0.020)	(0.054)
Initial characteristics				
SES in 2002	0.046	0.058	0.003	0.004
	(0.042)	(0.053)	(0.022)	(0.026)
% of free lunch in 2002	0.003^{*}	0.004	-0.001	-0.002
	(0.002)	(0.003)	(0.001)	(0.002)
% of minority in 2002	-0.002^{*}	-0.002	0.002^{**}	0.001
	(0.001)	(0.002)	(0.001)	(0.003)
State-fixed effects	Y	Y	Y	Y
District-fixed effects		Y		Y
N	520	520	2300	2300

Note: 1. Δ% of minority is the difference between categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses.
3. The regressions are unweighted.

Table 2.12: **The Policy Impact on Changes in Math Scores of Stayers**

Δ math	Black (1)	Black (2)	White (1)	White (2)
Black failed, White passed	-0.111	-0.027	-0.088	-0.065
	(0.067)	(0.086)	(0.054)	(0.113)
Black and White failed	0.197^{**}	0.137	0.165	0.138^{**}
	(0.094)	(0.099)	(0.196)	(0.070)
Black passed, White failed	-0.210	-0.461^{***}	-0.073	-0.400^{***}
	(0.177)	(0.112)	(0.067)	(0.029)
ΔSES	-0.211^{***}	-0.170^{***}	-0.019	-0.052
	(0.052)	(0.054)	(0.047)	(0.053)
Δ % of free lunch	-0.002^{**}	-0.002	0.000	0.004
	(0.001)	(0.001)	(0.002)	(0.003)
Δ % of minority	-0.036	-0.007	-0.060^{***}	-0.071^{*}
	(0.056)	(0.110)	(0.020)	(0.039)
State-fixed effects	Y	Y	Y	Y
District-fixed effects		Y		Y
N	420	420	1900	1900

Note: 1. Δ% of minority is the difference between categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses.
3. The regressions are unweighted.

Table 2.13: **The Policy Impact on Changes in Reading Scores of Stayers**

Δ reading	Black (1)	Black (2)	White (1)	White (2)
Black failed, White passed	-0.066	-0.056	0.030	0.043
	(0.073)	(0.108)	(0.057)	(0.053)
Black and White failed	-0.115	-0.120	0.150	0.217^{***}
	(0.119)	(0.103)	(0.127)	(0.066)
Black passed, White failed	-0.712^{***}	-1.1^{***}	0.342^{***}	0.116^{***}
	(0.347)	(0.101)	(0.105)	(0.026)
ΔSES	-0.095	-0.041	0.029	0.022
	(0.060)	(0.072)	(0.053)	(0.059)
Δ % of free lunch	-0.004^{**}	-0.002	-0.001	0.004
	(0.002)	(0.002)	(0.002)	(0.003)
Δ % of minority	-0.001	-0.079	-0.053^{**}	-0.008
	(0.070)	(0.100)	(0.022)	(0.063)
State-fixed effects	Y	Y	Y	Y
District-fixed effects		Y		Y
N	410	410	1890	1890

Note: 1. Δ% of minority is the difference between categorical variables in 2004 and 2002 for % of nonwhite students. 2. Standard errors clustered at school level are presented in parentheses.
3. The regressions are unweighted.

Appendix: Data Description

Table A1. The ECLS variables

Variable	Description
Individual-level variables	
Test scores	Fifth graders' math and reading scores in the spring of 2004; This paper used t-scores having a mean of 50 and variance 10 and standardized the scores to mean of zero and variance one in the raw entire sample
Prior test scores	Third graders' math and reading scores in the spring of 2002; Similarly, this paper standardized t-scores to mean of zero and variance one in the raw entire sample
Parental SES	This is a continuous variable, ranging from -2.48 to 2.54 To calculate the SES, the ECLS used parents' education, occupation, income and labor force status and normalized the each component to mean of zero and standard deviation of one. Then, they computed the SES score by averaging the normalized components.
Race	White/ Black/ Hispanic/ Asians/Other/Multi Race; This variable comes from a parent interview and FMS
School-level variables	
% of nonwhite students	less than 10% / 10-24% / 25-49% / 50-74%/ more than 75%; This variable indicates the percentage of nonwhites for the fifth graders in spring 2004. The ELCS summed the students with the exception of non-Hispanic Whites to compute the variable.
% of those eligible for free lunch	This is a continuous variable; This also comes from school administrator questionnaire and Common Core Data (CCD).

Note: the source of description comes from the confidential-use version manual of the ECLS provided by NCES, U.S. Department of Education.

Chapter 3

The Impact of School Accountability on Racial Diversity of Peer Groups

3.1 Introduction

School accountability aims to improve student test scores and public school performance and to narrow student achievement disparities. Numerous studies document that accountability policies have raised academic achievement of students who attend under performing schools (Ahn and Vigdor, 2014; Chiang, 2009; Deming et al, 2015; Reback et al, 2014; Rouse et al, 2013; and Springer, 2008). However, some researchers cast doubt on the impact of the school accountability, indicating that the policies result in unwanted consequences such as teaching to the test, academic gains of low achievers at the expense of high achieving students, and removal of low performing students from a testing pool (Chakrabarti, 2013; Figlio, 2006; Hanushek and Raymond, 2005; Jacob, 2005; Krieg, 2011; Neal and Schanzenbach, 2010; Reback, 2008). Although the studies show mixed evidence, it points out that student test scores are influenced by school accountability.

Student outcomes are also affected by students' peer groups in schools, which will have an impact on their long-run consequences such as labor earnings in adulthood. Prior studies point out that students are influenced by the racial composition of their peer groups. For example, relative to white students, black students in the same school may be adversely influenced by a high proportion of black peers who are engaged in school suspension activities (Billings, Deming and Rockoff

2014, Card and Rothstein 2007 and Hanushek, Kain and Rivkin 2009). There is evidence implying that school accountability alters students' peer groups. Accountability affects households' residential choices for children's schooling (Billings, Brunner and Ross, 2014; Figlio and Lucas, 2004). Hart and Figlio (2015) presents evidence that school accountability causes socioeconomic segregation across schools by increasing enrollments of kindergarteners from high socioeconomic families in good quality primary schools. However, few studies examine a relationship between accountability and the racial composition of schools.

How could school accountability change the racial composition of a student population in schools? If revealing information on school performance reduces a cost of gaining information on school quality, this could make it easy for parents to recognize which schools are low and high quality schools. Parents could send their children to a better quality school at a low cost of information on school quality. However, household residential choice and student mobility patterns may not be necessarily similar across race and ethnicity. If parents from a racial group are more responsive to the school ratings while others are less likely to move due to an income constraint, the racial composition within schools and between schools may change.

