UNIVERSITY OF CALIFORNIA
RIVERSIDE

Three Essays on the Economics of Education

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Economics

by

Yoon Jae Ro

June 2020

Dissertation Committee:

Professor Michael Bates, Co-Chairperson
Professor Anil Deolalikar, Co-Chairperson
Professor Ozkan Eren
Professor Carolyn Sloane

The Dissertation of Yoon Jae Ro is approved:

Committee Co-Chairperson

Committee Co-Chairperson

University of California, Riverside

Acknowledgments

I would like to express my sincere gratitude to my advisors for continued support during this journey. I am deeply indebted to Michael Bates, whose invaluable guidance and support made this dissertation possible. I profoundly appreciate the time and effort that he put into helping me to become a researcher. I would also like to thank Anil Deolalikar for his wisdom and advice that allowed me to broaden my perspective.

I am also very grateful to my dissertation committee members. Ozkan Eren has given me excellent and thoughtful advice on my research and presentations. I am genuinely grateful to Carolyn Solane for her wonderful feedback on my research, and for generous guidance, mentorship, and encouragement throughout the stressful and challenging phase. I have learned immensely and benefited from Joseph Cummins, Matthew Lang, Bree Lang, and other faculty members at the Department of Economics, University of California, Riverside.

I would like to send my sincere thanks to my colleagues and friends who have helped me improve my research and supported me. I received great assistance from my classmates, including Shiyun Zhang, Taghi Farzad, Ivan Sandoval; without them, this achievement would not have been possible.

I would not have come this far and pursued a Ph.D. in Economics if not for my parents and their fervent support of my ambitions. Their love and guidance are at the foundation of all of my pursuits, and I cannot thank them enough for always being there for me. Also, I am grateful to my little brother for his love and support.

To my parents for all the support.

ABSTRACT OF THE DISSERTATION

Three Essays on the Economics of Education

by

Yoon Jae Ro

Doctor of Philosophy, Graduate Program in Economics
University of California, Riverside, June 2020
Professor Michael Bates, Co-Chairperson
Professor Anil Deolalikar, Co-Chairperson

Chapter 1 examines the effects of providing teachers with performance measures on student achievement and how this policy may differentially impact students. In response to many policy initiatives, many states adopted new teacher evaluation systems comprised of multiple measures of teacher performance, including metrics based on students' performance on standardized tests: Student Growth Measures (SGMs). I construct an original information data set detailing each state's implementation policy and link it to the nation-wide data. Using the difference-in-differences and event studies, I find that releasing the SGMs to teachers negatively impacts students' math scores, and the impact becomes more prominent with time. By looking at the change in the distribution of scores, I find that this unexpected adverse effect of the policy is driven by the deterioration among previously high-performing districts and schools.

Chapter 2 examines the policy of providing Value-Added (VA) measures to teachers on student performance in Ohio and North Carolina. Using the within-state variation of the policy implementation, I find that the distribution of students' performance shifted

downward in schools with VA policy, suggesting that VA is detrimental to high-performing students. These results show that SGMs have unintended effects on student achievement, undermining students' performance at the top end of the distribution.

Chapter 3 discusses how students' exposure to drinking culture a year before reaching the legal drinking age affects their educational outcomes. I exploit a discrepancy in school cohort cutoffs in South Korea, which leads some students to be exposed to peers with the legal right to drink at an earlier age. I find that the students exposed to a peer of drinking age consume more alcohol, but this does not translate into a higher college dropout rate.

Contents

List of Figures **x**

List of Tables **xi**

1 Introduction **1**

2 The Effects of Teacher Performance Measures on Student Achievement **4**
- 2.1 Introduction . . . 4
- 2.2 Background and Data . . . 10
- 2.3 Empirical Strategy . . . 15
- 2.4 Results . . . 20
- 2.5 Conclusion . . . 23

3 The Effects of Teacher Performance Measures on Student Achievement **29**
- 3.1 Introduction . . . 29
- 3.2 Background and Data . . . 31
  - 3.2.1 Ohio . . . 31
  - 3.2.2 North Carolina . . . 34
- 3.3 Empirical Strategy . . . 36
  - 3.3.1 Ohio . . . 36
  - 3.3.2 North Carolina . . . 38
- 3.4 Results . . . 41
  - 3.4.1 Ohio . . . 41
  - 3.4.2 North Carolina . . . 42
- 3.5 Conclusion . . . 44

4 The Early Exposure to Drinking Culture and Impact on Educational Attainment **50**
- 4.1 Introduction . . . 50
- 4.2 Institutional Background: Policy Context In Korea . . . 53
- 4.3 Data and Sample . . . 54
- 4.4 Empirical Strategy . . . 56

4.5 Results 60
4.5.1 Evidence of Peer Effect 60
4.5.2 First-Stage Results 61
4.5.3 TSIV 62
4.6 Conclusion 64

5 Conclusion **69**

Reference **70**

Appendices **74**

A Appendix for Chapter 2 **75**

B Appendix for Chapter 3 **86**

List of Figures

2.1 Event Study - Math . . . 26
2.2 Event Study - Reading . . . 26

4.1 The Peer Effect on Drinking . . . 66
4.2 First-Stage Result . . . 66

A.1 Robustness Check: Sensitivity to Outliers . . . 75

B.1 Proficiency Level of Schools in Ohio . . . 87
B.2 Event Study Results for Ohio . . . 88
B.3 Math Score Trend in North Carolina . . . 89
B.4 Event Study Results for North Carolina . . . 90
B.5 Robustness Check: Randomization Test . . . 91

List of Tables

2.1 The Overall Effects of Releasing SGMs to Teachers on Student Math Scores 27
2.2 The Heterogeneous Effects of Releasing SGMs to Teachers on Student Math Scores . . . 28

3.1 The Overall Effects of Releasing VA to Teachers on Student Achievement . 46
3.2 The Distributional Effects of VA on Student Achievement . . . 47
3.3 The Heterogeneous Effects of VA on Student Achievement . . . 48
3.4 The Effect of VA on Student Achievement . . . 49

4.1 First-Stage Results . . . 67
4.2 The Effect of Teen Drinking on Educational Attainment . . . 68

A.1 State Policy Detail . . . 76
A.2 Summary Statistics of SEDA . . . 82
A.3 Event Study Model . . . 83
A.4 Event Study Model with Hetereogeneous Effects . . . 84
A.5 Robustness Check: Using Statewide Adoption Year . . . 85
A.6 Robustness Check: Using Statewide Adoption Year . . . 85

B.1 Summary Statistics: Ohio . . . 92
B.2 Summary Statistics: North Carolina . . . 93
B.3 Event Study Results for North Carolina . . . 94

Chapter 1

Introduction

Schools, teachers, and peers in our society play a crucial role in creating human capital. Thus, studying public policies that seek to enhance the impact on students human capital accumulation is essential. Three essays in this dissertation focus on the public policies that try to promote human capital accumulation: providing quality teachers to students and regulating teens access to substances. The first two chapters of this dissertation examine how teacher performance measures impact student achievement. Three different datasets and settings are used to answer the questions in these two chapters. In the last chapter, the focus is shifted to the policy that regulates teens substance use. This final chapter examines how early access to alcohol affects educational attainment.

Teacher evaluation is a constructive way to inform teachers of their performance and encourage them to improve their teaching. However, teacher evaluation systems have struggled with adequately differentiating effective teachers. Due to the many U.S. educational reforms that encourage using student growth data as a component in measuring teacher performance, many states have initiated policies to encourage their public schools to use student growth measures (SGMs) effectively.

The second and third chapters examine the effects of providing teachers with performance measures based on student achievement and how this policy may differentially impact students. Since the timing of SGM adoption in teacher evaluation is different across states and school districts, I use difference-in-differences and event studies to answer the question. In answering the question of how providing teachers, whose performance is calculated using student test scores, would be beneficial to students, I focus on the nationwide analysis in the second chapter. I construct an original information data set detailing each states implementation policy and link it to the nationwide data.

The third chapter examines the impact of this policy in two different settings and data sets to provide a complete analysis of the second chapter. I exploit within-state variation of the policy implementation in Ohio and North Carolina. In this chapter, I focus more on the heterogeneous effect of the policy as the effects of incorporating SGMs into teacher evaluation may not be uniform across students, schools, and districts. It is essential to know where contractions or expansions in the distribution of student achievement originates.

The last chapter examines the impact of regulated substances on student achievement. Teen alcohol consumption has always been a social problem, mainly because of the negative consequences of human capital accumulation. Thus, I examine the effect of alcohol consumption on attaining a college education by exploiting the quasi-experimental setting in South Korea. Two different policies–minimum legal drinking age (MLDA) and school attendance law–allow me to identify the peer effect on the decision to drink as well as the impact of drinking on educational outcomes.

Chapter 2

The Effects of Teacher Performance Measures on Student Achievement

2.1 Introduction

There is strong evidence that having a high-quality teacher is a critical factor in student achievement (Rockoff, 2004; Rivkin et al., 2005; Aaronson et al., 2007; Chetty et al., 2014). As a result, policymakers and educators have long been interested in finding accurate and efficient ways to identify effective teachers and to improve teaching quality. However, teacher evaluation systems have historically struggled to identify effective teachers. A study from the New Teacher Project (TNTP) in 2009 highlighted the discrepancy between formal teacher evaluation ratings and the true distribution of teacher effectiveness, noting that 99 percent of teachers are rated satisfactory when the districts use a binary setting (Weisberg et al., 2009). Federal interventions gave impetus to the focus on teacher evaluations. The first such intervention was the No Child Left Behind (NCLB) Act, which required states to set up a standardized assessment and to rate schools based on the proportion of students demonstrating proficiency. The federal Race to the Top (RTTT) competition and Elementary and Secondary Education Act waivers paved the way for deeper federal involvement in public education policy by creating strong incentives for states to require evidence of student learning in teacher evaluations.

In response to these incentives, many states rushed to adopt a new teacher evaluation systems comprised of multiple measures of teacher performance, including metrics based on students' performance on standardized tests: Student Growth Measures (SGMs).[1] Indeed, the use of SGMs in teacher evaluation has rapidly expanded over the past decade. In 2009, only 15 states required objective measures of student growth in teacher evaluations; this number had increased to 43 states as of 2015 (NCTQ, 2019).[2] Even among the remaining states without implementation, many large school districts have adopted the SGMs in their evaluation.

This study aims to systematically examine the nature and consequences of the inclusion of student performance data into teacher performance evaluation and how students may be differentially impacted by this policy. Although SGMs are now widely used, at times with considerable cost to states and school systems, little research has investigated

[1]There are several ways to measure student growth. Statistical methods such as student growth percentile (SGP) method, and value-added (VA) methods estimate a teacher's impact on student achievement using students' prior achievement. This paper treats these two approaches as functionally equivalent. An alternative way to measure the performance is student learning objectives (SLO) that sets a classroom-specific achievement growth targets set by individual teachers.

[2]In 2019, this number decreased to 34, as states dropped the requirement of the objective measures.

how informing teachers of their performance may affect students' outcomes across the prior distribution of district, school, or student performance. The lack of causal evidence on the impact of the adoption of SGMs on students is mainly owing to rigorous data requirements and the unavailability of a relevant policy rollout. For example, all school districts in a state often implement the new system at the same time. Moreover, until recently, there was no clear way to make comparisons across state-specific standardized tests.

This study overcomes these limitations by exploiting the differences in the timing of the adoption of SGMs across states and districts. Each state implemented SGMs in different years. For example, Pennsylvania started using value-added (VA) measures in 2006, while Florida began using the VA in 2011. I construct an original data set detailing each state's implementation policy. I gathered the information on whether the state had run a pilot program before the statewide implementation, when and how the SGMs were adopted and further included in teacher evaluation, and which type of SGMs are used. This study focuses on whether teachers received their performance measures based on SGMs rather than focusing on their inclusion in actual summative ratings on evaluation. This groundwork allows me to use the precise timing of the adoption of SGMs across states, which occurred between 2006 and 2015, in conjunction with nationwide data, to provide evidence on how the policy impacts student performance.

I use difference-in-differences (DID) and event studies to examine how monitoring teachers' contributions to student performance in standardized exams and releasing this information to teachers and administrators affects student achievement. However, the effects of incorporating SGMs into teacher evaluations may not be the same for all districts and all students. Examining only the average effect will mask the distributional effects of the policy. Thus, I examine the heterogeneity across the distribution of the previous district, school, and student performance. I start by exploring how the policy may differentially affect student achievement using the previous district and school's performance.

I examine these issues using the Stanford Education Data Archive (SEDA). The SEDA contains district-level average test scores in math and reading for 3rd to 8th graders of all states in the Unites States from the academic years 2005-2006 to 2014-2015. To examine the policys impact on student performance, I link the SEDA to the information data set, which includes the implementation details and the list of pilot districts, and exploit the different timings of districts adopting the SGMs between 2006 and 2015. Using the SEDA instead of state-specific data enables me to rely on consistent measures of student achievement that are more nationally representative.

The provision of teacher-level effectiveness measures, such as SGMs, introduces different incentives that may have a differential impact on students. First, SGMs may enhance student learning by providing the basis for more rigorous evaluation systems. Principal-agent theory suggests that if a supervisor monitors employee, an agent's work effort will increase. If schools are treated as firms, the education authorities want to implement a system that is designed to induce more effort from teachers. This claim is well supported in the literature as even subjective evaluations appear to improve teacher performance and student achievement (Taylor and Tyler, 2012; Jacob and Lefgren, 2008). Thus, as the stated aim of providing SGMs is to give valuable feedback to teachers and to improve their instruction quality, it is reasonable to expect that the policy could generate a positive impact on average students' performance.

However, it is possible that such expectations will not be met. There could be unintended consequences for students as teachers strategically react to the policy. For example, as the SGMs are necessarily tied to students' test results, teachers may redirect their effort level depending on the students' initial ability. Teachers may focus less on high-achieving students as there is less scope for achieving higher growth in their test scores. Neal and Schanzenbach (2010) found that after the NCLB, teachers would focus on the students who are at the margin of passing, rather than students who were already proficient or those far from becoming proficient. In addition, providing the VA measures to teachers increases the mobility of highly effective teachers, especially to high-performing schools, which could lead to a positive or negative impact on students. Thus, it is valuable to investigate how providing information about the effectiveness of individual teachers ultimately affects students' performance.

The results indicate that providing teachers with the performance score based on SGMs negatively impacts average student performance in math, and this is driven by the deterioration in the performance of students at the top. Students in school districts in which teachers could access their performance information performed significantly lower in math tests than those in control districts by 0.016 SD. Looking at the time-varying effects, by the fourth year of the implementation of the policy, the students' test scores decreased by 0.083 SD. This negative effect is due to the deterioration in students' performance from the previously high-performing school districts.

This study contributes to the existing literature by providing policy-relevant information on the impact of teacher accountability systems on student achievement. First, to the best of my knowledge, this is the first study to use nationally representative data to examine the causal effects of the adoption of SGMs on student performance. Many previous studies focus on the impact of the school accountability system, such as NCLB, on student performance (Dee and Jacob, 2011; Reback et al., 2014; Ladd, 2012). While there was an effort to isolate the causal effects of NCLB by using the comparison between own accountability systems and the national program, the overall test score effects of NCLB are inconclusive. For example, Dee and Jacob (2011) found that the NCLB led to an increase in math scores for 4th-grade students while Reback et al. (2014) and Ladd (2012) found that the significance disappeared as they manipulated the sample years. Neal and Schanzenbach (2010) revealed that there is an overall gain in student achievement while demonstrating that this effect originates from students close to the proficiency margin. This study provides the first causal evidence of the impact of providing teacher-level performance measures of students by exploiting regional variation and nationwide data.

