The Effects of National Standards-Based Reforms on Academically Vulnerable Students

By

Joshua Francis Bleiberg

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Leadership and Policy Studies

June 30, 2020

Nashville, Tennessee

Approved:

Jason A. Grissom, Ph.D.

Carolyn J. Heinrich, Ph.D.

Joseph F. Murph, Ph.D.

Andrew D. Ho, Ph.D.

ACKNOWLEDGEMENTS

Thanks to committee members Jason Grissom, Carolyn Heinrich, Joseph Murphy, and Andrew Ho. Each of you have made an enormous contribution to my progress as a scholar. Thanks for that effort, understanding, and kindness. Special thanks to Jason for guiding me throughout graduate school. I will always appreciate your mentoring and I am certain that it has improved my research enormously.

I have benefitted from fantastic collaborations with Jason Grissom, Joseph Murphy, Erica Harbatkin, Christopher Redding, Beth Schueler, Matthew Springer, and Darrell West. I have learned so much from each of you. I am thankful for my cohort and all of the students in Leadership, Policy, and Organization. We have built a strong student community that will have a long legacy.

I also want to thank my family, James, Susan, and Sarah. I am grateful for their continued support and encouragement. I would not have finished without it.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Standards-based reform describes a triumvirate of education policies: standards, assessments, and school accountability (Smith & O'Day, 1990). Standards set the expected level of performance for students. Assessments then determine if students made progress. Finally, accountability systems sanction schools for failing to meet goals.

I begin with an examination of the Common Core State Standards (CC). Beginning in 2007, the National Governors Association and the Council of Chief State School Officers partnered with the vast majority of states to adopt the CC. The CC reform effort involved many changes to education systems (i.e., curricula, contents standards, and assessments). This study focuses primarily on the effects of the content standards. I then examine No Child Left Behind (NCLB), the national school accountability law. NCLB scaled up state efforts and mandated that all states adopt both standardized assessments and school accountability systems. Finally, I study the NCLB Waivers. The NCLB Waivers released states from many of the school accountability requirements while maintaining the testing rules.

This dissertation contains three essays entitled:

1. Were Some of the Children Left Behind?
2. Did Revoking NCLB Regulations "Waive" in Better Student Outcomes?
3. Does the Common Core Have a Common Effect?: An Exploration of Effects on Academically Vulnerable Groups

Policymakers have sought to increase the rigor of content standards since the 1990s. However, the literature examining the effects of reforms to content standards on student outcomes is still developing. I examine the extent to which the Common Core State Content Standards (CC) affected student achievement and the size of achievement gaps. To identify the effect of CC, I compare early implementors of the CC to late implementors of the CC in a difference-in-differences framework. I conducted a document analysis to measure preparation for and implementation of the CC standards, which I merge together with the National Assessment of Educational Progress student-level data. I then exploit variation in the timing of state implementation of the CC to identify its effect on students overall and on academically vulnerable groups. I find that the CC has a positive effect on math scores in 4th and 8th grade, but not in reading. The CC had a large positive effect on economically advantaged students, but no detectable effect on economically disadvantaged students. Increasing the rigor of content standards without addressing the structural issues burdening economically disadvantaged students may result in unintended consequences.

The passage of No Child Left Behind (NCLB) was a watershed moment for American schools. NCLB required every state to test students and hold schools accountable for their performance and increased education spending by tens of billions of dollars. A robust literature has examined the effect of NCLB on student achievement. I contribute to that literature by examining the differential effects of NCLB across populations of academically vulnerable students and school types. I use the National Assessment of Education Progress (NAEP) student-level data (1990 to 2009) to compare student outcomes before and after the implementation of NCLB. In addition, I compare the difference in post-treatment outcomes between states that had consequential school accountability laws prior to NCLB to those that did not. I use a Comparative Interrupted Time Series dosage model to estimate the effect of NCLB in 2007 for states that did not have consequential accountability prior to NCLB relative to states that did have consequential accountability prior to NCLB. The dosage model weights the effect by the number of years a state had consequential school accountability prior to NCLB. I also examine whether school sanctions or changes to instruction spending mediate the effects of NCLB. NCLB appears to have increased achievement overall and for all academically vulnerable groups. The positive effect of NCLB did not vary across academically vulnerable students and did not influence the size of achievement gaps.

The No Child Left Behind (NCLB) waivers changed school accountability systems by giving states more control over school sanctions. Previous research on the effects of the sanctions has found a mix of positive and null results. I contribute to the literature on the waivers by analyzing their effect on students throughout the country and on academically vulnerable students. I use the National Assessment of Education Progress (NAEP) to estimate the effect of the waivers in a difference-in-differences framework. I exploit variation in the receipt of waivers to compare states that received waivers in 2013 to states that either received waivers in the next year (2014) or never received a waiver. One challenge is that waivers were granted in part based on the adoption of the Obama administration's preferred education reforms. To account for the threat of selection bias, I restrict the sample to include only states that adopted the policies incentivized by the waivers. I find no evidence of an average effect on student achievement. Analysis of heterogeneity suggests that the waivers were associated with a decrease in the size of racial achievement gaps. Increased test scores were isolated amongst Black and Hispanic students in reading.

CHAPTER 2

WERE SOME OF THE CHILDREN LEFT BEHIND

Introduction

A principal objective of No Child Left Behind (NCLB) was to close achievement gaps. President George W. Bush laid out the motivation for NCLB arguing, "too many American children are segregated into schools without standards…This is discrimination, pure and simple, the soft bigotry of low expectations" (2000). The law's goal was not simply to raise test scores for students overall, but, to lift academic outcomes for students of all backgrounds. Under NCLB schools did not escape sanctions if academically vulnerable students struggled to meet achievement targets.

Previous research has shown that NCLB caused modest increases in average student test scores (Cronin et al., 2005; T. Dee & Jacob, 2011; Neal & Schanzenbach, 2010; Reback et al., 2014; M. Wong et al., 2009). The same studies found that NCLB improved average outcomes for academically vulnerable groups. Better outcomes for students overall and for subgroups imply that NCLB shrank achievement gaps or the relative difference between these groups. However, the literature does not test whether NCLB influenced for achievement gaps between advantaged and academically vulnerable students. I am the first to use the student-level data to estimate the effect of NCLB on achievement gaps. I examine how NCLB influenced academically vulnerable students (e.g., race/ethnicity, socio-economic status (SES)) and the intersection of these characteristics relative to privileged students.

I employ the CITS dosage approach used by Dee and Jacob (2011) in the student level National Assessment of Education Progress (NAEP) to identify the causal effect of NCLB on student achievement overall and achievement gaps. I identify the effect of NCLB by comparing students in states that did not have consequential accountability prior to NCLB to students in states that did have consequential accountability prior to NCLB. The CITS dosage model compares the difference in the deviations from pre-treatment trends for both groups in 2007. My estimation strategy does not identify the effect of the other policies that were enacted as part of NCLB (e.g., Reading First, Supplemental Education Services, Transfer Options). My approach tests whether the implementation of NCLB's test-based school sanction regime (i.e., school accountability) changed student outcomes overall and achievement gaps.

I find that the effect of NCLB varies little across student and school groups. NCLB had significant positive effects on 4th grade math and reading outcomes. After NCLB, the white-Black achievement gap closed slightly. The positive effect of NCLB was inappreciably smaller for students from high-SES families. NCLB appears to have raised student achievement for students across a wide variety of subgroups by about the same level, which left achievement gaps intact.

NCLB seems to have achieved its objective of increasing student test scores on average. NCLB also improved student achievement for academically vulnerable students relative to their pre-treatment baseline. But, due to the homogeneity of the effect, achievement gaps remained approximately the same size after NCLB. School sanctions were in theory the most potent NCLB intervention for helping to boost outcomes for academically vulnerable students. The policies that states were required to implement under the sanctions escalated with repeated consecutive years of failing to meet AYP. After 1 year of failing to meet AYP, districts had to develop improvement plans and allow students to transfer to other schools. After 2 years, districts had to provide Supplemental Education Services (i.e., tutoring). After subsequent years of not meeting AYP districts were required to implement their improvement plans and eventually restructure the school. Descriptive analyses suggest that the development and implementation of improvement plans did not have a positive effect. Embedded in the waivers were policies (i.e., Supplemental Educational

Services, transfers) that sought to leverage market forces to improve education systems. However, the available research suggests neither were associated with changes in student outcomes (Heinrich et al., 2010; Murphy & Bleiberg, 2018). NCLB sanctions changed school processes, but my analysis is not able to examine within the "black box" of schools. It is possible that without school sanctions outcomes for academically vulnerable students would have declined even further. Another possibility is that time-varying education reforms (e.g., other standards-based reforms) confound the overall effect of NCLB or the differential effect on academically vulnerable students.

No Child Left Behind

The federal role in education increased considerably after NCLB (Viteritti, 2011). The 2001 law scaled up state efforts at school accountability to the entire country (Manna, 2010). NCLB required states to administer standardized tests that were aligned with state standards each year. The law required that, by the conclusion of the 2013-2014 school year, all students would reach proficiency in reading and math. States used the annual assessments to identify whether students were making Adequate Yearly Progress (AYP) towards the goal of universal proficiency. NCLB also mandated that schools meet AYP targets for several subgroups including: major racial/ethnic groups, economically disadvantaged students, and students with disabilities. Schools that did not meet AYP goals were sanctioned.

NCLB's theory of action involves some stark tradeoffs. The national accountability system created stronger organizational links across levels of government. The tighter coupling between organizations limited the flexibility of school districts to respond to local problems and rendered them less resilient to change (Spillane et al., 2011). Additionally, NCLB's approach to school accountability is quite punitive. The lack of attention to pastoral care and the student perspective could hinder school reform efforts (Murphy & Bleiberg, 2018).

NCLB's accountability systems also created incentives to engage in "educational triage" (Booher-Jennings, 2005). Teachers may have focused on helping students close to the border of achieving proficiency or schools may have increased efforts to instruct students in tested subjects and grades. NCLB also runs amok of Campbell's law (1979) a phenomenon where increasing the use of standardized tests encouraged strategic behaviors that corrupts their value as a measure. NCLB encouraged teachers to spend more classroom time on tested subjects, working to improve instruction, and less usefully "teach to the test" (T. Dee et al., 2012; Grissom et al., 2014; Murnane & Papay, 2010; Reback et al., 2014). To some extent, educator focus on tested content was a goal of NCLB. Alternatively, teachers may have narrowed their instruction to materials assessed on tests that have little use in a real-world setting (Figlio & Loeb, 2011; D. Koretz, 2017; D. Koretz & Hamilton, 2006; D. M. Koretz & Barron, 1998). In extreme cases, accountability pressures could result in compelling teachers to cheat by changing student answers in an attempt to evade sanctions (Jacob & Levitt, 2003).

Conceptual Framework

The underlying policy logic of NCLB relies on principal-agent theory. The framework employs the metaphor of a contract between a principal or boss and an employee or agent (Manna, 2010). In the context of NCLB, the contract is government mandated action by district administrators and principals. According to principal agent theory, accountability policies have the greatest effect when sanctions and/or rewards for violating the rules are made clear. The system breaks down when there are information asymmetries or when the agent has hidden information that is valuable to the principal (Bendor et al., 2001; Moe, 2006). In the context of NCLB, the principal (or boss) are state or federal policymakers and the agent (or employee) are teachers.

The state needs the information about student performance that schools hold to determine if they are meeting AYP requirements. To provide states data on student outcomes, the law included mechanisms for transparent reporting of student performance and school sanctions (Manna, 2010). NCLB required students to take tests in math and reading every year from 3rd to 8th grade and at least once during high school. The law also required test score data to be publicly available for schools and disaggregated by student subgroups (e.g., race, economic advantage). Transparent data was designed to fix two information asymmetries. The open data improved within school information sharing among principals and teachers. It also provided robust data to state and federal policymakers about schooling outcomes.

According to NCLB's theory of action, student achievement will improve because states will monitor outcomes, set goals, and hold schools accountable (Smith & O'Day, 1990). NCLB's testing mandate enables monitoring of student outcomes. The law also set achievement targets for students overall and sub-groups so that gains for privileged students (e.g., white, high SES) do not obscure achievement gaps. If schools do not meet achievement targets they will face sanctions. To help schools improve NCLB provides additional resources (e.g., Title I, Comprehensive School Reform). The synergy of NCLB's reforms and infusion of resources is designed to change the incentives for teachers and school leaders to increase the quality of instruction. Improved instruction ought to then improve student achievement overall and close achievement gaps.

Impacts of NCLB on Students and Schools

Previous studies of NCLB have found small to large positive effects on student achievement. Cronin et al. (2005) used Northwest Education Association longitudinal data from the year before and after NCLB was implemented and found non-persistent gains in math and reading. Wong and coauthors (2009) found substantively large gains for students in public schools on the NAEP exam when compared to students in Catholic schools. Neal and Schanzenbach (2010) compared Chicago students who took a high-stakes NCLB test with students who took the same test but under low stakes the previous year and find small positive effects on reading scores and slightly larger effects on math. Dee and Jacob (2011) also used the state-level NAEP data and compare students in states that previously had consequential accountability prior to NCLB to those that did not and similarly find positive effects in math but not reading. Using data from the Early Childhood Longitudinal Program, Reback, Rockoff, and Schwartz (2014) found that NCLB had a small positive effect on reading scores but did not have a significant effect on math or science scores during the first 2 years of implementation. States that more stringently implemented NCLB had higher average eighth grade math NAEP scores (V. C. Wong et al., 2019).

Several studies have examined whether the implementation of NCLB is correlated with changes in state average test scores for academically vulnerable students. Dee and Jacob (2011) find NCLB caused significant increases in 4th grade math scores for Black (0.47 SD), Hispanic (0.32 SD), and FRPL eligible students (0.36 SD). They also find significant effects of NCLB on 8th-grade math achievement for Hispanic (0.22 SD) and FRPL eligible students (0.41 SD), but insignificant effects for Black students. Wong and colleagues (2019) find NCLB stringency is positively and significantly correlated with 8th grade math outcomes for FRPL eligible students. These studies examine changes in average scores for academically vulnerable groups, which only allow for tangential inferences about NCLB's effect on achievement gaps (i.e. comparing effects on white students to effects on Black students).

A few studies have estimated the impact of NCLB on state or school average achievement gaps. Gaddis and Lauen (2014) used administrative data from North Carolina to compare school level outcomes after the implementation of NCLB. They found that white-Black achievement gaps were reduced in size by increasing Black test scores and not by lowering white test scores. Reardon and co-authors used state achievement test and state NAEP results to examine race and gender achievement gaps before and after NCLB (Reardon et al., 2012). They found that after NCLB there is a significant decrease in the size of white-Black and white-Hispanic achievement gaps, but the magnitude of the decrease is quite small (0.01-0.02 SD). Lee (2006) found little evidence that state white-Black achievement gaps decreased in size after the passage of NCLB. In a subsequent study, Lee & Reeves (2012) found no evidence of changes in the size of white-Black achievement gaps (difference in state averages) associated with the implementation of NCLB. None of these studies directly estimate student level achievement gaps. The effects they find of NCLB could be a function of the number of academically vulnerable students in a school or state. For example, if a state with few Black students like New Hampshire succeeded in closing achievement gaps but a state with many Black students like Georgia did not, then it would erroneously appear that achievement gaps were closing on average. A weakness of all three studies is that state or school level datasets cannot accommodate the estimation of intersectional achievement gaps.

A relatively smaller literature examines the effects of NCLB on different types of schools (e.g., Title I, majority non-white). There are no national studies of how NCLB's effects differed across schools. The available research does suggest that NCLB had a significant and positive effect on economically disadvantaged students (Ballou & Springer, 2017; T. Dee & Jacob, 2011; Lauen & Gaddis, 2012; V. C. Wong et al., 2019). There is some evidence NCLB coincided with an increase in test scores for North Carolina schools in the lowest quartile of poverty (Gaddis & Lauen, 2014). Schools that met overall achievement targets and also failed to meet one or more academically vulnerable group target tended to see scores improve (by about 3 to 6 percentage points) for the failing subgroup in subsequent years (Hemelt, 2011), which implies that the accountability systems raised school-level test scores for academically vulnerable students.

Contribution

My primary contribution is to estimate differential effects of NCLB on students and schools. Previous studies either used state-level NAEP data (T. Dee & Jacob, 2011; Lee & Reeves, 2012; V. C. Wong et al., 2019) or data from several states (Ballou & Springer, 2017; Cronin et al., 2005;

Gaddis & Lauen, 2014; Neal & Schanzenbach, 2010). My approach to estimating the differential effects of NCLB is unique in two ways. First, the student level data allows me to estimate differential effects across student and school characteristics that are not available in the state-level data (i.e., Title I, SES composite, urbanicity, school percent Black/Hispanic). For example, the FRPL indicator is not available in every year in the public data, but in the student-level data I am able to estimate a measure of SES for every year. Similarly, the operationalization of urbanicity changes during the period of study (1990-2009), and access to the restricted data allows me to construct a valid measure across all years. Second, I can estimate intersectional achievement gaps where I test whether the effect of NCLB differs for students that belong to multiple academically vulnerable groups (i.e., Black and economically disadvantaged), which is not possible in the public data. NCLB's effect on intersectional achievement gaps matter because they inform whether NCLB achieved its objective of closing achievement gaps.

I also test whether additional resources (i.e., per pupil instructional spending) and school sanctions mediate the effect of NCLB. Dee and Jacob (2010) find that NCLB significantly increased instructional spending, but do not test whether these increases mediate the effect of NCLB on student outcomes. NCLB's school sanctions were targeted towards under resourced schools that served academically vulnerable students. Research has found that the effects of NCLB were largest for Black and FRPL eligible students (T. Dee & Jacob, 2011; M. Wong et al., 2009). The sanctions (e.g., restructuring and corrective action) were designed to boost outcomes for academically vulnerable students. But no study has examined whether NCLB's school sanctions mediate the effect of NCLB on student outcomes. Improving our understanding of which NCLB policies were effective is important because it informs future federal education reform efforts.

The differences between my analysis and Dee & Jacob (2011) also contributes to my contribution. Dee and Jacob (2011) estimate the effect of NCLB on average outcomes for various student groups (e.g., race, FRPL eligibility). Implicitly, Dee and Jacob's approach tests achievement gaps because a precondition for the closure of those gaps would be relatively better outcomes for academically vulnerable students. I estimate the effect of NCLB on achievement gaps or more specifically whether NCLB changed the difference in outcomes between Black and white students, which speak directly to NCLB's effect on achievement gaps. Dee & Jacob (2011) use a Weighted Least Squares (WLS) model to estimate the effect of NCLB on average outcomes for various student groups instead of the CITS dosage model, which I employ to estimate differential effects. Dee and Jacob (2011) include covariates that are publicly available through the Common Core of Data and the Census (i.e., race or gender). However, there are numerous characteristics included in the NAEP microdata that are not readily available at the state level. For example, state-level time varying measures of social and cultural capital (e.g., books in home, social networks) or modal age for grade are not available in public use data sets. State-level NAEP models will typically include controls for the percent of students in race/ethnicity categories. A student-level regression can accommodate both student-level race variables and the percent of student in race/ethnic groups in the school. A state-level regression cannot include both because the state average of both measures has a nearly perfect correlation. The exclusion of student and school characteristics will bias the estimated effects of policies using the state average NAEP data.

Research Questions

I endeavor to answer 4 questions:

1. To what extent did NCLB affect student achievement?
2. To what extent did NCLB differentially affect academically vulnerable groups of students?
3. To what extent did NCLB affect students in different types of schools?
4. Did specific NCLB policies (i.e., changes to spending, school sanctions) mediate the effect of NCLB overall?