In this paper, I study whether the test-based school accountability changes the racial diversity of a student population in schools and causes racial segregation across schools. Specifically, I test if there is an increase in black student enrollment and a decrease in white student enrollment in underperforming schools. For analysis, I use the North Carolina administrative data. As North Carolina has administered its own accountability called the "ABCs program" since 1996-1997, longitudinal records at student-level and school-level under accountability have been kept. Using the data, I estimate the policy effect of an inaugural year instead of exploiting time variation in school ratings. School grades and student test scores are inherently volatile over time (e.g. Kane and Staiger; 2002, Kane et al; 2002, and Figlio and Lucas, 2004). Parents may not trust accountability results. Then, exploiting time variation in school grades may be an inaccurate measure of how households respond to the information on school quality. Exploiting the cross-sectional variation in school grades in the first year, I estimate a difference-in-differences model. I compare schools recognized as low-performing schools in 1996-1997 as a treatment group and No Recognition schools in the same year as a control group over time. The control group of schools would face no stigma unlike the low-performing schools do. Thus, I regard changes in the outcome variable in the control group as what the treatment group would have experienced in the absence of the ABC accountability pressure in 1997. I also take a group-specific trend into account in the model to allow for a pre-existing trend in the treatment group which differs from the control group.

Controlling for the group-specific trend, I find that both black and white student enrollment statistically insignificantly change after the policy began. On the other hand, when using the enrollment shares as a dependent variable to increase statistical power, I find that, in the treatment schools, the share of black students statistically falls statistically significantly relative to the control schools over time while the share of white students rises statistically significantly relative to the control schools during the post-policy period. It may be plausible that parents of black students recognize a relatively good quality school at the cheap cost under the policy and move to another school, and this could be one reason why the enrollment share of black students is relatively low. However, to understand why white students' enrollment share in the treatment group rises relative to the control schools is somehow puzzling because low-performing schools in North Carolina exhibited a high proportion of black students prior to the policy. One possible explanation is that the treatment group of schools might invest in school quality to attract additional white students after the policy began. My findings imply that the low initial share of whites in low-performing schools means that the rise serves to integrate schools racially, considering that most of a student population in low-performing schools is black.

3.2 Background

As a part of public school reform, North Carolina passed its test-based accountability framework, called the ABCs program, in June 1996.[1] The state has implemented the ABCs since academic year 1996-1997. In the first year of implementation, students in elementary and middle schools took End of Grade (EOG) math, reading, and writing exams, and high school students took tests since 1997-1998.[2]

[1] The ABCs program emphasizes Accountability, Basic Skills with high educational standards, and local Control.
[2] High school students took End of Course exams.

Based upon the student assessment scores, the ABCs program has constructed two measures - an achievement performance composite and a growth composite. The performance measure represents the percentage of students whose test scores are at or above a certain level of achievement, and the growth measure represents whether schools met the expected growth which is determined by statewide average score growth. School growth scores are calculated using two years of reading and math, and three years of writing exam scores (Department of Public Instruction, 1997 [3]). In 1996-1997, the growth composite used only one previous year of EOG reading and math exams, and two previous years of Grade 4 writing exam while the performance composite employed scores of all students in math, reading, and writing exams at the current year.

Public schools are evaluated based upon the two measures. In 1996-1997, there were four school ratings: Exemplary schools, Expected schools, Schools with No Recognition, and Low Performing schools. Table 3.1 presents this classification. If schools met the expected growth target, they were classified as Exemplary schools when their growth composite scores were at least 10% higher than the target. Otherwise, they were recognized as Expected schools. Among schools that did not meet the growth measure, they became No Recognition schools if 50% of students were above the performance standard and schools were identified as Low Performing schools if their performance is 50% below the standard. In addition to the four categories, schools were recognized as Excellence schools if they met the expected growth target and their performance composite scores were higher than 90%. If schools met the growth target and their performance scores were higher than 80%, they were recognized as Distinguished schools. The special school recognitions of excellence and distinction are not exclusive. Also, standards and categories of the school ratings have changed over time. In academic year 2001-2002, the standard of performance measure for No Recognition increased from 50% to 60%, and a new category was added to the existing four-categories. If the level of performance measure was between 50%-59%, schools were recognized as priority schools irrespective of having make the score growth.

Once schools were recognized as Exemplary or Excellence schools, teachers in the schools received monetary rewards. In the meantime, low-performing schools must notify parents of the school rating while other schools were not required to do so, which created school stigma for the

[3] Executive Summary http://www.ncpublicschools.org/abc_results/results_97/

low-performing school performance. In addition, the state offered assistance teams to the low-performing schools to help them change their curriculum and practice.

ABCs school ratings highly fluctuated over time as Table 3.2 indicates. In the first year of the policy, 122 schools were identified as low-performing schools but only about twenty schools received a low-performing grade. On the other hand, while 517 schools received an Exemplary grade in 1997, more than 1,200 schools received the highest grade in 1998.

3.3 Data

I use the administrative longitudinal data of North Carolina public schools provided by North Carolina Education Research Data Center (NCERDC). This data contains school-level and student-level information. Public School Universe (PSU) collects data of all public elementary and secondary schools, including numbers of enrollments by grade and ethnicity, a percentage of students who are eligible for free lunch, the number of full time teachers, pupil-teacher ratio, and more. School Report Card (SRC) releases public school performance under the ABCs accountability since 1996-1997. In SRC, performance composite scores, growth composite scores, and school ratings are available. Using the unique school identifier, I link SRC to PSU from 1996-1997 and 1999-2000. Then, I merge this school-level data to the individual-level data. This combined data covers student information from 1994-1995 to 1999-2000 who are in grades 3 through 8 in schools. The data of year 1996 includes students only in grades 3 and 4. So, I drop the year from the sample. I use regular public schools that received a rating of Low-performing or of No Recognition in the inaugural year. Next, as Table 3.3 shows, there is no standardized rule of serving student grades across North Carolina public schools. Due to heterogeneity of sample, I need to define consistent standards about student grades, and so I use elementary schools only. I define elementary schools as a group of schools that serve students in either grades 3, 4, or up to 5 as a highest student grade. Then, I create a sample, which contains 53 Low-performing schools and 240 No Recognition schools in each year as Table 3.4 indicates. All non-regular public schools including charter schools are out of my sample.

3.3.1 Variables

The main outcome of interest is student enrollment by race. To construct the student racial enrollment, I aggregate the number of students who enrolled in a school for each racial group.[4] The aggregated racial enrollment only counts students in grades 3 through 5 because the individual-level data does not have any information about students in grades K through 2 even if they serve those students.