Second, my study is closely related to several studies that focus on how information and evaluation influence the performance of teachers and, consequently, students. Bergman and Hill (2018) and Pope (2019) examined the effect of LA Times ratings on teachers. Both studies find that teachers with low ratings improved their performance when informed of their rating based on VA scores. However, the first study found that the public rating did not affect students overall test scores, while the latter found a beneficial impact. It should be noted that positive student and teacher sorting drive these mixed results of the effects.

Unlike these studies, I focus on the teacher-level VA information given to teachers rather than to parents.

2.2 Background and Data

The quality of a teacher is one of the most important factors found to promote students' immediate learning and even their long-term outcomes, such as job earnings (Rockoff, 2004; Rivkin et al., 2005; Aaronson et al., 2007; Chetty et al., 2014). Also, there is a substantial variation in teacher quality in raising students' achievement on standardized tests (Rivkin et al., 2005). As a result, policymakers and educators have long been interested in finding accurate and efficient ways to identify effective teachers and to improve teaching quality. In theory, teacher evaluations are used to inform teachers of their performance and guide their professional development. However, teacher evaluation systems relied heavily on subjective measures and failed to differentiate the heterogeneity in teachers' effectiveness. Recognizing the teachers' role in education function, and failure of the teacher evaluation system, the education policy in the U.S. moved toward focusing on how to discern the quality teachers effectively.

Moreover, federal government intervention encouraged the states to focus on developing more rigorous evaluation systems. After the NCLB laid the foundation for states to reform teacher evaluation systems, the federal Race to the Top competition created strong incentives for states to make specific changes. Among the directed changes was the requirement to develop the high-stakes system comprised of multiple measures of teacher performance, including metrics based on students' performance on standardized tests and increasing the frequency of the evaluation.

Driven by those incentives and growing recognition of the importance of considering teachers' contributions to students, states rushed to adopt the new system with student growth data. In 2009, only 15 states required objective measures of student growth in teacher evaluations; by 2015, this number increased to 43 (NCTQ, 2019). Among the many methods used for teacher effectiveness measurement, Value-Added (VA) models, and Student Growth Percentile (SGP) models are widely used. While conceptually similar, the two models differ in the estimation method. The VA models compare a student's predicted performance to the average performance of a given teacher's students. The SGP models compare students' progress to that of other students with similar past performance.[3] In this analysis, I do not discern between the VA measures and the SGP since both use the student test scores in measuring teacher effectiveness.

Yet, there has been some push to delay linking teacher evaluations with test scores. Many states that passed such legislation faced delay due to many reasons, including push-back from teacher unions, and technical difficulty in developing the new rubrics. Thus, the actual implementation date can be different from the time of passage of the law. Also, some states even dropped the requirement of objective measures of student growth since 2015. Thus, currently, as of 2019, 34 states require teacher evaluations to include objective measures of student growth, down from a high of 43 in 2015(NCTQ, 2019).

To confirm the exact policy implementation year and its detail, I compiled data on the state's teacher evaluation system by a systematic search and outreach process. The

[3]In 2015, 15 states used VA measures, and 19 states use SGP models.

details of this policy are shown in Table A.1 which presents the year of actual implementation for each state as well as the set of states without such policy.[4] I began by reviewing the State Teacher Policy Database from the National Council on Teacher Quality that contains the detail information on state laws, rules, and regulations of the teaching profession. I then reviewed information on state education agency websites to verify policy implementation details. In addition, I searched for research papers, reports, and news articles to research whether the implementation details are different from the passed legislation. Lastly, for the states where I couldn't find reliable information, I directly contacted the agency to request such information. My rigorous search produced data on the detailed information of a complete set of teacher evaluation systems of 50 states with great attention to student data use. This information is crucial in conducting this research because omitting this could bias the results.

I want to point out several finding that this research produced. First, the research revealed that many states stalled the implementation of the new teacher evaluation systems. This means that even after the passage of legislation to link student performance on standardized test scores and teacher performance evaluation, the actual use of these measures was put on hold for some states. Second, many states ran a pilot of a new teacher evaluation system before the statewide implementation. Even though the pilot does not necessarily use the student growth measure in calculating the summative rating, many pilot states provide the measure to teachers. Third, there is vast heterogeneity in types, and the percentage of SGMs are used.

[4]Note that Washington, DC is excluded both from Table A.1 and from this analysis.

I link this information data set to the SEDA to conduct nationwide analysis. The SEDA includes a range of detailed data on educational outcomes in school districts and counties across the United States. Mainly, SEDA contains the district-level average test score in math and reading of 3rd to 8th graders of all states in the U.S. from 2008-2009 to the 2014-2015 school year. The most useful feature of the SEDA comes from providing the test scores in common metrics across states and districts, which allows the researcher to compare the test scores across the state, district, and year. In addition to the test score, the data includes a rich set of district characteristics such as gender and ethnicity/race composition, and percentage of students who have English Learner status, Special Education status, and Free Lunch status.

In this paper, I refer to state or district adopting policy when teachers start to receive their performance measures that use SGMs. Many states have started distributing the performance measures that are tied to the students' test score growth before incorporating those measures in their final ratings. I focus on the provision of the performance information to teachers rather than the summative ratings based on those measures. Also, I only consider the treatment status when a state or a district adopted the policy by 2015. There are two reasons for this. First, the SEDA contains the test scores only till the 2014-2015 school year and allows me only to examine the policy until these years. More importantly, many states repealed the use of SGMs in their teacher evaluation after 2015.

Although I focus on releasing the information on teacher effectiveness to teachers, some states use those performance measures in the final evaluation rating, and further tie it to the reward and sanction. The most common way to reward teachers is by performance pay for teachers. Teachers' effort level can change with the monetary incentive. To control this, I link the information regarding the rewards for teachers from the Schools and Staffing Survey (SASS). The SASS is conducted by the Department of Education every few years and surveys a stratified random sample of teachers who provide information on their background, compensation, attitudes, school activities, and teaching methods. I am particularly interested in one variable the data contains: how much percentage of school districts indicated that they used financial incentives to reward excellence in teaching. Unfortunately, SASS is available in three waves: 2003,2007 and 2011. Thus, I use 2003 as a baseline and use the other two years to control how much growth in this percentage has happened in the states.

Descriptive statics for the sample by policy adoption status are reported in Table A.2. Data includes 11994 school districts from 50 states, which results in 330443 of observation. Note that none of the non-adopted states have won the Race to the Top, and districts in adopted states are more likely to use teacher compensation than non-adopted states. One caveat of SEDA is that test scores for the grade-subject pairs are missing. For example, Arizona's 8th-grade math scores are not reported in 2009, 2010, and 2015 due to the technical issues. However, as I am using district-level average test scores across grades for each subject, the bias coming from the missingness is minimal. Still, it would be a concern if this is a non-random missing for certain states. To check whether the states missing test scores for some grades is causing bias, I check the sensitivity of the coefficient estimates as I drop each grade from 4 to 8. The results are robust to this sensitivity test.

2.3 Empirical Strategy

As stated, this paper aims to evaluate the impact of releasing student growth measures on student performance by exploiting the variation across states, districts, and schools. All analyses mainly use the differences-in-difference (DID) method utilizing the different timing of adoption with different data sources. In this section, I describe the empirical strategies for each data set.

Mainly, I exploit within-state (or within-district), and cross-cohort differences in exposure to the policy driven by cross-state (or cross-district) variation in the timing of when or whether states (or districts) adopted the policy in a difference-in-difference framework. This involves comparing the differences in average student outcomes before and after the adoption of the individual teacher student-outcome based performance measures within states (or districts) that adopt the policy against changes over the same time frame in states (or districts) that do not adopt the policy. I estimate:

$$Y_{ist} = \beta_o + \alpha_i + \lambda_t + \theta_g + \beta_1 D_{ist} + X'_{ist}\gamma + \epsilon_{ist} \tag{2.1}$$

I also consider the district-level variation within state. Since some districts adopt the objective teacher evaluations ahead of the rest of the state, considering this variation gives more power to identify the causal effects. Y_{ist} is the mean student achievement in district i in state s in year t. In the SEDA, the student achievement is given as district average by each grade and subject. D_{ist} is an indicator for whether a district currently release the student growth measures to teachers. The district fixed effects control for variation in outcomes that are common across students within a district, and the year fixed effects account for national shocks that impact all students in the same year. I also control for the proportion of students of ethnicity/race (Black, Asian, Hispanic), free lunch status, English limited status in each district. These controls are in the vector X in equation (2.1). All standard errors are clustered at the district level as the level of treatment assignment (Abadie et al., 2017).

Adopting the individual teacher effectiveness measure can change teachers' effort as teachers become more aware of the policy over time. This can generate a time-varying treatment effect based on the length of exposure to the policy. The source of variation comes from the fact that each state adopted the student growth measure in different years. Thus, I employ an event study model that examines how outcomes changed among students who were differentially exposed to the policy that had been in place for different lengths of time based on which state, district and in which grade they were in.

Equally important, this event study model allows me to inspect the evidence of the key difference-in-differences assumption. Conditional on the controls in the model, the variation in policy exposure comes from two sources. The first is within-state (or district) differences in exposure over time driven by the year of the policy adoption. The second is cross-state (or district) variation in the timing of when or whether states adopted the policy. The assumption underlying the identification of parameters is that the policy should not be endogenous to unobserved state-level shocks. That is, the decision of whether and when to adopt the policy must be uncorrelated with any prior trends in outcomes. For example, if a state adopted policy after having a negative trend in student test scores, the policy estimates would spuriously capture the positive impact even if the policy did not have causal impact on students. I estimate an event study model as follows:

$$Y_{ist} = \beta_o + \alpha_i + \lambda_t + \theta_g + \sum_{\tau=-k}^{K} \beta_\tau D_{ist}^{\tau} + X_{ist}'\gamma + \epsilon_{ist} \tag{2.2}$$

The variables used are same as previous equation except that the DID estimator is replaced with the event study indicators. The variable $D_{ist}^{\tau} = I\left(t - t_{0i} = \tau\right)$ is an indicator equal to one for being τ time periods relative to i's initial treatment (t_{0i}) with τ ranges from -7 to 7 and $\tau = 0$ being the year of initial treatment. For example, if a district adopted the VA measure in year 2012, it will have a relative time of -1 for the year 2011 and 1 for the year 2013. This variable takes value of zero in states that have never been had a VA measure adoption. β_τ can be interpreted as estimates for pre-trends (for $\tau \leq 0$) as well as time-varying treatment effects (for $\tau > 0$). I omit the estimate for the $D_{ist}^{0} = I\left(\tau = 0\right)$ such that all β_τ estimates are relative to the year of adoption. Equation (2.2) also includes grade (θ_g), district (α_i), and year (λ_t) fixed effects. Standard errors are clustered at district level. The parameters of interest in equation (2) are β_1 to β_7, which show the time-varying effects of the policy among students who are first exposed to this policy in relative years 1 to 7. Also, the β_{-7} to β_{-1} estimates in equation (2.2) serves as a test of the assumption that there is no selection.

I also conduct several robustness checks. First, the existence of alternative policies that were implemented concurrently with the adoption of student growth measures can be a threat to the identification. One policy that can directly affect teachers' behavior and thus have impact on student outcome is Teacher Incentive Pay. Financial incentives may have positive impact on students' achievement by improving teacher's effort level. However, it can have no impact if teachers were teaching at their highest effort level, or not knowing how to increase student achievement and a negative impact if teachers cheat. I control this by using the Schools and Staffing Survey (SASS). I include an additional control variable that indicates the percentage of teachers receiving performance pay tied to the student test scores. I am also including the indicator of the Race to the Top winner states where 18 states and D.C. won awards that ranged from \$17 million to \$700 million. Race to the Top promotes states to develop and adopt standard assessment system with a statewide longitudinal data, evaluation system of teachers and principal based on performance. Since the Race to the Top encourages states to have performance based and standardized assessment system, winning the award can show the states' interest in improving their student achievement. Indeed, winners implemented more education related policies.

In addition, I examine the sensitivity of the results to outliers: whether a particular state is driving the effect. I estimate equation (2.1) 50 times, each time excluding a different state from the sample. Lastly, I use the state-level adoption years instead of using the district-level variation. The impact would be different to teachers when the state officially adopts the student growth measures instead of the local education agency adopting the policy. I provide the results of all robustness checks in the next section, along with the results from the main analyses.

The effects of providing student growth measures to teachers may not be the same for all students. In order to examine whether the differential impacts on student performance, I relax the linearity assumption in equation (2.1) and (2.2) and consider the heterogeneous achievement of districts. I explore whether the districts have differentially impacted by the policy depending on their initial performance. The policy may differentially affect the schools that were previously high or low performing. Teachers in different schools might exert different levels of effort to improve their students' performance. The teachers in high performing schools were already having a student with relatively higher test scores where the additional growth coming from students might be small. This can make teachers put less effort, which led to the null effect of the policy. In contrast, low-performing schools can be experiencing higher growth in students' test scores since most of the students have a large room to increase. I estimate the following specification:

$$Y_{ist} = \beta_o + \alpha_i + \lambda_t + \theta_g + \beta_1 D_{ist} + \beta_2 D_{ist} \times I(H_{ist-1}) + X'_{ist}\gamma + \epsilon_{ist} \tag{2.3}$$

Where $I(H_{ist-1})$ is indicator of initial district achievement; and all other variables are defined as in equation (2.1). Indicator variable of previously high-achieving districts uses districts' performance of one year before the policy adoption year. The parameter of interest is β_2 which measures the differential impacts of the VA policy on students' test scores of initially high-performing districts relative to that of low-performing districts. In addition to the heterogeneity analysis using difference-in-differences method, I also present the results using the event study models. These are estimated using equation (2.3) except $I(H_{ist-1})$ is interacted with the event study indicators.

2.4 Results

This section describes the results from all the analyses, including the test for the identification strategy, main results, and robustness checks. Figure 2.1 and Figure 2.2 show the full set of estimates using the event study model of equation (2.1) with including all control variables. I also overlay a linear fit for the pre- and post-treatment periods to see if there are differential pre-trends and if there are time-varying treatment effects. The visual evidence in Figure 2.1 supports the identification strategy: there is no evidence of differential trends in test scores in pre-treatment periods in math. The point estimates of pre-treatment periods are small and insignificant. As the school districts adopt the SGMs, the effect on math scores is small and remains unchanged for the first three years. From the fourth year of adoption, the math scores decline as a function of exposure time. However, the presence of pre-trends is detected in Figure 2.2. Reading scores linearly declines as the exposure time to the policy while exhibiting the significant positive pre-trends in reading scores. It can be interpreted as the policy change appears to have an effect on the outcome before it is implemented. Thus, I only report the results for math scores in nationwide analysis.

The estimates shown in Table 2.1 confirms the adverse effects on student achievement in math shown in Figure 2.1. The estimates in column (2.3) that includes control variables as well as alternative policy controls indicate that there is negative impact on students with the adoption SGMs. Attending the schools in a district with SGMs decrease students' math score by 0.016 SD. The different timing of policy adoption can generate a time-varying treatment effects based on the length of the exposure to the policy.

Table A.3 reports the estimates from event study model. The results indicate that there is no immediate effect on the students as the estimates are insignificant in the early years of policy adoption. The effect grows and becomes significant, beginning four years of adoption. These results indicate that attending school in a state with this policy reduces students' math achievement by approximately 0.083 SD after four years of adoption. The effect grows to 0.175 SD after seven years of adoption.

These results are somewhat striking since it shows the opposite results from the previous literature which estimates the effects of teacher evaluation on student performance. For example, two pieces of evidence from Cincinnati Public Schools show that subjective measures of teacher effectiveness promote student achievement (Kane et al., 2011; Taylor and Tyler, 2012). More recent evidence from Chicago public schools shows that schools that participated in a teacher evaluation program designed to improve classroom instruction through structured principal-teacher dialogue performed better in reading (Steinberg and Sartain, 2015). One reason for these contrasting results is that teachers with high performance scores may reduce their effort in teaching, and thus their influence on student achievement gains decreases. Also, teachers' strategic reaction to the policy may lag since it requires time for teachers to understand the policy better. This can explain maybe the fact that there are no effects on test scores by the third year of implementation but starts declining from the 4th year of the implementation. After teachers learn about their productivity, teachers may reduce effort, and this translates into a negative impact on students' performance.