Data, Measures, and Sample

I use NAEP data from four grade-subjects (4th grade math, 8th grade math, 4th grade reading, and 8th grade reading) across 11 waves of the NAEP (1990, 1992, 1994, 1996, 1998, 2000, 2002, 2003, 2005, 2007, 2009). The NAEP uses a complex three-stage sampling design to allow for valid inferences about student achievement in each state and the nation overall (A. Rogers et al., 2014). In the context of studying NCLB, the NAEP data has some unique strengths. The NAEP oversamples students from academically vulnerable populations (A. Rogers et al., 2014). The NAEP assesses a broader set of skills than the average state summative assessment. NAEP relies on committees of subject matter experts, practitioners, researchers, educators, business leaders, and policymakers to write the frameworks used to develop the NAEP test items. The NAEP's sampling design provides the statistical power to detect outcomes for diverse groups of students. Another important strength is the low-stakes nature of the assessment. Accountability pressures on students and teachers could induce measurement error in tests that states use to evaluate schools (D. Koretz, 2017; D. Koretz & Hamilton, 2006; D. M. Koretz & Barron, 1998). NAEP's purpose is to inform policy and practice, mitigating the incentive for cheating or gaming.

I merged the NAEP with school-level data on AYP and school sanctions. NCLB went into effect in the 2002 school year. States designated schools that "need to improve" in 2002 and 2003, but sanctions did not go into effect until 2004 (Murphy & Bleiberg, 2018). I use the Consolidated State Performance Reports (CSPR) (U.S. Department of Education, 2018) to determine which schools were facing sanctions in 2007 and 2009. In 2005, no CSPR report was produced.[1] To impute 2005 sanctions, I used the 2004 CSPR report combined with 2003 and 2004 AYP data (Reback et

[1] The National Adequate Yearly Progress and Identification (NAYPI) database includes school sanctions for the year 2005. However, the data is not available online and AIR was not able to locate it after multiple requests.

al., 2013). Schools under sanction in 2004 that then failed to meet AYP in that same year would remain under sanction in 2005. Schools that met AYP targets in both 2003 and 2004 would not be under sanction in 2005. In 2004 and 2007 the CSPR reports do not include state or federal school identifiers. I used the STATA package reclink to "fuzzy name match" NCES identifiers for using school, district, and state names. In 2007, 95.7% of public schools were matched to NCES identifiers and in 2004, 88.5%were matched. Once merged into the NAEP data about 4 percent of schools were under sanction in 2005 and about 15 percent in 2007, which is similar to national figures. The procedure I use to identify school sanctions appears to have slightly undercounted the total number by about 1 or 2 percent. I will not identify sanction status if a school has a common name (e.g., Lincoln Elementary School). Missing data could induce measurement error in the models where I estimate the effect of the sanctions and make it more difficult to detect significant effects. I strongly suspect that the measurement error is random because the uniqueness of a school name is likely uncorrelated with student outcomes.

Into the NAEP I merge data on education reforms adopted during the period of study including school accountability system features and standards-based reforms. See Appendix Table B1 for the full list of education policies and source information. I also merge in average state per-pupil expenditures and average state per-pupil instructional expenditures from the Common Core of Data (U.S. Department of Education, 2020).

Dependent Variables

To construct my outcomes of interest, I rely on test score information from eleven waves of the NAEP. The NAEP is a matrix-based assessment in which each student completes a sample of test items. NAEP provides plausible values that are created through an Item Response Theory (IRT) procedure. The NAEP includes multiple plausible values to allow the analyst to account for the uncertainty that a student would have received a specific score if they took the entire exam. NAEP then transforms the plausible values into scale scores. I then standardized the scale scores within grade-subject to have a mean of 0 and a standard deviation of 1. I use the first standardized plausible value as the dependent variable. Using the first plausible value should produce results similar to other approaches (e.g., multiple imputation framework, averaging plausible values) (Jerrim et al., 2017) . The results are robust to these approaches because variation in each plausible value is approximately the same.

Independent Variables

The accountability provisions of NCLB went into effect in 2003 for every state. To create a treatment and comparison group I use the strategy proposed by Dee and Jacob (2009). They place states into categories based on their school accountability policies prior to NCLB. To create the school accountability measure Dee and Jacob consulted previous studies (Carnoy & Loeb, 2002; Hanushek & Raymond, 2005; Lee & Wong, 2004), media reports (Edweek, 1999), and Lexus Nexus searches.

Covariates

The NAEP student survey contains a robust set of student characteristics. I control for exogenous student characteristics including gender, whether the student has an Individualized Education Plan (IEP), Limited English Proficiency (LEP) status, eligibility for Free or Reduced-Price Lunch (FRPL), and race/ethnicity. I also add measures for whether the student is at, above, or below the modal age for their grade level. I also use school characteristics including an indicator for whether a school is eligible for Title I funding, the percent of a school's students that are Black and Hispanic, and the schools urbancity.[2] Following Dee and Jacob (2011), I control for the percent of

[2] In 1990 and 1992 there is no Title I indicator. In these years I have imputed the Title I variable based on a categorical variable for the percent of student eligible for FRPL in the school (0%, 1-5%, 6-10%, 11-25%, 26-50%, 51-75%, 76-90%, >90%). The cutoff for Title I eligibility is 40 percent. In 1990 and 1992, schools that were 51 percent or greater

students in a state that were excluded from the NAEP to control for bias from selection into the sample. I include student and school covariates in my main models to control for differences between states that had consequential accountability prior to NCLB and states that did not, which are also correlated with student achievement (Institute of Education Sciences, 2017).

In models that control for FRPL, after list-wise deletion there is a paucity of pre-treatment data (either one or two years). To create an SES composite I conducted Confirmatory Factor Analysis (CFA) using Structural Equation Modeling (SEM) with the available measures of social, cultural, and economic capital available in the NAEP.[3] I examine the intersectional effects of membership in race/ethnic groups and SES. A challenge here is that no student level FRPL indicator is available in either 1990 or 1992. To create a variable that is available in every year I created an SES composite for each NAEP grade-subject where I assumed the data were missing at random. Every CFA model includes measures for whether the student attended a Title I school, eligibility for FRPL, availability of cultural capital in the home (i.e., newspapers, encyclopedias, books, magazines), and urbanicity. In grade 8, I also included parent's highest level of education (i.e., no high school degree, high school degree, some college, college degree or more). 4th grade students frequently did not provide answers about their parent's level of education or cultural capital (i.e., whether there is a globe in the home). In 4th grade, parent's level of education is missing for at least 50 percent of observations. A likely explanation is that 4th grade students do not know their parent's level of education. In my preferred specification, I add binary variables indicating membership in quintiles of SES to control for differences between treatment and comparison states that are correlated with student outcomes. There are several benefits to using the SES composite including

FRPL eligible students were considered Title I schools, which undercounts the true number of Title I schools. NCES uses three urbancity measures during the period of study. Based on the labels assigned to these variables (i.e., urban, suburban, town, rural) I created three categories: urban, suburban/town, and rural.

[3] For the full list of measures use to estimate the latent SES variables by grade-year see the note in Appendix Table B2.

greater reliability and measuring a fuller range of SES factors when compared to a binary FRPL variable (Cowan et al., 2012). The main effect of NCLB is robust to the inclusion of measures of SES (See Appendix Table B2). Table 1 includes descriptive statistics for each NAEP grade and subject.

Sample

Table 2 describes the number of states that were included in the NAEP sample for the 11 waves from 1990 to 2009. From 1990 to 2002 the NAEP was administered for a subset of grade-subjects and from 2003 to 2009 it was administered in all four grade-subjects. Table 2 also includes the number of students included in the NAEP sample in a specified year, pooled across grade-subjects. Prior to 2003, students from at least 38 states are sampled which enables valid national inferences. In my preferred specification I employ Dee and Jacob's (2011) sample restriction, which removes states that are observed for fewer than two pre-treatment years. See Appendix Table 1 for which states were included in each of the NAEP grade-subjects after the sample restriction. Restricting the sample strengthens panel balance and ensures states without pre-treatment outcomes are not used to identify the effect of NCLB.

Estimation Strategy

Following Dee and Jacob (2011) I estimate the causal effect of NCLB using a series of CITS dosage models separately for each NAEP grade and subject that assume the following form:

$$(1)\ y_{icst} = \beta_1 Year_t + \beta_2 NCLB_t + \beta_3 (Yr\_Since\_NCLB_t) + \beta_4 (T_s \times Year_t) + \beta_5 (T_s \times \mathrm{NCLB}_t) + \beta_6 (T_s \times Yr\_Since\_NCLB_t) + \rho \boldsymbol{F}'_{it} + \delta \boldsymbol{G}'_{ct} + \gamma \boldsymbol{H}'_{st} + \alpha_s + \mu_{icst}$$

Where *y* is a NAEP test score (standardized within subject and grade) for student *i*, school *c*, state *s*, and in year *t*. $Year_t$ is a trend variable equal to the year the NAEP is administered minus 1989. $NCLB_t$ is a dummy variable equal 1 for all states starting in the 2003 school year. $Yr\_Since\_NCLB_t$ is defined as the year the NAEP is administered minus 2002, equal to 1 for the 2003 year which corresponds to the 2003 NAEP test. *Ts* is a time-invariant variable that measures the treatment imposed by NCLB. *Ts* equals the number of years during the period of study that a state did not have school accountability. For the comparison group *Ts* varies depending on the first time a state had school accountability. For the treatment group, *Ts* equals 11, which is the spring of the first year NCLB is implemented (2003) minus fall of the first year a state implemented consequential school accountability (1992). $\boldsymbol{F}$, $\boldsymbol{G}$, and $\boldsymbol{H'}$ are vectors of time-varying student, school, and state characteristics, respectively. α_s is a vector of state fixed effects. μ is an idiosyncratic error term clustered by state. In the CITS dosage framework, the effect of NCLB is 6 times the intercept shift ($Ts \times NCLB_t$) plus 30 times the slope shift ($Ts \times Yr\_Since\_NCLB_t$) where 6 is the number of years from 2007 (the last year of outcomes) minus 2001 the last year a school adopted school accountability prior to NCLB and 5 is 2007 minus 2002 (the last year prior to treatment). The effect of NCLB is equal to $(6 \times \beta_5)+(5 \times 6 \times \beta_6)$. The estimates from the CITS dosage framework are interpretable as the difference in the deviations from the pre-treatment trends of NCLB in 2007 for states that did not have consequential accountability prior to NCLB relative to states that did have consequential accountability prior to NCLB. Additionally, the models that include state fixed effects compare outcomes within treated states or those that did not have consequential accountability prior to NCLB. The CITS dosage model weights the effect by the number of years a state had consequential school accountability prior to NCLB. If a state implemented their school accountability policy one year earlier then the dosage of NCLB decreases by one unit. The CITS dosage model I use here to estimate the main effect of NCLB is quite similar to Dee and Jacob (2011). I supplement their approach by adding available student and school covariates.

$$(2)\ y_{icst} = \tau_1 Year_t + \tau_2 NCLB_t + \tau_3(\text{Yr\_Since\_NCLB}_\text{t}) + \tau_4(T_s \times Year_t) + \tau_5(T_s \times \text{NCLB}_t) + \tau_6(T_s \times Yr\_Since\_NCLB_t) + \tau_7(Academ\_Vuln_{icst} \times T_s \times NCLB_t) + \tau_8(\text{Academ\_Vuln}_\text{icst} \times T_s \times Yr\_Since\_NCLB_t) + \tau_9 Academ\_Vuln_{icst} + \rho \boldsymbol{F'}_{it} + \delta \boldsymbol{G'}_{ct} + \gamma \boldsymbol{H'}_{st} + \alpha_s + e_{icst}$$

Equation 2 describes the framework that I use to estimate the differential effects of NCLB. I supplement equation 1, by interacting *Academ_Vuln* which is a binary variable indicating membership in an academically vulnerable group (i.e., student or school characteristic) with $Ts \times NCLB_t$ (τ_7) and $Ts \times Yr\_Since\_NCLB_t$ (τ_8). I report the main effect of NCLB, which has the same interpretation as the estimates from equation 1. I also report the differential effect of NCLB on membership in an academically vulnerable group relative to a privileged group (τ_7+τ_8). For example, I report whether the effect of NCLB differs for Black students relative to white students. I report the difference in the estimated effect rather than the overall effect of NCLB on a sub-group for two reasons. First, the differences in the estimated effect allow for an inference about whether achievement gaps are growing or shrinking. Second, it is useful to report whether NCLB's effects on different groups are different from each other rather than whether the overall effect of NCLB is significantly different from zero. I find that the differential effects of NCLB are quite small and the main effect is large and significant. Reporting the overall effect of NCLB for a subgroup would obscure the significant differences that I do find.

My approach to estimating the differential effects of NCLB diverges from Dee and Jacob (2011). Using the state-level data, they restrict the sample to a sub-group of interest (e.g., Black, FRPL eligible) and use Ordinary Least Squares (OLS) or WLS to estimate the effect of NCLB on a subgroup. Comparing the effects of NCLB on Black and White students from WLS models allows for an implicit test of the effect of NCLB on achievement gaps. For example, they find that NCLB has a larger effect for Black students than white students. A key strength of my approach is that I test whether those differences are significant in the CITS dosage framework. The WLS estimates are biased because they do not account for systematically different pre-treatment trends.

Threats to Causal Inference

The key assumption for identifying causal effects in the CITS framework is that outcomes for students in treated states (no accountability prior to NCLB) would have followed the same trajectory as students in comparison states (accountability prior to NCLB) if NCLB had never occurred, after controlling for the pre-treatment trend. The CITS estimates remain unbiased if the treatment and comparison groups have systematically different pre-treatment trends and those trends are linear (St. Clair et al., 2016). However if the pre-treatment trends are clearly not linear (e.g., the curve of the trend is V shaped) then the CITS estimate is biased.

Figure 2 visually describes the pre-trends for each grade and subject. For math, the pre-trends increase over time, but appear to be parallel and not systematically different. For 4th grade reading, there is visual evidence of non-linear pre-treatment trends, which implies the CITS estimates are biased. For 8th grade reading there are only two pre-treatment years. Limited pre-treatment data makes it difficult to detect non-linearities in the pre-treatment trends and to judge validity of the CITS estimator for 8th grade reading. The results in Figure 2 are consistent with the empirical test of trends in the Appendix Table A2.

Another challenge in estimating the effect of NCLB is that all states were treated at the same time (T. S. Dee et al., 2010). As a consequence, it is not possible to observe outcomes for students in states that are not treated. I follow Dee and Jacob (2011) and argue that states with consequential school accountability prior to NCLB can serve as a control group, while states that did not can serve as the treated group. NCLB borrowed many policies from previous state accountability systems (Manna, 2010). Dee and Jacob (2009) conducted an extensive analysis of pre-NCLB school accountability policies. Their labeling of state accountability systems is consistent with research they did not consider in their review (Snyder & Hoffman, 1998). Pre-NCLB school accountability systems evolved over time. If the differences in those pre-NCLB school accountability systems led some states to struggle with adapting to NCLB then it could lead to flat or decreasing outcomes for comparison states, which would bias the effect of interest. Dee and Jacob (2009) find the evolution of school accountability systems prior to NCLB does not appear to bias the effect of NCLB. Their results are robust to excluding states that adopted school accountability from before and after 1998. Using the NAEP teacher survey I conduct one additional check. In the 2002 NAEP survey (4th and 8th grade reading), teachers were asked if their state had an accountability system. Responses included, "Yes my state has an accountability system that monitors performance in at least one subject" or "No, my state has no accountability system for any subject." Teachers in states that Dee and Jacob describe as having consequential accountability prior to NCLB are 24 points ($p<0.01$) more likely to report their state has an accountability system that monitors performance. The relatively small difference suggests the contrast between the treatment and control groups is weak. If states in the comparison group had weak accountability systems (e.g., performance standards but no accountability mechanism) rather than no accountability system at all, then the estimates here would understate the true effect of NCLB.

Another concern is that the states that chose to implement school accountability prior to NCLB may differ significantly from states that chose not to implement school accountability prior to NCLB. Treated states (those without school accountability prior to NCLB) may have served relatively fewer academically vulnerable students, which could have reduced the pressure to implement school accountability prior to NCLB. If privileged students were better equipped to deal with accountability pressures under NCLB then it would positively bias the effect of NCLB. Treated states (those without school accountability prior to NCLB) may have chosen to not pursue a broader range of education reforms beyond school accountability. A general antipathy towards education reform could have suppressed student achievement. NCLB exposed students to a variety of education policies (e.g., Reading First, HQT requirements) that may then confound the effect of NCLB. Comparison states (those that implemented school accountability prior to NCLB) may have had a higher baseline interest for implementing education reforms. The synergy of these education reforms may have blunted the effect of NCLB on student achievement. To validly identify effects, the similarity between the comparison group's implementation of school accountability and school accountability under NCLB must explain the relatively flatter outcomes for the comparison group. If the effect of NCLB was relatively weaker for the comparison group due to some other set of state characteristics then it would positively bias the effect of NCLB.

I construct my treatment and comparison group using Dee and Jacob's measure of whether states adopted school accountability prior to NCLB. If states chose to implement school accountability due to low levels of student achievement, characteristics of the students they serve, adoption of other education policies (e.g., standards-based reforms) then it could bias the effect of NCLB if these characteristics were related to student outcomes. To test for observable differences between the treatment and comparison groups I collected student and school characteristics from the NAEP and state education policies from the accountability reform period (1980 to 2002). See Appendix Table B1 for a list of policies and detailed source information. I ran a series of bivariate regressions in the NAEP student level data where the outcome is whether or not a state had school accountability prior to NCLB at baseline (See Appendix Table A3). The baseline year was 1992 for 4th grade math, 8th grade math, and 4th grade reading and 1998 for 8th grade reading. At baseline, there are no significant differences in NAEP outcomes across the treatment and control groups for student and school characteristics. States in the math sample that had consequential accountability prior to NCLB appear to have also been more likely to have implemented school finance reforms (SFR). The lack of balance is likely not a source of bias because states are balanced at baseline on per-pupil expenditures. All other baseline differences between the treatment and control group are either statistically insignificant, quite small (less than 0.05 standard deviations), or fall within the range (0.05 SDs to 0.25 SDs) where covariate adjustment is an appropriate solution (Institute of Education Sciences, 2017).

A remaining issue is the possibility of unobserved education policies that are correlated with student achievement and implemented contemporaneously with NCLB (T. Dee & Jacob, 2009). For example, NCLB included a grant program called Reading First. Reading First provided almost $1 billion to bolster state literacy programs. If the grants boosted 4th grade reading scores then the effect of that unobserved policy would confound the effect of NCLB. It is also possible that states in the treatment or comparison groups may have been systematically more likely to implement education policies contemporaneous with NCLB that were not changed under the omnibus education law. For example, states may have made changes to teacher evaluation or voucher policies. The estimates are robust to the inclusion of state-specific linear trends and controlling for broad set of time varying education policies (See Appendix Tables A4 and A5). All available evidence suggests that unobserved time varying education policies do not bias the effect of NCLB.

Results

Effect of NCLB

Table 3 includes models replicating Dee and Jacob's (2010) results and the estimated effects of NCLB in the student level data. The first row of results replicates Dee and Jacob's (2010, Table 1, Row 1) results using the state-level NAEP data, where the outcome is scale score points. In the second row of results, I run the same model in the student-level NAEP data. The difference between the results in the state and student data is approximately 1 scale score point or less (SD=~35 scale score points). In the student level data, the effect of NCLB on 4th grade math is about the same (7.4 scale score points), the effect of NCLB on 4th grade reading is now significant (2.6 scale score points) and the effect of NCLB on 8th grade reading approaches conventional levels of statistical significance (-3.1 scale score points). The main difference between the state and student level results is the sample. The flag indicating the reporting sample was not available in the Math 1990 results (8th grade math). The 1998 reading NAEP (4th and 8th grade) is unique in that NAEP uses a split sampling design to test whether the scores for students with and without accommodations were comparable (A. M. Rogers et al., 2000). The flags indicating the split sample were also not available. The sample differences along with a slight improvement in power account for the different results.