For the main independent variable, I generate a dummy variable indicating that a school was recognized as a low-performing school or not in 1996-1997, the inaugural year of the accountability policy. Although schools received different grades over time, the value of the indicator is same over time to identify the first receipt of the low-performing grade.

Table 3.5 displays a data structure.

3.3.2 Descriptive Statistics

To check the validity of a common trend assumption, I examine trends in black and white enrollment in both treatment and control groups. In Figure 3.1, a top dash line represents an average of black student populations in treatment schools and a bottom dash line represents a trend in average white student enrollment in the group of schools. A top solid line represents a trend in white student population in control schools and a bottom solid line represents a trend in black student enrollment in the same group. This displays different patterns in the trends by race in each group of schools. In the treatment group, there exists a pre-existing trend in racial enrollment. Relative to the control group, black student enrollment was on an upward trend while white student enrollment was on a downward trend before the policy started. Therefore, it is important to control for the pre-trend in the treatment group. Otherwise, this will bias the results.

Table 3.6 presents summary statistics for the sample. On average, the total enrollment in the treatment schools is smaller than the control schools each year. On the other hand, the average black enrollment in the treatment schools is relatively big to the control schools. In 1995, the size of the average black enrollment was about 118 and about 131 in 1997. On the other hand, the size in

[4]The enrollment status from the student-level indicates that a student attended a school in Spring semester because the data collected during the statewide End of Grade exam testing period in May.

the control schools was about 82 in 1995 and about 88 in 1997 respectively. The white enrollments on average in the treatment schools were less than half of the enrollments in the control schools. In 1995, the white enrollment was about 58, but the size kept decreasing over time, and the enrollment was about 42 in 2000. In the control schools, the average white enrollment was about 163 in 1995, about 165 in 1997, and about 155 in 2000 respectively. In the meanwhile, the average enrollment of the other racial groups was quite small in both groups of schools.

I also present average outcomes of the difference-in-differences in Table 3.7 to see how the variables change in the treatment and control groups before and after the policy introduced. In Panel A, I use level enrollment variables. Column (1) shows enrollment values for 1997, and column (2) for 2000: 1997 represents the pre-policy period, and 2000 represents the post-policy period. Row (1) - (4) show total enrollment values, black enrollment values, white enrollment values, and other enrollment values for the treatment schools respectively. Row (5) - (8) present the values for the control schools. I difference each enrollment values for both groups and present the results in column (3): the total enrollment, the black enrollment, the white enrollment fall for the treatment schools by 3.25, by 1.03, by 6.93 and other enrollment for the treatment schools rises by 4.72. For the control schools, the total enrollment rises by 1.26, the black enrollment rises by 3.7, but the white enrollment falls by 9.3, and the other enrollment rises by 6.86. I assume that the total enrollment in the treatment group would have risen by 1.26 in the absence of the policy, the black enrollment by 3.7, the other enrollment by 6.86, and that the white enrollment in the treatment group would have fallen by 9.3. Therefore, the difference in these differences in column (5) represents the effect of the policy. The results are -4.51 for the total enrollment, -4.74 for the black enrollment, 2.36 for the white enrollment, and -2.13 for the other enrollment in the treatment schools.

In Panel B, I use log enrollment variables. Row (1) - (4) in Panel B show log enrollment values for the treatment schools and row (5) - (8) present the values for the control schools. Between the two periods, for the treatment schools, the total enrollment rises by 0.4 log points, the black enrollment falls by 0.3 log points, the white enrollment falls by 7.3 log points, and the other enrollment rises by 32.7 log points. In the absence of the policy, for the treatment schools, the total enrollment would have risen by 0.2 log points, the black enrollment would have risen by 4.1 log points, the white enrollment would have fallen by 12.2 log points, and the other enrollment would have risen by 38.3 log points. Thus, due to the policy impact, the total enrollment rises by 0.1 log points, the black enrollment falls by 4.4 log points, the white enrollment rises by 4.8 log points, and the other enrollment falls by 5.6, relative to the control schools. For these values, log point changes and percent changes are approximately the same. Also, these results are comparable to the regression results which are in logs.

3.4 Empirical Framework

I estimate the effect of being recognized as low-performing schools using the difference-in-differences model. Figlio and Lucas (2004) indicates massive variation in the school grade in the years after the introduction of accountability. Table 2 shows that there exists huge fluctuation in North Carolina school grades under accountability. By following Figlio and Lucas (2004), I use school ratings receiving at the inaugural year of the ABCs rather than changes in school evaluation over time. I compare low-performing schools in 1996-1997 as a treatment group and No Recognition schools in the same year as a control group over time. The control group of schools face no stigma; the low-performing schools do. I regard changes in the outcome variable in the control group as what the treatment group would have experienced in the absence of the ABCs accountability pressure. The empirical framework is as follows.

$$Y_{s,t} = \sum_{j=t} \beta_j \mathrm{Yr}_j Low_s + \sum_{j=t} \gamma_j \mathrm{Yr}_j + \mu_s + \varepsilon_{s,t} \tag{3.1}$$

Let s and t represent a school and year. Let Low_s be equal to one if s was recognized as a low-performing school in Summer 1997, or to zero otherwise. Let Yr_t denote a vector of year dummies where $t = (1995, 1997, 1998, 1999, 2000)$. Year dummies control for a common trend in the outcome variable. A reference year is the academic year 1996-1997, the inaugural year of the ABCs. To control for observed and unobserved school characteristics, I add a set of school dummies to the equation. I use the proportion of black students and of white students in a student population as well as logged enrollment variables as a dependent variable. I also use the share of black student enrollment and of white student enrollment for a dependent variable.

In Equation (1), the coefficients of interest are the β_{98}, β_{99}, and β_{00}. The β_{98}, which is the coefficient of the most interest, is an estimate of the effect on the outcome variable one year after schools were recognized as low-performing schools in Summer 1997. μ_s is the school-fixed effects to control for unobserved time invariant effects. I test if the coefficients of interest are positive, which implies that low-performing schools experience a decrease in the racial diversity of a student population even more than No Recognition schools do.