While I cannot directly examine the change in the effort level of the teachers, I investigate the possible mechanism through examining the heterogeneous effects of the policy: how the policy differentially affects the school districts depending on their previous performance. Column (2) in Table 2.2 presents the estimates of heterogeneous treatment effects from equation (2.3). The students in relatively high-performing school districts experience a stronger negative impact on their math scores. The policy reduces math scores by 0.0461 SD in high-performing school districts relative to the low-performing school districts. This heterogeneous effect of policy on student performance further confirmed by the event study model, where the same specification is used except the indicator of high-performing districts is interacted with a series of relative time indicators in equation (2.2). Table A.4 presents the full set of estimates from a single estimation equation. The estimates in column (4) represent the time-varying effects of the policy on high-performing districts relative to low-performing districts. Notably, there is an immediate negative effect on students' math scores for the high-performing school districts. Attending school districts in states with the SGMs reduces the students' math score by 0.038 SD in the first year. The effect grows to 0.115 SD three years after the policy adoption. However, the negative impact on student becomes small and insignificant from year 5, and it becomes positive in year 7.

The results are robust to a range of alternative specifications. First, the negative effect on the student test score could be spurious if one or many early-adopted states are driving the results. Thus, I examine the sensitivity of my findings to outliers by reestimating equation (2.1) 50 times for the outcome, each time dropping a different state from the analysis sample. The results of this exercise are reported in Figure A.1. The estimates are insensitive to excluding each one state. Second, to see whether early-adopted districts within the states are driving the result, I estimate equation (2.1) by using the state adoption year instead of the district's adoption year. Both estimations of overall and heterogeneous effects are similar to the main analysis. The results of using state-level adoption year of SGMs are presented in Table A.5 and Table A.6.

2.5 Conclusion

Many states in the U.S. initiated policies to encourage their public schools to effectively use any form that incorporates student growth data calculated from test scores. With the federal intervention NCLB and Race to the Top program, states rushed to use student growth measures as one component that measures teacher performance. The use of SGMs in teacher evaluation rapidly expanded over the past decade: from 15 states in 2009 to 43 states in 2015. Despite the importance of this policy, there have been very few opportunities to evaluate the impact of this policy. The rigorous data requirements and the unavailability of relevant policy rollout are the main reasons. This paper overcomes the limitation by exploiting different timing in the adoption of SGMs with three different settings and data and examines how teachers receiving their performance measures calculated from student test score growth affects student achievement and how students may be differentially impacted by this policy.

I first examine the impact of providing objective performance measures to teachers on overall student achievement. Using variation in the timing of policy adoption across states and districts, I find that providing SGMs to teachers leads to a decrease in average math scores. That is, attending schools in a district with SGMs decrease students' math score by 0.016 SD. Exploring the time-varying treatment effects of the policy, I find that attending school in a district with VA policy reduces students' math achievement by approximately 0.083 SD after four years of adoption, and it grows to 0.175 SD after seven years of adoption.

These results are surprising since the aim of providing objective measures of performance is encouraging teachers to put more effort into improving their performance, which could lead to an increase in student achievement. One possible explanation of opposite empirical findings from the expectation is that teachers may strategically react to the policy by focusing on students with certain initial academic achievement.

It also implies that the impact of this policy may not be the same for all districts and all students. Thus, I examine the heterogeneity across the distribution of prior district. Investigating the heterogeneous effects of policy across the distribution of districts, I find that previously high-achieving districts suffer from a bigger negative impact on math scores. These adverse effects are larger for the upper end of the distribution of district performance.

These results show that SGMs have an unwanted effect on student achievement, undermining the performance of students in a top portion of the distribution. The contraction of the distribution of student achievement suggests that policy may alter the equity of education. Providing VA to teachers make it easier for teachers to move to other schools since the VA serves as a signal of productivity of teachers. Indeed, Bates (2020) showed that highly effective teachers move to higher-performing schools. When VA is provided, increased teacher mobility and generated transition costs from these movements can harm student achievement and even exacerbate the gaps between districts (Boyd et al., 2008). However, if the VA measures are targeted for improving the teaching quality and effort level of the low effective teachers rather than rewarding the highly effective teachers, the discrepancy in added effort level with VA measures can cause the contraction of distribution of student achievement.

Figures and Tables

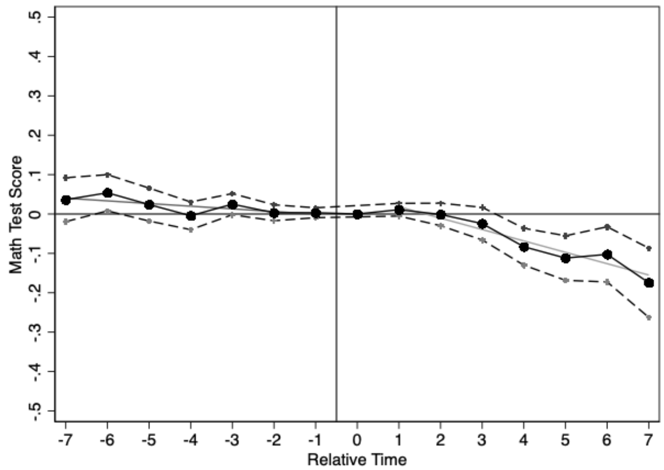


Figure 2.1: Event Study - Math

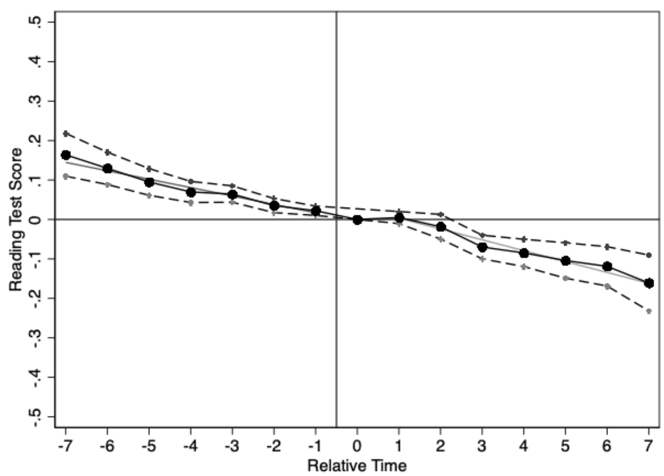


Figure 2.2: Event Study - Reading

Table 2.1: The Overall Effects of Releasing SGMs to Teachers on Student Math Scores

	(1)	(2)	(3)
Relative Years to Policy Adoption	-0.0092**	-0.0059	-0.0063
	(0.0039)	(0.0041)	(0.0042)
Treated	0.0264***	0.0261***	0.0257***
	(0.0088)	(0.0083)	(0.0084)
Relative Years * Treated	-0.0088	-0.0159***	-0.0161***
	(0.0054)	(0.0052)	(0.0053)
Control variables	N	Y	Y
Policy controls	N	N	Y
Grade FE	Y	Y	Y
Year FE	Y	Y	Y
District FE	Y	Y	Y
R-squared	0.8634	0.8644	0.8644
N	328,215	328,215	328,215

* Note: The table presents the estimates from equation (2.1) using the SEDA. Grade, Year, District FE are included in all specifications. Column (2) uses the control variables including student charactriscits of the districts such as gender, race/ethnicity, special education status, limited english learners, and free lunch status. Column (3) adds the policy control such as RTTT winner states and percentage of teachers receiving performance pay in the districts. Standard errors clustered at district level are shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Table 2.2: The Heterogeneous Effects of Releasing SGMs to Teachers on Student Math Scores

	(1)	(2)	(3)
DID	-0.0161***	0.0178	-0.0283***
	(0.0053)	(0.0116)	(0.0050)
DID*High		-0.0461***	
		(0.0123)	
DID*Low			0.0461***
			(0.0123)
Control variables	Y	Y	Y
Policy controls	Y	Y	Y
Grade FE	Y	Y	Y
Year FE	Y	Y	Y
District FE	Y	Y	Y
R-squared	0.8644	0.8660	0.8660
N	328,215	328,215	328,215

[*] Note: The table presents the estimates from equation (2.3) using the SEDA. Grade, Year, District FE as well as control variables are included in all specifications. Column (1) shows the estimates of high-performing districts and Column (2) shows the estimates of low-performing districts. Standard errors clustered at district level are shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Chapter 3

The Effects of Teacher Performance Measures on Student Achievement

3.1 Introduction

Teacher evaluation is one of the constructive ways to inform teachers of their performance and encourage them to improve their teaching. However, teacher evaluation systems had struggled in adequately differentiating effective teachers. Due to the federal Race to the Top program, which encourages the use of student growth data as one component that measures teacher performance, many states in the U.S. have initiated policies to encourage their public schools to use value-added (VA) measures effectively.

This paper examines how the adoption of teacher VA impacts student performance using the data from two states. First, I exploit a variation introduced by the Ohio Teacher Evaluation System (OTES), which piloted the incorporation of student growth measures as a part of teacher evaluations. The pilot involved 139 of 611 school districts and ran during the 2011 and 2012 school years before being incorporated statewide in 2013. Second, I exploit a variation introduced in North Carolina, where two large school districts adopted VA measures before the statewide adoption.

The purpose of the present research is to systematically examine the nature and consequences of the VA measure. The stated aim of distributing VA information to teachers is that it may motivate teachers to improve their performance, thereby improving student learning. It is imperative to investigate the effect of this policy on student achievement gain. However, the effects of incorporating student growth into teacher evaluations may not be the same for all students. These implications may differ depending on where students stand regarding their initial test scores. For example, it would be valuable to know if the new evaluation system increases test scores at the bottom of the distribution. Therefore, it is important to evaluate not only the effects on the mean of student achievement but also the effects of the policy across the distribution of students.

In order to evaluate the average effects of teacher VA adoption on students' performance, I employ a differences-in-differences (DID) model that compares student achievement in treatment schools relative to control schools. The 4th- to 8th-grade teachers in treated schools received the VA information while teachers in control schools did not.

In Ohio, I find no evidence of the policy increasing student achievement in various outcomes that demonstrate student learning. Overall, I find a negative effect on school-level proficiency in mathematics, although it is not statistically significant, and a null result in reading. In order to gain more insight into the overall effect, I examine how the policy affected the proportion of students in each proficiency level category. The schools report the number of students in ordered category levels: advanced, accelerated, proficient, basic, and limited. Employing the same identification strategy mentioned above, I find a significant shift in the share of students at the more advanced levels towards the proficiency margin for both reading and mathematics. This result shows the evidence of student distribution is being shifted downward or compressed to the mean.

In North Carolina, I find that the VA policy is detrimental to the top half of the distributin, even with a positive impact on the average test scores. However, these results should be interpreted with caution since the statistical inference might be biased.

3.2 Background and Data

3.2.1 Ohio

Ohio is one of the states that started a new teacher evaluation system, under the Race to the Top initiative, called the Ohio Teacher Evaluation System (OTES). The significant alteration in teacher evaluation in Ohio was shifting the teacher evaluation process from looking at how teachers do to what students learn in the classroom. Previously the teacher evaluation focused on the component where the supervisors conduct formal (and informal) observations on teacher performance in the classroom. With the introduction of

OTES, teachers are now partially evaluated based on student performance measured by student growth.

The OTES is comprised of two categories of measurements: Teacher Performance on Standards and VA score. Each category makes up 50% of the teacher's Final Summative Rating. The Teacher Performance on Standards is an assessment by administrators through classroom observation. The VA score is provided to the teachers by SAS through their Education Value-Added Assessment System (henceforth EVAAS). Until the 2013-2014 school year, EVAAS calculates teacher-level VA based on Ohio Achievement Assessments (OAA) results for grades 4 through 8 in reading and mathematics. For teachers whose VA is not available (for example, for teachers not teaching reading or mathematics for grade 4-8, or simply data not available), the school districts may use other assessments provided by national testing vendors and approved for use in Ohio.

This paper explores a variation introduced by a pilot program where 139 of 611 school districts participated during the 2011-2012 school year before the statewide implementation of OTES in 2013. The purpose of the pilot was mainly to inform the teachers and principals of the new components of OTES. Each pilot district could choose one of four approved teacher evaluation models they planned to implement. The four models either implemented OTES or developed the evaluation system to align to the OTES, which both either include a student growth measure or not. However, regardless of the evaluation models that they choose to pilot, the major part of the pilot was providing teacher-level VA (before only providing school-level VA was provided). Teachers teaching grades 4 to 8 in either Reading or Mathematics had access to their VA information during the pilot year. The information on the pilot is gathered through personnel email correspondence with the Ohio Department of Education. Through this process, a list of pilot participants was provided, and I confirmed that the teachers in pilot schools actually received the VA score. It is reasonable to use the pilot program as the policy variation the same as previous analysis, and I am focusing on the provision of the VA scores to teachers.

To evaluate the VA policy impact on student performance in Ohio, I mainly use the school report cards that provide the public records of each school's and district's performance information. From the publicly available data in the Ohio Department of Education (ODE) websites, I compiled a school-level panel from the 2005-2006 school year through the 2016-2017 school year. The data contains various measures of student performance as well as school characteristics. The primary outcome I use in the analysis is the school-level performance of students in math and reading. For each grade and subject, the percentage of students who are proficient in each school is provided. This proficiency level can be further broken down into five ordered categories: advanced, accelerated, proficient, basic, and limited. This last feature of the data makes it possible to examine the distributional impact of the policy by exploring the within-school variation.

To avoid any bias in estimates confounded by the effect of schools closing or opening, I require schools to have the proficiency level reported throughout the entire sample year. If a school has been closed due to their poor performance, including it in the analysis would bias the coefficient of interest. Thus, I limit the sample to those who have such information on each year of the sample period.[1] In the end, there are 597 pilot schools out of

[1] I exclude charter schools from the analysis since charter schools operate teacher evaluations differently than public schools.

the total of 2108 schools in the analysis sample. The data also includes student characteristics of schools such as the composition of students' gender and race by each grade, as well as the percentage of limited English proficiency students and economically disadvantaged students.

Summary statistics of certain key variables for the pilot and the control school in Ohio samples are shown in Table B.1. The table presents mean pre-treatment school-level characteristics for the pilot schools and the rest of the schools in Ohio. On average, pilot schools have a higher proportion of black students, a higher percentage of students with limited English proficiency, and economically disadvantaged students than the rest of Ohio. However, this will not be a problem in implementing the empirical method I use in the analysis as the difference between the two groups stays stable throughout the year.

3.2.2 North Carolina

Before the statewide implementation of individual teacher VA into the teacher evaluation system in 2013, two school districts started releasing VA information to teachers and principals: Guilford in 2000 and Winston-Salem in 2008. Same as the case from Ohio, teacher effectiveness of each teacher in this report is estimated using the EVAAS. SAS calculates the VA measure for each given academic year, and this information is presented in the EVAAS teacher report.

North Carolina requires end-of-grade (EOG) assessments for math and reading for 3rd to 8th graders. The VA information is estimated for teachers who teach subjects and grades that require EOG assessments. Once the VA score is estimated using the student test scores, teachers and principals can access the VA information. Also, the EVAAS teacher report presents the teacher effectiveness across the student performance distribution. Teachers having access to this information might give incentives for teachers to strategically react to the VA score disclosure by focusing their efforts on students at different levels of initial ability. Thus, I exploit two school districts' adoption of individual VA measures ahead of the state adoption to examine the differential effects on students.

I use the student records covering the period from 1997-2011 of North Carolina Education Research Data Center (NCERDC). The data from the NCERDC contains the EOG math and reading test scores of 3rd to 8th graders and a rich set of students, school, teacher characteristics. The student characteristics include grade, gender, race, and exceptional status, including the academically gifted.