The third row of results is identical to the second except the outcome is now standardized within grade and subject. The fourth row of results adds student and school covariates. The effect of NCLB on 4th grade outcomes is significant and positive (MG4=0.3 SD, RG4=0.18) while the effects on 8th grade outcomes are not robust to the inclusion of controls. In the sixth row of results, I replace the state fixed effects with district fixed effects. The district fixed effects control for district-specific and time-invariant responses to NCLB. A tradeoff with the district fixed effects models is that school districts are frequently sampled fewer than twice in the pre-treatment period and the panel balance for districts is quite poor. If the outcomes for districts with fewer than two pre-treatment observations systematically differed from other districts then it could bias the results. The results follow the same general pattern with significant effects in 4th grade-subjects (MG4=0.42 SD, RG4=0.34 SD) but not 8th grade. In the last row of results, I run the same model adding in an additional year of data from 2009. The effects (MG4=0.50 SD, RG4=0.31 SD) are larger mostly as a function of the CITS dosage model, which assumes that the effect increases along with the time states are treated (i.e., dosage) (See Hoxby in T. S. Dee et al., 2010). Subsequently, I use only results through 2007 to improve the clarity of the comparison with Dee and Jacob (2010).

Differential Effects

Table 4 describes the differential effects of NCLB across student race/ethnicity. The first row includes the estimated effect of NCLB from a CITS dosage model that includes interactions with student's race/ethnicity. The subsequent rows describe the difference between the overall effect of NCLB and the effect of NCLB for students of different race/ethnicities (i.e., Black, Hispanic, Asian). For example, the effect of NCLB on 4th grade Math for Black students relative to the effect of NCLB for white students is 0.302 SDs. The effect of NCLB on Hispanic and Asian students is statistically indistinguishable from white students. The effect of NCLB on Black students is slightly larger for 4th grade math and reading (MG4=0.023 SD, RG4=0.015 SD). The size of those effects implies that NCLB shrunk the Black-white achievement gap by 2.3 percent of an SD in 4th grade math. Dee and Jacob's OLS estimates imply that NCLB had an insignificant effect and the WLS estimates imply the gap closes by 10 scale score points on 4th grade math (Dee & Jacob, 2010, See Table 2). The 10-point closure is quite a large effect considering that the white-Black 4th grade achievement gap closes from 34 points in 1992 to 26 points in 2007 (8-point gap closure). The WLS estimates do not account for systematic pre-treatment differences in outcomes between the treatment and comparison groups. The results in Table 4 use Dee and Jacob's preferred strategy that controls for pre-treatment trends (CITS dosage) and suggest that the WLS estimates are biased.

Table 5 describes the differential effects of NCLB across levels of student's SES. The effects of NCLB on students from higher SES families (quintiles 3, 4, and 5) are significantly smaller, but substantively indistinguishable. Across grades and subjects the effect of NCLB was about 0.01 to 0.02 SD smaller for students in the third and fourth quintile of SES relative to the first quintile of SES. The effect of NCLB in the fifth quintile of SES is about 0.01 SD smaller than the effect of NCLB on students in the first quintile of SES in math. The results suggest that NCLB had approximately the same effect on students from families across levels of SES. The sign and size of the effects are approximately similar to what Dee and Jacob find when restricting the sample to students that are eligible and ineligible for FRPL.

Table 6 describes the intersectional effects of NCLB on race/ethnicity and SES. I add interactions between student race/ethnicity (i.e., Black, Hispanic, Asian) and SES quintiles. Each estimate is the effect of NCLB for students in a specified race/ethnicity relative to white students in the first SES quintile. The effect of NCLB on 4th grade math and reading scores for Black students is slightly more positive (1 percent of a SD) for students from the first quintile of SES. NCLB's effect on higher SES Hispanic students in math is slightly smaller (0.01-0.03 SD) than the effect for white students in the first SES quintile. There are few differences in the effect of NCLB for Asian students across SES quintiles relative to white students in the first SES quintile. There appears to be more variation in the effects of NCLB across race/ethnicity than across SES quintiles.

Table 7 describes how NCLB's effect varies across urbancity. The first row of results describes the main effects, which is the effect of NCLB on urban schools. The second and third row of results describe the effect of NCLB in suburban/town and rural schools relative to the effect of NCLB in urban schools. The results in Table 7 suggest that the effect of NCLB was approximately the same on urban, suburban, and rural districts. The estimates are quite precise and can rule out very small (0.01 SD) differences in the effect of NCLB across urbanicity).[4]

Table 8 describes whether the effect of NCLB differed for Title I relative to non-Title I schools. The first row of results is the main effect of NCLB on non-Title I schools and the second row of results is the difference in the effect of NCLB on Title I schools relative to non-Title I schools. The estimated effect of NCLB on 4th grade reading on non-Title I schools is close to zero.

[4] The definition of urbancity changes twice during the period of study. Some of the changes do not clearly map onto the categories described here (e.g., urban, suburban, rural). For example, urban fringe could describe an area that is either urban, suburban, or rural. I do not suspect the changes in the definition of urbancity bias the result.

The effect of NCLB on Title I schools in 4th grade math and 8th grade reading is significantly larger than non-Title I schools but substantively the same size (0.01 SD). The effect of NCLB on Title-I schools relative to non-Title I schools on 4th grade reading is quite large (0.13 SD). It seems that the positive effects of NCLB on 4th grade math scores are isolated in non Title-I schools, but the effect in 4th grade reading are isolated in Title-I schools. One possible explanation is that school accountability programs (both pre- and post-NCLB) could have also included literacy interventions that were targeted at Title I schools.

Mediation Analysis

Table 9 tests whether school sanctions mediate the effect of NCLB (Baron & Kenny, 1986). The first row replicates the findings in the CITS dosage model. The second, includes results from models where I regress a binary measure of whether a student's school was under sanction on NAEP scores. The third row of results estimates the effect of school sanctions on student outcomes. Finally, the fourth row estimates the effect of NCLB and school sanctions on student outcomes. NCLB did increase the number of schools under sanction by about 3-5 percent. There are no detectable effects of the school sanctions in 8th grade math or reading. NCLB sanctions appear to have a large negative effect on 4th grade math outcomes (-0.2 SD), but no effect on the outcomes in other grade subjects. The estimated effect of NCLB is virtually the same after controlling for school sanctions, which suggests that the sanctions did not mediate the effect of NCLB. School sanctions were intended to improve schools that served academically vulnerable students. But, it appears they neither explain NCLB's positive effect overall nor benefitted the students in sanctioned schools.

Table 10 tests whether state average per-pupil instructional expenditures mediate the effect of NCLB using a similar approach to Table 9. NCLB significantly increased instructional spending for the states in the math samples ($532-$638 per pupil). The effects of NCLB on instructional spending for states in reading were positive but did not reach conventional levels of statistical significance. The positive effect of NCLB on per-pupil instructional expenditures is consistent with Dee and Jacob (2010). The third row of results shows that higher instructional spending is correlated with higher student outcomes in math and 4^{th} grade reading. The estimated effect of NCLB is robust to controlling for instructional expenditures, which implies that state average per-pupil instructional expenditures do not mediate the effect of NCLB.

Robustness Checks

State Specific Linear Trends

As an additional robustness check, I add state-specific linear trends to the CITS dosage model (Angrist & Pischke, 2008). If the inclusion of the state specific trends does not change the results then it provides additional evidence that I can rule-out the possibility of unobserved state characteristics that could bias the effect of NCLB. Appendix Table A4 includes the CITS Dosage model with covariates and state specific linear trends. The sign, size, and significance of the estimates in the state specific linear trends model and the main results in Table 3 are quite similar.

Confounding Education Policies

A related concern are endogenous education policies (i.e., state or federal). Education policies would bias the estimated effect of NCLB if they were time-varying, implemented contemporaneously with NCLB, and correlated with student outcomes. To test whether endogenous education policies explained the effects of NCLB I ran a series of models where I add controls for a broad range of state and federal education policies. If controlling for these policies changed the effect of NCLB then it would suggest that the results suffered from Omitted Variable Bias. In addition to the policies included in Appendix Table A5, I added several policies adopted from 2003 to 2009 including: growth waivers (Hoffer et al., 2011), proficiency standards rigor

(Erpenbach et al., 2003; Erpenbach & Forte, 2005, 2007; Fast & Erpenbach, 2004; Forte & Erpenbach, 2006), teacher evaluation from Bleiberg and Harbatkin (2018), and a broad selection of state education policies (Howell & Magazinnik, 2017). The sign, size, and significance of the effect of NCLB after controlling for education policies (See Appendix Table A5) is quite similar to the effects estimated in Table 3. The robustness of the results suggests that omitted education policies are not biasing the effect of NCLB.

Discussion

NCLB had the complementary goals of achieving universal proficiency and closing achievement gaps. If measured against the goal of closing achievement gaps, NCLB did not achieve its objectives. Across populations of students (e.g., race/ethnicity, SES) and school types (e.g., urbancity, Title I) the differential effects of NCLB are either statistically insignificant or substantively quite small (less than 3 percent of a SD). NCLB's school accountability system was designed to identify and incentivize school improvement activities that would benefit academically vulnerable students. NCLB's targeted approach was theorized to help academically vulnerable students. Although achievement increase for academically vulnerable students in 2007 there were still sizable inequities in student achievement outcomes. In 2009, 59 percent of 4th grade white students had NAEP scores that were proficient or advanced, compared to 16 percent of black students. The evidence from these analyses suggests that the sanctions did not have their intended effect and achievement gaps remained about the same size.

Whether NCLB increased student achievement over all is more ambiguous. I find that the effects of NCLB are larger for 4^{th} grade than for 8^{th} grade. One possible explanation is that 4^{th} graders received a larger dose of NCLB that 8^{th} graders. 8^{th} graders from each post-treatment wave would have attended school prior to NCLB. Whereas only students who attended 4^{th} grade in 2003 would have participated in schools prior to NCLB. The effect of NCLB in the CITS dosage model is much larger than the canonical CITS and event study estimates. The validity of the CITS dosage approach depends on the assumption that the effects of NCLB increase linearly over time. In her commentary on Dee and Jacob's analysis, Hoxby explains that the trend in the CITS dosage model is constructed to be, "linear in year of implementation, so that if a state's accountability program was implemented one year earlier, its NCLB dosage decreases by one unit. This specification does not match up with reality" (T. S. Dee et al., 2010, p. 199). If students did not continue to improve every year under NCLB the CITS dosage estimates are too large.

The continued presence of achievement gaps suggests that the approach Congress took when replacing NCLB with the Every Student Succeeds Act (ESSA) was warranted. The law gave states more flexibility in how states could demonstrate their pursuit of closing achievement gaps (McGuinn, 2016). The role that the federal government could play in closing achievement gaps remains unclear. But altering the strategies used under NCLB is a useful first step. States could also benefit from the opportunity to develop innovative solutions that the federal government could scale up in the future.

The results suggest that school sanctions do not work and could decrease student achievement. The mediation analysis is not a sufficient approach for identifying the causal effects of school sanctions on student achievement. Sanctioned schools serve disproportionately large populations of academically vulnerable students and the descriptive models here do not address that source of bias. In addition, I treat school sanctions as being a component of the NCLB policy. But, several states did have school sanctions prior to NCLB. The results here are consistent with a growing body of research showing that school sanctions/turnaround does not work in every context (Atchison, 2020; Dragoset et al., 2017; Heissel & Ladd, 2018; Henry & Harbatkin, 2019). In future work, more detailed pre-treatment data could cleanly identify the effects of school sanctions.

It remains possible that selection bias might partially explain the effects of NCLB. States in the comparison group were more likely to have implemented school finance reform. However, average state per pupil expenditures overall and average state instructional spending were not significantly different. In addition, I do not find evidence the per pupil expenditures mediated the effect of NCLB. In future work I will explore whether using a district level measure of expenditures explains more variation in student outcomes.

Dee and Jacob also find that NCLB increased instructional time using data from the Schools and Staffing Survey. The NAEP teacher survey also includes a measure of time spent on instruction. In future work, I plan to replicate their finding and test if instructional time mediates the effect of NCLB on student achievement.

I attribute the effect of NCLB to the school accountability provisions. The omnibus law made many changes to education policies that I am not able to measure. For example, Reading First, Highly Qualified Teacher requirements, Comprehensive School Reform, and Supplemental Education Services (Heinrich et al., 2010) were all components of NCLB and may have mediated its effect on student achievement. In future work I hope to investigate whether these policies mediate the effect of NCLB. For example, the NAEP includes information about teacher certification and education, which could be used to test the effect of the Highly Qualified Teacher provisions.

The results here are consistent with the hypothesis that NCLB improved student achievement overall. NCLB seems to have been a rising tide that lifted achievement for all students. NCLB is successful in so far as the benefits were equally shared amongst students of all backgrounds. After NCLB schools produced outcomes that were about as inequitable as before the law. NCLB failed to achieve its stated goal of closing achievement gaps. For education reformers a new approach to school improvement is warranted to help academically vulnerable students.

Chapter 2 References

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Atchison, D. (2020). The Impact of Priority School Designation Under ESEA Flexibility in New York State. *Journal of Research on Educational Effectiveness*, *13*(1), 121–146. https://doi.org/10.1080/19345747.2019.1679930

Ballou, D., & Springer, M. G. (2017). Has NCLB encouraged educational triage? Accountability and the distribution of achievement gains. *Education Finance and Policy*, *12*(1), 77–106.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173.

Bendor, J., Glazer, A., & Hammond, T. (2001). Theories of delegation. *Annual Review of Political Science*, *4*(1), 235–269.

Bleiberg, J., & Harbatkin, E. (2018). Innovation and Diffusion of Teacher Evaluation Reform: A Convergence of Federal and Local Forces. *Educational Policy*.

Booher-Jennings, J. (2005). Below the bubble:"Educational triage" and the Texas accountability system. *American Educational Research Journal*, *42*(2), 231–268.

Bush, G. W. (2000, August 3). *Address Accepting the Presidential Nomination at the Republican National Convention in Philadelphia*. http://www.presidency.ucsb.edu/ws/index.php?pid=25954&st=&st1=

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, *2*(1), 67–90. https://doi.org/10.1016/0149-7189(79)90048-X

Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, *24*(4), 305–331.

Cavell, L., Blank, R. K., Toye, C., & Williams, A. (2005). *Key state education policies on PK-12 education: 2004*.

Census Bureau, & National Center for Education Statistics. (2018). *Annual Survey of School System Finances*. https://www.census.gov/programs-surveys/school-finances.html

Cowan, C. D., Hauser, R. M., Kominski, R. A., Levin, H. M., Lucas, S. R., Morgan, S. L., & Chapman, C. (2012). Improving the measurement of socioeconomic status for the national assessment of educational progress: A theoretical foundation. *Washington: National Center for Education Statistics*.

Cronin, J., Kingsbury, G. G., McCall, M. S., & Bowe, B. (2005). The Impact of the No Child Left Behind Act on Student Achievement and Growth: 2005 Edition. *Northwest Evaluation Association*.

Dee, T., & Jacob, B. (2009). The Impact of No Child Left Behind on Student Achievement. *National Bureau of Economic Research Working Paper Series*, *No. 15531*. https://doi.org/10.3386/w15531

Dee, T., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, *30*(3), 418–446.

Dee, T., Jacob, B., & Schwartz, N. L. (2012). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis*, 0162373712467080.

Dee, T. S., Jacob, B. A., Hoxby, C. M., & Ladd, H. F. (2010). The impact of No Child Left Behind on students, teachers, and schools [with Comments and Discussion]. *Brookings Papers on Economic Activity*, 149–207.

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., Boyle, A., Upton, R., Tanenbaum, C., Giffin, J., & Wei, T. E. (2017). *School Improvement Grants: Implementation and Effectiveness*. US Department of Education.

Edweek. (1999). *Quality Counts*. https://www.edweek.org/ew/qc/index.html

Erpenbach, W. J., & Forte, E. (2005). Statewide Educational Accountability Under the No Child Left Behind Act—A Report on 2005 Amendments to State Plans. *Council of Chief State School Officers.*

Erpenbach, W. J., & Forte, E. (2007). Statewide Educational Accountability Systems Under the NCLB Act—A Report on 2007 Amendments to State Plans. *Council of Chief State School Officers.*

Erpenbach, W. J., Forte-Fast, E., & Potts, A. (2003). *Statewide Educational Accountability under NCLB. Central Issues Arising from An Examination of State Accountability Workbooks and US Department of Education Reviews under the No Child Left Behind Act of 2001.*

Fast, E. F., & Erpenbach, W. J. (2004). Revisiting Statewide Educational Accountability Under NCLB: A Summary of State Requests in 2003-2004 for Amendments to State Accountability Plans. *Council of Chief State School Officers.*

Figlio, D., & Loeb, S. (2011). School accountability. In *Handbook of the Economics of Education* (Vol. 3, pp. 383–421). Elsevier.

Forte, E., & Erpenbach, W. J. (2006). Statewide Educational Accountability Under the No Child Left Behind Act: A Report on 2006 Amendments to State Plans. A Summary of State Requests in 2005-06 for Amendments to Their Educational Accountability Systems Under NCLB. *Council of Chief State School Officers.*

Gaddis, S. M., & Lauen, D. L. (2014). School accountability and the black–white test score gap. *Social Science Research*, *44*, 15–31.

Grissom, J. A., Nicholson-Crotty, S., & Harrington, J. R. (2014). Estimating the effects of No Child Left Behind on teachers' work environments and job attitudes. *Educational Evaluation and Policy Analysis*, *36*(4), 417–436.

Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management: The Journal of the Association for Public Policy Analysis and Management*, *24*(2), 297–327.

Heinrich, C. J., Meyer, R. H., & Whitten, G. (2010). Supplemental education services under No Child Left Behind: Who signs up, and what do they gain? *Educational Evaluation and Policy Analysis*, *32*(2), 273–298.

Heissel, J. A., & Ladd, H. F. (2018). School turnaround in North Carolina: A regression discontinuity analysis. *Economics of Education Review*, *62*, 302–320.

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

Henry, G. T., & Harbatkin, E. (2019). *Turnover at the Top: Estimating the Effects of Principal Turnover on Student, Teacher, and School Outcomes*. Annenberg Institute at Brown University. http://edworkingpapers.com/ai19-95

Hoffer, T. B., Hedberg, E. C., Brown, K. L., Halverson, M. L., Reid-Brossard, P., Ho, A. D., & Furgol, K. (2011). Final Report on the Evaluation of the Growth Model Pilot Project. *US Department of Education*.

Howell, W. G., & Magazinnik, A. (2017). Presidential Prescriptions for State Policy: Obama's Race to the Top Initiative. *Journal of Policy Analysis and Management*, *36*(3), 502–531.

Institute of Education Sciences. (2017). *What works clearinghouse: Standards handbook (Version 4.0).*

Jackson, C. K., Johnson, R. C., & Persico, C. (2015). *The effects of school spending on educational and economic outcomes: Evidence from school finance reforms*. National Bureau of Economic Research. http://www.nber.org/papers/w20847

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, *118*(3), 843–877.

Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, *61*, 51–58.

Koretz, D. (2017). *The Testing Charade: Pretending to Make Schools Better*. University of Chicago Press.

Koretz, D., & Hamilton, L. S. (2006). *Testing for accountability in K-12*. https://www.rand.org/pubs/external_publications/EP20060030.html

Koretz, D. M., & Barron, S. I. (1998). *The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)*.

Lauen, D. L., & Gaddis, S. M. (2012). Shining a light or fumbling in the dark? The effects of NCLB's subgroup-specific accountability on student achievement. *Educational Evaluation and Policy Analysis*, *34*(2), 185–208.

Lee, J. (2006). *Tracking Achievement Gaps and Assessing the Impact of NCLB on the Gaps: An In-depth Look into National and State Reading*.

Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources: State NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis*, *34*(2), 209–231.

Lee, J., & Wong, K. K. (2004). The impact of accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, *41*(4), 797–832.