The validity of estimating Equation (1) relies on the assumption of a common trend. If both groups exhibit a common linear trend in the dependent variable prior to the policy began, an interaction of Year 1995 with the treatment variable would be insignificant. However, as Figure 1 indicates, the treatment group seems to exhibit a different pre-existing trend in the enrollment variables compared to the control group. Thus, to allow for the group-specific trend, I replace the interaction variable with an interaction variable, Low_sTime_t, in Equation (1) where $Time_t = (0,2,3,4,5)$ for each t. Then, I estimate Equation (2).

$$Y_{s,t} = \sum_{j=t} \beta_j \mathrm{Yr}_j Low_s + \sum_{j=t} \gamma_j \mathrm{Yr}_j + \delta Low_s Time_t + \mu_s + \varepsilon_{s,t} \tag{3.2}$$

3.5 Findings

Table 3.8 presents results of the policy effect on student log enrollment. To test whether there exists the pre-existing trend, I estimate Equation (1) and report the results in the first four columns. In column (5)-(8), I present the results of estimating Equation (2). In column (1), total student enrollment in the treatment group is 0.4% higher than the control group prior to the policy. There seems to be no pre-existing trend because the estimate is small and statistically insignificant. After the policy introduced, there is a small and insignificant increase in total enrollment in the treatment schools.

In column (2), the coefficient of the pre-trend in the black student enrollment is not statistically significant but is large: the black enrollment in the treatment schools is 4.8% lower than the control schools in 1995, which means that there may exist a pre-existing trend the treatment schools that I am unable to estimate precisely. After the policy started, the black student enrollment is 1% higher than the control schools in 1998, and then it was 3.4% and 3.6% lower than the control schools in 1999 and 2000 respectively but are not statistically significantly so.

In column (3), white enrollment is statistically insignificantly 14.5% higher than the control schools before the policy was introduced, implying that there may be a pre-trend. In 1998, the white student enrollment is 1.3% higher than the control schools. In 1999, the white enrollment in the treatment schools is 3.6% lower than the control schools and is 4.3% higher than the control schools in 2000, but are not statistically significantly so. In column (4), the enrollment for the other racial groups is 0.7% lower than the control schools in 1995, and there seems no pre-trend because the coefficient is small and insignificant. The enrollment of the other racial group is insignificantly lower than the control schools after the policy introduced.

Taking evidence of a pre-existing trend into account, I allow for the group-specific trend. The interaction of Treatment with Year 1995 is no longer identified in this specification, and thus it is dropped. In column (5), looking at the coefficient of an interaction of Treatment with time, total enrollment in the treatment schools falls by 0.2% per year compared to the control schools. The size of the coefficient is not significant and small and this means there may be no pre-existing trend. The total student enrollment after the policy is 2.9% higher than the control schools in 1998, 0.8% and 1.3% higher than the control schools in 1999 and 2000 respectively, but are not statistically significant. This suggests that the policy had no effect on total enrollment.

Results for black students in column (6) are different. Black student enrollment in the treatment schools rises 2.4% per year compared to the control schools as hinted at in Figure 1 but the coefficient is not statistically significant. After the policy began, the black student enrollment in the treatment schools falls relative to control schools but not statistically significantly so. The size of the coefficient is now large, however. After one year passed, black student enrollment is 1.4% lower than the control schools and 8.1% and 10.7% lower than the control schools in 1999 and 2000 respectively.

In column (7), the white student enrollment in the treatment schools falls by 7.2% per year relative to the control school. Though it is not statistically significant, the large coefficient suggests that there may be a downward trend as we saw in Figure 1. The policy appears to lead to rapidly rising white enrollment though the rises are statistically insignificant. The white student enrollment is 8.5% higher than the control schools in 1998. In 1999, the white enrollment is 10.9% higher than the control schools, and it is 26.1% higher than the control schools in 2000.

In Table 3.9, I examine the policy impact on the enrollment share by race. Black and white enrollment in Table 8 appears to respond oppositely to the policy and to have opposite trends, so using the shares allows me to gain precision. Columns (1) - (3) show the results without the group-specific trend. For black enrollment in column (1), the coefficient for an interaction of Treatment with Year 1995 is statistically significant at the level of 10%. In 1995, the black enrollment share is 2.7 percentage points lower than the control schools. After the policy starts, the share of black student enrollment falls relative to the control schools over time. The share is lower by 0.01 percentage points, 1.2 percentage points, and 2 percentage point than the control schools in 1998, 1999, and 2000 respectively, but none of coefficients of interest are statistically significant.

In column (2), the share of white enrollment is 2.9 percentage points higher than the control schools at the significance level of 10% during 1995, suggesting a pre-existing upward trend. After the policy is introduced, the share of white student enrollment falls and is lower by 0.8 percentage points, 1.2 percentage points, and 2.5 percentage points higher than the control schools in 1998, 1999, and 2000 respectively. The magnitudes of the coefficients are not large and not statistically significant. In column (3), the share of the other racial group enrollment fluctuates over time after the policy, but the magnitude of the coefficients is quite small, and the estimates are not statistically significant.

I add an interaction of Treatment with Time to control for the group-specific trend and present the results in columns (4) - (6) in Table 9, which changes the results. In column (4), the treatment school's share of black students rises 1.3 percentage points per year compared to the control schools at the significance level of 10% and exhibits an upward trend as Figure 1 shows. After the new policy, the share of black students in the treatment schools is on a downward trend relative to the control schools. In 1998, the share of black students is 1.4 percentage points lower than the control schools but is statistically insignificant. In 1999, the share is 3.9 percentage points lower than the control schools at the significance level of 10%. The effect in 2000 is large because the share is 6 percentage points lower than the control schools at the significance level of 10%.

In column (5), the coefficient on Treatment times Time indicates a statistically significant downward trend in the share of white students of 1.5 percentage points per year for the treatment schools relative to the control schools. After the policy is introduced, the share of white students in the treatment schools is on a statistically significant upward trend relative to the control schools. In 1998, the share in the treatment schools is statistically significantly 2.3 percentage points higher than the control schools, 4.1 percentage points and 6.9 percentage points higher than the control schools in 1999 and in 2000 respectively. In column (6), the treatment-group specific trend is small and not statistically significant. After the policy is introduced, the share of the other racial groups seems to be on a downward trend in the treatment schools relative to the control group. But none of estimates are statistically significant over time and the size of the coefficients is relatively smaller than the other two racial groups show.

Additionally, since there is evidence that of a pre-existing trend, I add one more year, Year 1996, to allow for group-specific trend. However, this year includes students in grade 3 and 4 only, and this is why I do not consider the year for the main results. Results are displayed in Appendix Table B1 and Table B2. The results still yield similar results.

To summarize the results of Table 3.9, school accountability changes the share of enrollment differently by race. In the treatment schools, the black student enrollment share falls relative to the control schools, but the white student share rises statistically significantly in the treatment schools relative to the control schools. The results are different from what I expected, but it may be possible that the policy changes the treatment group of schools' education environments and so the schools could attract additional white students.