One main advantage of using the data from the NCERDC is the ability to define students in the same classroom. The primary objective of this part is to determine whether and how differentially students' learning is influenced when a teacher receives VA information. Thus, it is important to identify the students that correspond to the test scores form each classroom. The data from NCERDC contains the identifiers that attempt to link students in the same classroom, which allows me to identify a student's performance relative to the entire test score distribution in that classroom. I used the student's first attempt test score, thus excluded the re-attempt record for the analysis. I also excluded charters schools from the sample. In the end, there are 116 school districts in the sample.

Summary statistics of certain key variables for North Carolina, including Guilford and Winston-Salem, are shown in Table B.2. Table B.2 compares the means and standard deviation of Guildford (Winston-Salem) and the rest of the districts for the pre-treatment periods, respectively. The two districts that adopt VA do have a higher percentage of black students than the rest of the state. However, the averages in achievement in these districts are not different from the state average.

3.3 Empirical Strategy

3.3.1 Ohio

I examine the effects of policy in more detail by exploiting the cross-school district variation in the adoption of VA measures in Ohio. As described earlier, Ohio ran a pilot of adopting VA measures before the statewide implementation. The 4th- to 8th-grade teachers in pilot schools received the VA information while teachers in control schools did not. I again use difference-in-differences approach to compare the student performance in the treatment schools relative to the control schools after the policy was implemented. In order for this approach to be valid, the policy should not be endogenous to unobserved district-level shocks. In addition to the graphical evidence of the parallel trend in pre-pilot periods in Figure B.1, I estimate an event study specification that allows for a complete set of interactions between the indicator of treatment status and years. The result is represented in Figure B.2 which confirms the assumptions in using the specification, I proceed to the main empirical model as follows:

$$Y_{ist} = \beta_o + \alpha_i + \lambda_t + \theta_g + \beta_1 D_{is,1t} + \beta_2 D_{is,2t} + X'_{ist}\gamma + \epsilon_{ist} \tag{3.1}$$

Since the teachers teaching in grades 4 to 8 received the VA score, analysis is limited to 4th to 8th grade students. In order to accommodate the two-step roll-out of the policy including pilot and statewide implementation, there are two dummy indicators in the equation. $D_{is,1t}$ is a dummy variable that takes the value one for the treated schools after the OTES pilot in year 2011-2012, while $D_{is,2t}$ equals one for the treated schools after the OTES statewide implementation in year 2013-2014. Since the treated schools here are defined for the pilot participant schools, the initial pilot participants remain the same even after the statewide implementation. Thus, β_1 represents the immediate effect of the pilot and β_2 represents the long-term effect of the pilot. X_{ist} is the vector of student characteristics of schools that includes the share of limited English learners, economically disadvantaged students, gender, and race/ethnicity. In addition, the model includes a set of grade (θ_g), school (α_i), and year(λ_t) fixed effects. I provide the standard errors clustered at the district level as the pilot was assigned at district level.

Y_{ist} is the student performance variable of school i in district s and in year t. There are two sets of outcomes that I use in the analysis. First, to see the overall effect of the policy on student performance, I use the percentage of students that are proficient for each school, which I retrieved from the Ohio Report Card. Using this outcome allows me to examine the overall policy impact on student outcome. Second, to gain more insight into the overall effect, I examine how the policy affected the proportion of students in each performance level category. The schools report the number of students in ordered category levels: advanced, accelerated, proficient, basic, and limited. Examining only the change in percentage of students who pass the proficiency level of performance can mask the distributional effects of the policy. Thus, I estimate the same DID model (equation (3.1)) for five different outcomes, respectively: the percentage of students in each ordered category. Since I am estimating five separate regressions with the outcomes that are correlated with each other, the equation errors would be correlated. That is, the set of five equations has contemporaneous cross-equation error correlation since each proficiency category is summed to one and change in one part is accompanied by a change in other parts by definition. For example, an increase in one group must be followed by a decrease in another group. To address this issue, I adjust the standard error of each regression by implementing a seemingly unrelated regression (SUR).

In addition to exploring the heterogeneous impact of policy by examining how distribution of students change, I investigate the heterogeneous impact of the policy across schools. To explore this heterogeneity, I ask to what extent the evaluation pilot differentially impacted achievement in pilot schools with different levels of prior achievement. I divide the schools into two groups (high- and low-performing) based on the average performance during the pre-policy period. Similar to the previous specification, I estimate the following:

$$Y_{ist} = \beta_o + \alpha_i + \lambda_t + \theta_g + \beta_1 D_{is,1t} + \beta_2 D_{is,2t} + \beta_3 D_{is,1t} \times I(High_{ist-k}) + X'_{ist}\gamma + \epsilon_{ist} \quad (3.2)$$

3.3.2 North Carolina

I examine the overall effect of VA policy on student achievement by implementing difference-in-differences model and event study model. First, I use the event study model to examine whether policy has a mean effect on student performance.

$$Y_{ist} = \beta_o + \alpha_s + \lambda_t + \theta_g + \sum_{\tau=-k}^{K} \beta_\tau D_{ist}^{\tau} + X_{ist}'\gamma + \epsilon_{ist} \tag{3.3}$$

With the availability of the individual-level data from NCERDC, Y_{ist} becomes the achievement in of student i in district s in year t. The test scores are normalized with the pre-treatment periods. The variable $D_{ist}^{\tau} = I(t - t_{0i} = \tau)$ is an indicator equal to one for being τ time periods relative to i's initial treatment (t_{0i}). τ ranges from -3 to 3 for Guildford and Winston-Salem with $\tau = 0$ being the year of initial treatment. The control districts are randomly divided into two groups and attached to each of the treated districts for the analysis. Thus, the sample is the stack of two groups: 1997-2003 years for Guilford and 2005-2011 years for Winston-Salem, and corresponding control districts.

There are two reasons why I define the estimation sample in this way. First, I need to have a student's records in pre-treatment periods and after-treatment periods, as I want to examine the distributional effect of the policy. Second, Winston-Salem displays the negative trend in 2002-2004 years (shown in Figure B.3), so including these years in the analysis will spuriously estimate the positive effect for Winston-Salem. I also control for the student characteristics, such as gender, race/ethnicity, and gifted status. The standard errors are clustered at school district level. The parameters of interest in equation (3.3) are β_1 to β_3, which show the time-varying effects of the policy among students who are first exposed to this policy in relative years 1 to 3. I show a full set of beta estimates in the Figure B.4 and Table B.3. Since the estimates β_{-3} to β_{-1} in equation (3.3) reprensents the pre-exisitng trends before the policy adoption, it can be also served as a test of the parallel trend assumption.

I also examine the effects with difference-in-differences model to investigate the average effect of policy on student test scores. For this I estimate the model as follows:

$$Y_{ist} = \beta_o + \alpha_s + \lambda_t + \theta_g + \beta_1 D_{ist} + X'_{ist}\gamma + \epsilon_{ist} \tag{3.4}$$

Only difference in this equation to the equationa (3.3) is that D_{ist} is an indicator for whether a district currently release the VA to teachers. I further investigate the heterogeneous effect in North Carolina.

$$Y_{ist} = \beta_o + \alpha_s + \lambda_t + \theta_g + \beta_1 D_{ist} + \beta_2 D_{ist} \times I(H_{ist-1}) + X'_{ist}\gamma + \epsilon_{ist} \tag{3.5}$$

$I(H_{ist-1})$ is indicator of initial district achievement and all other variables are defined as in equation (3.3). In order to investigate how the effect of VA varies across the distribution of test scores, I generate the indicator of students' previous achievement status in two ways. First, to see where a student stands relative to the state's test score distribution, I compare the student's prior achievement to the state median. Second, to see whether effect differs on student's achievement status within the classroom, I compare the student's performance to the class average. In addition to the heterogeneity analysis using difference-in-differences method, I also present the results using an event study model. These are estimated using the same equation except the indicator of previous achievement is interacted with the event study indicators.

3.4 Results

3.4.1 Ohio

In order to further investigate the possible mechanism driving the results found in the nationwide analysis, I exploit the variation in the timing of adoption of VA in Ohio. The event study estimates shown in Figure B.2. not only confirm the parallel trend assumption in DID specification but also provide the preview of the results. There are no significant effects of VA on school performance in both subjects. Table 3.1 summarizes the average impact of the VA policy on student achievement from equation (3.1). Again, the results confirm that there is no impact on student performance measured in the percentage of students who are proficient or above.

Although I did not find any VA impact on the average performance level of schools, exploring the heterogeneous effects of the policy could give us more insight into the overall effect. Examining only the change in the percentage of students who meet the proficiency level of performance can mask the distributional effects of the policy. There are two ways to examine the heterogeneous effects of the policy. First, I examine how the policy affected the proportion of students in each performance level category: advanced, accelerated, proficient, basic, and limited. Treating these ordered categories as a separate measure of school performance, I estimate the DID model (equation (3.1)) for five different outcomes, respectively.

Table 3.2 reports the full set of results for both subjects. For both subjects, there is a decline in the percentage of students in the advanced category, while there is an increase in the proficient category. The percentage of the most advanced performance level decreased by 1.6 percentage points, and 0.9 percentage points decline in the accelerated performance level in Math. At the same time, there is a 0.7 percentage point increase in proficient level and basic level, respectively. The results from reading scores exhibit a similar pattern to those from math scores, while the magnitude is much smaller. The percentage of students in advanced performance levels decreased by 0.53 percentage points, while the percentage of students in proficient performance level increased by 1.2 percentage points. Although it is difficult to identify whether a decrease in top category increases the middle category, the results confirm that the student distribution is compressed toward the middle. The fact that there is a significant decrease in the percentage of students who are ranked at the highest performance level serves as a piece of suggestive evidence that the VA is detrimental to the previously high-performing students.

Second, I examine the heterogeneous effect across schools by comparing the policy impact between high-performing schools and low-performing schools. The results from equation (3.2) are shown in Table 3.3. As shown in Table 3.3, there are no differential impacts detected for high-performing schools relative to low-performing schools in both overall proficiency level and five categories of performance.

3.4.2 North Carolina

Table B.3 shows the results of the event study model for math score. I only report the results for the math scores due to the pre-trends in reading scores. As shown in the Figure B.4, reading scores in pre-policy periods show positive trends. Thus, the result of negative estimates in the event study model might be representing the test scores reverting to the mean. First, the estimates in column (2) shows the evidence of VA positively affecting average test scores in math. The math scores immediately increase by 0.052 SD, and the size of the estimate grows to 0.154 SD after two years and 0.183 SD after three years. Column (1) in Table 3.4 shows the results from DID model. Similar to the results from event study model, students attending schools in districts with VA policy have 0.075 SD gain in math score. This result is similar (although slightly small in magnitude) to findings from (Lee, 2019): he found a 0.096 SD increase in math in Guildford.

The estimates from the heterogeneous analysis show that the VA is detrimental to the top half of the distribution, even with a positive impact on the average test scores. Column (3) and (4) from Table B.3 and column (2) from Table 3.4 shows the negative impact on high-performing students. These results indicate that VA harms the previously high-performing students, which is one possible explanation as to why my results differ from those found in the literature.

However, these results should be interpreted with caution since the statistical inference might be biased. Clustering standard errors with only a small number of clusters can underestimate the true standard errors. To address this issue, I conduct a randomization test. I randomly draw two treatment districts from the sample for 500 times and generate the distribution of t-statistics from the same regression. Comparing with true t-statistics from the results shows that there is possibility of over-rejection. The results from randomization test is presented in Figure B.5.

3.5 Conclusion

With the increasing use of VA measure and a greater push for not only school but also teacher accountability, informing educators and policymakers about how teachers and students respond to VA is essential. The school accountability reforms that required states to have a standardized test has made it easier for schools to adopt the new policy that uses student growth measure in teacher evaluation. In addition to the traditional observational components, VA information becomes one critical component of the teacher evaluation system. Providing VA measures may encourage teachers to put more effort in improving their performance which could lead to an increase in student achievement.

Despite the importance of this policy, there have been very few opportunities to evaluate the impact of this policy. By using a natural experiment in which Ohio and North Carolina informed teachers of their VA scores, I am able to analyze the effect of the policy on student achievement. Before implementing VA policy statewide, Ohio had run a pilot program that uses the teachers' individual VA on their evaluation and North Carolina had two school districts using the VA.

I found no evidence that the policy is beneficial to the students in Ohio. However, the overall null effects that I find in Ohio do not necessarily suggest that the policy has no impact on Ohio students. Teachers might respond to the new information and exert their effort strategically. Hence, I investigate whether providing VA information affects the distribution of students' achievement. Since the academic growth measure of each student will be counted in teachers' evaluation, the policy can change the academic distribution of students. By further investigating the differential effects, I find that the evidence of the detrimental impact of the VA for the students at the top of the performance distribution in schools. In Ohio, I find the distribution of student performance in schools with VA shifts toward the mean. In both subjects, the percentage of the students in the top performance category decreases as the schools adopt VA. This pattern is confirmed in North Carolina as well. In North Carolina, students who were at the top of the score distribution got a negative impact. The math score of top students in the classroom declined as the schools adopt VA.

Most likely, the design of the pilot program was a main reason for not finding a significant effect on students' performance in Ohio. The pilot program was preliminary in the sense that there was no actual evaluation tied to high-stakes personnel decisions, such as tenure or promotion decisions. Thus, future work is needed to understand whether the policy affects students where the VA information is tied to the high-stake decisions.

Figures and Tables

Table 3.1: The Overall Effects of Releasing VA to Teachers on Student Achievement

	Math		Reading	
	(1)	(2)	(3)	(4)
Pilot	-0.014	-0.018	0.0023	-0.0014
	(0.013)	(0.011)	(0.0086)	(0.0047)
Statewide	-0.020	-0.019	-0.0061	-0.0076
	(0.014)	(0.012)	(0.012)	(0.0092)
Control variables	N	Y	N	Y
Grade FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
School FE	Y	Y	Y	Y
Observations	48,769	48,769	48,769	48,769
R-squared	0.785	0.751	0.807	0.784

* Note: The table presents the estimates from equation (3) using Ohio data. Grade, Year, District FE are included in all specifications. Column (2) and Column (4) use the control variables including student charactriscits of the schools such as gender, race/ethnicity, limited english learners, and economically disadvantage status. Standard errors clustered at district level are shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Table 3.2: The Distributional Effects of VA on Student Achievement

VARIABLES	(1) advanced	(2) accelerated	(3) proficient	(4) basic	(5) limited
Panel A. Math					
pilot	-0.016**	-0.0089**	0.0072**	0.0071**	0.011
	(0.0074)	(0.0039)	(0.0034)	(0.0035)	(0.0082)
statewide	-0.0090*	-0.012*	0.0020	0.0030	0.016
	(0.0053)	(0.0066)	(0.0059)	(0.0040)	(0.013)
R-squared	0.711	0.517	0.504	0.575	0.685
Panel B. Reading					
pilot	-0.0053**	-0.0048	0.012**	0.0028	-0.0051
	(0.0024)	(0.0046)	(0.0052)	(0.0030)	(0.0081)
statewide	-0.0056	0.0046	-0.0051	0.0058*	0.00030
	(0.0052)	(0.0044)	(0.0051)	(0.0034)	(0.014)
R-squared	0.587	0.634	0.599	0.645	0.673
Control variables	Y	Y	Y	Y	Y
Grade FE	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y
School FE	Y	Y	Y	Y	Y
N	48,769	48,769	48,769	48,769	48,769

[*] Note: The table presents the estimates from equation (4) using Ohio data. Grade, Year, District FE, and control variables are included in all specifications. Standard errors are adjusted with SUR and shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Table 3.3: The Heterogeneous Effects of VA on Student Achievement