Manna, P. (2010). *Collision course: Federal education policy meets state and local realities*. CQ Press.

McGuinn, P. (2016). From No Child Left behind to the Every Student Succeeds Act: Federalism and the Education Legacy of the Obama Administration. *Publius: The Journal of Federalism*, pjw014.

Moe, T. M. (2006). Political Control and the Power of the Agent. *Journal of Law, Economics, and Organization*, *22*(1), 1–29.

Murnane, R. J., & Papay, J. P. (2010). Teachers' views on No Child Left Behind: Support for the principles, concerns about the practices. *Journal of Economic Perspectives*, *24*(3), 151–66.

Murphy, J., & Bleiberg, J. (2018). *School Turnaround Policies and Practices in the US - Learning from Failed School Reform*. Springer International Publishing.

National Alliance for Public Charter Schools. (2016). *Charter School Data Dashboard*. http://www.publiccharters.org/

NCES. (2016). *Digest of Education Statistics* (States Requiring Testing for Initial Certification of Elementary and Secondary Teachers, by Skills or Knowledge Assessment and State). https://nces.ed.gov/programs/digest/

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

Reardon, S. F., Greenberg, E., Kalogrides, D., Shores, K. A., & Valentino, R. A. (2012). Trends in Academic Achievement Gaps in the Era of No Child Left Behind. *Society for Research on Educational Effectiveness*.

Reback, R., Rockoff, J., & 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.

Reback, R., Rockoff, J., Schwartz, H. L., & Davidson, E. (2011). *Barnard/Columbia No Child Left Behind Database, 2002-2003 and 2003-2004*. http://www.gsb.columbia.edu/nclb

Reback, R., Rockoff, J., Schwartz, H. L., & Davidson, E. (2013). *Barnard No Child Left Behind Database, 2002-2003 and 2003-2004*. Barnard Columbia NCLB Data Project. http://www.gsb.columbia.edu/nclb

Rogers, A. M., Kokolis, G. A., Stoeckel, J. J., & Kline, D. L. (2000). *1998 Reading Assessment Secondary-Use Data Files, Data Companion*. NCES.

Rogers, A., Tarsitano, C., & Sikali, E. (2014). *National Assessment of Educational Profress (NAEP) 2013 Mathematics and Reading Grades 4 and 8 Assessments Restricted-Use Data Files Data Companion*. National Center for Education Statistics.

Smith, M. S., & O'Day, J. (1990). Systemic school reform. In S. H. Fuhrman & B. Malen (Eds.), *The politics of curriculum and testing*. Consortium for Policy Research in Education.

Snyder, T. D., & Hoffman, C. M. (1998). *State Comparisons of Education Statistics: 1969-70 to 1996-97*.

Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational routines as coupling mechanisms: Policy, school administration, and the technical core. *American Educational Research Journal*, *48*(3), 586–619.

St. Clair, T., Hallberg, K., & Cook, T. D. (2016). The validity and precision of the comparative interrupted time-series design: Three within-study comparisons. *Journal of Educational and Behavioral Statistics*, *41*(3), 269–299.

The Council of State Governments. (2015). *Klarner Book of the States*. http://knowledgecenter.csg.org/kc/category/content-type/content-type/book-states

U.S. Department of Education. (2009). *Data reflecting the number of schools in improvement, corrective action, and restructuring in school year (SY) 2008-2009 (based on SY 2007-2008 assessments) reported by States in the SY 2007-2008 Consolidates State Performance Report (CSPR)*. https://www2.ed.gov/programs/statestabilization/schooldata.pdf

U.S. Department of Education. (2018, February 15). *Consolidated State Performance Reports* [Data Collection Instruments; Application Materials]. https://www2.ed.gov/admins/lead/account/consolidated/index.html

U.S. Department of Education. (2020). *Common Core of Data*. https://nces.ed.gov/ccd/ccddata.asp

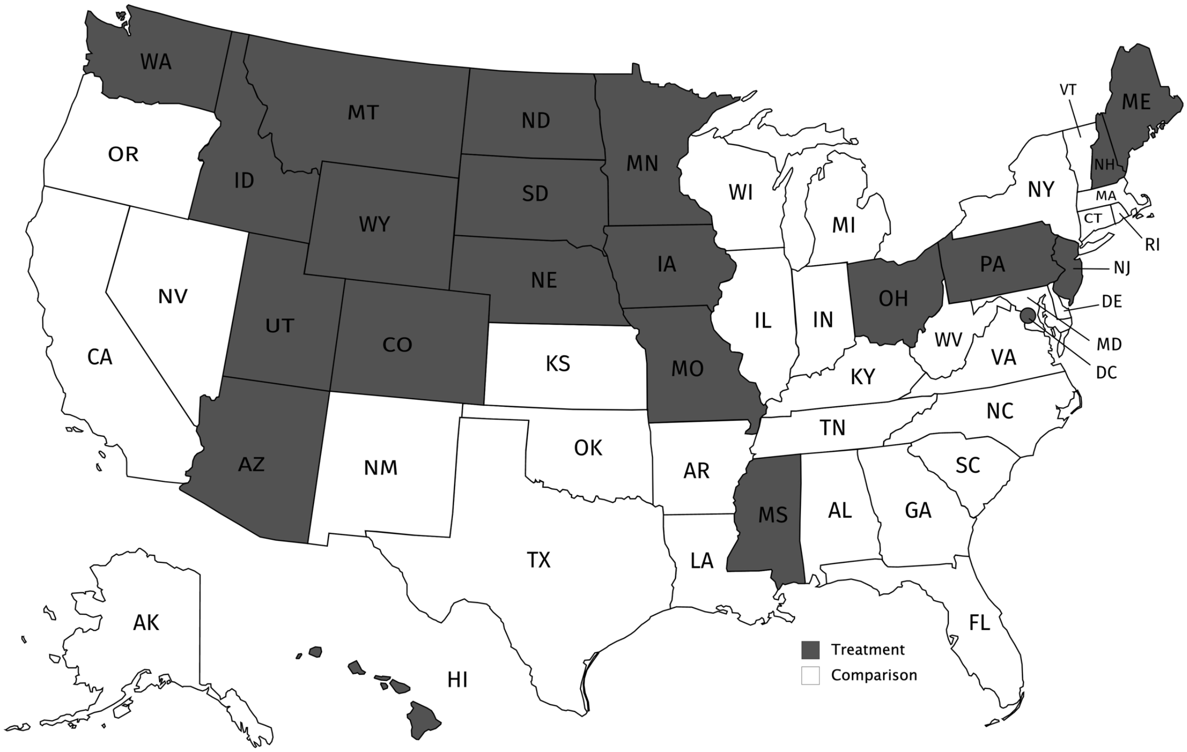
Viteritti, J. P. (2011). Federal Role in School Reform: Obama's Race to the Top, The. *Notre Dame L. Rev.*, *87*, 2087.

Warren, J. R., & Kulick, R. B. (2007). Modeling states' enactment of high school exit examination policies. *Social Forces*, *86*(1), 215–229.

Wong, M., Cook, T. D., & Steiner, P. M. (2009). No Child Left Behind: An interim evaluation of its effects on learning using two interrupted time series each with its own non-equivalent comparison series. *Institute for Policy Research*, *18*.

Wong, V. C., Wing, C., Martin, D., & Krishnamachari, A. (2019). *The Impact of Intensifying State Accountability Pressures on Student Achievement under No Child Left Behind*. University of Virginia.

Figure 1. Map of School Accountability Prior to NCLB (Treatment and Comparison Groups)



Note: The treatment group is states that did not have school accountability prior to NCLB. The comparison group is states that did have school accountability prior to NCLB. Previous consequential school accountability measure from Dee and Jacob (2011).

Figure 2. Trends in NAEP Scores for Treatment and Comparison Groups

Note: Group means estimates with analytic sample that includes states observed in 2 or more pre-treatment waves. The blue line is the comparison group (states with school accountability prior to NCLB) and the red line is the treatment group. Treatment centered on 2002, the last year prior to the implementation of No Child Left Behind. Sample size rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP= National Assessment of Educational Progress test score standardized within grade-subject. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007, 2008-2009.

Table 1. Descriptive Statistics

	4th Grade Math			8th Grade Math			4th Grade Reading			8th Grade Reading		
	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N
NAEP Outcome	0.0000	1.0000	981960	0.0000	1.0000	995040	0.0000	1.0000	1130760	0.0000	1.0000	803870
IEP	0.0961	0.2947	981220	0.0806	0.2722	994640	0.0854	0.2795	1130160	0.0896	0.2856	803360
LEP	0.0604	0.2382	980820	0.0321	0.1763	994160	0.0558	0.2295	1129870	0.0369	0.1884	802960
Gender	0.4935	0.5000	981960	0.4978	0.5000	995040	0.4966	0.5000	1130760	0.5002	0.5000	803870
White	0.5370	0.4986	981960	0.5964	0.4906	995040	0.5389	0.4985	1130760	0.5572	0.4967	803870
Black	0.1465	0.3537	981960	0.1443	0.3514	995040	0.1461	0.3532	1130760	0.1491	0.3562	803870
Hispanic	0.2116	0.4084	981960	0.1625	0.3689	995040	0.2105	0.4077	1130760	0.1847	0.3881	803870
Asian/PI	0.0395	0.1947	981960	0.0429	0.2027	995040	0.0409	0.1981	1130760	0.0460	0.2095	803870
American Indian	0.0310	0.1734	981960	0.0218	0.1460	995040	0.0323	0.1769	1130760	0.0230	0.1498	803870
Other Race	0.0344	0.1823	981960	0.0320	0.1761	995040	0.0312	0.1738	1130760	0.0400	0.1959	803870
At Modal Age	0.6014	0.4896	981960	0.5890	0.4920	995040	0.6043	0.4890	1130760	0.5928	0.4913	803870
Below Modal Age	0.0024	0.0493	981960	0.0039	0.0624	995040	0.0025	0.0504	1130760	0.0032	0.0568	803870
Above Modal Age	0.3961	0.4891	981960	0.4071	0.4913	995040	0.3932	0.4884	1130760	0.4039	0.4907	803870
Parent Education; No HS	0.0612	0.2396	261930	0.0821	0.2745	879600	0.0555	0.2289	344830	0.0784	0.2688	703400
HS	0.1842	0.3877	261930	0.2321	0.4222	879600	0.1752	0.3801	344830	0.2077	0.4057	703400
Some College	0.1186	0.3233	261930	0.2003	0.4002	879600	0.1417	0.3487	344830	0.2041	0.4030	703400
College	0.6360	0.4811	261930	0.4855	0.4998	879600	0.6277	0.4834	344830	0.5098	0.4999	703400
FRPL Eligible	0.5367	0.4987	843200	0.6038	0.4891	759100	0.5388	0.4985	975270	0.5985	0.4902	774190
SES Composite	0.0016	0.2694	981960	0.0023	0.3962	995040	0.0008	0.2539	1130760	0.0013	0.4314	803870
SES Q1	0.1712	0.3767	981960	0.2427	0.4287	995040	0.1481	0.3552	1130760	0.2545	0.4356	803870
SES Q2	0.2492	0.4325	981960	0.1415	0.3486	995040	0.2515	0.4339	1130760	0.1333	0.3399	803870
SES Q3	0.1746	0.3796	981960	0.1884	0.3910	995040	0.2421	0.4283	1130760	0.1797	0.3840	803870
SES Q4	0.2715	0.4447	981960	0.1410	0.3480	995040	0.2353	0.4242	1130760	0.1387	0.3456	803870
SES Q5	0.1336	0.3402	981960	0.2864	0.4521	995040	0.1230	0.3284	1130760	0.2939	0.4555	803870
Title I School	0.4891	0.4999	919220	0.3055	0.4606	962610	0.5084	0.4999	1100650	0.3134	0.4639	767410
School Pct Black	17.2117	26.9251	935490	16.1305	25.3757	983890	16.8205	26.4622	1125940	16.6616	25.7812	803710
School Pct Hispanic	12.2485	21.7836	935470	10.6169	19.5466	983890	12.0534	21.5780	1125910	11.9938	20.6427	803710
Urban	0.5169	0.4997	978770	0.4589	0.4983	991990	0.4992	0.5000	1127120	0.5268	0.4993	802290
Suburban	0.2580	0.4375	978770	0.2918	0.4546	991990	0.2683	0.4431	1127120	0.2321	0.4222	802290
Rural	0.2251	0.4177	978770	0.2493	0.4326	991990	0.2324	0.4224	1127120	0.2410	0.4277	802290
NAEP Exclusion Rate	0.0374	0.0207	981960	0.0407	0.0190	995040	0.0590	0.0237	1130210	0.0494	0.0195	803290

Note: Sample size rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP= National Assessment of Educational Progress test score standardized within grade-subject, IEP=Individualized Education Plan, LEP=Limited English Proficiency, PI=Pacific Islander, HS=High School, FRPL=Free and Reduce Price Lunch, SES=Socio Economic Status, Pct=Percent. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007, 2008-2009.

Table 2. NAEP Sample Characteristics

Year	Math G4	Math G8	Reading G4	Reading G8	States	Students
1990		X			38	97,900
1992	X	X	X		42	321,120
1994			X		39	100,150
1996	X	X			45	202,980
1998			X	X	41	184,890
2000	X	X			42	185,630
2002			X	X	46	243,160
2003	X	X	X	X	51	657,290
2005	X	X	X	X	51	623,070
2007	X	X	X	X	51	648,380
2009	X	X	X	X	51	647,070

Note: Sample size rounded for the number of districts, schools, and students in accordance with National Center for Education Statistics nondisclosure rules. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007, 2008-2009.

Table 3. Replication of NCLB Effects in Public and Restricted Data

	Math G4	Math G8	Reading G4	Reading G8
	(1)	(2)	(3)	(4)
Public State NAEP (Scale Scores)	7.244**	3.704	2.297	-2.101
	(2.240)	(2.464)	(1.441)	(2.070)
N	227	220	249	170
F	126.54	21.14	13.82	8.09
	(5)	(6)	(7)	(8)
Restricted NAEP (Scale Scores)	7.414**	2.363	2.592*	-3.084+
	(2.413)	(3.454)	(1.191)	(1.549)
N	662,320	590,420	749,780	471,710
F	492.30	72.66	34.52	8.26
	(9)	(10)	(11)	(12)
Restricted NAEP (Standardized)	0.247**	0.066	0.071*	-0.090+
	(0.080)	(0.096)	(0.033)	(0.045)
N	662,320	590,420	749,780	471,710
F	492.30	72.66	34.52	8.26
	(13)	(14)	(15)	(16)
Restricted NAEP with Covariates (Standardized)	0.300***	0.095	0.181***	-0.006
	(0.068)	(0.079)	(0.043)	(0.035)
N	612,700	567,290	725,710	453,030
F	4,701.36	4,804.93	3,051.20	2,2079.56
	(17)	(18)	(19)	(20)
Restricted NAEP with Covariates and District FE (Standardized)	0.422*	-0.006	0.341**	0.121
	(0.178)	(0.082)	(0.114)	(0.091)
N	611,520	566,000	724,440	451,580
F	4,513.39	6,032.33	2,217.16	4,043.74
	(21)	(22)	(23)	(24)
Restricted NAEP with Covariates and 2009 Outcomes (Standardized)	0.495***	0.183	0.308***	0.0001
	(0.130)	(0.122)	(0.079)	(0.067)
N	730,830	674,750	846,920	548,820
F	5,419.86	9,694.18	6,542.93	4,983.53

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. The author replicated the estimates from Dee and Jacob (T. S. Dee et al., 2010) Table 1 Row 1. NAEP restricted data outcomes standardized within grade and subject. All models include state fixed effects and standard errors clustered by school. Covariates includes Female, Individual Education Plan, Limited English Proficiency, race/ethnicity, modal age for grade, SES composite, school Title I eligibility, school percent Black, school percent Hispanic, Urbanicity, and NAEP state exclusion proportion. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind, FE=Fixed Effect. +$p < 0.1$, *$p < 0.05$, **$p<0.01$, ***$p<0.001$. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007, 2008-2009.

Table 4. Heterogenous Effects of NCLB by Race/Ethnicity

	Math G4	Math G8	Reading G4	Reading G8
	(1)	(2)	(3)	(4)
NCLB	0.279***	0.060	0.133**	-0.019
	(0.064)	(0.074)	(0.038)	(0.035)
Black X NCLB	0.023***	0.007	0.015**	0.003
	(0.006)	(0.004)	(0.004)	(0.003)
Hispanic X NCLB	-0.001	-0.002	-0.0001	0.002
	(0.005)	(0.005)	(0.003)	(0.008)
Asian X NCLB	0.008	0.001	-0.002	-0.011+
	(0.005)	0.002	(0.004)	(0.006)
N	612,700	567,290	725,710	453,030
F	2,369.70	2,382.25	1,910.02	1,567.53

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Standard errors clustered by state. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. +p < 0.1, *p < 0.05, **p<0.01, ***p<0.001.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Table 5. Heterogenous Effects of NCLB by Quintiles of SES

	Math G4	Math G8	Reading G4	Reading G8
	(1)	(2)	(3)	(4)
NCLB	0.351***	0.147+	0.231***	0.028
	(0.074)	(0.079)	(0.046)	(0.035)
SES Q2 X NCLB	-0.005	-0.003	-0.000	-0.004*
	(0.004)	(0.002)	(0.002)	(0.002)
SES Q3 X NCLB	-0.018**	-0.010***	-0.009*	-0.004+
	(0.006)	(0.003)	(0.004)	(0.002)
SES Q4 X NCLB	-0.024***	-0.008**	-0.012*	-0.008***
	(0.005)	(0.003)	(0.005)	(0.002)
SES Q5 X NCLB	-0.017**	-0.006*	-0.001	-0.002+
	(0.005)	(0.003)	(0.006)	(0.001)
N	612,700	567,290	725,710	453,030
F	2359.41	2389.01	1929.58	1579.15

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Standard errors clustered by state. See Appendix Table B2 for the procedure used to construct the SES quintiles. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. +p < 0.1, *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Table 6. Heterogenous Effects of NCLB by Quintiles of SES and Race

	Math G4	Math G8	Reading G4	Reading G8
	(1)	(2)	(3)	(4)
NCLB	0.334***	0.095	0.145**	0.008
	(0.065)	(0.074)	(0.041)	(0.040)
Black X SES Q2 X NCLB	0.013*	0.005	0.017*	0.004
	(0.005)	(0.006)	(0.007)	(0.008)
Black X SES Q3 X NCLB	0.010**	-0.002	0.013**	0.003
	(0.003)	(0.005)	(0.004)	(0.006)
Black X SES Q4 X NCLB	0.006	-0.006	0.012**	0.002
	(0.004)	(0.005)	(0.004)	(0.009)
Black X SES Q5 X NCLB	0.013**	0.003	0.018**	0.005
	(0.005)	(0.003)	(0.005)	(0.011)
Hispanic X SES Q2 X NCLB	-0.011*	0.002	0.016	-0.007
	(0.005)	(0.009)	(0.009)	(0.012)
Hispanic X SES Q3 X NCLB	-0.029***	-0.016***	-0.006	0.001
	(0.005)	(0.004)	(0.004)	(0.008)
Hispanic X SES Q4 X NCLB	-0.013*	-0.009	-0.006	0.001
	(0.007)	(0.008)	(0.006)	(0.015)
Hispanic X SES Q5 X NCLB	-0.016*	-0.016***	-0.003	-0.005
	(0.008)	(0.004)	(0.007)	(0.011)
Asian X SES Q2 X NCLB	0.001	-0.027***	-0.002	-0.008
	(0.006)	(0.007)	(0.008)	(0.017)
Asian X SES Q3 X NCLB	-0.0001	0.005	0.006	-0.001
	(0.010)	(0.009)	(0.006)	(0.007)
Asian X SES Q4 X NCLB	-0.015+	0.009	-0.010	-0.009
	(0.008)	(0.011)	(0.007)	(0.016)
Asian X SES Q5 X NCLB	-0.006	-0.009	0.007	-0.025*
	(0.010)	(0.012)	(0.009)	(0.010)
N	612,700	567,290	725,710	453,030
F	1,257.52	1,254.84	1,004.96	806.28