3.6 Discussion and Concluding Remarks

There is an extensive literature on how school accountability affects student test scores. Although the policy could affect parents' school choices and therefore the compositions of students' peer groups, there are few studies examining this relationship. With a focus on the racial composition of a student population in schools, I examine whether the accountability alters the racial demographics of students and causes increased racial segregation across schools.

Hart and Figlio (2015) finds that school accountability in Florida increases income segregation among incoming students across schools. Extending this to examine racial segregation, I find evidence that school accountability could reduce racial segregation. The results differ from what

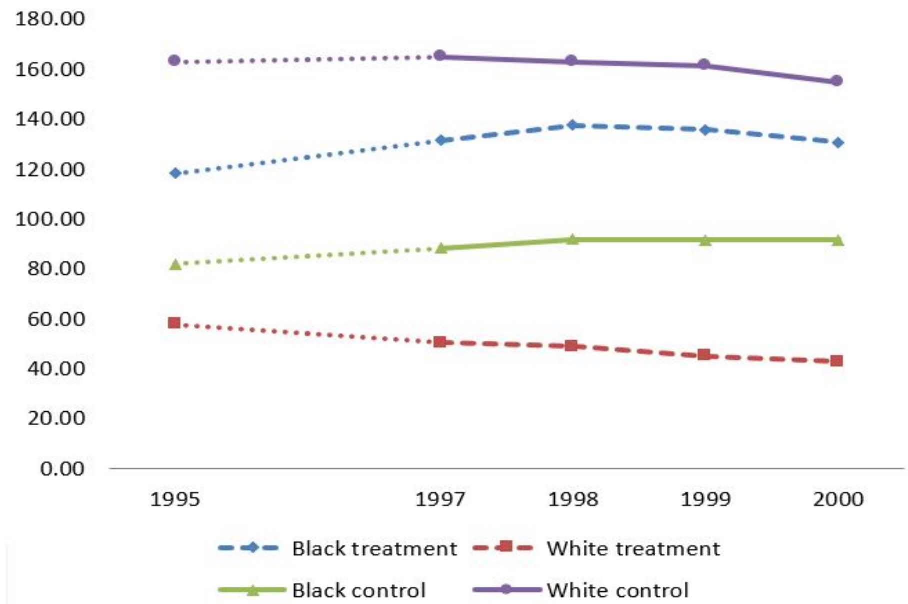
I expected. Using the enrollment shares as a dependent variable to increase statistical power, I find that, in the treatment schools, the share of black students statistically falls statistically significantly relative to the control schools over time while the share of white students rises statistically significantly relative to the control schools during the post-policy period.

It may be possible that parents of black students recognize their children schools' quality and which school is a relatively quality school at the cheapest cost under the policy. However, it is a bit puzzling why white students' share increases in the treatment group during the post-policy period. Parents of white students may believe that the schools improve the quality of schools, staffs, or education environment more than the control group does. It could be also plausible that the treatment group invests in school quality to attract additional white students after they are recognized as low-performing schools. Further analysis necessitates more data information prior to the introduction of the policy, which is a limit of this paper.

It is important to see that the two racial groups in the treatment schools show different responses to the school accountability policy. From the results, I conclude that the low initial share of whites in low-performing schools means that the rise serves to integrate schools racially, considering that the majority of a student population in low-performing schools is black. Together, my findings imply that the accountability policy reduces racial segregation.

Figures and Tables

Figure 3.1: **Trends in Student Populations by Race**



Note: 1. 53 observations for treatment and 240 observations for control group are used per year. 2. Schools are elementary serving students in either grades 3, 4, or 5 as a highest grade. 3. Schools are regular public schools, non-charter schools. 4. enrollment variables are level variable, averaged by year. 5. Panel B includes year 1996 but only grades 3 and 4 enrollment variables.

Table 3.1: **ABCs Recognition**

Growth Measure	Performance Measure (%)	Rating
Met exemplary growth[1]	Not available	Exemplary school
Met expected growth	Not available	Expected school
Not meet expected growth	Greater than 50%	School of No Recognition
Not meet expected growth	Less than 50%	Low performing school
Not available	50%-59%	Priority school[2]

Note: 1. Exemplary growth differs from the expected growth. If a school's growth composite score is 10% higher than the expected growth standard, the statewide average growth score, the school is recognized as an exemplary school while the expected school meets the standard, but its growth composite is not above the additional 10%. These two ratings are irrespective of the performance composite measure. 2. In academic year 2001-2002, a new category, Priority school, was added to existing four-categories. From the academic year 2001-2002, the standard of performance measure for No Recognition changed from 50% to 60%.

Table 3.2: **ABCs Ratings over Time**

Year	Exemplary	Expected	Low Performing	No Recognition
1997	517	383	122	557
1998	1,217	366	23	281
1999	1,123	432	7	329
2000	904	458	21	552

Note 1. The number of observations comes from a raw data including all public primary and secondary schools.

Table 3.3: **Grade Range**

Grade range	Obs	Percent
02-05	6	0.51
03-05	33	2.82
04-05	30	2.56
K-03	19	1.62
K-04	14	1.19
K-05	654	55.80
PK-03	19	1.62
PK-04	28	2.39
PK-05	369	31.48
Total	1,172	100

Note: 1. Samples are primary schools that operate over all five years. 2. Sample schools received either Low-performing or No Recognition grade in 1997. 3. K represents kindergarteners and PK represents pre-kindergarteners. 4. Grade range variable comes from the school-level data, which does not have year 1995 information. Thus, the total number of schools is not the same as the total number of sample school for analysis.

Table 3.4: **Numbers of Sample Schools**

Year	Low Performing	No Recognition
1995	53	240
1997	53	240
1998	53	240
1999	53	240
2000	53	240
Total	265	1200

1. Schools are elementary serving students in either grades 3, 4, or 5 as a highest grade. 2. Schools are regular public schools, non-charter schools. 3. Schools are operating over all five years.

Table 3.5: **Data Structure**

	1994-1995	1995-1996	1996-1997	1997-1998	1998-1999	1999-2000
school-level enrollment						
PK, KG, 1∼8	-	-	x	x	x	x
Total student	-	-	x	x	x	x
Total blacks	-	-	x	x	x	x
Total whites	-	-	x	x	x	x
Total asians	-	-	x	x	x	x
Total Hispanics			x	x	x	x
student-level data						
student race	x	x	x	x	x	x
student grade	3∼8	3∼4	3∼8	3∼8	3∼8	3∼8

Note: 1. School-level grade enrollment variables are not disaggregated by race. 2. Student grade from the student-level data ranges between 3 and 5 in my sample and while it is available to use information of students in grades 3 through 8 from the raw data.