VARIABLES	(1) overall	(2) advanced	(3) accelerated	(4) proficient	(5) basic	(6) limited
Panel A: Math						
DID	-0.023	-0.017*	-0.010*	0.0052	0.013***	0.0094
	(0.017)	(0.0097)	(0.0063)	(0.0052)	(0.0047)	(0.014)
DID*High	0.014	-0.0013	0.0087	0.0069	-0.0055	-0.0087
	(0.017)	(0.0082)	(0.0083)	(0.0073)	(0.0051)	(0.017)
R-squared	0.752	0.712	0.519	0.505	0.578	0.688
Panel B: Reading						
DID	0.00056	-0.0055	-0.0063	0.012**	0.0051	-0.0056
	(0.0091)	(0.0035)	(0.0063)	(0.0048)	(0.0039)	(0.0078)
DID*High	0.0021	0.0065	-0.00076	-0.0036	-0.0038	0.0016
	(0.012)	(0.0060)	(0.0060)	(0.0057)	(0.0039)	(0.011)
R-squared	0.784	0.588	0.635	0.600	0.645	0.674
Control variables	Y	Y	Y	Y	Y	Y
Grade FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y
School FE	Y	Y	Y	Y	Y	Y
N	48,769	48,769	48,769	48,769	48,769	48,769

[*] Note: The table presents the heterogeneous effects in Ohio. Grade, Year, District FE, and control variables are included in all specifications. Column (1) shows the impact on overall proficiency level of the school. Column (2)-(3) present the heterogeneous effects. Standard errors are adjusted with SUR and shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Table 3.4: The Effect of VA on Student Achievement

	(1)	(2)
DID	0.0752***	0.0987***
	(0.0242)	(0.0174)
DID*High		-0.0576***
		(0.0075)
Control variables	Y	Y
Grade FE	Y	Y
Year FE	Y	Y
District FE	Y	Y
R-squared	1,732,693	1,732,693
N	0.3089	0.5778

[*] Notes: The table presents the results from DID model using the NCERDC. All specifications include Grade, District, Year FE and control variables. Column (1) shows the overall effect of VA on student math score. Column (2) shows the heterogeneous effect using student achievement relative to the state performance.. Standard errors clustered at district level are shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Chapter 4

The Early Exposure to Drinking Culture and Impact on Educational Attainment

4.1 Introduction

A large stream of literature links alcohol consumption to its adverse outcomes, such as risky behavior, traffic fatalities, criminal activities, low educational attainment, and worse labor market outcomes.[1] Since the habit of drinking has a tendency to be developed in the early stage of life which has significant implications for future decisions, much of

[1]For example, there are studies investigates the effect of alcohol consumption on mortality (Dee, 1999; Carpenter and Dobkin, 2009), crime (Markowitz and Grossman, 1998; Carpenter, 2007; Carpenter and Dobkin, 2010; Cook and Durrance, 2013), risky sexual behavior (Chesson et al., 2000; Rees et al., 2001; Rashad and Kaestner, 2004; Carpenter, 2005; Waddell, 2012), employment (Mullahy and Sindelar, 1989; Terza, 2002; Dave and Kaestner, 2002; MacDonald and Shields, 2004; Renna, 2008), teen pregnancy (Dee, 2001; Fertig and Watson, 2009), educational attainment (Mullahy and Sindelar, 1989; Cook and Moore, 1993; Dee and Evans, 2003; Chatterji and DeSimone, 2006).

research on drinking focused on the behavior of teens. Especially, several studies conclude that youthful consumption of alcohol inhibits the educational attainment by decreasing the number of years of schooling and the likelihood of completing school (Mullahy and Sindelar, 1989; Cook and Moore, 1993; Chatterji and DeSimone, 2006). In addition, recent studies also provide strong evidence that alcohol consumption impacts the academic development of college students (Carrell et al., 2011; Lindo et al., 2013).

The most popular way to control alcohol access to younger individuals is the Minimum Legal Drinking Age (MLDA), and its' effect has also been studied a lot by researchers. The earliest study that estimates the causal relationship between alcohol consumption and education is by Cook and Moore (1993). The authors estimate the impact of drinking on schooling years by using the variation in the beer tax and the MLDA by states. The authors find that a higher number of drinks consumed per week reduces years of education completed. On the other hand, Dee and Evans (2003) use within-state variation in the MLDA to show that teen drinking does not affect college entrance, completion of both high-school and college.

Understanding the effectiveness of MLDA as a policy is also crucial because the debate over lowering the MLDA from 21 is still ongoing in the United States (Carpenter and Dobkin, 2011). Even with the evidence of the MLDA being effective in regulating alcohol-related incidences, the opposing argument is that a majority of young adults under the age of 21 still obtain alcohol illegally, causing them to learn irresponsible behaviors in alcohol consumption. Thus, lowering the MLDA would help young adults drink more safely and responsibly.

In this paper, I examine the causal link between alcohol consumption and educational attainment by exploiting the quasi-experimental setting in South Korea. Two different policies, the MLDA, and school entrance cutoff, provide a unique setting to identify the causal effect of drinking on educational attainment. First, I evaluate the strength of peer influences on alcohol consumption using a Regression Discontinuity Design (RDD). This Regression Discontinuity design utilizes the fact that these two policies produce a difference in the availability of alcohol for young adults. Using data on drinking behavior from the Korean Labor and Income Panel Study (KLIPS), I demonstrate that underage students who have a drinking-eligible peer are more likely to drink than those who do not have one.

Further, I examine the effect of alcohol consumption on education attainment in college. Since KLIPS has insufficient information on educational attainment, I use additional data and two-sample instrumental variables (TSIV) method (Angrist and Krueger, 1992, 1995). Using detailed information on education from Young Panel (YP), I find that underage drinking does not have any impact on long-term educational attainment.

This paper makes two main contributions to the existing literature. First, using a quasi-experimental setting in South Korea, this study provides new estimates of the effect of having a peer group that has eligibility to drinking on alcohol consumption behavior of young adults who does not have the legal right to drink. Secondly, I examine the impact of having early access to alcohol on college students' educational attainments. The discrete jump in alcohol consumption for the non-eligible group should have adverse spillover effects on educational attainment, which proved to be not in this study. The results indicate that early access to alcohol does not have an impact on the college dropout rate and years of education.

4.2 Institutional Background: Policy Context In Korea

In South Korea, the legal access to alcohol is allowed from the 1st day of the year that an individual becomes 19 no matter their actual day of birth by the Juvenile Protection Act. This naturally results in variation in biological age among the birth cohorts who freshly gain the legal right. That is, there always will be a group of 18 years old who are eligible to drink before they become 19. And this discrepancy in drinking eligibility is further complicated by the school attendance law: Framework Act on Education. South Korean education system is fairly similar to those from the U.S. except two features. First, the elementary school attendance cutoff is in March, resulting in children born in January and February going to school with their peers nearly a year older than themselves. Second, the standard age of entering elementary school is seven instead of six as in the U.S. Thus, after 12 years of formal education, students complete high school and eligible to enter college at the age of 19.[2] The combination of these two laws results in a peer group with both drinking-eligible and drinking-ineligible members. That is, individuals born in January and February will be 18 years old when they finish high school but will never become 19 during that year, resulting in them being ineligible to drink for the entire year. However, students born in March through December will become 19 at some point during that year, which gives them the legal right to drink from the beginning of the year. Following these policies,

[2]The entire educational system in Korea consists of six years of elementary school, three years of middle school and three years of high school. Kindergarten is not a required public education.

those January and February born students will have one year of early exposure to drinking culture while they don't have the legal right to drink.

4.3 Data and Sample

This study uses two data sources: the Korean Labor and Income Panel Study (KLIPS) and Youth Panel (YP). For the first part of the analysis, where I examine the peer effect on alcohol consumption, I use KLIPS for the years of 1998-2015. The KLIPS is a longitudinal survey of urban households in South Korea that has modeled after the National Longitudinal Surveys (NLS) and PSID in the United States. The survey started in 1998 by the Korean Labor Institute with the nationally representative sample of 5000 urban households and their members aged 15 years or older. KLIPS collects detailed information on individuals, including their employment, earnings, education, health behaviors, and other demographic characteristics. The strongest advantage of KLIPS is the availability of the exact date of birth. Due to its confidentiality and relation to the social security number, most of the public data in South Korea does not provide the precise date of birth. KLIPS also contains information regarding drinking and smoking behavior. For the analysis, I need both months of birth and drinking behavior information of individuals. KLIPS is the only survey that provides these sets of information in the same data.

In the survey, underage drinking status is defined as a one-time drinking incident during year 18; that is, an individual is considered to be a drinker if they had more than one alcohol consumption in 18. Individuals answered the frequency and participation in alcohol consumption in the survey. I defined a drinker as one who reports having had at least one drink during the survey year. In order to alleviate the concern of under-reporting the drinking behavior, I limited the sample of those who self-answered the survey.

To evaluate the impact of alcohol consumption on educational attainment, I additionally use the Youth Panel (YP) for the year of 2001-2006. The Youth Panel is also a longitudinal survey conducted by the Korean Employment Information System (KEIS). The survey started in 2001 with a nationally representative sample of 5956 Korean youth ages 15 to 29. YP followed the individuals in the original sample annually for a broad range of information on education and employment until 2006. YP contains detailed information about the individuals' education that can be used to construct the educational attainment variables. The sample for this study reduces to individuals aged between 17 to 30. After dropping observations that were unavailable due to missing values, 16,515 individuals in total were used for analysis. These two data are particularly useful for this study because they contain information on the birthday, which permits the examination of the impact of drinking on educational attainment through investigating the peer effect of alcohol consumption.

The dropout status of all individuals in the sample is obtained from the survey question regarding the current enrollment status. The status code indicates whether the person paused schooling (dropped out), completed the current grade, or finished with a diploma. Based on this information, I defined a college dropout as any student who had a status code indicating dropout, and I set the age of dropout for that student as his or her age as of the effective year. The survey has insufficient information on the college entry date or grade, making it difficult to construct an accurate indicator for an immediate dropout (after the first year of college). Thus, I define the indicator variable for an immediate dropout as the student's current enrollment status at age 19 and 20, reflecting as a dropout. In addition, I define a college dropout as any student who quit college and never returned to complete the degree. I determine the dropout status measured at different ages: dropout status measures by age 24, 27, and 30. Lastly, I look at the years of education completed by age 30. One potential issue with the use of dropout is that it provides information only about extreme cases. In South Korea, dropping out from college is a rare event compared to the U.S. Thus, I use years of completed schooling as a measure of academic performance to minimize the potential bias introduced by using dropout.

4.4 Empirical Strategy

Using this information on self-reported underage drinking and subsequent educational attainment, I found the correlation between underage drinking and schooling decisions. The results indicates that is no effect of underage drinking on dropout for a younger age, but becomes significant for older ages. These naive results are most likely biased due to potential endogeneity. Students who place little value on potential earnings in the future are more likely to drop out of school since schooling decision reflects the personal cost of extra education. Also, such students may be more prone to engage in risky behaviors that might reduce the years of schooling. Since these decisions are made simultaneously, presented estimates may overestimate the actual effect of underage drinking on schooling outcomes. Researchers have established a way to avoid bias. Earlier studies have used state-level variation in beer taxes or MLDA as exogenous determinations of youthful drinking. Most of the states in the U.S. had set drinking age of 18, 19, and 20 before all states were required to adopt the MLDA of 21 years old by federal legislation in 1984. Thus, using these state policy changes as natural experiments finds evidence of alcohol consumption by young adults and negative social outcomes, including educational attainment, motor vehicle accidents, crime, and so on.[3]

This study exploits the variation in alcohol availability at the age of 18 to identify the effects of underage drinking on educational attainment. However, the traditional instrumental variables (IV) estimator that adopts this approach would require that we have a panel that contains data both on teens' drinking and their subsequent schooling decisions. Unfortunately, no survey has all the necessary information. To overcome the lack of data, I relied on the method that allows me to generate instrumental variables estimates using the exact birth date information in two data sets.

To illustrate how TSIV estimates are generated, consider the following structural equation of educational attainment:

$$Education_i = \beta_0 + \beta_1 Drinking_i + \gamma X_i + e_i \tag{4.1}$$

$Education_i$ is an indicator of educational attainment by person i; $Drinking_i$ is an indicator of drinking of person i; X_i is a vector of individual characteristics; e_i is a random error. The potentially endogenous variable is an indicator for drinking, $Drinking_i$.

[3]Cook and Moore (1993) and Dee and Evans (2002) use these policies as the instrument and investigate the impact of teen drinking on educational attainment. However, even with this instrument, the endogeneity problem is not entirely solved. The primary concern is the unobserved characteristics that possibly correlated with drinking behavior. Earlier studies also used RD design to investigate the effect of the MLDA on alcohol consumption and related outcomes, such as student performance or mortalities (Carpenter and Dobkin 2009; Carrell et al. 2011; Lindo et al. 2013). They find that gaining legal access to alcohol at age 21 leads to an increase in alcohol consumption and results in relevant adverse outcomes.

Since the birth month of an individual determines the drinking eligibility, the instrumental variable for $Drinking_i$ will be an indicator $BirthMonth_i$ for whether a teen was exposed to a drinking environment.

Unfortunately, the data that provides information on teen drinking does not record the individuals' ultimate level of education. As a result, I do not have $Education_i$, $Drinking$, and $BirthMonth_i$ in the same data set.[4] The TSIV procedure requires only one dataset with the information on $Education_i$ and $BirthMonth_i$ and second data set with the information on $Drinking_i$ and $BirthMonth_i$ for the same cohort. The first-stage data set, which has information on teen drinking and birth month is KLIPS. The second data set, which has information on educational attainment and birth month, is YP.

The first step in calculating the TSIV estimate is to fit the models for alcohol use of 18 years old with the KLIPS data and to use those parameters to predict the drinking behavior of the contemporaneous respondents. I investigate how the group of individuals born in January and February differs in drinking behavior from the group of individuals born in March through December. Since only 18-years-old birth cohort has a difference in alcohol consumption eligibility, I limit the sample to the 18-year-olds. Every individual who shares the same birth year, the biological age, is the same across the birth month. However, individuals born in January and February enter elementary school one school-year ahead. This results in the difference in eligibility in drinking within the same birth cohort at the age of 18. This RD design uses the fact that the MLDA and early entrance produces differences in drinking eligibility for young adults on either side of March of the year of high school

[4]Even though I have some information in the KLIPS since it must follow the individual from 18 to their later age, the sample size is insufficient for the estimation.

graduation and college entrance. Then, the TSIV estimate is generated by a regression of the educational outcomes of the YP respondents on the cross-sample fitted value for their alcohol use. More specifically, from the KLIPS data set, I obtain an estimate of the first-stage relationship between teen drinking and alcohol availability (depending on the birth month) by estimating the equation:

$$\begin{aligned} Drinking_i &= \beta_0 + \beta_1 age_i + \beta_2 1(age > 0) + \beta_3 age_i^2 \\ &+ \beta_4 age_i * 1(age > 0) + \beta_5 age_i^2 * 1(age > 0) + \gamma X_i + e_i \end{aligned} \tag{4.2}$$

$Drinking_i$ represents drinking incidence for the individual i at the age of 18. Drinking outcome measure is a dummy variable indicating whether or not the individual drank in the survey year. Age is calculated in days: using 1st of March as cutoff, positive numbers are assigned to January and February in day order, while negative numbers are assigned to the rest of the month. Thus, $1(age > 0)$ is an indicator variable equal to one of the student born in January and February. X_i represents a set of measured characteristics of individual i at age 18: the only gender is included. I estimate the model with a second-order polynomial in age to address the age profile of the outcome. Empirical evidence of positive β_2 supports the existence of peer effect on drinking.

Then from the YP data set, I can obtain an estimate of the relationship between educational attainment and drinking by estimating the following equation:

$$Education_i = \beta_0 + \beta_1 \widehat{Drinking}_i + \gamma X_i + \epsilon_i \tag{4.3}$$

In the second stage, the estimated coefficient from the first stage that predicts the drinking behavior of 18 years old ($\widehat{Drinking_i}$) is included in the model of educational outcomes using YP data.