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Standard errors clustered by state. See Appendix Table B2 for the procedure used to construct the SES quintiles. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. +p < 0.1, *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Table 7. Heterogenous Effects of NCLB by Urbanicity

	Math G4	Math G8	Reading G4	Reading G8
	(1)	(2)	(3)	(4)
NCLB	0.307***	0.106	0.184***	-0.007
	(0.069)	(0.081)	(0.043)	(0.035)
NCLB X Suburban/Town	-0.004	-0.003+	-0.004	-0.007*
	(0.003)	(0.002)	(0.003)	(0.003)
NCLB X Rural	-0.002	0.004	-0.006***	-0.001
	(0.003)	(0.003)	(0.001)	(0.002)
N	612,700	582,000	742,130	469,800
F	2,400.069	2,516.59	1,996.64	1,683.75

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Standard errors clustered by state. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p<0.01$, $^{***}p<0.001$.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Table 8. Heterogenous Effects of NCLB by School Title I Status

Outcome	Math G4	Math G8	Reading G4	Reading G8
	(1)	(2)	(3)	(4)
NCLB	0.155*	0.059	-0.001	-0.030
	(0.072)	(0.069)	(0.004)	(0.036)
NCLB X Title I	0.007*	0.004	0.133**	0.004*
	(0.004)	(0.004)	(0.043)	(0.002)
N	612,700	567,290	725,710	453,030
F	2456.65	2547.68	2041.08	1693.51

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Standard errors clustered by state. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p<0.01$, $^{***}p<0.001$. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Table 9. Mediation Analysis, NCLB Sanctions

Outcome: NAEP	Math G4	Math G8	Reading G4	Reading G8
NCLB	0.300***	0.095	0.181***	-0.006
	(0.068)	(0.079)	(0.043)	(0.035)
Outcome: Sanctions	Math G4	Math G8	Reading G4	Reading G8
NCLB	0.028***	0.054***	0.032***	0.053***
	(0.007)	(0.010)	(0.008)	(0.008)
Outcome: NAEP	Math G4	Math G8	Reading G4	Reading G8
Sanctions	-0.204***	0.008	-0.028	-0.008
	(0.030)	(0.025)	(0.029)	(0.016)
Outcome: NAEP	Math G4	Math G8	Reading G4	Reading G8
Sanctions	-0.214***	0.019	-0.020	0.004
	(0.031)	(0.021)	(0.030)	(0.016)
NCLB	0.303***	0.094	0.182***	-0.006
	(0.071)	(0.079)	(0.043)	(0.035)
N	725,710	725,710	453,030	453,030

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Sanction data collected from Reback and colleagues (2011) for the 2004 academic year and Comprehensive School Performance Reports for 2004 and 2007 (U.S. Department of Education, 2009). Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. $^{+}p < 0.1$, $^{*}p < 0.05$, $^{**}p<0.01$, $^{***}p<0.001$. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Table 10. Mediation Analysis

Outcome: NAEP	Math G4	Math G8	Reading G4	Reading G8
NCLB	0.300***	0.095	0.181***	-0.006
	(0.068)	(0.079)	(0.043)	(0.035)
Outcome: State Instructional PPE	Math G4	Math G8	Reading G4	Reading G8
NCLB	0.532*	0.682*	0.119	0.471
	(0.205)	(0.273)	(0.176)	(0.381)
Outcome: NAEP	Math G4	Math G8	Reading G4	Reading G8
State Instructional PPE	0.174***	0.120***	0.056***	0.009
	(0.024)	(0.012)	(0.005)	(0.007)
Outcome: NAEP	Math G4	Math G8	Reading G4	Reading G8
State Instructional PPE	0.008	0.019+	0.030+	-0.001
	(0.012)	(0.010)	(0.016)	(0.009)
NCLB	0.296***	0.082	0.178***	-0.006
	(0.071)	(0.081)	(0.042)	(0.034)
N	725,710	725,710	453,030	453,030

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Standard errors clustered by state. State Instructional Per Pupil Expenditures collected from the F33 Survey (Census Bureau & National Center for Education Statistics, 2018). Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. +p < 0.1, *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Appendix Table A1. Analytic Sample by State and Dataset

State	Comparison	Math G4	Math G8	Reading G4	Reading G8
Alabama	X	X	X	X	X
Alaska	X				
Arizona		X	X	X	X
Arkansas	X	X	X	X	X
California	X	X	X	X	X
Colorado					
Connecticut	X	X	X	X	X
Delaware	X			X	X
DC		X	X	X	X
Florida	X			X	X
Georgia	X	X	X	X	X
Hawaii		X	X	X	X
Idaho		X	X		
Illinois	X		X		
Indiana	X	X	X		
Iowa		X		X	
Kansas	X			X	X
Kentucky	X	X	X	X	X
Louisiana	X	X	X	X	X
Maine		X	X	X	X
Maryland	X	X	X	X	X
Massachusetts	X	X	X	X	X
Michigan	X	X	X	X	
Minnesota		X	X	X	
Mississippi		X	X	X	X
Missouri		X	X	X	X
Montana		X	X	X	X
Nebraska		X	X		
Nevada	X	X		X	X
New Hampshire					
New Jersey					
New Mexico	X	X	X	X	X
New York	X	X	X	X	X
North Carolina	X	X	X	X	X
North Dakota		X	X		
Ohio		X	X		
Oklahoma	X	X	X	X	X
Oregon	X	X	X	X	X
Pennsylvania					
Rhode Island	X	X	X	X	X
South Carolina	X	X	X	X	X
South Dakota					
Tennessee	X	X	X	X	X
Texas	X	X	X	X	X
Utah		X	X	X	X
Vermont	X	X	X		

Virginia	X	X	X	X	X
Washington				X	X
West Virginia	X	X	X	X	X
Wisconsin	X				
Wyoming		X	X	X	X

Note: Dee and Jacob (2011) restrict the sample to include only states where pre-treatment outcomes are available for at least two waves within each grade and subject, except for 8^{th} grade reading pre-treatment outcomes where only available in one year. States with accountability prior to NCLB are in the comparison group (T. Dee & Jacob, 2009). DC=District of Colombia.

Appendix Table A2. Event Study

	(1)	(2)	(3)	(4)
Pre-Treatment Effect 1992	0.103	0.044	0.130**	
	(0.067)	(0.041)	(0.038)	
Pre-Treatment Effect 1994			0.132**	
			(0.045)	
Pre-Treatment Effect 1996	0.069*	0.043		
	(0.030)	(0.031)		
Pre-Treatment Effect 1998			0.051	-0.017
			(0.036)	(0.035)
2003 Effect	-0.027	-0.026	0.015	-0.001
	(0.038)	(0.025)	(0.026)	(0.023)
2005 Effect	-0.034	-0.055	-0.002	0.006
	(0.046)	(0.037)	(0.025)	(0.025)
2007 Effect	0.042	-0.071	0.024	-0.022
	(0.040)	(0.043)	(0.024)	(0.020)
N	612,700	567,290	725,710	453,030
F	30788.41	11869.01	15158.94	25287.24

Note: Reference category is the last year that data is available prior to NCLB (i.e., 2000 for math and 2002 for reading). See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Standard errors clustered by state. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. +p < 0.1, *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Appendix Table A3. Balance on Baseline Covariates

Characteristic	Math G4 1992	Math G8 1992	Reading G4 1992	Reading G4 1998
NAEP Score	-0.02720	-0.03030	-0.01550	-0.00770
IEP	0.0001*	0.01570	0.00610	0.00840
LEP	0.143*	0.1439*	0.07510	0.06630
Gender	0.00090	0.0078*	-0.00240	0.00300
White	-0.1075*	-0.1145*	-0.06010	-0.06900
Black	0.07860	0.06880	0.04540	0.05950
Hispanic	0.1122*	0.1323*	0.07170	0.07700
Asian/PI	0.03970	0.04710	-0.01970	-0.03110
Native American	-0.05780	-0.09350	-0.06040	-0.08510
Other Race	-0.06730	0.14340	0.04070	0.02860
Title I	0.04820	0.04890	0.04310	0.06530
At Modal Age	0.05720	0.03210	0.04280	0.03630
Below Modal Age	0.0894*	0.07390	0.04710	0.00840
Above Modal Age	-0.0587*	-0.03430	-0.04360	-0.03660
School Percent Black	0.04270	0.04420	0.02700	0.04110
School Percent Hispanic	0.0762*	0.0756*	0.04700	0.04430
Urban	0.06150	0.05630	0.05660	0.02670
Town/Suburban	0.02370	0.01300	0.00810	0.00030
Rural	-0.1243*	-0.0997*	-0.09890	-0.05000
NAEP Exclusion Rate	0.0997*	0.2069*	0.0754*	0.09970
SFR	0.322*	0.3114*	0.218*	-0.05460
Charter Law	0.09120	0.08900	0.00160	0.19160
Tch Testing Reqs	0.35800	0.36380	0.39480	0.0001
Tch Cert Standards	0.00590	0.00080	-0.01840	0.01430
Math Content Standards	0.06580	0.08090	0.05300	0.00010
ELA Content Standards	-0.11020	-0.08710	-0.869*	-0.1521*
HS Exit Exams	0.07570	0.03710	0.14020	0.05110
HS Graduation Reqs	0.01930	0.00050	-0.03750	-0.19670
State Inst PPE	0.23060	0.22740	0.16270	0.15380

Note: All estimates standardized within grade and subject. Estimates adjusted using NAEP student-level probability weights. NAEP= National Assessment of Educational Progress test score standardized within grade-subject, IEP=Individualized Education Plan, LEP=Limited English Proficiency, PI=Pacific Islander, HS=High School, FRPL=Free and Reduce Price Lunch, SES=Socio Economic Status, Pct=Percent, SFR=School Finance Reform, Tch=Teacher, Reqs=Requirements. Policy data collected from various sources: SFR (Jackson et al., 2015), Charter Law (National Alliance for Public Charter Schools, 2016), Teacher Certification Testing (NCES, 2016), Math Content Standards (NCES, 2016), ELA Content Standards (NCES, 2016), Teacher Certification Standards (Cavell et al., 2005), High School Exit Exam (Warren & Kulick, 2007), HS graduation requirement (The Council of State Governments, 2015). Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Appendix Table A4. State Specific Linear Trends

	(1)	(2)	(3)	(4)
NCLB	0.305***	0.044	0.151*	-0.001
	(0.061)	(0.071)	(0.056)	(0.038)
N	662,320	590,420	749,780	471,710

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects, state specific linear trends, and covariates. Standard errors clustered by state. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. +p < 0.1, *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Appendix Table A5. Robustness to Time Variant State Education Policies

Policy	Math G4	Math G8	Reading G4	Reading G4
Growth Score Waiver	0.2808*	0.1105	0.1808*	0.0066
Proficiency Standard	0.3017*	0.0894	0.1317*	-0.0368
SFR	0.2994*	0.0937	0.1818*	-0.0077
Charter Law	0.2993*	0.0936	0.1851*	-0.0027
Tch Testing Reqs	0.2968*	0.0973	0.182*	-0.0036
Tch Cert Standards	0.3293*	0.1091	0.185*	-0.0204
Math Content Standards	0.2973*	0.0925	0.1811*	-0.0036
ELA Content Standards	0.3022*	0.094	0.1834*	-0.0046
Fully Day K	0.3012*	0.0948	0.1821*	-0.0032
HS Exit Exams	0.2965*	0.0922	0.1682*	-0.01
HS Graduation Reqs	0.313*	0.1004	0.1874*	-0.0111
Annual Tch Eval	0.3012*	0.0945	0.1874*	-0.0068
Common Assessments	0.304*	0.0942	0.1832*	-0.0039
Statewide Data System	0.2988*	0.0966	0.1836*	0.0048
Data System with Identifiers	0.3014*	0.0854	0.1807*	-0.008
Evaluation Firing	0.2995*	0.0943	0.1834*	-0.0023
Eval PD	0.2866*	0.1066	0.1965*	0.007
Eval compensation	0.2981*	0.0925	0.1837*	-0.0034
Eval Responsibility	0.2912*	0.069	0.1827*	-0.0023
Eval Grant Tenure	0.2995*	0.0943	0.1834*	-0.0023
Charter Authorizer	0.2958*	0.0922	0.1842*	-0.0052
Charter Building Funds	0.2758*	0.0934	0.159*	0.0045
Charter Cap	0.2947*	0.0835	0.1816*	-0.0055
School Turnaround	0.2973*	0.0977	0.1575*	-0.0358
Evaluation Growth Targets	0.3139*	0.1028	0.1814*	-0.006
Alt Preparation Programs	0.3033*	0.0946	0.1848*	-0.0032
Vouchers	0.3084*	0.0922	0.1869*	0.0114
High School Exit Exams	0.3269*	0.0696	0.1607*	-0.0086
Testing Grades 3-8	0.303*	0.0922	0.2*	0.0274
State Takeover of Districts	0.3008*	0.0925	0.1743*	0.019
Universal Pre-School	0.2992*	0.1006	0.1857*	-0.0073
School Choice	0.282*	0.0979	0.174*	-0.0054
Education Strategic Plan	0.2927*	0.077	0.1745*	-0.0027

Note: All estimates standardized within grade and subject. Estimates adjusted using NAEP student-level probability weights. NAEP= National Assessment of Educational Progress test score standardized within grade-subject, IEP=Individualized Education Plan, LEP=Limited English Proficiency, PI=Pacific Islander, HS=High School, FRPL=Free and Reduce Price Lunch, SES=Socio Economic Status, Pct=Percent, SFR=School Finance Reform, Tch=Teacher, Reqs=Requirements. Policy data collected from various sources: SFR (Jackson et al., 2015), Charter Law (National Alliance for Public Charter Schools, 2016), Teacher Certification Testing (NCES, 2016), Math Content Standards (NCES, 2016), ELA Content Standards (NCES, 2016), Teacher Certification Standards (Cavell et al., 2005), High School Exit Exam (Warren & Kulick, 2007), HS graduation requirement (The Council of State Governments, 2015), all other policies from Howell and Magazinnik (2017). Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Appendix Table A6. CITS Effects Estimates without Dosage Framework

	(1)	(2)	(3)	(4)
NCLB	0.045**	0.011	0.012+	-0.019*
	(0.014)	(0.019)	(0.006)	(0.009)
N	662,320	590,420	749,780	471,710
F	4,701.36	4,804.93	3,051.20	2,2079.56

Note: See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Standard errors clustered by state. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. +p < 0.1, *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Appendix Table B1. Merged Education Policies

Policy	Description	Citation
Growth Waivers	Waiver to use growth measures instead of proficiency measures for school accountability under NCLB	Hoffer et al., 2011
Proficiency Standards Rigor	The cuts score for proficiency mapped onto the NAEP scale	Erpenbach et al., 2003; Erpenbach & Forte, 2005, 2007; Fast & Erpenbach, 2004; Forte & Erpenbach, 2006
ESEA Waivers	Flexibility under NCLB	Center on Education Policy, 2018
School Finance Reforms	Based on first school finance reform order	Jackson et al., 2015
Charter School	Authorizing charter schools	National Alliance for Public Charter Schools, 2016
Teacher Certification Testing	State requires test for teacher certification	NCES, 2016
Teacher Certification Standards	State outlines requirements for teacher certification process	Cavell et al., 2005
ELA Content Standards	State has ELA content standards	NCES, 2016
Math Content Standards	State has math content standards	NCES, 2016
High School Exit Exam	State requires students to exam to graduate high school	Warren & Kulick, 2007
High School Graduation Requirement	State specifies any high school graduation requirements	Council of State Governments, 2015
Teacher Evaluation	Teacher evaluation that includes student data	Bleiberg & Harbatkin, 2018
Bush/Obama Reforms	Popular state education reforms from 2000 to 2014	Howell and Magazinnik, 2017

Appendix Table B2. Procedure for Constructing the Socio-Economics Status Composite

Estimates	Math G4	Math G8	Reading G4	Reading G8
NCLB effect controlling for FRPL	0.145+	0.209*	-0.027	-0.032
	(0.078)	(0.094)	(0.058)	(0.043)
NCLB effect controlling for SES Factor	0.208**	0.191*	0.004	-0.002
	(0.074)	(0.094)	(0.059)	(0.045)
NCLB effect controlling for FRPL and other covariates	0.265***	0.178+	0.030	-0.005
	(0.073)	(0.093)	(0.073)	(0.033)
NCLB effect controlling SES Factor and other covariates	0.264***	0.180+	0.033	0.002
	(0.072)	(0.097)	(0.069)	(0.032)
N	504,910	463,680	611,490	611,490

Note: The SES factor was estimated separately for each grade-subject. In each model whether the student attended a Title I school, eligibility for FRPL, availability of cultural capital (i.e., newspapers, encyclopedias, books, magazine), and urbanicity. In grade 8, I also included parent's highest level of education: no high school degree, high school degree, some college, college degree or more. See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. NAEP restricted data outcomes standardized within grade and subject. See Table 3 for a full list of covariates. All models include state fixed effects and covariates. Standard errors clustered by state. See Appendix B for the procedure used to construct the SES quintiles. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. FRPL=Free and Reduced-Price Lunch, NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. +p < 0.1, *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

Appendix Table C1. CITS Estimates with Dosage Framework and Covariates

	Math G4	Math G8	Reading G4	Reading G8
	(1)	(2)	(3)	(4)
Individualized Education Plan	-0.703***	-0.974***	-0.817***	-0.977***
	(0.019)	(0.021)	(0.023)	(0.021)
Limited English Proficiency	-0.414***	-0.625***	-0.543***	-0.728***
	(0.020)	(0.030)	(0.018)	(0.032)
Derived Sex	-0.110***	-0.111***	0.150***	0.241***
	(0.006)	(0.005)	(0.004)	(0.006)
Black	-0.552***	-0.616***	-0.383***	-0.518***
	(0.017)	(0.018)	(0.020)	(0.014)
Hispanic	-0.336***	-0.399***	-0.266***	-0.358***
	(0.020)	(0.016)	(0.024)	(0.019)
Asian/Pacific Islander	0.133***	0.230***	0.093*	0.032
	(0.027)	(0.035)	(0.040)	(0.026)
American Indian	-0.280***	-0.367***	-0.254***	-0.347***
	(0.018)	(0.028)	(0.019)	(0.029)
Other Race	-0.042**	-0.166***	0.003	-0.117***
	(0.014)	(0.012)	(0.024)	(0.018)
Below modal age	0.223**	0.310***	0.114***	0.175***
	(0.072)	(0.020)	(0.021)	(0.028)
Above modal age	-0.081***	-0.186***	-0.077***	-0.159***
	(0.016)	(0.015)	(0.012)	(0.018)
Title 1	-0.269***	-0.174***	-0.248***	-0.148***
	(0.016)	(0.023)	(0.019)	(0.017)
School Lunch	0.392***	0.320***	0.359***	0.306***
	(0.013)	(0.016)	(0.013)	(0.015)
School Percent Black	-0.003***	-0.004***	-0.005***	-0.003***
	(0.000)	(0.000)	(0.000)	(0.000)
School Percent Hispanic	-0.001	-0.002*	-0.002**	-0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Urban/City	0.087***	0.094***	0.074***	0.053**
	(0.015)	(0.014)	(0.013)	(0.016)
Town/Suburban	0.061***	0.073***	0.068***	0.015
	(0.011)	(0.020)	(0.013)	(0.011)
Year	0.058***	0.065***	0.036+	0.013*
	(0.011)	(0.011)	(0.020)	(0.005)
NCLB	0.170**	-0.021	-0.128**	-0.116+
	(0.060)	(0.028)	(0.042)	(0.068)
Years Since NCLB	-0.027	-0.007	0.000	0.009
	(0.024)	(0.024)	(0.026)	(0.012)
T × Year	-0.004*	-0.005***	0.002	0.001
	(0.001)	(0.001)	(0.002)	(0.001)
T ×NCLB	0.016+	0.018**	0.013*	0.012
	(0.009)	(0.005)	(0.005)	(0.008)

T × Years Since NCLB	0.006*	0.002	-0.002	-0.003+
	(0.003)	(0.003)	(0.003)	(0.001)
NAEP Exclusion Rate	-0.106	-0.298	3.062+	2.160
	(1.331)	(1.471)	(1.697)	(1.336)
NAEP Exclusion Rate^2	-10.853+	3.048	-22.717+	-13.268
	(5.429)	(13.360)	(12.482)	(11.802)
Constant	-0.355*	-0.181+	-0.559***	-0.228**
	(0.154)	(0.094)	(0.095)	(0.069)
N	612,700	567,290	725,710	453,030
F	4,701.36	4,804.93	3,051.20	2,2079.56

Note: See Dee and Jacob (2009; 2010) for information on the construction of the variables used to estimate the Comparative Interrupted Time Series. In the CITS dosage framework the effect of NCLB is 6 times the intercept shift plus 30 times the slope shift where 6 is the number of years from 2007 (the last year of outcomes) minus 2001 the last year a school adopted school accountability prior to NCLB and 5 is 2007 minus 2002 (the last year prior to treatment): NCLB Effect=(6 × [T × NCLB])+(5×6×[T × Years Since NCLB]). See Appendix Table A1 for the states assigned to the treatment group and sample restrictions by grade and subject. The author replicated the estimates from Dee and Jacob (T. S. Dee et al., 2010) Table 1 Row 1. NAEP restricted data outcomes standardized within grade and subject. All models include state fixed effects and standard errors clustered by school. Covariates includes Female, Individual Education Plan, Limited English Proficiency, race/ethnicity, modal age for grade, SES composite, school Title I eligibility, school percent Black, school percent Hispanic, Urbanicity, and NAEP state exclusion proportion. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress, NCLB=No Child Left Behind. +$p < 0.1$, *$p < 0.05$, **$p<0.01$, ***$p<0.001$. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 1989-1990, 1991-1992, 1993-1994, 1995-1996, 1997-1998, 1999-2000, 2001-2002, 2002-2003, 2004-2005, 2006-2007.