Table 3.6: **Summary Statistics**

	Treatment Group		Control Group	
Variable	Mean	Std. Dev.	Mean	Std. Dev.
Year 1995				
total enrollment	191.64	101.24	256.66	106.72
black enrollment	117.92	64.81	81.77	63.11
white enrollment	57.58	51.13	162.55	97.34
others enrollment	16.13	20.96	12.33	29.13
Year 1997				
total enrollment	200.81	99.23	268.04	106.74
black enrollment	131.15	67.66	88.09	69.18
white enrollment	50.09	45.73	164.65	99.45
others enrollment	19.57	23.82	15.30	28.17
Year 1998				
total enrollment	205.51	97.82	271.12	108.24
black enrollment	137.25	68.95	91.63	70.61
white enrollment	48.66	44.08	162.77	102.41
others enrollment	19.60	20.78	16.73	27.31
Year 1999				
total enrollment	202.98	99.88	272.15	108.89
black enrollment	135.42	73.07	91.37	68.35
white enrollment	44.77	42.20	161.26	105.06
others enrollment	22.79	26.26	19.52	29.25
Year 2000				
total enrollment	196.64	80.10	268.22	107.82
black enrollment	130.38	61.91	91.26	69.79
white enrollment	42.47	34.79	154.55	103.81
others enrollment	23.79	26.03	22.41	30.12

1. 53 observations for treatment and 240 observations for control group are used. 2. Schools are elementary serving students in either grades 3, 4, or 5 as a highest grade. 3. Schools are regular public schools, non-charter schools. 4. All enrollment variables are aggregated from the individual-level data, only including students in grades 3 through 5.

Table 3.7: **Difference-in-Differences**

Panel A : Levels	1997	2000	2000-1997	(2000-1997)/1997	DD
Treatment					
total enrollment	199.492	196.246	-3.246	-0.016	-4.507
black enrollment	133.787	132.754	-1.033	-0.008	-4.736
white enrollment	47.475	40.541	-6.934	-0.146	2.369
other enrollment	18.230	22.951	4.721	0.259	-2.139
Control					
total enrollment	270.846	272.107	1.261	0.005	-
black enrollment	90.807	94.511	3.704	0.041	-
white enrollment	163.836	154.532	-9.304	-0.057	-
other enrollment	16.204	23.064	6.861	0.423	-
Panel B : Logs	**1997**	**2000**	**2000-1997**	**(2000-1997)/1997**	**DD**
Treatment					
total enrollment	5.187	5.190	0.004	0.001	0.001
black enrollment	4.758	4.755	-0.003	-0.001	-0.044
white enrollment	3.422	3.349	-0.073	-0.021	0.048
other enrollment	2.352	2.679	0.327	0.139	-0.056
Control					
total enrollment	5.525	5.528	0.002	0.000	-
black enrollment	4.037	4.078	0.041	0.010	-
white enrollment	4.892	4.771	-0.122	-0.025	-
other enrollment	2.372	2.755	0.383	0.162	-

1. 61 observations for treatment and 280 observations for control group are used. 2. Schools are elementary serving students in either grades 3, 4, or 5 as a highest grade. 3. Schools are regular public schools, non-charter schools. 4. All enrollment variables are aggregated from the individual-level data, only including students in grades 3 through 5. 5. Results of level variables are displayed in Panel A and results of log variables are presented in Panel B. 6. I add one to enrollment variables before taking natural logarithm. 7. Column DD presents first difference of treatment minus a first difference of treatment. 8. Column DD in% presents treatment in fourth column minus control in the same column.

Table 3.8: **The Effect on Log Enrollment in Grade 3 through 5**

	Column (1)	Column (2)	Column (3)	Column (4)	Column (5)	Column (6)	Column (7)	Column (8)
Variable	All	Blacks	Whites	Others	All	Blacks	Whites	Others
Year 1995*Treatment	0.004	-0.048	0.145	-0.007	-	-	-	-
	(0.036)	(0.044)	(0.091)	(0.116)				
Year 1998*Treatment	0.027	0.01	0.013	-0.099	0.029	-0.014	0.085	-0.102
	(0.031)	(0.037)	(0.049)	(0.078)	(0.038)	(0.047)	(0.058)	(0.104)
Year 1999*Treatment	0.004	-0.034	-0.036	-0.113	0.008	-0.081	0.109	-0.12
	(0.039)	(0.051)	(0.070)	(0.092)	(0.053)	(0.070)	(0.129)	(0.150)
Year 2000*Treatment	0.006	-0.036	0.043	-0.135	0.013	-0.107	0.261	-0.146
	(0.047)	(0.058)	(0.097)	(0.099)	(0.076)	(0.094)	(0.182)	(0.214)
Year 1995	-0.050^{***}	-0.061^{***}	-0.015	-0.226^{***}	-0.050^{***}	-0.061^{***}	-0.015	-0.226^{***}
	(0.009)	(0.019)	(0.018)	(0.045)	(0.009)	(0.019)	(0.018)	(0.045)
Year 1998	0.012^{*}	0.052^{***}	-0.030^{***}	0.138^{***}	0.012^{*}	0.052^{***}	-0.030^{***}	0.138^{***}
	(0.007)	(0.019)	(0.009)	(0.034)	(0.007)	(0.019)	(0.009)	(0.034)
Year 1999	0.017	0.059^{**}	-0.063^{***}	0.274^{***}	0.017	0.059^{**}	-0.063^{***}	0.274^{***}
	(0.012)	(0.023)	(0.020)	(0.043)	(0.012)	(0.023)	(0.020)	(0.043)
Year 2000	0.001	0.047^{*}	-0.130^{***}	0.419^{***}	0.001	0.047^{*}	-0.130^{***}	0.419^{***}
	(0.015)	(0.026)	(0.025)	(0.048)	(0.015)	(0.026)	(0.025)	(0.048)
Treatment*time	-	-	-	-	-0.002	0.024	-0.072	0.004
					(0.018)	(0.022)	(0.045)	(0.058)
school dummy	Y	Y	Y	Y	Y	Y	Y	Y
N	1465	1465	1465	1465	1465	1465	1465	1465

Note: 1. Standard errors clustered at school level are presented in parentheses. 2. The regressions are unweighted. 3. The reference year is 1997. 4. The dependent variable is a logged enrollment. 5. Others represents the other racial and ethnic groups of students. 6. Enrollment variables are aggregated from the individual-level data, which include students in grades 3 through 5 only.