4.5 Results

4.5.1 Evidence of Peer Effect

I first present the difference in the probability of drinking depending on their birth month. I analyzed a group of individuals who were in the same school cohort, but not necessarily of the same age, although eligibility for drinking comes from the biological age. By comparing the drinking instances between individuals with varying eligibility to drink, I find that having a peer who has legal access to alcohol leads to an increase in alcohol consumption among the persons who are not eligible to drink. Figure 4.1 depicts the percentage of drinkers by birth month for two different age groups, which shows that drinking behavior might be influenced by having an eligible peer. While the 18-year-olds show a relatively lower rate of drinking instances than 19-year-olds, January and February born 18-year-olds show higher drinking cases than the rest of the birth months born 18-year-olds. 40% of ineligible individuals who have peers of drinking age in their cohort consume alcohol, though the average drinking rate for the eligible group is 60%. When a peer group only has ineligible drinkers, however, a mere 10% of these individuals drink. Figure 4.1 shows clear evidence of a jump in alcohol consumption for the group of ineligible members with eligible drinkers.

4.5.2 First-Stage Results

Table 4.1 presents the key results from the first stage, which explores the probability of drinking for 18-years-old. I included the results from the global linear model and the second-order-polynomial model. There is evidence of a positive association between students born in January and February and their alcohol consumption. Among 18-year-olds, individuals born in January and February are associated with a 26-percentage point increase in the probability of participation in drinking. Figure 4.2 shows the graphical result of the estimation. There is a clear jump in the likelihood of drinking for the group of students born in January and February, which indicates that having a drinking-eligible peer has a significant impact on alcohol consumption of the ineligible drinkers. Thus, the month of birth of 18-year-olds does appear to provide a valid source of exogenous variation for identifying the educational outcomes of underage drinking.

One concern here is that there is a possibility of parents manipulating the timing of enrolling their children to school. Parents may have an incentive to delay their children's enrollment due to the potential benefits. Several papers documented the adverse effects on academic performance of being the youngest student in a classroom in U.S.[5] Thus, there is substantial interest in the choice that parents face as they decide at what age to enroll their children in kindergarten. Addressing this potential bias is necessary for the estimation, I calculated the Wald estimate for the early entrance. The computed Wald estimate is 0.52. Wald estimate indicates the effect of compilers: the actual effect on alcohol consumption

[5] Bedard and Dhuey (2006) show that the youngest members of fourth and eighth-grade classes have lower test results than the oldest students in the same cohort for OECD countries. Similarly, Datar (2006) shows that children who start kindergarten later get higher test results by using variation in school entry cutoff dates.

of January and February born individuals who entered the school one year earlier than the birth cohort. The early entered January and February born students show a 52 percentage point higher probability of drinking. The 2SLS that January and February born students show a 43 percentage point higher probability of drinking.

4.5.3 TSIV

The results from the first stage show that the two policies had a significant impact on 18-year-olds' drinking participation. It follows that if teen drinking had an independent effect on schooling decisions, then having eligible drinking peers should have also influenced educational attainment. The results presented in this section address this question directly by estimating the effect of exposure to drinking culture during the age of 18 on various educational attainment indicators.

Teenage alcohol consumption is associated with a range of individual and family factors that directly affect educational attainment. This paper builds on recent efforts to better use identifying variables, like a birth month, as a good predictor of underage drinking; however, identifying variables should not directly affect educational attainment. Since only the January and February born students who already are at the stage of high school graduation are subject to underage alcohol use, I focus on college educational attainments. Three different indicators for education levels are of interest: (1) immediate college dropout (2) college dropout by age 24, and 27 (3) years of education completed by age 30.

I estimated the effects of early exposure to a drinking environment on educational attainment using linear probability models and bootstrap corrected standard errors. The results show the impact of having an eligible drinking peer while in 18 on college dropout and years of education at the age of 30. College dropouts are measured at ages 19 and 20 for the immediate dropout from college, 24, and 27 for the ultimate dropout from the college. Since dropping out is a rare event in South Korea, I also included years of education measured at age 30. The results of these evaluations are reported in Table 4.2, which reports the effect of having a drinking-eligible peer at age 18 on different educational attainment measures.

Exposure to an early-drinking culture has positive and statistically significant effects on immediate dropouts. The results suggest that underage drinking while age 18 may increase the probability of dropping out of college during the earlier stage of a college education. For example, they indicate that underage drinkers are 4 percentage points more likely to drop out of college at age 19 and 3.6 percentage points more likely to drop out of college by age 20.

However, the construction of the outcome variables of dropout incidence raises questions about the interpretation of our estimates. Looking at dropouts at age 19, most of the students could have had a maximum of one year of education. In contrast, January and February born students could have had two years of a college education at maximum. These differences in the possible schooling year can overestimate the effect of drinking on the immediate dropout. There could be many reasons that increase the probability of immediate dropouts, such as longer time spent in college, or encountering difficult subjects in their junior year of college. While it is unfortunate that I cannot untangle these mechanisms given the available data, it is reasonable to say there is an impact on the immediate dropout. The estimation results of dropout outcomes by age 24 and 27 indicate that alcohol availability had small and insignificant effects on overall college dropout. The one-year early exposure to drinking culture that results in a higher level of alcohol consumption among the treated group does not have a significant impact on the level of education.

4.6 Conclusion

Although there has been substantial research on the effect of alcohol consumption, existing literature has faced several significant limitations. First, attempts to use state-level law, MLDA, have been questioned about the validity of the instrument. Second, the papers using the RD method focused on specific groups' outcomes, which is hard to interpret as universal results. Thus, it is unclear whether the consumption of alcohol has effects on a broader set of individuals.

In this paper, I expect two different contributions to the existing literature. Mainly, by using the quasi-experimental setting in South Korea, I can estimate the effect of exposure to a drinking environment. Having a peer group whose members have eligibility to drink is associated with higher instances of drinking. The peer effect exists as the young group whose member has eligible peer shows higher drinker instances than others. For the policy purpose, the potential existence and magnitude of peer effects can be of interest by itself since peer effects on alcohol consumption may serve to amplify the impact of interventions.

Secondly, I demonstrated that teen drinking does not affect the ultimate level of education. Even with the statistically significant effect of drinking on immediate dropout, the final level of education is not affected by drinking at the age of 18. This result can be a contribution to the literature where the youthful consumption of alcohol is believed to inhibit the accumulation of schooling.

Figures and Tables

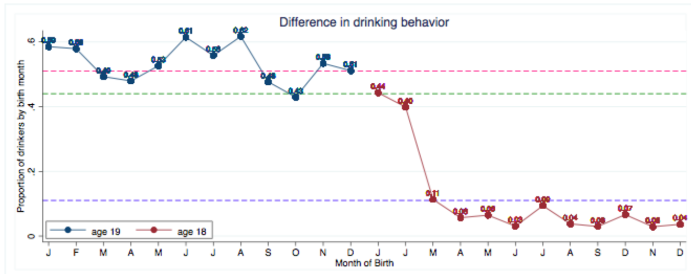


Figure 4.1: The Peer Effect on Drinking

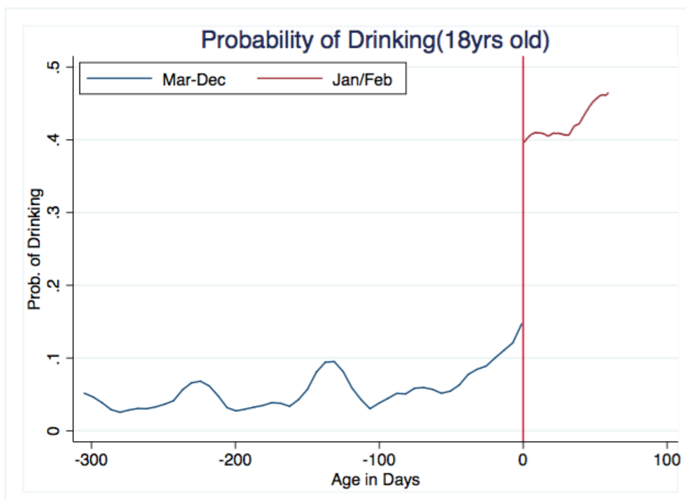


Figure 4.2: First-Stage Result

Table 4.1: First-Stage Results

Drinking	(1)	(2)
Jan&Feb	0.285***	0.263***
	(0.0560)	(0.0896)
Age	0.000202***	0.000744**
	(7.21e-05)	(0.0003)
Age^2		1.69e-06* (9.66e-07)
Age x J&F	0.00133	0.000348
	(0.002)	(0.007)
Age^2 x J&F		-1.77e-07 (0.0001)
Gender	0.0762***	0.0771***
	(0.0147)	(0.0228)
Observation	1,729	1,729

Table 4.2: The Effect of Teen Drinking on Educational Attainment

Education	Dropout by age				Years of Educ.
	19	20	24	27	
Drinkers	0.0402***	0.0362**	-0.00459	0.000380	0.0281
	(0.0142)	(0.0172)	(0.0235)	(0.0271)	(0.0614)
Gender	0.00941***	0.0149***	0.0247***	0.0281***	0.0661***
	(0.00132)	(0.00149)	(0.00212)	(0.00228)	(0.00462)
Observations	2,979	2,995	3,014	3,015	2,435

Note. - Bootstrap corrected standard errors are reported in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Chapter 5

Conclusion

In our society, most wealth is held in the form of human capital. Thus, public policy that aims to produce a positive impact on students' human capital accumulation is important. The overall purpose of this research is to understand the implications of such policies on students' performance. More particularly, the essays examine the effect of education reform on teachers and students, and the impact of regulating substances, such as alcohol on students' achievement.

In the second chapter, I exploit the different timings of the adoption of SGMs in teacher evaluation across states and districts. Using the difference-in-differences and event studies, I find that providing student growth measures to teachers leads to a decrease in average math scores of students. Furthermore, exploring the time-varying treatment effects of the policy, I find that attending school in a district with this policy decreases students' math achievement after four years of adoption. These results are surprising when we think about the aim of the policy. Providing objective measures of teacher performance is to encourage them to improve their performance, which could lead to an increase in student achievement. One possible explanation of contradictory findings is that teachers may strategically focusing on students.

Thus, in the next chapter, I examine the heterogeneity across the distribution of prior student achievement by using the within-state variation in the timing of policy adoption in Ohio and North Carolina. In both states, I find a negative impact on students at the top of the performance distribution in schools and classrooms. In Ohio, the percentage of the students in the top performance category decreases as the schools use the value-added (VA) score. In North Carolina, math scores of previously high-performing students fall as the schools adopt the VA policy. These results confirm that this policy might have an unwanted effect on student achievement, undermining students' performance in a top portion of the distribution, altering the education equity.

The last chapter analyzes the impact of public policy focused on substance regulation and its impact on educational outcomes. I find that the students who are exposed to a peer of drinking age consume more alcohol, but this does not translate into a higher college dropout. My findings can alleviate the fear of relaxing the minimum drinking age surrounding current public debates in countries such as the United States.

Bibliography

Aaronson, D., Barrow, L., and Sander, W. (2007). Teachers and student achievement in chicago public schools. *Journal of Labor Economics*, 25(1):95–135.

Abadie, A., Athey, S., Imbens, G. W., and Wooldridge, J. (2017). When should you adjust standard errors for clustering? Technical report, National Bureau of Economic Research.

Angrist, J. D. and Krueger, A. B. (1992). The effect of age at school entry on educational attainment: an application of instrumental variables with moments from two samples. *Journal of the American statistical Association*, 87(418):328–336.

Angrist, J. D. and Krueger, A. B. (1995). Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics*, 13(2):225–235.

Bates, M. (2020). Public and private employer learning: Evidence from the adoption of teacher value added. *Journal of Labor Economics*, 38(2):375–420.

Bedard, K. and Dhuey, E. (2006). The persistence of early childhood maturity: International evidence of long-run age effects. *The Quarterly Journal of Economics*, 121(4):1437–1472.

Bergman, P. and Hill, M. J. (2018). The effects of making performance information public: Regression discontinuity evidence from los angeles teachers. *Economics of Education Review*, 66:104–113.

Boyd, D., Grossman, P., Lankford, H., Loeb, S., and Wyckoff, J. (2008). Who leaves? teacher attrition and student achievement. Technical report, National Bureau of Economic Research.

Carpenter, C. (2005). Youth alcohol use and risky sexual behavior: evidence from underage drunk driving laws. *Journal of health economics*, 24(3):613–628.

Carpenter, C. (2007). Heavy alcohol use and crime: evidence from underage drunk-driving laws. *The Journal of Law and Economics*, 50(3):539–557.

Carpenter, C. and Dobkin, C. (2009). The effect of alcohol consumption on mortality: regression discontinuity evidence from the minimum drinking age. *American Economic Journal: Applied Economics*, 1(1):164–82.

Carpenter, C. and Dobkin, C. (2010). Alcohol regulation and crime. In *Controlling crime: Strategies and tradeoffs*, pages 291–329. University of Chicago Press.

Carpenter, C. and Dobkin, C. (2011). The minimum legal drinking age and public health. *Journal of Economic Perspectives*, 25(2):133–56.

Carrell, S. E., Hoekstra, M., and West, J. E. (2011). Does drinking impair college performance? evidence from a regression discontinuity approach. *Journal of public Economics*, 95(1-2):54–62.

Chatterji, P. and DeSimone, J. (2006). High school alcohol use and young adult labor market outcomes. Technical report, National Bureau of Economic Research.

Chesson, H., Harrison, P., and Kassler, W. J. (2000). Sex under the influence: the effect of alcohol policy on sexually transmitted disease rates in the united states. *The Journal of Law and Economics*, 43(1):215–238.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Cook, P. J. and Durrance, C. P. (2013). The virtuous tax: lifesaving and crime-prevention effects of the 1991 federal alcohol-tax increase. *Journal of health economics*, 32(1):261–267.

Cook, P. J. and Moore, M. J. (1993). Drinking and schooling. *Journal of Health Economics*, 12(4):411–429.

Datar, A. (2006). Does delaying kindergarten entrance give children a head start? *Economics of Education Review*, 25(1):43–62.

Dave, D. and Kaestner, R. (2002). Alcohol taxes and labor market outcomes. *Journal of Health Economics*, 21(3):357–371.

Dee, T. S. (1999). State alcohol policies, teen drinking and traffic fatalities. *Journal of public Economics*, 72(2):289–315.

Dee, T. S. (2001). The effects of minimum legal drinking ages on teen childbearing. *Journal of Human Resources*, 36(4).

Dee, T. S. and Evans, W. N. (2003). Teen drinking and educational attainment: evidence from two-sample instrumental variables estimates. *Journal of Labor Economics*, 21(1):178–209.

Dee, T. S. and Jacob, B. (2011). The impact of no child left behind on student achievement. *Journal of Policy Analysis and management*, 30(3):418–446.

Fertig, A. R. and Watson, T. (2009). Minimum drinking age laws and infant health outcomes. *Journal of Health Economics*, 28(3):737–747.

Jacob, B. A. and Lefgren, L. (2008). Can principals identify effective teachers? evidence on subjective performance evaluation in education. *Journal of labor Economics*, 26(1):101–136.

Kane, T. J., Taylor, E. S., Tyler, J. H., and Wooten, A. L. (2011). Identifying effective classroom practices using student achievement data. *Journal of human Resources*, 46(3):587–613.

Ladd, H. (2012). Commentary on dee and jacob. *Brooking Papers on Economic.*

Lee, H. (2019). The effect of releasing teacher performance information to schools: Teachers? response and student achievement. Unpublished manuscript, Michigan State University.

Lindo, J. M., Swensen, I. D., and Waddell, G. R. (2013). Alcohol and student performance: Estimating the effect of legal access. *Journal of health economics*, 32(1):22–32.

MacDonald, Z. and Shields, M. A. (2004). Does problem drinking affect employment? evidence from england. *Health Economics*, 13(2):139–155.