CHAPTER 3

DID REVOKING NCLB REGULATIONS "WAIVE" IN BETTER STUDENT OUTCOMES?

Introduction

In 2002, No Child Left Behind (NCLB) set the ambitious goal that students in every school would reach proficiency in math and reading by 2014. NCLB was scheduled for reauthorization in 2007, but Congress remained gridlocked over education reform issues (Saultz et al., 2016). Congressional inattention left every state on a trajectory to have very high numbers of schools failing to meet Adequate Yearly Progress (AYP) (K. K. Wong, 2015). In the 2010-2011 school year, 52 percent of schools were failing to meet AYP and about 1 in 5 schools were facing sanctions (U.S. Department of Education, 2018a). The Obama administration developed an innovative policy solution. The Education Department (ED) offered waivers allowing states to develop their own testing goals. In return for receiving the waivers states committed to a slate of education reforms and continued focus on the closure of achievement gaps. The effect of providing states greater flexibility over school accountability on student achievement remains unknown (McGuinn, 2016). I examine the waiver's deregulatory approach to national standards-based reform on student outcomes. I also study whether a larger state role benefitted academically vulnerable students.

Previous studies have focused on school sanctions under the waivers. There is mixed evidence that school sanctions under the waivers (i.e., priority, focus) influenced student outcomes overall (Bonilla & Dee, 2020, Dee & Dizon-Ross, 2019; Dougherty & Weiner, 2017; Hemelt & Jacob, 2018). No study has tested the effect of the waivers on achievement gaps. The literature on the waivers has also not examined the effects of the waivers on schools that did not face sanctions.

I estimate the causal effect of waiver receipt on National Assessment of Education Progress (NAEP) scores in a difference-in-differences framework. I identify the effect of waiver receipt on student achievement by comparing the first group of states to receive waivers (2012-2013) to states that either received waivers in the next year (2013-2014) or never received a waiver. During the period of study (2003 to 2013) treated states that received waivers in 2013 were given the flexibility to decrease the number of schools facing sanctions, while comparison states (i.e., received waiver in 2014 or never received waivers) operated under the NCLB school accountability system that mandated mass-sanctioning of schools. My strategy compares the less intensive or deregulatory approach to school accountability under the waivers to the more stringent or rules burdened approach under NCLB. The major barrier to obtaining causal effects is the selection bias resulting from the Obama administration granting waivers to states that adopted their preferred education reforms. I account for the resulting selection bias by restricting the sample to include only states that adopted the policies incentivized by the waivers. My approach identifies the causal effect of relaxing school accountability requirements on student achievement overall and achievement gaps.

I find no evidence for the hypothesis that the waivers influenced student achievement overall. Although, the waivers appear to shrink white-Black and white-Hispanic achievement gaps in reading. The outcomes for FRPL eligible students decreased slightly as a result of the waivers. The waivers were the best available option for the Obama administration given the Congressional inattention to NCLB. If NCLB had remained in place, the number of schools under sanction would have overwhelmed state capacity (Jochim & Murphy, 2013). The waivers succeeded in decreasing the number of sanctions while keeping national school accountability in place. Although, the evidence that relieving accountability pressures influenced student achievement overall is weak, states did continue to make progress on the closure of race-based achievement gaps. The waiver model of school accountability where the federal government provides states flexibility to target schools appears to work better than the stricter NCLB model.

Conceptual Framework

Federal agencies and departments use waivers to provide flexibility to local governments (Barron & Rakoff, 2013). States vary in their administrative, political, and fiscal capacity to improve schools. In the past the ED has provided flexibility in the form of waivers when it was clear that states were not able to comply with the law. For example, in 2005, the ED granted waivers to states allowing the use of growth models because of criticisms about the accuracy of NCLB's testing requirements (Weiss & May, 2012).

Waivers also increase the level of control that states have over policy (Bowling & Pickerill, 2013; K. K. Wong, 2015). From a governance perspective, waivers are an interactional form of federalism where the President takes power from Congress and returns it to the states (Gais & Fossett, 2005). Waivers can only remove requirements. Without state compliance waivers will have no effect on reforms, which makes waivers somewhat ill-suited for compelling changes to sub-national education policies. The removal of regulations could enable states to adopt innovative education reforms. State education leaders that the waivers would improve the identification of schools needing improvement (McMurrer & Yoshioka, 2013). If federal regulations were preventing states from implementing effective education policies then waivers freeing states from requirements could benefit students.

To receive a waiver states also had to comply with four principles. These principles included: (1) College- and Career-Ready Expectations for All; (2) State-Developed Differentiated Recognition, Accountability, and Support; (3) Supporting Effective Instruction and Leadership; (4) Reducing Duplication and Unnecessary Burden (U.S. Department of Education, 2011). The easiest way for states to demonstrate compliance with the waiver principle was adopting the Obama administration's preferred education reforms (e.g., content standards, teacher evaluation, principal evaluation, and school turnaround/accountability).

The clearest plausible pathway for the waivers to influence student outcomes is through changes to accountability systems. By 2011, NCLB required states to set test score goals (i.e., Annual Measurable Objectives (AMO)) at close to universal proficiency. The unreasonably high AMOs may have put undue pressure on schools or caused them to ignore the targets all together. Under the waivers, states were again permitted to choose goals that were high but achievable. States could also focus improvement efforts on a smaller percentage of schools. The capacity of states to turnaround failing schools is extremely limited (Jochim & Murphy, 2013; Murphy & Bleiberg, 2018). The waivers allowed states to greatly reduce the number of schools that received sanctions (Hyslop, 2013). Under the waivers, states issued two types of sanctions. Priority schools scored in the bottom 5% of achievement and focus schools were either in the bottom 10% of achievement or had large achievement gaps. The waivers let states develop their own formulas to assess school achievement. Both new policies gave states more control over sanctions than was permitted under NCLB. The improved sanction targeting may have benefitted students under state waiver accountability systems.

Accountability systems under the waivers differed from NCLB in terms of how academically vulnerable students were considered. Under NCLB, to meet AYP, schools had to meet proficiency targets for every sub-group. Under the waivers, schools had to identify which groups were meeting targets, but interventions were not automatic if sub-groups failed to meet targets. Under the waivers without accountability requirements for sub-groups there is a concern that states would lower expectations for academically vulnerable students. Chubb and Clark argued, "The evidence indicates…that left to their own devices, the states will exacerbate the nation's achievement gap between haves and have-nots" (2013).

Another reason to remain skeptical of the waivers is the school improvement strategies implemented within sanctioned schools. Some states made changes for how to identify failing schools (Polikoff et al., 2014). However, states frequently used the same school improvement strategies to turnaround failing schools under both NCLB and the waivers (Hyslop, 2013). The essence of school turnaround under the waivers looked very similar to previous efforts under NCLB.

Effects of Waivers

Several studies have examined the effect of school turnaround under the waivers. These studies rely on the scores that states used to assign sanctions as the forcing variable in a Regression Discontinuity (RD) design. In Rhode Island, Louisiana, and Michigan, schools that were sanctioned under the waiver system had indistinguishable differences from comparable schools that were not sanctioned (Dee & Dizon-Ross, 2019; Dougherty & Weiner, 2017; Hemelt & Jacob, 2018). In Kentucky, school sanctions led to sizable positive improvements in both math and reading test scores (Bonilla & Dee, 2020).

Hopkins (2019) uses district-level data from the Stanford Education Data Archive (SEDA) to examine whether the waivers increased achievement overall and reduced racial achievement gaps. He finds that the waivers had no effect on average. Achievement among Black and Hispanic students increased in states that received waivers, but the change in district level achievement gaps was insignificant.

Marsh and colleagues (2016) examined how the waivers changed school accountability systems in California. They found that school leaders thought the more holistic approach to identifying struggling schools was an improvement over measures that were exclusively test based. School and district leaders also approved of the emphasis on supporting rather than sanctioning schools. The reports from school leaders help to contextualize why the transition from NCLB to the waivers may have benefitted students.

Contribution

I make two primary contributions to the growing literature on the waivers. First, I employ an approach with a strong claim to external validity and second, I estimate differential effects on academically vulnerable students. Previous studies have used Regression Discontinuity designs to examine school turnaround under the waivers. The generalizability of the RD studies is limited to schools that received sanctions under the waivers (i.e., focus, priority) and similar schools that almost received sanctions. In the RD studies, schools that were not close to receiving sanctions do not contribute to the Local Average Treatment Effect estimates. A weakness of the RD studies is that the waivers also influenced schools that had average outcomes and were not close to receiving sanctions. Without a waiver the preponderance of schools in the middle two quartiles would have received sanctions under NCLB. I estimate the effect of the waivers on all schools. My approach allows for an inference about relinquishing accountability pressure, which is an important component of the waiver reform. The RD studies use administrative data and their inferences are limited to the confines of those states (i.e., Kentucky, Louisiana, Michigan, Rhode Island). Studying the average effect of the waivers across all states is valuable because the waivers influenced students in every state. The waivers were also a less stringent policy than NCLB and provided considerable flexibility to states in the administration of school accountability systems. The proportion of schools that states sanctioned and rewarded varied considerably (Hyslop, 2013). The states that received waivers (including Kentucky, Louisiana, Michigan, Rhode Island) have a history of intensive education reforms. The unique approach to school accountability under the waivers in these 4 states in part explains their findings.

Finally, my study is the first to estimate student level achievement gaps. Previous studies of the waivers have either not examined differential effects across populations of vulnerable students (Dee & Dizon-Ross, 2019; Dougherty & Weiner, 2017; Hemelt & Jacob, 2018) or studied aggregate school or district level differences (Bonilla & Dee, 2020; Hopkins, 2019). Bonilla and Dee (2020) examine school average heterogenous effects for a "super group" of academically vulnerable students. My approach allows me to test whether the effects of the waivers varied across populations of academically vulnerable students. I highlight the differential effects of the waivers on academically vulnerable students because the primary success of NCLB was calling attention to achievement gaps. The waivers do not require states to sanction schools if a single subgroup fails to meet achievement targets. Estimating differential effects for each academically vulnerable group allows me to test if the new approach to school accountability influenced the size of achievement gaps.

My study has a few important differences with Hopkins (2019). His analysis uses a district-level threat index to estimate the likelihood that schools in a given district would have faced sanctions under the waivers. In the NAEP I observe each students school and whether or not the school received a sanction. Observing school-level sanction data allows me to directly estimate the effect of school sanctions and improves the precision of my estimates. The SEDA is missing for a variety of reasons including low test participation, changes to assessments, and states not submitting test data to ED (Fahle et al., 2019). For example, data is missing for 15 states in 2014 due to low participation. It is possible that states which received waivers may have also had lower test participation rates because of changes to accountability systems which could have also influenced student outcomes. Fortunately, the NAEP includes data from students in every state during the period a study.

Research Questions

I ask the following questions:

1. To what extent did waiver receipt influence student outcomes overall and the size of achievement gaps?
2. Did waiver receipt change school accountability systems?
3. To what extent did the waiver sanctions (i.e. priority, focus) influence student outcomes?

Data, Measures, and Sample

I use data from the National Assessment of Educational Progress (NAEP). I use data from four subject/grade datasets (4th grade math, 8th grade math, 4th grade reading, and 8th grade reading) over six odd numbered years from 2003 to 2013. The NAEP study uses a complex three-stage sampling design to allow for valid inferences about student achievement outcomes for each state (Rogers et al., 2014). The NAEP assesses a broader set of skills than the average state summative assessment. NAEP relies on committees of subject matter experts, practitioners, researchers, educators, business leaders, and policymakers to write the frameworks used to develop the NAEP test items. The NAEP data have several unique advantages within the context of studying the waivers. The NAEP is designed to measure trends in state-level education outcomes across time and enables the study of state-level education reforms like the waivers. The NAEP includes academically vulnerable students (e.g., race/ethnicity, class, students with disabilities), which allows me to estimate the effect of the NCLB waivers on a broad set of intersectional achievement gaps. Cheating or gaming may influence summative tests used for accountability purposes (D. Koretz, 2017; D. Koretz & Hamilton, 2006; D. M. Koretz & Barron, 1998). Fortunately, the NAEP is not used for accountability purposes by states and is likely unaffected.

I merged school-level data on AYP and school sanctions into the NAEP. I use the Comprehensive School Performance Reports (CSPR) (U.S. Department of Education, 2018b) to determine which schools were facing sanctions. In 2003, schools were notified of their sanction status but not required to make changes (Murphy & Bleiberg, 2018). To impute sanctions in 2005, I used the 2004 CSPR report combined with 2003 and 2004 AYP data (Reback et al., 2013).[5] Schools under sanction in 2004 that then failed to meet AYP in that same year would remain under sanction in 2005. Schools that met AYP targets in both 2003 and 2004 were not under sanction in 2005. I use the CSPR to identify school sanctions in 2007, 2009, 2011, 2013, and 2015.[6]

I merge in binary measures of state accountability system features, including state Annual Measurable Objectives (AMO) from Ed Data Express (2018a), Growth waiver receipt (Hoffer et al., 2011), NCLB accountability system features (e.g., safe harbor provision, AMO timeline structure) (Erpenbach et al., 2003; Erpenbach & Forte, 2005, 2007; Fast & Erpenbach, 2004; Forte & Erpenbach, 2006) and a continuous measure of proficiency standard rigor (U.S. Department of Education, 2020). I also merge in data on Obama era education policies that states were incentivized to adopt via the waivers (i.e., content standards, teacher evaluation, principal evaluation, and school turnaround/accountability) (Bleiberg & Harbatkin, 2018; Howell & Magazinnik, 2017), in addition to pre-waiver school accountability policies (Dee & Jacob, 2009; Edweek, 2011).

Dependent Variable

To construct my outcomes of interest, I rely on test score information from nine waves of the NAEP (2003-2015). The NAEP is a matrix-based assessment in which each student completes a sample of test items. The NAEP provides plausible values that are created through an Item Response Theory (IRT) procedure. The NAEP includes multiple plausible values to allow the analyst to account for the uncertainty that a student would have received a specific score if they took the entire exam. I use the first standardized plausible value as the dependent variable. Using the first plausible value should produce results similar to other approaches (e.g., multiple imputation

---

[5] No CSPR was produced for 2005. The National Adequate Yearly Progress and Identification (NAYPI) database includes school sanctions for the year 2005. However, the data is not available online and AIR was not able to locate it after multiple requests.

[6] Appendix C describes the process for identifying school sanctions in greater detail.

framework, averaging plausible values) (Jerrim et al., 2017). The results are robust to these approaches because variation in each plausible value is approximately the same. I standardize the NAEP scores within grade, subject, and year to have a mean of 0 and a standard deviation of 1.[7]

Independent Variables

To construct my treatment indicator, I determined the month that states first received their waivers based on data from the Center on Education Policy (2018), ED (2016), and CSPR (U.S. Department of Education, 2018b). The ED documents include all waiver applications and subsequent response (e.g., approval, termination, warning status). The CSPR lists whether schools in each state received NCLB sanctions (e.g., School Improvement, Corrective Action) or NCLB waiver sanctions (i.e., priority, focus, reward) in a given school year. The Center for Education Policy (2018) analyzed the ED documents and published the dates when states received their waiver. I replicated the Center for Education Policy analysis with the ED documents. I assumed that any state which received a waiver before the start of the school year in that state had implemented the waiver for that year. In the NAEP, the first year in which I observe states under the waivers is 2013. Among the group of treated states only Idaho had not received their waiver before the start of the 2013 school year. Idaho received their waiver early in the 2013 school year (October) and I consider them treated because they were operating under the waiver for the vast majority of the school year.[8] Once received implementing the waivers was relatively simple because no changes to state laws or additional resources were required (McMurrer & Yoshioka, 2013). The ED documents operationalize waiver adoption and the CSPR reports measure waiver implementation. Each data source independently implies the same waiver start date.

---

[7] The means of the NAEP test scores are different than zero in the analytic sample due to listwise deletion.
[8] The assumption of instantaneous treatment is consistent with how waivers were implemented in Tennessee and other states (Evan Kramer, personal communication, 2019).

The waiver program was announced in September 2011 and ED issued the first waivers to states in February 2012. waivers were then issued in rounds several times a year (Center on Education Policy, 2018). ED started to issue waiver renewals in 2014. In December, 2015 Congress passed the Every Student Succeeds Act (ESSA) replacing both NCLB and the waivers. In the NAEP, which includes only odd numbered years, the first year I observe treated states is 2013. Thirty-four states receive waivers prior to the start of the 2012-13 school year. 9 states received waivers prior to the 2013-14 school year and 8 states never receive a waiver. I consider the 34 states that have waivers by 2013 to be treated and the remaining 17 serve as my comparison group. Appendix Table 1 describes the treatment and comparison groups by state.[9]

I primarily use a binary measure of waiver receipt as my independent variable, but I also test the robustness of my findings to a continuous independent variable (i.e., percent of schools sanctioned). The key tradeoff between these measures is that the binary variable does not measure the full variation in school accountability. Under NCLB some states sanctioned far more schools than others (V. C. Wong et al., 2018). One strength of the continuous independent variable is that it measures changes in the level of school accountability for the comparison group.

I estimate some models using a sample that includes only states that implemented policies incentivized by the Obama waiver program, which helps address a plausible source of selection bias. After the sample restriction there are 20 states in the treatment group and 6 in the comparison group (i.e. Hawaii, Illinois, Maine, Pennsylvania, West Virginia, and Wyoming). In the matched comparison sample, all treated states received waivers in 2013 and all comparison states received waivers by 2014 except for Wyoming (See Appendix Table A1). I also examine the effects of the waivers in 2015. A serious tradeoff with estimating the effect of the waivers in 2015 is that the states

[9] All results are robust to the exclusion of 8 states that never receive a waiver.

that never receive waivers were systematically less likely to implement the policies incentivized by the waivers and make a poor counter-factual.