Table 3.9: **The Effect on Enrollment Share**

	Column (1)	Column (2)	Column (3)	Column (4)	Column (5)	Column (6)
Variable	Blacks	Whites	Others	Blacks	Whites	Others
Year 1995*Treatment	-0.027^{*}	0.029^{*}	-0.002	-	-	-
	(0.016)	(0.015)	(0.008)			
Year 1998*Treatment	-0.0001	0.008	-0.008	-0.014	0.023^{**}	-0.009
	(0.007)	(0.006)	(0.005)	(0.010)	(0.009)	(0.007)
Year 1999*Treatment	-0.012	0.012	0.000002	-0.039^{*}	0.041^{*}	-0.002
	(0.012)	(0.012)	(0.006)	(0.023)	(0.022)	(0.010)
Year 2000*Treatment	-0.02	0.025	-0.005	-0.060^{*}	0.069^{**}	-0.009
	(0.015)	(0.016)	(0.008)	(0.031)	(0.031)	(0.013)
Year 1995	-0.009^{**}	0.019^{***}	-0.010^{***}	-0.009^{**}	0.019^{***}	-0.010^{***}
	(0.004)	(0.004)	(0.002)	(0.004)	(0.004)	(0.002)
Year 1998	0.011^{***}	-0.017^{***}	0.006^{***}	0.011^{***}	-0.017^{***}	0.006^{***}
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Year 1999	0.016^{***}	-0.030^{***}	0.015^{***}	0.016^{***}	-0.030^{***}	0.015^{***}
	(0.004)	(0.004)	(0.002)	(0.004)	(0.004)	(0.002)
Year 2000	0.020^{***}	-0.046^{***}	0.026^{***}	0.020^{***}	-0.046^{***}	0.026^{***}
	(0.005)	(0.005)	(0.003)	(0.005)	(0.005)	(0.003)
Treatment*time	-	-	-	0.013^{*}	-0.015^{*}	0.001
				(0.008)	(0.008)	(0.004)
school dummy	Y	Y	Y	Y	Y	Y
N	1465	1465	1465	1465	1465	1465

Note: 1. Standard errors clustered at school level are presented in parentheses. 2. The regressions are unweighted. 3. The reference year is 1997. 4. The dependent variable is a proportion of enrollment. 5. Others represent the other racial and ethnic groups of students. 6. Enrollment variables are aggregated from the individual-level data, which include students in grades 3 through 5 only.

Appendix: Tables

Table B1.The Effect on Log Enrollment in Grades 3-4 Including 1996

	Column (1)	Column (2)	Column (3)	Column (4)	Column (5)	Column (6)	Column (7)	Column (8)
Variable	All	Blacks	Whites	Others	All	Blacks	Whites	Others
Year 1995*Treatment	-0.008	-0.052	0.133	-0.031	-	-	-	-
	(0.037)	(0.043)	(0.084)	(0.108)				
Year 1996*Treatment	0.046	0.038	0.053	0.076	0.050^{*}	0.064^{*}	-0.013	0.092
	(0.028)	(0.036)	(0.056)	(0.087)	(0.029)	(0.033)	(0.047)	(0.079)
Year 1998*Treatment	0.032	0.015	0.011	-0.044	0.028	-0.011	0.077	-0.059
	(0.029)	(0.034)	(0.044)	(0.075)	(0.036)	(0.043)	(0.054)	(0.100)
Year 1999*Treatment	0.007	-0.028	-0.034	-0.051	-0.002	-0.081	0.099	-0.082
	(0.036)	(0.046)	(0.064)	(0.087)	(0.051)	(0.065)	(0.118)	(0.143)
Year 2000*Treatment	0.003	-0.042	0.031	-0.082	-0.010	-0.120	0.230	-0.128
	(0.043)	(0.053)	(0.088)	(0.093)	(0.074)	(0.087)	(0.166)	(0.200)
Year 1995	-0.014	-0.033^{*}	0.021	-0.166^{***}	-0.014	-0.033^{*}	0.021	-0.166^{***}
	(0.011)	(0.018)	(0.018)	(0.041)	(0.011)	(0.018)	(0.018)	(0.041)
Year 1996	-0.432^{***}	-0.429^{***}	-0.413^{***}	-0.426^{***}	-0.432^{***}	-0.429^{***}	-0.413^{***}	-0.426^{***}
	(0.009)	(0.018)	(0.013)	(0.036)	(0.009)	(0.018)	(0.013)	(0.036)
Year 1998	0.008	0.043^{**}	-0.031^{***}	0.128^{***}	0.008	0.043^{**}	-0.031^{***}	0.128^{***}
	(0.008)	(0.017)	(0.011)	(0.031)	(0.008)	(0.017)	(0.011)	(0.031)
Year 1999	0.013	0.049^{**}	-0.063^{***}	0.255^{***}	0.013	0.049^{**}	-0.063^{***}	0.255^{***}
	(0.012)	(0.021)	(0.018)	(0.039)	(0.012)	(0.021)	(0.018)	(0.039)
Year 2000	-0.003	0.041^{*}	-0.130^{***}	0.387^{***}	-0.003	0.041^{*}	-0.130^{***}	0.387^{***}
	(0.015)	(0.024)	(0.023)	(0.045)	(0.015)	(0.024)	(0.023)	(0.045)
Treatment*time	-	-	-	-	0.004	0.026	-0.066	0.015
					(0.018)	(0.021)	(0.042)	(0.054)
school dummy	Y	Y	Y	Y	Y	Y	Y	Y
N	1974	1974	1974	1974	1974	1974	1974	1974

Note: 1. Standard errors clustered at school level are presented in parentheses. 2. The regressions are unweighted. 3. The reference year is 1997. 4. The dependent variable is a logged enrollment. 5. Others represents the other racial and ethnic groups of students. 6. Enrollment variables are aggregated from the individual-level data, which include students in grades 3 through 5 only.