Markowitz, S. and Grossman, M. (1998). Alcohol regulation and domestic violence towards children. *Contemporary Economic Policy*, 16(3):309–320.

Mullahy, J. and Sindelar, J. (1989). Life-cycle effects of alcoholism on education, earnings, and occupation. *Inquiry*, pages 272–282.

NCTQ (2019). Measures of professional practice national results. *State Teacher Policy Database.*

Neal, D. and Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, 92(2):263–283.

Pope, N. G. (2019). The effect of teacher ratings on teacher performance. *Journal of Public Economics*, 172:84–110.

Rashad, I. and Kaestner, R. (2004). Teenage sex, drugs and alcohol use: problems identifying the cause of risky behaviors. *Journal of health economics*, 23(3):493–503.

Reback, R., Rockoff, J., and Schwartz, H. L. (2014). Under pressure: Job security, resource allocation, and productivity in schools under no child left behind. *American Economic Journal: Economic Policy*, 6(3):207–41.

Rees, D. I., Argys, L. M., and Averett, S. L. (2001). New evidence on the relationship between substance use and adolescent sexual behavior. *Journal of Health Economics*, 20(5):835–845.

Renna, F. (2008). Alcohol abuse, alcoholism, and labor market outcomes: looking for the missing link. *ILR Review*, 62(1):92–103.

Rivkin, S., Hanushek, E., and Kain, J. (2005). Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2):417–458.

Rockoff, J. (2004). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review: Papers and Proceedings of the One Hundred Sixteenth Annual Meeting of the American Economic Association*, 94(2):247–252.

Steinberg, M. P. and Sartain, L. (2015). Does teacher evaluation improve school performance? experimental evidence from chicago's excellence in teaching project. *Education Finance and Policy*, 10(4):535–572.

Taylor, E. S. and Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102(7):3628–51.

Terza, J. V. (2002). Alcohol abuse and employment: a second look. *Journal of Applied Econometrics*, 17(4):393–404.

Waddell, G. R. (2012). Gender and the influence of peer alcohol consumption on adolescent sexual activity. *Economic inquiry*, 50(1):248–263.

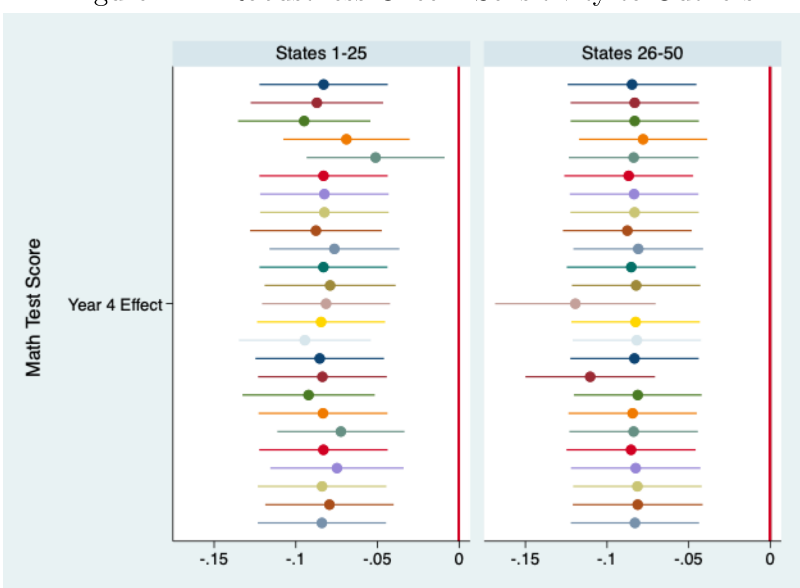
Weisberg, D., Sexton, S., Mulhern, J., Keeling, D., Schunck, J., Palcisco, A., and Morgan, K. (2009). The widget effect: Our national failure to acknowledge and act on differences in teacher effectiveness. *New Teacher Project*.

Appendix A

Appendix for Chapter 2

Figure

Figure A.1: Robustness Check: Sensitivity to Outliers



Notes: The graphs present the result of robustness check from estimating equation (1) as described in the text using the SEDA. Each point represents a point estimate excluding a given state from the regression. The 95% confidence intervals are calculated using standard errors that are clustered at the district level.

Tables

Table A.1: State Policy Detail

	AL	AK	AZ	AR
Pilot				
Year of Partial Pilot			2012-2014	2012-2013
Year of Full Pilot	2009-2010	2015-2016		2013-2014
Student Test Score Used	N	Y	Y	N
Summative Rating Provided	Y	N	Y	N
Percentage			20	
Statewide				
New Teacher Eval. System				
Year of VA provided				
Year of the New Teacher Eval. System	2015	2016	2013	2014
Student Test Score Used	N	N	Y	N
Type of Measure			SG	
Student Growth in Eval.				
Year of Inclusion in Summative Rating			2013	
Percentage			33	

	CO	CT	DE	FL
Pilot				
Year of Partial Pilot	2011-2012	2012-2013		
Year of Full Pilot			2011-2012	
Student Test Score Used	Y	Y	Y	
Summative Rating Provided	N	N	N	
Percentage				
Statewide				
New Teacher Eval. System				
Year of VA provided				2011
Year of the New Teacher Eval. System	2013	2014	2012	2012
Student Test Score Used	Y	Y	Y	Y
Type of Measure	SG	SLO	SLO	VA
Student Growth in Eval.				
Year of Inclusion in Summative Rating	2013	2014	2012	2012
Percentage	50	45	50	50

	GA	HI	ID	IL
Pilot				
Year of Partial Pilot	2011-2014	2011-2013		
Year of Full Pilot				2015-2016
Student Test Score Used	Y	N		Y
Summative Rating Provided	N	N		N
Percentage				
Statewide				
New Teacher Eval. System				
Year of VA provided				2012
Year of the New Teacher Eval. System	2014	2013	2014	2016
Student Test Score Used	Y	Y	Y	Y
Type of Measure	SGP	SGP	SG	SLO
Student Growth in Eval.				
Year of Inclusion in Summative Rating	2014	2013	2014	2016
Percentage	50	25	33	30

	IN	KS	KY	LA
Pilot				
Year of Partial Pilot	2011-2012	2011-2012	2012-2013	2009-2012
Year of Full Pilot		2013-2014	2013-2014	
Student Test Score Used	Y	N	Y	Y
Summative Rating Provided	Y	N	N	Y
Percentage	25-50			50
Statewide				
New Teacher Eval. System				
Year of VA provided				2011
Year of the New Teacher Eval. System	2012	2014	2014	2012
Student Test Score Used	Y	Y	Y	Y
Type of Measure	SGP	SG	SLO	VA
Student Growth in Eval.				
Year of Inclusion in Summative Rating	2016	2014	2014	2012
Percentage	no info	no info	no info	50

	ME	MD	MA	MI
Pilot				
Year of Partial Pilot		2011-2013		
Year of Full Pilot	2014-2015			2011-2013
Student Test Score Used	Y	Y		Y
Summative Rating Provided	Y	Y		Y
Percentage		50		
Statewide				
New Teacher Eval. System				
Year of VA provided				
Year of the New Teacher Evaluation System	2015	2013	2013	2015
Student Test Score Used	Y	Y	Y	Y
Type of Measure	SLO	SLO	SGP	SGP
Student Growth in Eval.				
Year of Inclusion in Summative Rating	2015	2013	2014(lv4) 2015(all)	2015
Percentage	20	50	no info	25

	MN	MS	MO	NV
Pilot				
Year of Partial Pilot	2013-2014	2011-2012	2012-2013	
Year of Full Pilot				2013-2014
Student Test Score Used	Y	Y	Y	N
Summative Rating Provided	Y	N	N	N
Percentage	35			
Statewide				
New Teacher Eval. System				
Year of VA provided				
Year of the New Teacher Evaluation System	2014	2014	2014	2015
Student Test Score Used	Y	Y	Y	Y
Type of Measure	VA	SGP	SLO	SG
Student Growth in Eval.				
Year of Inclusion in Summative Rating	2014	2014	2014	2016
Percentage	35	30	no info	20

	NH	NJ	NM	NY
Pilot				
Year of Partial Pilot	2012-2013	2011-2013	2012-2013	
Year of Full Pilot				
Student Test Score Used	Y	Y	Y	
Summative Rating Provided	N	Y	Y	
Percentage	0-25	0-10		
Statewide				
New Teacher Eval. System				
Year of VA provided				
Year of the New Teacher Evaluation System		2013	2013	2011
Student Test Score Used		Y	Y	Y
Type of Measure		SGP	VA	SGP
Student Growth in Eval.				
Year of Inclusion in Summative Rating		2013	2013	2011
Percentage		45	50	20

	ND	NC	OH	OK
Pilot				
Year of Partial Pilot		2008-2009	2011	
Year of Full Pilot		2009-2012		2012-2013
Student Test Score Used		Y		Y
Summative Rating Provided		N		Y
Percentage				35
Statewide				
New Teacher Eval. System				
Year of VA provided				2012
Year of the New Teacher Evaluation System	2015	2011	2013	2013
Student Test Score Used	Y	Y	Y	Y
Type of Measure	student growth	VA	VA	VA
Student Growth in Eval.				
Year of Inclusion in Summative Rating	2015	2011	2013	2015
Percentage	no info	15	50	35

	OR	PA	RI	SC
Pilot				
Year of Partial Pilot	2012-2013			2013-2014
Year of Full Pilot	2013-2014			
Student Test Score Used	Y			Y
Summative Rating Provided	N			N
Percentage				30
Statewide				
New Teacher Eval. System				
Year of VA provided		2006		
Year of the New Teacher Evaluation System	2014	2013	2012	2015
Student Test Score Used	Y	Y	Y	Y
Type of Measure	SLO	student growth	SGP	VA
Student Growth in Eval.				
Year of Inclusion in Summative Rating	2014	2015	2013	2016
Percentage	no info	15	30	20

	SD	TN	TX	UT
Pilot				
Year of Partial Pilot	2013-2014	2010-2011	2014-2016	2012-2013
Year of Full Pilot				
Student Test Score Used	Y	N	Y	N
Summative Rating Provided	N	N	Y	N
Percentage				
Statewide				
New Teacher Eval. System				
Year of VA provided		1996		
Year of the New Teacher Evaluation System	2014	2011	2016	2015
Student Test Score Used	Y	Y	Y	Y
Type of Measure	SLO	VA	VA	SGP
Student Growth in Eval.				
Year of Inclusion in Summative Rating	2015	2011	2017	2015
Percentage	no info	35	20	20

	VA	WA	WV
Pilot			
Year of Partial Pilot	2011-2012	2010-2012	2011-2012
Year of Full Pilot			
Student Test Score Used	Y	N	N
Summative Rating Provided	Y	N	N
Percentage	40		20
Statewide			
New Teacher Eval. System			
Year of VA provided			
Year of the New Teacher Evaluation System	2012	2015	2015
Student Test Score Used	Y	N	N
Type of Measure	SGP	SLO	
Student Growth in Eval.			
Year of Inclusion in Summative Rating	2012	2015	2015
Percentage	40	no info	

	WI	WY
Pilot		
Year of Partial Pilot	2012-2013	2014-2015
Year of Full Pilot		
Student Test Score Used	Y	N
Summative Rating Provided	N	N
Percentage	50	
Statewide		
New Teacher Eval. System		
Year of VA provided		
Year of the New Teacher Evaluation System	2014	2016
Student Test Score Used	Y	Y
Type of Measure	SLO	SGP
Student Growth in Eval.		
Year of Inclusion in Summative Rating	2014	2016
Percentage	50	no info

Table A.2: Summary Statistics of SEDA

	Adopted		Non-adopted	
	Mean	SD	Mean	SD
Math Score	256.2807	20.93127	261.2401	20.34998
ELL	0.069	0.129	0.033	0.064
Free Lunch	0.399	0.213	0.381	0.210
Special Ed	0.119	0.058	0.142	0.043
Asian	0.029	0.070	0.019	0.041
Black	0.066	0.160	0.090	0.178
Hispanic	0.170	0.248	0.109	0.177
White	0.716	0.302	0.755	0.265
RTTT	0.000	0.000	0.450	0.503
Performance Pay	0.500	0.527	0.725	0.452
N	57,460		272,983	

[*] Notes: The table presents the descriptive statistics for the sample used in nationwide analysis using the SEDA.

Table A.3: Event Study Model

Event Study Model	(1)	(2)	(3)
Year -7	0.061**	0.034	0.036
	(0.026)	(0.027)	(0.028)
Year -6	0.073***	0.052**	0.054**
	(0.021)	(0.023)	(0.023)
Year -5	0.036*	0.021	0.024
	(0.020)	(0.021)	(0.021)
Year -4	0.009	-0.007	-0.005
	(0.017)	(0.017)	(0.018)
Year -3	0.034***	0.023*	0.025*
	(0.013)	(0.013)	(0.014)
Year -2	0.013	0.002	0.003
	(0.010)	(0.010)	(0.010)
Year -1	0.006	0.003	0.003
	(0.007)	(0.006)	(0.006)
Year 1	0.018**	0.012	0.011
	(0.008)	(0.008)	(0.008)
Year 2	0.014	-0.000	-0.001
	(0.014)	(0.014)	(0.015)
Year 3	-0.006	-0.023	-0.024
	(0.021)	(0.020)	(0.021)
Year 4	-0.061***	-0.083***	-0.083***
	(0.023)	(0.023)	(0.024)
Year 5	-0.085***	-0.109***	-0.112***
	(0.026)	(0.026)	(0.029)
Year 6	-0.080***	-0.097***	-0.102***
	(0.030)	(0.033)	(0.036)
Year 7	-0.150***	-0.169***	-0.175***
	(0.039)	(0.042)	(0.045)
Control variables	N	Y	Y
Policy controls	N	N	Y
Grade FE	Y	Y	Y
Year FE	Y	Y	Y
District FE	Y	Y	Y
R-squared	0.864	0.865	0.865
N	328,215	328,215	328,215

[*] Notes: The table presents the estimates from equation (2.2) using the SEDA. Grade, Year, District FE are included in all specifications. Column (2) uses the control variables including student charactriscits of the districts such as gender, race/ethnicity, special education status, limited english learners, and free lunch status. Column (3) adds the policy control such as RTTT winner states and percentage of teachers receiving performance pay in the districts. Standard errors clustered at district level are shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Table A.4: Event Study Model with Hetereogeneous Effects

	Event Study Estimator			Event Study Estimator * High			
	(1)		(2)		(3)		(4)
Year -7	0.113***	Year 1	0.029**	Year -7	-0.161***	Year 1	-0.038***
	(0.037)		(0.012)		(0.030)		(0.013)
Year -6	0.112***	Year 2	0.038*	Year -6	-0.144***	Year 2	-0.067***
	(0.028)		(0.020)		(0.021)		(0.023)
Year -5	0.047*	Year 3	0.039	Year -5	-0.072***	Year 3	-0.115***
	(0.027)		(0.029)		(0.023)		(0.031)
Year -4	0.022	Year 4	-0.031	Year -4	-0.074***	Year 4	-0.080**
	(0.023)		(0.039)		(0.020)		(0.041)
Year -3	0.034*	Year 5	-0.085**	Year -3	-0.035**	Year 5	-0.037
	(0.017)		(0.043)		(0.016)		(0.041)
Year -2	-0.006	Year 6	-0.108**	Year -2	0.008	Year 6	0.009
	(0.013)		(0.055)		(0.013)		(0.053)
Year -1	-0.023***	Year 7	-0.227***	Year -1	0.044***	Year 7	0.085
	(0.009)		(0.067)		(0.010)		(0.065)
Policy controls	Y						
Grade FE	Y						
Year FE	Y						
District FE	Y						
R-squared	0.865						
N	328,215						