Covariates

The NAEP student survey contains a robust set of student characteristics. I control for exogenous student characteristics including gender, whether the student has an Individualized Education Plan (IEP), Limited English Proficiency (LEP) status, eligibility for Free or Reduced-Price Lunch (FRPL), and race/ethnicity. I also add measures for whether the student is at, above, or below the modal age for their grade level. The covariates control for observable differences between the students in states that received waivers by 2013 and students in comparison states (i.e. states that received waivers in 2014 and never received a waiver) that are correlated with student outcomes. I also include a baseline measure of school achievement (AYP status in 2003) to control for pre-treatment differences in student outcomes.

Sample

Table 1 describes the states, districts, schools and students in the analytic sample. The 4th grade (math and reading) datasets include about 900,000 students for all years from 2003 to 2013 and the 8th grade datasets (math and reading) include about 780,000 students. I observe about 140,000 students for each grade, subject, and year. The analytic sample includes students from every state and the District of Columbia and samples students from about 3,000 districts and 5,000 schools in each grade, subject, and year.

Table 2 describes mean characteristics for each year during the period of study. Student outcomes and other observable characteristics change relatively little across time, which implies that cohort composition is unlikely to bias the effect of the waivers. In addition, the overall trend in NAEP outcomes is fairly flat from 2003 to 2013. There was an increase in the number of students eligible for FRPL, likely due to the Great Recession. The sample is also about 10 points less white in 2013 than in 2003 which is consistent with the national change in the demographics of students during the period of study.

Estimation Strategy

I estimate the causal effect of receiving an NCLB waiver on student achievement in a difference-in-differences framework. I identify effects by comparing students in states that received waivers for the 2012-2013 school year to students in states that either received waivers in the next year (2013-2014) or never received a waiver. I begin by estimating a series of models that assume the following general form:

$$(1)\quad y_{icst} = \beta_1 Waiver_s \times\ 2013_t + \rho \boldsymbol{F'}_{it} + \tau \boldsymbol{G'}_{ct} + \alpha_s + \pi_t + e_{icst}$$

Where *y* is a NAEP test score (standardized within subject, grade, and year) for student *i*, school *c*, state *s*, and in year *t*. *waiver ×2013* is a binary variable equal to 1 if a state has received a waiver in 2013. β_1 is the coefficient of interest, the effect of receiving a waiver on NAEP scores within states that received waivers in 2013. ***F*** and ***G'*** are vectors of time-varying student and school covariates. α_s is a vector of either state or school district fixed effects. π_t is a year fixed effect and *e* is an idiosyncratic error term clustered by school.[10] I estimate each model 4 times using each of the NAEP datasets (4th grade math, 8th grade math, 4th grade reading, 8th grade reading). I also add interactions between the treatment indicator, membership in race/ethnic groups, and eligibility for FRPL. These models identify the effects of the waivers for academically vulnerable groups of students.

---

[10] Following Abadie, Athey, Imbens, & Wooldridge (2017) I cluster my standard errors at the school level because the errors of students in schools are correlated due to the IRT procedure employed by NAEP. In addition, clustering at the school level is appropriate because there are schools in the population that I do not observe in the sample.

To answer the second research question, I explore differences between school sanctions under NCLB and the waivers. I modify equation 1 to test whether the waivers were correlated with changes in the number of schools that received sanctions and rewards. In addition, I specify a series of models following equation 2, where I restrict the sample to include only sanctioned schools. I then dynamically estimate the effect of the waiver sanctions (i.e., focus, priority).

$$(2)\quad y_{icst} = \beta_1 Focus_c \times 2013_t + \beta_2 Priority_c \times 2013_t + \rho \boldsymbol{F}'_{it} + \tau \boldsymbol{G}'_{ct} + \alpha_s + \pi_t + e_{icst}$$

Focus × *2013* and *Priority* × *2013* are binary variables equal to 1 if a school received a specified waiver sanction in 2013. β_1 is the effect of focus schools on student outcomes relative to students in NCLB sanctioned schools and β_2 is the effect of priority schools on student outcomes relative to students in NCLB sanctioned schools. The comparison group for these models includes schools in any stage of the NCLB improvement process (i.e., school improvement, corrective action, restructuring).

Threats to Causal Inference

The assumption required for estimating a causal effect in the difference-in-differences approach is that outcomes for students in states that received waivers in 2013 would have followed the same trajectory as students in comparison states that received waivers in 2014 and never received waivers in the absence of treatment. If the pre-treatment NAEP outcomes for the treatment and comparison groups appear to have different trends then the assumption of parallel trends is violated. Figure 2 presents the trends for the treatment and comparison groups for each of the 4 grade-subjects. The pre-treatment trends for the treatment and control groups are all approximately flat. There is a slight downward trend in 8th grade math outcomes for the treated group and a slight upward trend for the comparison group. The trend in 8th grade math scores is consistent with the Granger test of the pre-treatment trends in Appendix Table A2. These models test whether the outcomes for the treatment and comparison groups are significantly different, prior to treatment.

The significant pre-treatment coefficients for the 8th grade math sample, imply that the treatment and comparison groups have different trajectories prior to any state receiving a waiver. The visual evidence from Figure 2 and the Granger test suggests there are no clear violations of parallel trends for 4th grade math, 4th grade reading, or 8th grade reading.

The most salient barrier to obtaining unbiased estimates stems from the process the Obama administration used to grant waivers. The Obama administration explicitly considered the education reforms that states had adopted when granting waivers (K. K. Wong, 2015). If the policies that the waivers incentivized were correlated with student outcomes, then it will bias the effect of the waivers. For example, the Obama administration considered whether states had rigorous content standards and assessments. If those reforms influenced student achievement they would confound the effect of the waivers.

To account for the threat of selection bias, I restrict the sample to include only states that implemented policies that were incentivized by the NCLB waivers. More specifically, I restrict the sample to include only states that belong to a testing consortium, had committed to adopting College and Career Ready Standards, were developing high quality assessments, were providing assistance to low-performing schools (non-Title I), and evaluated teachers and principals (using multiple measures including student data). Restricting the sample is conceptually similar to exact matching on education policies. After the sample restriction the treatment and comparison states have adopted the same waiver incentivized policies. In Appendix Tables 3, 4, and 5, I show that using the matched comparison sample with state fixed effects eliminates the selection bias from the incentives to adopt the waiver policies and mitigates the selection bias from differences in student characteristics.

Balance on Standards-Based Reforms

Appendix Table 3 includes the results from a series of state-level bivariate models. I regress a binary measure of whether a state had a waiver in 2013 on school accountability characteristics and other education policies measured prior to treatment. The results show significant differences between the teacher and principal evaluation systems of states that received waivers in 2013 and those in the comparison group. Changes to teacher and principal evaluation were one of the policies that were incentivized by the waivers. The threat of selection bias motivates my use of the matched comparison sample. Teacher and principal evaluation were two of the variables that I used to construct my matched comparison sample. Within the matched comparison sample, there are no baseline differences in standards-based reforms. States that had waivers in 2013 were also more likely to have consequential school accountability prior to NCLB. But, the differences in school accountability policies occurred prior to the period of study and are time-invariant. The state fixed effects account for the selection bias from pre-NCLB differences in school accountability systems.

Balance on Waiver Characteristics and Receipt

Appendix Table A4 test for pre-treatment outcome differences between states that received waivers at different times. In these state-level models I regress state average NAEP test scores at baseline (2003) on whether states ever received a waiver, whether a state was ever at high risk of losing their waiver, months until states received their waiver (after waiver announcement), and states that never received a waiver. The differences between each group of states is statistically insignificant. The lack of significant effects suggests that the timing of the waiver is exogenous and there were no differences between these groups of states prior to treatment.

Balance on Student Characteristics

Appendix Table A5 describes the baseline (measured in 2003) differences between the treatment and comparison groups in the full sample and the matched comparison sample. In the full sample there are significant and large differences between the treatment and comparison groups. The observable differences suggest that the students in the treatment and comparison groups could also differ on unobservable characteristics. If those uncontrolled for differences were also correlated with student outcomes then it would bias the effect of the waivers. The balance on baseline characteristics is considerably improved in the matched comparison sample. A few significant differences between the treated and comparison groups remain. But, the size of the differences is small enough (less than 0.25 SDs) that the inclusion of covariates and state fixed effects are a reasonable strategy for addressing the remaining selection bias.

A potential source of bias are unobserved state education reforms that were implemented when treated states received their waivers in 2013. For example, if a state court ordered school finance reform in 2013 then the effect of that policy would confound the effect of the waivers. Adding state fixed effects addresses the potential bias in part by accounting for time-invariant education reforms. The effect of the waivers remains about the same when I control for time varying education policies (See Appendix Tables 6 and 7). A related concern are changes in the administration of the waiver policy itself. For example, Washington had their waiver revoked in April 2014 (Klein, 2014). In that same year several of the largest districts in California received a specialized waiver. Washington and California may have reacted to these events by making changes to school accountability policies that could have influenced student outcomes. The potential bias from state reactions to changes in policies motivates my exclusion of the 2015 data from the main models.

A final issue is relevant for the models where I regress NAEP scores on waiver sanctions. The results are descriptive because states select schools for both the treatment and comparison groups based on prior student achievement. I attempt to account for the selection bias by restricting the sample to include only sanctioned schools and controlling for school AYP status in 2003. However, it remains likely that other unobservable characteristics of students are correlated with their enrollment in sanctioned schools and student outcomes. Another concern is that schools selected for sanctions could bias the effect of the waiver sanctions. As the number of schools that receive sanctions increases over time, average outcomes for sanctioned schools increase. Changes in how schools were selected for sanctions under NCLB could bias the effect of the waiver sanctions.

Results

Descriptively there is weak evidence that the waivers affected student outcomes overall. Figure 2 visualizes the trends in average outcomes for the treatment and control groups. Figure 2 shows very small increases for 4th grade math and reading (about 1 percent of an SD) after the waivers. Eighth grade math and reading outcomes decline for students in states that received waivers by about 1 percent of an SD. In 2015, outcomes for the treatment and comparison groups are about flat relative to 2013 (See Appendix Figure 2).

Regressions

Table 3 contains the results for the estimated effects of the waivers from several specifications for each of the four NAEP grade-subjects. The first four rows describe results from state and district fixed effects models with and without covariates for the full sample. There appears to be a small positive effect of the waivers on 4th grade math scores (about 4 percent of an SD). In the models with state and district fixed effects the effect of waiver receipt is significant and negative (-0.04 to -0.05 SDs). But after adding covariates for observable student characteristics and baseline school achievement the effect of waiver receipt is indistinguishable from zero. The fifth row includes a continuous version of the independent variable that is equal to the percent of schools sanctioned for each state and year. The coefficients in the fifth-row capture variation in how school accountability was administered across states and changes in the waivers across time. A one unit increase in the percent of schools sanctioned by the state is correlated with a 0.001 SD increase in NAEP scores and is indistinguishable from zero. The last two rows of results restrict the sample to include only states that implemented policies that were incentivized by the waivers. In the matched comparison sample I do not find evidence that waiver receipt had an effect. The exception to the pattern is 8th grade math where I find a significant negative effect (-0.059 SDs). However, a pre-treatment decline in outcomes for the treated group may explain the negative effect of the waivers on 8th grade math outcomes. The pattern from these models suggests that there is no detectable effect of the waivers on NAEP scores.

Table 4 includes the results from an event study specification for the full sample with and without covariates. The more conservative event study controls for effects in all pre-treatment years. Models 1 through 4 include state fixed effects and I add covariates to models 5 through 8. The pre-treatment estimates also allow for an inference about whether outcome trends prior to the waivers explain the results. Each pre-treatment estimate is insignificant with the exception of 8th grade math in 2003, which is consistent with the descriptive pre-treatment trends (See Figure 2). After adding covariates to the model there are no detectable effects of the waivers on student achievement. However, the standard errors in the models with covariates are fairly imprecise and I cannot rule out even small effects (0.02 SD to 0.1 SD).

Differential Effects

Tables 5, 6, and 7 describe the effects of the waivers on academically vulnerable students. Table 5 provides weak evidence that the waivers had a negative effect on FRPL eligible students.

Table 5, models 1 through 8, include results from the full sample and models 9 through 16 include the results from the matched comparison sample. In the full sample, the effect of the waivers on 8th grade outcomes on FRPL eligible students is small and negative (-0.04 SD) before the inclusion of covariates. The sign and significance of the waivers on FRPL eligible 8th grade math scores is robust to controlling for covariates and using the matched comparison sample. I do not detect an effect of the waivers for FRPL eligible students on 4th grade math and reading. Considering the pre-treatment trend in 8 grade math outcomes the results are consistent with the finding that the waivers do not have an effect on FRPL eligible students.

Table 6 includes differential effects of the waivers by race/ethnicity. The effect of the waivers on white students appears to be negative for 8th grade math. In the matched comparison sample, test scores for white students decline by about 8 percent of an SD. Fourth grade reading scores for Black students increased after the waivers. Hispanic students appear to have benefitted the most from the waivers. The positive effect of the waivers on Hispanic students in the matched comparison sample for 4th grade math, 4th grade reading and 8th grade reading is about 5 to 7 percent of an SD. Students from academically vulnerable race/ethnicities appear to have benefitted from the waivers.

Table 7 describes the intersectional effects of the waivers by FRPL eligibility and race/ethnicity. Consistent with Tables 5 and 6, the effect of the waivers on white students is negative. The negative effect for 8th grade math is about 10 percent of a SD for FRPL eligible white students and 7 percent of a SD for FRPL ineligible white students. The waivers increased reading test scores by about 10 percent of an SD for FRPL eligible students in reading and about 11 percent for FRPL ineligible students. The waivers also appear to have increased the outcomes for FRPL ineligible Asian students by about 11 percent of an SD. The waivers had a negative effect on white students in 8th grade math and increased scores for non-white students. The positive effects of the waivers are slightly larger for economically advantaged non-white students.[11]

Changes to School Accountability

Table 8 examines how the waivers changed school accountability systems. Using state-level data, I regress the number of sanctioned schools and the number of reward schools on waiver receipt. States with waivers were required to identify "reward schools" or Title I schools that either increased test scores overall or closed achievement gaps. States were also encouraged to provide incentives to reward schools. Table 8 shows that states with waivers, sanctioned substantively fewer schools (about 19 percent less in 2013). The results here are consistent with Hyslop (2013) who also finds that the waivers decreased the number of schools under sanction. States gave the reward designation to about 5 percent of schools. Taken together waiver accountability systems were less punitive in that fewer schools were sanctioned and more were lauded.

Sanctions

Table 9 includes the descriptive effects of the sanctions. These models include only students in schools that were under sanction. I compare students in waiver sanctioned schools (i.e., focus, priority) to NCLB sanctioned schools (e.g., school improvement, corrective action). Models 1 through 4 include state fixed effects and models 5 through 8 add covariates. The effects of attending a focus school when compared to a NCLB sanctioned schools are negative, but statistically indistinguishable from zero after the inclusion of controls. Without controls, the effect of attending a priority school relative to an NCLB sanction school is quite large and negative (-0.5 to -0.6 SDs).

---

[11] These results are robust to the inclusion of policies incentivized by the waivers and observable school accountability system characteristics (i.e., percent of schools sanctioned, percent of reward schools). The robustness of the results suggests that the neither mediate the effect of the waivers.

After adding covariates the effect of attending a priority school remains large and negative (-0.15 to -0.22 SDs).

The estimates in Table 9 are biased because of procedures that states used to select schools for sanctions under both NCLB and the waivers. Under the waivers, priority schools are in the lowest 5 percent of test scores and focus schools have either lower than average scores (6th to 15th percentile) or large achievement gaps. The selection of schools with systematically worse outcomes explains why the negative effect of the priority schools is larger than the focus schools. The NCLB sanctioned schools in the comparison group are also selected based on their outcomes. NCLB raised the AMOs that schools need to reach each year. The criteria used for selecting schools for NCLB sanctions increases across time. In 2004 —the first year of sanctions under NCLB—only the lowest performing schools received sanctions (~bottom 5 percent). Midway through NCLB (2007) the schools that received sanctions were no longer among the lowest performing in the country. Schools in the 20th through 50th percentile were routinely sanctioned. Each year NCLB selected schools for sanctions that had higher test scores, which increased outcomes for the comparison group. Outcomes for waiver sanctioned schools were systematically lower than NCLB sanctioned schools prior to treatment. The selection bias likely explains the negative effects of the waiver sanctions. With that caveat in mind, controlling for whether a school failed to meet AYP in 2003 ought to partially account for the selection bias. The results suggest that the waiver sanctions were not significantly better than the NCLB sanctions.

Long-term Outcomes

In Tables 10 and 11, I employ different approaches to estimate the effect of the waivers in 2015. Each approach involves important tradeoffs. The results may be biased by how states reacted to changes in how the waivers were administered (i.e. waiver termination, CORE waiver). In Table 10, I estimate the effect of the waivers dynamically across time. The first row of results compares outcomes in states that received waivers in 2013 relative to states that never received waivers. The second row of results compares states that received waivers in either 2013 or 2014 to states that never received waivers. After the inclusion of covariates, the results imply the waivers had a positive effect on 4th grade math scores and a negative effect on 8th grade math scores. The effect of the waivers on reading scores is insignificant. States that received waivers were significantly more likely to implement the policies incentivized by the waivers. Those policies confound the effect of the waivers and explain the pattern of effects. For example, states that received waivers were more likely to adopt College and Career Ready Standards like the Common Core. The Common Core had a positive effect on 4th grade math scores (in 2013) that is approximately the same size as the estimated effect of the waivers in (Model 5) (See essay herein). Overall the long-term results are consistent with the conclusion that the waivers did not have a detectable effect on student outcomes.

In Table 11, I compare states that received waivers in 2013 to states that received waivers in 2014. I exclude states that never received a waiver from these models. The results in Table 11 suggest the waivers had no detectable effect on student outcomes. Appendix Figure A1 shows that the scores for neither treatment nor the comparison groups change much in 2015 relative to 2013. The null effects could mean the waivers did not affect student outcomes. Alternatively, it could be that the waivers had the same sized positive effect on states which received waivers in 2013 and 2014. In that scenario there would be no difference between the groups in 2015.

In Table 12, I use a different approach to estimate the effect of facing sanctions. I compare students in schools that were subject to sanctions under NCLB but would not have been subject to sanctions under the waivers. My estimation strategy isolates the effect of facing a school sanction rather than conflating it with the effect of the sanction itself. Understanding the effect of the sanction threat is essential to understanding how accountability systems influence student outcomes. The waiver sanction criteria and the NCLB sanction criteria from 2005 through 2009 are quite similar. In these years the schools that would have been sanctioned under NCLB would almost certainly have been sanctioned under the waivers. The similarity in sanction criteria means there are very few schools available to estimate the difference between pre and post waiver outcomes. Instead I examine descriptive differences between a subset of schools in the post-waiver period (2013 and 2015).

Under the waivers, states sanctioned schools in the bottom 15 percent of achievement. To escape sanctions under NCLB, schools would need to reach approximately 85 percent proficiency (i.e., estimated AMO) in reading and math. States did not publish AMOs for 2011 through 2015 in a centralized location. I estimate the AMO in non-waiver states in 2013 is 85 percent. I Imputed the AMO in 2015 using the predicted values from a model where I regressed AMOs (from 2003 to 2011) on a linear year trend. I assume that schools where between 15 and 85 percent of students were proficient or better on both math and reading tests would have failed to meet AYP under NCLB and would not have faced sanctions under the waivers.