Table B2.The Effect on Enrollment Share, Grades 3-4 Including 1996

	Column (1)	Column (2)	Column (3)	Column (4)	Column (5)	Column (6)
Variable	Blacks	Whites	Others	Blacks	Whites	Others
Year 1995*Treatment	−0.025*	0.029**	-0.003			
	(0.014)	(0.014)	(0.007)			
Year 1996*Treatment	0.001	0.0001	-0.001	0.013*	−0.014**	0.001
	(0.010)	(0.009)	(0.006)	(0.008)	(0.007)	(0.005)
Year 1998*Treatment	-0.002	0.006	-0.004	-0.015	0.021**	-0.006
	(0.006)	(0.006)	(0.005)	(0.009)	(0.008)	(0.007)
Year 1999*Treatment	-0.014	0.010	0.004	−0.039*	0.038*	0.001
	(0.011)	(0.011)	(0.006)	(0.021)	(0.020)	(0.009)
Year 2000*Treatment	-0.022	0.024*	-0.001	−0.060**	0.066**	-0.006
	(0.014)	(0.014)	(0.007)	(0.028)	(0.028)	(0.012)
Year 1995	−0.011***	0.019***	−0.008***	−0.011***	0.019***	−0.008***
	(0.003)	(0.004)	(0.002)	(0.003)	(0.004)	(0.002)
Year 1996	−0.007**	0.007**	-0.0003	−0.007**	0.007**	- 0.0003
	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)	(0.002)
Year 1998	0.011***	−0.016***	0.005***	0.011***	−0.016***	0.005***
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)
Year 1999	0.015***	−0.029***	0.014***	0.015***	−0.029***	0.014***
	(0.004)	(0.004)	(0.002)	(0.004)	(0.004)	(0.002)
Year 2000	0.020***	−0.045***	0.025***	0.020***	−0.045***	0.025***
	(0.004)	(0.005)	(0.003)	(0.004)	(0.005)	(0.003)
Treatment*time				0.013*	−0.014**	0.002
				(0.007)	(0.007)	(0.004)
school dummy	Y	Y	Y	Y	Y	Y
N	1974	1974	1974	1974	1974	1974

Note: 1. Standard errors clustered at school level are presented in parentheses. 2. The regressions are unweighted. 3. The reference year is 1997. 4. The dependent variable is a proportion of enrollment. 5. Others represent the other racial and ethnic groups of students. 6. Enrollment variables are aggregated from the individual-level data, which include students in grades 3 through 5 only.

Bibliography

Ahn, Thomas and Jacob Vigdor (2014), The Impact of No Child Left Behind's Accountability Sanctions on School Performance: Regression Discontinuity Evidence from North Carolina, NBER Working Paper, No. 20511.

Billings, Stephen B., Eric Brunner and Stephen L. Ross (2014), The Housing and Educational Consequences of the School Choice Provisions of NCLB: Evidence from Charlotte, NC, *Working Paper*

Billings, Stephen B., David J. Deming and Jonah Rockoff (2014), School Segregation, Educational Attainment, and Crime: Evidence from the End of Busing in Charlotte-Mecklenburg, *Quarterly Journal of Economics*, 129(1), pp.435-476.

Card, David and Jesse Rothstein (2007), Racial Segregation and the Black-White Test Score Gap, *Journal of Public Economics*, 91(11-12), pp. 2158-2184.

Center on Education Policy (2008), A Call to Restructure Restructuring: Lessons from the No Child Left Behind Act in Five States, News Release, http://www.cep-dc.org/pages/News.cfm (Visited on October 27, 2016)

Chakrabarti, Rajashri (2013), Vouchers, Public School Response and the Role of Incentives: evidence from Florida, *Economic Inquiry*, 51(1), pp.500-526.

Chakrabarti, Rajashri (2014), Incentives and Responses under No Child Left Behind: Credible Threats and the Role of Competition, *Journal of Public Economics*, vol.110., pp.124-146.

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

Deming, David, Sarah Cohodes, Jennifer Jennings and Christopher Jencks (2015), School Accountability, Postsecondary Attainment and Earnings, *NBER working paper*

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

Figlio, David (2006), Testing, Crime and Punishment, *Journal of Public Economics*, 90(4-5), pp.837-851.

Figlio, David, Cecilia Elena Rouse (2006), Do Accountability and Voucher Threats Improve Low-performing Schools?, *Journal of Public Economics*,90, pp.239- 255.

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

Hanushek, Eric, John Kain and Steven 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), pp.349-383.

Hart, Cassandra and David Figlio (2015), School Accountability and School Choice: Effects on Student Selection across Schools, *National Tax Journal*, 68(3S), pp.875-900.

Jacob, Brian (2005), Accountability, Incentives and Behavior: the Impact of High-stakes Testing in the Chicago Public Schools, *Journal of Public Economics*, 89(5-6), pp.761-796.

Kane, Thomas and Douglas Staiger (2002), Volatility in School Test Scores: Implications for Test-Based Accountability Systems, *Brookings Papers on Education Policy*, No.5, pp.235-283.

Kane, Thomas, Douglas Staiger, and Jeffrey Geppert (2002), Randomly Accountable: Test Scores and Volatility, *Education Next*, 2(1), pp.56-61.

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

Krieg, John (2011), Which Students Are Left Behind? The Racial Impacts of the No Child Left Behind Act, *Economics of Education Review*, 30(4), pp.654-664.

National Center for Education Statistics (2004), Early Childhood Longitudinal Study, Kindergarten Class of 1998-99: a Manual for a Confidential-use, U.S. Department of Education.

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

Obama White House, "https://obamawhitehouse.archives.gov/issues/education/k-12/reforming-no-child-left-behind" (Visited September 12, 2018)

Reback, Randall (2008), Teaching to the Rating: School Accountability and the Distribution of Student Achievement, *Journal of Public Economics*, 92(5-6), pp.1394-1415.

Reback, Randall, Jonah Rockoff, Heather Schwartz and Elizabeth Davidson (2011), "Barnard and Columbia No Child Left Behind Database, 2002-2003 and 2003-2004,"
"http://www.gsb.columbia.edu/nclb" (Visited June 15, 2016)

Reback, Randall, Jonah Rockoff and Heather Schwartz (2014) Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind, *American Economic Journal: Economic Policy*, 6(3), pp.207-241.

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), pp.251-281.

Sims, David (2013), Can Failure Succeed? Using Racial Subgroup Rules to Analyze the Effect of School Accountability Failure on Student Performance, *Economics of Education Review*, vol.32, pp.262-274.

Springer, Matthew (2008), The Influence of an NCLB Accountability Plan on the Distribution of Student Test Score Gains, *Economics of Education Review*, 27, pp.556-563.

U.S. Department of Education (2002), The No Child Left Behind Act of 2001, Public Law 107-110, http://www2.ed.gov/policy/elsec/leg/esea02/index.html (Visited May 23, 2016)

U.S. Department of Education (2005), No Child Left Behind: A Road Map to State Implementation.

U.S. Department of Education, Every Student Succeeds Act (ESSA) “https://www.ed.gov/ESSA” (Visited September 12, 2018)

U.S. Department of Education, The Purpose of Title 1, http://www2.ed.gov/programs/titleiparta/index.html (Visited October 27, 2016)