[*] Notes: The table presents the estimates from equation (2.2) using the SEDA. Grade, Year, District FE are included in all specifications. Column (2) uses the control variables including student charactriscits of the districts such as gender, race/ethnicity, special education status, limited english learners, and free lunch status. Column (3) adds the policy control such as RTTT winner states and percentage of teachers receiving performance pay in the districts. Standard errors clustered at district level are shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Table A.5: Robustness Check: Using Statewide Adoption Year

	(1)	(2)	(3)
Relative Years to Policy Adoption	-0.0101**	-0.0070*	-0.0075*
	(0.0040)	(0.0041)	(0.0042)
Treated	0.0234**	0.0230**	0.0224**
	(0.0103)	(0.0092)	(0.0094)
Relative Years * Treated	-0.0083	-0.0155***	-0.0158***
	(0.0056)	(0.0054)	(0.0055)
Control variables	N	Y	Y
Policy controls	N	N	Y
Grade FE	Y	Y	Y
Year FE	Y	Y	Y
District FE	Y	Y	Y
R-squared	0.8634	0.8644	0.8644
N	328,215	328,215	328,215

* Notes: The table presents the estimates from equation (2.1) using the statewide adoption year as a treatment. Standard errors clustered at state level are shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Table A.6: Robustness Check: Using Statewide Adoption Year

	(1)	(2)	(3)
DID	-0.0158***	0.0089	-0.0374***
	(0.0055)	(0.0120)	(0.0047)
DID*High		-0.0463***	
		(0.0127)	
DID*Low			0.0463***
			(0.0127)
Control variables	Y	Y	Y
Policy controls	Y	Y	Y
Grade FE	Y	Y	Y
Year FE	Y	Y	Y
District FE	Y	Y	Y
R-squared	0.8644	0.8649	0.8649
N	328,215	328,215	328,215

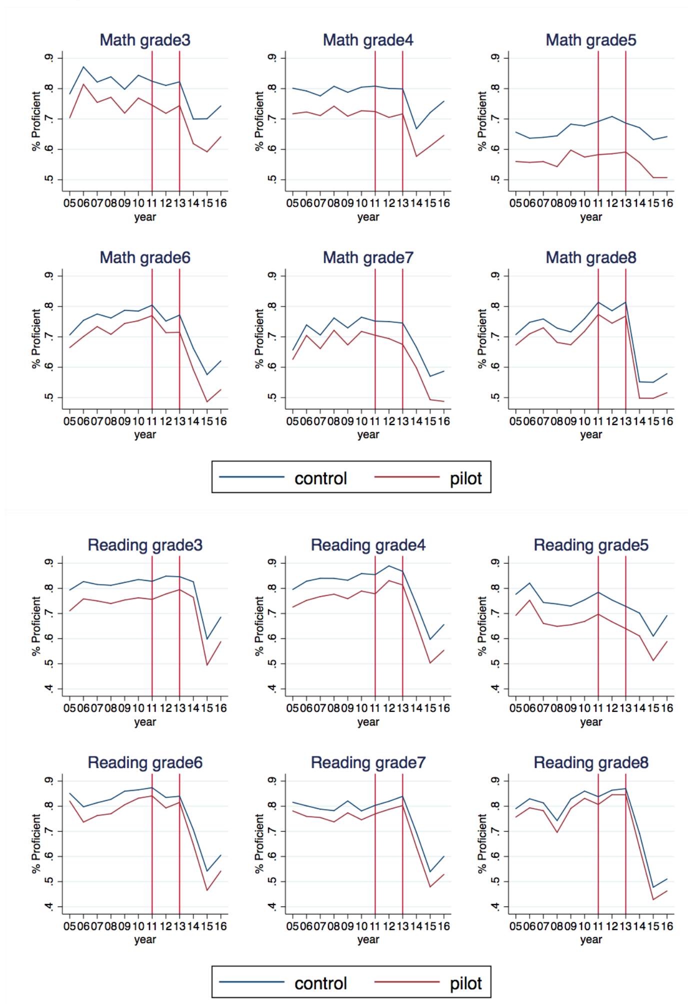
* Notes: The table presents the estimates from equation (2.2) using the statewide adoption year as a treatment. Standard errors clustered at state level are shown in parentheses. Statistically significant at ***1%, **5%, and *10%

Appendix B

Appendix for Chapter 3

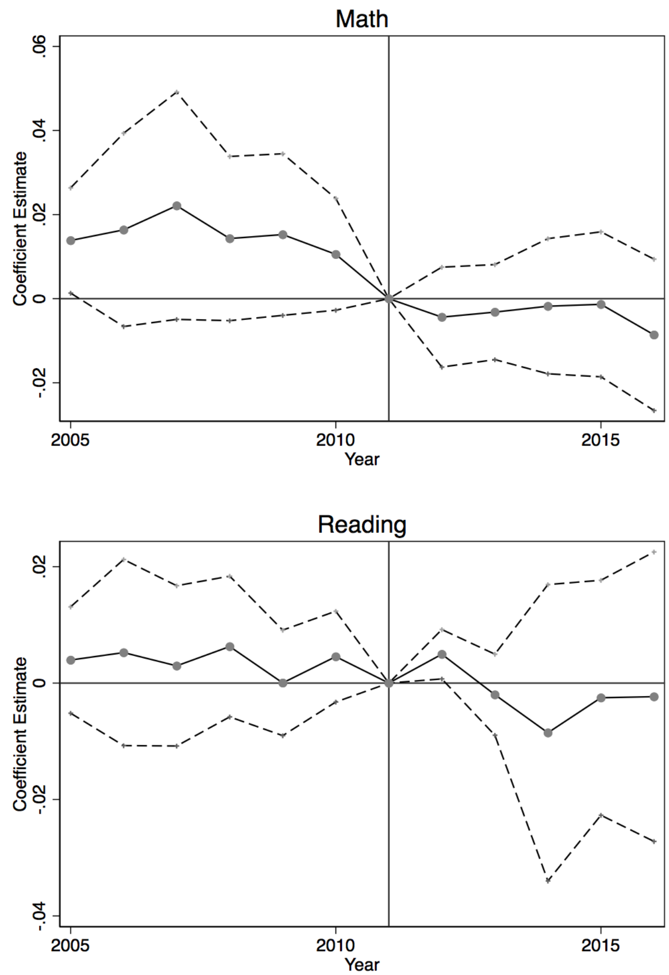
Figures

Figure B.1: Proficiency Level of Schools in Ohio



Notes: The graphs show the percentage of students who are proficient by control and treatment groups. The teachers in pilot schools recieved VA scores in 2011. Ohio implemented the VA measures statewide in 2013. The significant drop in proficiency level since 2014 is due to the change in test system in Ohio.

Figure B.2: Event Study Results for Ohio



Notes: The graphs present the result of event study model for Ohio. The point estimates show the effect of providing VA scores to teachers on school proficiency level relative to year of implementation.

Figure B.3: Math Score Trend in North Carolina

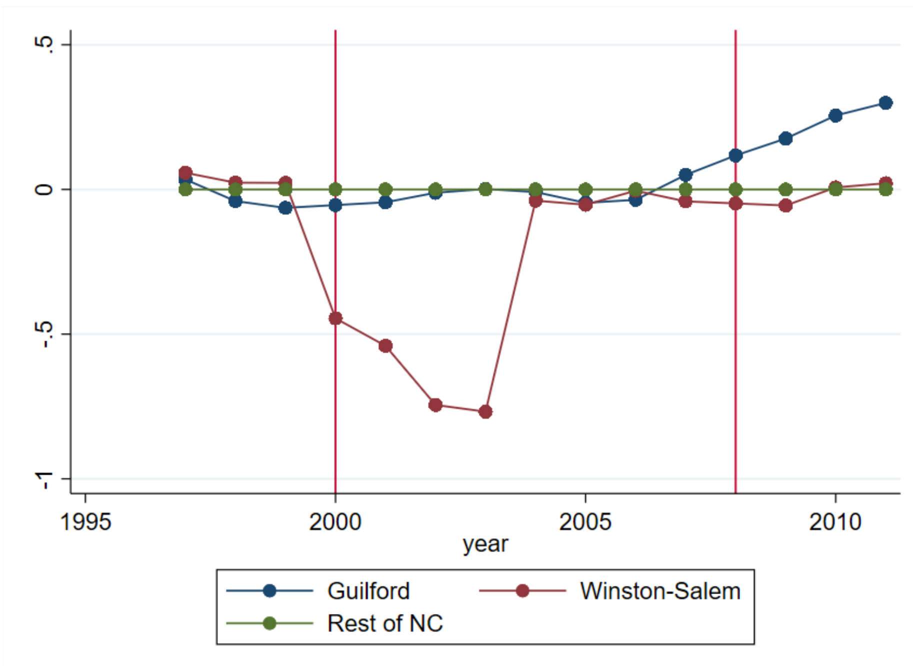


Figure B.4: Event Study Results for North Carolina

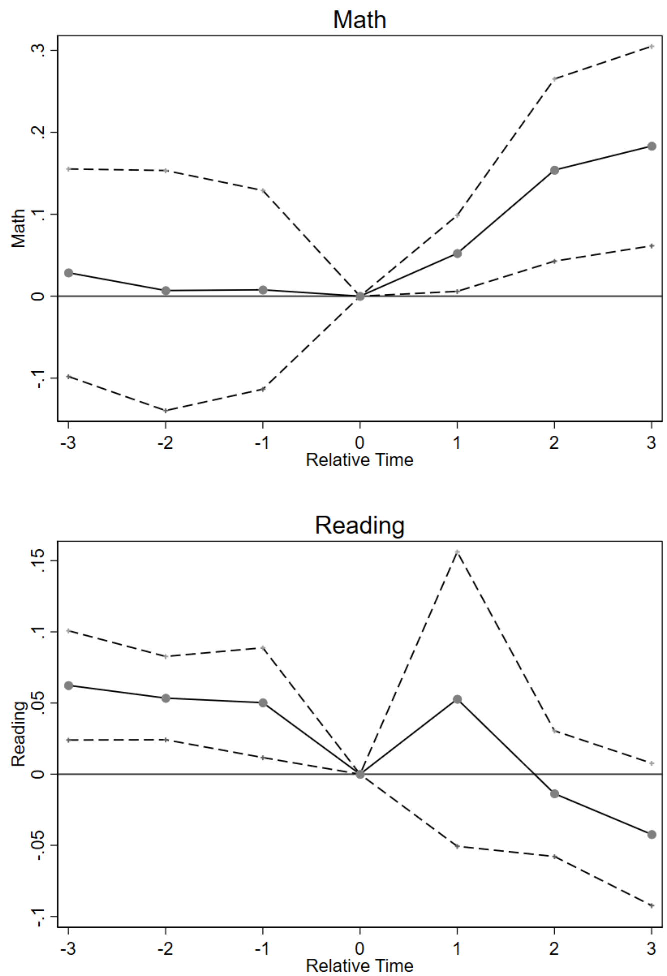
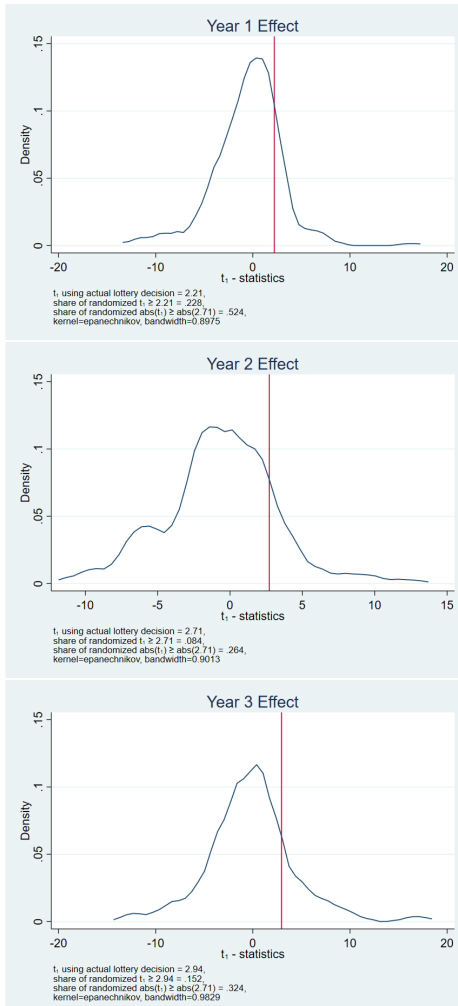


Figure B.5: Robustness Check: Randomization Test



Tables

Table B.1: Summary Statistics: Ohio

	Pilot (Treated)		Control	
	Mean	SD	Mean	SD
unit of observations: school-year (2005-2010)				
Female	0.485	(0.036)	0.485	(0.021)
Limited English	0.042	(0.082)	0.026	(0.080)
Econ Disadvantaged	0.545	(0.270)	0.401	(0.258)
Asian	0.025	(0.060)	0.018	(0.039)
Black	0.248	(0.307)	0.127	(0.241)
Hispanic	0.046	(0.080)	0.033	(0.075)
White	0.651	(0.330)	0.791	(0.269)
N	3582		9066	

* Notes: The table presents the descriptive statistics for the sample from ODE.

Table B.2: Summary Statistics: North Carolina

	Guilford		Rest of NC		Winston-Salem		Rest of NC	
	1997-1999				2005-2007			
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Black	0.34	0.22	0.27	0.24	0.36	0.26	0.24	0.24
Hispanic	0.03	0.07	0.02	0.04	0.07	0.08	0.08	0.10
White	0.62	0.23	0.68	0.26	0.47	0.29	0.61	0.28
Asian	0.01	0.03	0.01	0.04	0.05	0.07	0.02	0.04
Gifted	0.19	0.00	0.14	0.05	0.26	0.01	0.15	0.05
Female	0.51	0.12	0.51	0.11	0.50	0.11	0.50	0.12
N	36161		1273999		64660		1309826	

* Notes: The table presents the summary statistics from the NCERDC.

Table B.3: Event Study Results for North Carolina

	(1)	(2)	(3)	(4)
Year -3	-0.00498	0.0286	0.0186	0.0437
	(0.0186)	(0.0646)	(0.0204)	(0.0714)
Year -2	-0.0278	0.00689	-0.0191	0.00186
	(0.0433)	(0.0747)	(0.0337)	(0.0800)
Year -1	-0.0298	0.00777	-0.127***	-0.109**
	(0.0291)	(0.0619)	(0.00964)	(0.0467)
Year 1	-0.00994	0.0523**	0.0327	0.0607***
	(0.0220)	(0.0237)	(0.0554)	(0.0144)
Year 2	0.0291	0.154***	0.199***	0.181***
	(0.0255)	(0.0567)	(0.0156)	(0.0206)
Year 3	0.0140	0.183***	0.247***	0.232***
	(0.0535)	(0.0621)	(0.0157)	(0.0210)
Year -3 * High			-0.0135	-0.0690
			(0.0220)	(0.0914)
Year -2 * High			0.0225	-0.0170
			(0.0168)	(0.0814)
Year -1 * High			0.202***	0.166***
			(0.0338)	(0.0513)
Year 1 * High			-0.0426	-0.0800***
			(0.0627)	(0.0223)
Year 2 * High			-0.170***	-0.122***
			(0.0137)	(0.0187)
Year 3 * High			-0.203***	-0.164***
			(0.0135)	(0.0188)
Control variables	N	Y	Y	Y
Grade FE	Y	Y	Y	Y
Year FE	Y	Y	Y	Y
District FE	Y	Y	Y	Y
R-squared	0.028	0.309	0.578	0.532
N	1,732,693	1,732,693	1,732,693	1,732,693

* Notes: The table presents the results from event study model using the NCERDC. All specifications include Grade, District, Year FE. Column (1) includes no control variables. Column (3) and (4) show the heterogeneous effects. Column (3) use student achievement relative to the state performance, and column (4) use student achievement relative to the classroom in generating high-performing indicator. Standard errors clustered at district level are shown in parentheses. Statistically significant at ***1%, **5%, and *10%