An additional challenge is that the AYP data that I use to create my sanction threat measure is not missing at random. In 2013 and 2015 about 36 percent of AYP data is missing for primary schools and 12 percent is missing for secondary schools. Critically the AYP data are not missing at random because they are correlated with both treatment and NAEP outcomes. The Education Department has the AYP data, but does not make them publicly available when states make reforms to summative assessments, school accountability systems, or receive waivers (i.e., treatment) (U.S. Department of Education, 2019). In addition, NAEP outcomes are correlated with missing AYP data.

The first row of results in Table 12 includes unadjusted mean differences. In the second row of results I add covariates and the third row restricts the sample to states in the matched comparison sample. I interpret the mean differences in Table 12 as the descriptive effect of the diminished threat of sanctions under the waivers. With the caveats I discuss above in mind, I find schools which were not at risk of sanctions under the waivers had significantly higher 4th grade outcomes than schools that were at risk for sanctions under NCLB (0.16 SD in 4th grade math and 0.13 SD in 4th grade reading). I find similar effects after controlling for student characteristics and restricting the sample to only states that adopted waiver incentivized policies. Schools that were threatened by sanctions had worse outcomes on average in 4th grade than schools which faced no such threat.

Robustness Checks

Endogenous Education Reforms

Policies that were implemented contemporaneously with the waivers and influenced student outcomes would bias the estimated effect of the waivers. In Appendix Table A6, I test whether the main effect of the waivers is sensitive to adding controls for time-varying education reforms. I created a database of 24 education policies popular with states and the Bush/Obama administrations during the period of study (Bleiberg & Harbatkin, 2018; Howell & Magazinnik, 2017; Jackson et al., 2015; McMaken, 2008). Appendix Table A6 contains the estimated effect of the waivers after the inclusion of a control for an education policy. Changes in the sign, size, and significance of the effect of waivers after controlling for education reforms suggests that the omitted policy may bias the estimate of interest. The effects for 4th grade math and 8th grade reading are robust to controlling for education policies. The estimates for 8th grade math and 4th grade reading increase in size enough to move across the conventional threshold for statistical significance. However, the magnitude of the changes is so small that the estimated effect of the waivers is insignificantly different from the model without policy controls.

I also test the robustness of the main findings to the inclusion of state specific linear trends (Angrist & Pischke, 2008). If the effect of the waivers does not change after adding the state specific trends then it provides evidence that omitted education policies and other unobserved characteristics are not biasing the effect of the waivers. Appendix Table A7 includes results from models with state fixed effects. The estimates are precise enough to rule out effects larger than 0.025 SDs. The results here support my conclusion that the waivers did not have an effect on student achievement overall.

Discussion

In 2012, the Obama administration faced a difficult situation. If no action was taken approximately 85 percent of schools would face sanctions in the next school year. An equally unacceptable outcome was deprioritizing NCLB's focus on closing achievement gaps. The waiver policy was intended to solve both issues by letting states sanction fewer schools and focus on schools with large achievement gaps. The waivers succeeded in so far as they avoided the untenable situation of mass sanctions for schools. The redesigned sanctions (i.e., focus, priority) were intended to increase average achievement and close achievement gaps. The evidence that the waivers help to increase student achievement overall is quite weak. The waivers do not have a detectable effect after accounting for the education policies that states were incentivized to adopt. But, the waivers did contribute to small but notable decreases in the size of achievement gaps for Black and Hispanic students. The waivers also appeared to decrease 8^{th} grade math outcomes for FRPL eligible white students. Overall given the policy context that demanded quick action the waivers were a positive step after NCLB.

Policy Implications

The waivers transferred power from the federal government to the states by allowing states flexibility in how to identify schools for sanctions (McGuinn, 2016). The federal government has historically been more aggressive about protecting civil rights in schools relative to states (Peterson, 1995; Peterson et al., 1986). Empowering states could have resulted in lower expectations for Black and Brown students and widened achievement gaps. However, race-based achievement gaps were smaller after the waivers. The smaller achievement gaps suggest that states maintained their focus on improving outcomes for academically vulnerable students, even after federal pressure decreased.

The changes to the cohorts of schools that received sanctions over time make it difficult to identify the effect of the school sanctions under the waivers. With that caveat in mind, I find descriptive effects that are alarmingly negative. They suggest not just that the efforts to turnaround schools under the waivers did not work, but that they actually decreased student outcomes. Under the waivers far fewer schools were sanctioned than under NCLB. Other research that employs causal estimation strategies find the schools sanctions had null effects on student outcomes (Dee & Dizon-Ross, 2019; Dougherty & Weiner, 2017; Hemelt & Jacob, 2018). If these studies reflect the true effect of school sanctions then the benefit of the waivers is that it decreases the number of sanctioned schools overall. But, the negative effects of the waiver sanctions relative to the NCLB sanctions appears to have cancelled out any potential improvement from decreasing the number of schools under sanction. States have limited capacity to focus on school improvement (Tanenbaum et al., 2015). Careful and judicious targeting of schools for improvement is likely a prerequisite for success. A possible explanation for the negative effects is that states were transitioning from mass administration of school sanctions to a more targeted approach. After years under NCLB it may take states time to implement best practices for school improvement.

Under NCLB sanctioned schools directed additional resources towards Black and Hispanic students (Krieg, 2011; Springer, 2008). Waivers required states to focus on race-based achievement gaps and gave them additional discretion for how to target schools for sanctions. Greater levels of state autonomy may have enabled states to shift even more within school resources towards Black and Hispanic students and away from white students. If true it would in part explain the negative effect of the waivers on 8th grade math outcomes for white students.

Limitations

Identifying the mechanisms through which the waivers influenced academically vulnerable students is challenging. There are two main pathways through which NCLB could influence student outcomes: (1) the policies that were incentivized by the waivers and (2) the changes waivers made to school accountability systems. An acute challenge when trying to determine the mechanisms through which the waivers influences student outcomes are the largely null effects on students overall. Restricting the sample to include only states that implemented waiver incentivized policies removes a source of selection bias and also the variation that could identify the effect of those policies. Without observing whether the waivers caused states to implement education policies it is not possible to disentangle their effect from the waivers themselves.

One potential source of unaccounted for Omitted Variable Bias are the School Improvement Grants (SIG). The grants were administered by states to turn around schools during the period of study. Many schools that face sanctions under NCLB and the waivers received SIG. The only study on the nationwide effect of SIG found largely negative effects (Dragoset et al., 2017). If the effect of SIG was negative and the waivers had a positive effect then it would explain the pattern of results found here. It is also plausible that states anticipated the waivers. State policymakers could have predicted that Congressional gridlock would eventually force President Obama to take executive action. If true, then states may have started planning or enacting changes to school accountability systems prior to the announcement of the waivers. Anticipation of treatment could explain the pre-waiver trend in outcomes for 8th grade math.

Future Research

I hope to extend my research on the waivers by examining the effect of the rewards schools. Disentangling the causal effect of the reward schools on student achievement is quite challenging because reward schools were chosen based on their improved performance. It is possible that the effects of the reward schools could help to explain the positive effects on academically vulnerable students. I also hope to study how the waivers influenced schools and teachers. Years of consequential school accountability changed the relationship between government and schools (Spillane et al., 2011). The shift in intergovernmental relations may have either increased or decreased the strength of the relationship between schools and government.

The waivers were replaced in 2015 by ESSA, but the Obama era policy changed school accountability systems. States sanctioned about the same number of schools under ESSA as they did under the waivers. The waivers were the beginning of the end for NCLB and the lessons from this transition can inform a new era of school accountability.

Chapter 3 References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When Should You Adjust Standard Errors for Clustering?* (Working Paper No. 24003). National Bureau of Economic Research. https://doi.org/10.3386/w24003

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Barron, D. J., & Rakoff, T. D. (2013). In defense of big waiver. *Columbia Law Review*, 265–345.

Bleiberg, J., & Harbatkin, E. (2018). Innovation and Diffusion of Teacher Evaluation Reform: A Convergence of Federal and Local Forces. *Educational Policy*.

Bonilla, S., & Dee, T. S. (2020). The Effects of School Reform Under NCLB Waivers: Evidence from Focus Schools in Kentucky. *Education Finance and Policy*, *15*(1), 75–103.

Bowling, C. J., & Pickerill, J. M. (2013). Fragmented federalism: The state of American federalism 2012–13. *Publius: The Journal of Federalism*, *43*(3), 315–346.

Center on Education Policy. (2018). *NCLB/ESEA Waivers*. https://www.cep-dc.org/displayTopics.cfm?DocumentSubTopicID=48

Chubb, J., & Clark, C. (2013). The new state achievement gap: How federal waivers could make it worse—or better. *Washington, DC: Education Sector*.

Dee, T., & Dizon-Ross, E. (2019). School performance, accountability and waiver reforms: Evidence from Louisiana. *Educational Evaluation and Policy*, *41*(3), 316–349.

Dee, T., & Jacob, B. (2009). The Impact of No Child Left Behind on Student Achievement. *National Bureau of Economic Research Working Paper Series*, *No. 15531*. https://doi.org/10.3386/w15531

Dougherty, S. M., & Weiner, J. M. (2017). The Rhode to Turnaround: The Impact of Waivers to No Child Left Behind on School Performance. *Educational Policy*, 0895904817719520.

Dragoset, L., Thomas, J., Herrmann, M., Deke, J., James-Burdumy, S., Graczewski, C., Boyle, A., Upton, R., Tanenbaum, C., Giffin, J., & Wei, T. E. (2017). *School Improvement Grants: Implementation and Effectiveness*. US Department of Education.

Edweek. (2011). *Quality Counts*. https://www.edweek.org/ew/qc/index.html

Erpenbach, W. J., & Forte, E. (2005). Statewide Educational Accountability Under the No Child Left Behind Act—A Report on 2005 Amendments to State Plans. *Council of Chief State School Officers*.

Erpenbach, W. J., & Forte, E. (2007). Statewide Educational Accountability Systems Under the NCLB Act—A Report on 2007 Amendments to State Plans. *Council of Chief State School Officers*.

Erpenbach, W. J., Forte-Fast, E., & Potts, A. (2003). *Statewide Educational Accountability under NCLB. Central Issues Arising from An Examination of State Accountability Workbooks and US Department of Education Reviews under the No Child Left Behind Act of 2001*.

Evan Kramer. (2019). *Phone Discussion on School Accountability and Waivers* [Personal communication].

Fahle, E. M., Shear, B. R., Kalogrides, D., Reardon, S. F., Chavez, B., & Ho, A. D. (2019). *Stanford Education Data Archive Technical Documentation Version 3.0 July 2019*.

Fast, E. F., & Erpenbach, W. J. (2004). Revisiting Statewide Educational Accountability Under NCLB: A Summary of State Requests in 2003-2004 for Amendments to State Accountability Plans. *Council of Chief State School Officers*.

Forte, E., & Erpenbach, W. J. (2006). Statewide Educational Accountability Under the No Child Left Behind Act: A Report on 2006 Amendments to State Plans. A Summary of State Requests in 2005-06 for Amendments to Their Educational Accountability Systems Under NCLB. *Council of Chief State School Officers*.

Gais, T., & Fossett, J. (2005). Federalism and the executive branch. *The Executive Branch*, 486–524.

Hemelt, S. W., & Jacob, B. A. (2018). How Does an Accountability Program that Targets Achievement Gaps Affect Student Performance? *Education Finance and Policy*, 1–68.

Hoffer, T. B., Hedberg, E. C., Brown, K. L., Halverson, M. L., Reid-Brossard, P., Ho, A. D., & Furgol, K. (2011). Final Report on the Evaluation of the Growth Model Pilot Project. *US Department of Education.*

Hopkins, B. (2019). *The Impact of NCLB Waivers on Student Achievement Gaps*. Association of Education Finance and Policy.

Howell, W. G., & Magazinnik, A. (2017). Presidential Prescriptions for State Policy: Obama's Race to the Top Initiative. *Journal of Policy Analysis and Management*, *36*(3), 502–531.

Hyslop, A. (2013). It's All Relative: How NCLB Waivers Did—and Did Not—Transform School Accountability. *New America Foundation.*

Jackson, C. K., Johnson, R. C., & Persico, C. (2015). *The effects of school spending on educational and economic outcomes: Evidence from school finance reforms*. National Bureau of Economic Research. http://www.nber.org/papers/w20847

Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, *61*, 51–58.

Jochim, A., & Murphy, P. (2013). The capacity challenge: What it takes for state education agencies to support school improvement. *Seattle, Wash.: Center on Reinventing Public Education.*

Klein, A. (2014, April 24). Arne Duncan Revokes Washington State's NCLB Waiver. *Education Week*. http://blogs.edweek.org/edweek/campaign-k-12/2014/04/washington_state_loses_waiver_.html?cmp=SOC-SHR-FB

Koretz, D. (2017). *The Testing Charade: Pretending to Make Schools Better*. University of Chicago Press.

Koretz, D., & Hamilton, L. S. (2006). *Testing for accountability in K-12*. https://www.rand.org/pubs/external_publications/EP20060030.html

Koretz, D. M., & Barron, S. I. (1998). *The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)*.

Krieg, J. M. (2011). Which students are left behind? The racial impacts of the No Child Left Behind Act. *Economics of Education Review*, *30*(4), 654–664. https://doi.org/10.1016/j.econedurev.2011.02.004

Marsh, J. A., Bush-Mecenas, S., Hough, H. J., Park, V., Allbright, T., Hall, M., & Glover, H. (2016). At the Forefront of the New Accountability Era: Early Implementation Findings from the CORE Waiver Districts. *Policy Analysis for California Education, PACE*.

McGuinn, P. (2016). From No Child Left behind to the Every Student Succeeds Act: Federalism and the Education Legacy of the Obama Administration. *Publius: The Journal of Federalism*.

McMaken, J. (2008). *State Statutes Regarding Kindergarten: Policies concerning district offering of and student attendance in full- and half-day kindergarten programs*. Education Commission of the States.

McMurrer, J., & Yoshioka, N. (2013). States' Perspectives on Waivers: Relief from NCLB, Concern about Long-Term Solutions. *Center on Education Policy*. Washington, DC.

Murphy, J., & Bleiberg, J. (2018). *School Turnaround Policies and Practices in the US - Learning from Failed School Reform*. Springer International Publishing.

Peterson, P. E. (1995). *The Price of Federalism*. Brookings Institution Press.

Peterson, P. E., Rabe, B. G., & Wong, K. K. (1986). *When federalism works*. Brookings Institution Press.

Polikoff, M. S., McEachin, A. J., Wrabel, S. L., & Duque, M. (2014). The waive of the future? School accountability in the waiver era. *Educational Researcher*, *43*(1), 45–54.

Reback, R., Rockoff, J., Schwartz, H. L., & Davidson, E. (2013). *Barnard No Child Left Behind Database, 2002-2003 and 2003-2004*. Barnard Columbia NCLB Data Project. http://www.gsb.columbia.edu/nclb

Rogers, A., Tarsitano, C., & Sikali, E. (2014). *National Assessment of Educational Profress (NAEP) 2013 Mathematics and Reading Grades 4 and 8 Assessments Restricted-Use Data Files Data Companion*. National Center for Education Statistics.

Saultz, A., McEachin, A., & Fusarelli, L. D. (2016). Waivering as governance: Federalism during the Obama administration. *Educational Researcher*, *45*(6), 358–366.

Spillane, J. P., Parise, L. M., & Sherer, J. Z. (2011). Organizational routines as coupling mechanisms: Policy, school administration, and the technical core. *American Educational Research Journal*, *48*(3), 586–619.

Springer, M. G. (2008). The influence of an NCLB accountability plan on the distribution of student test score gains. *Economics of Education Review*, *27*(5), 556–563. https://doi.org/10.1016/j.econedurev.2007.06.004

Tanenbaum, C., Boyle, A., Graczewski, C., James-Burdumy, S., Dragoset, L., Hallgren, K., & others. (2015). *State capacity to support school turnaround.* Mathematica Policy Research. https://ideas.repec.org/p/mpr/mprres/98202aa167224a8a9e0ad4b449900a5b.html

U.S. Department of Education. (2011). *ESEA flexibility: Frequently asked questions*.

U.S. Department of Education. (2016, May 12). *Index Page for the ESEA Flexibility Page* [Letters (Correspondence); Reference Materials; Reports]. US Department of Education (ED). https://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

U.S. Department of Education. (2018a). *Education Data Express*. https://eddataexpress.ed.gov

U.S. Department of Education. (2018b, February 15). *Consolidated State Performance Reports* [Data Collection Instruments; Application Materials]. https://www2.ed.gov/admins/lead/account/consolidated/index.html

U.S. Department of Education. (2019). *EDFacts Data Files*. https://www2.ed.gov/about/inits/ed/edfacts/data-files/index.html

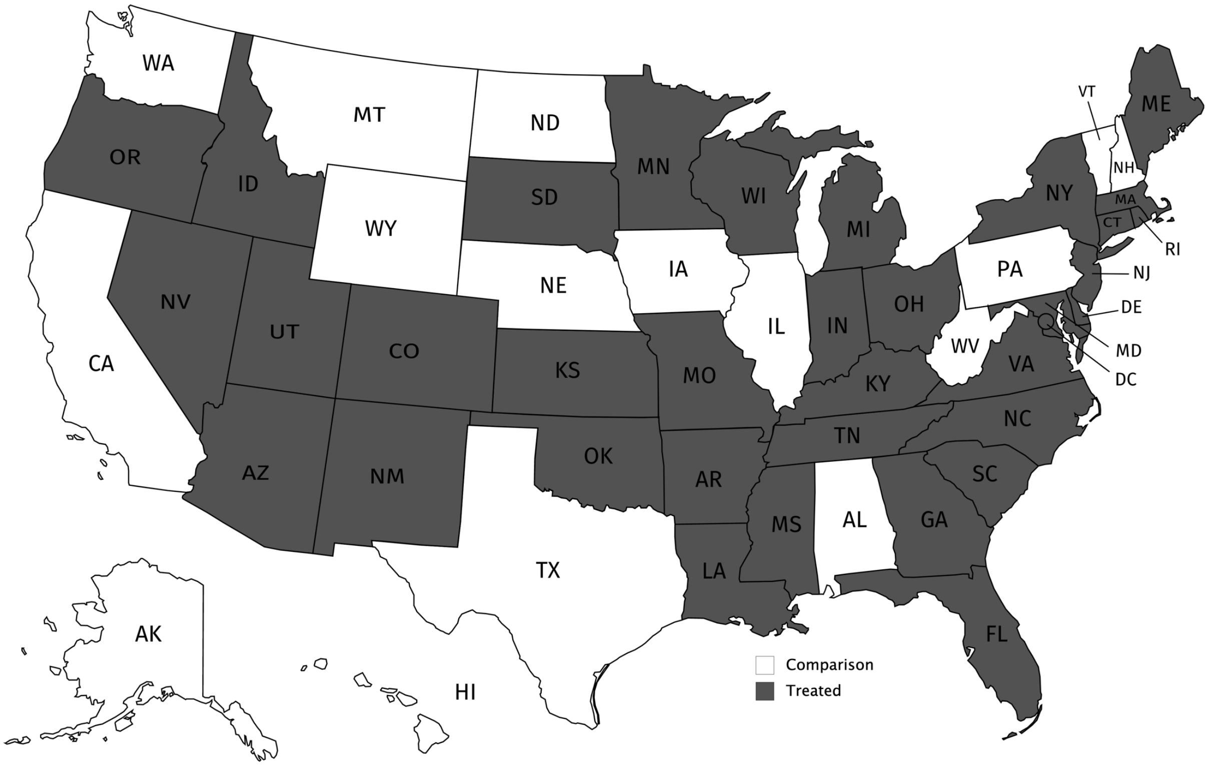
U.S. Department of Education. (2020). *NAEP State Mapping*. https://nces.ed.gov/nationsreportcard/studies/statemapping/

Weiss, M. J., & May, H. (2012). A policy analysis of the federal growth model pilot program's measures of school performance: The Florida case. *Education Finance and Policy*, *7*(1), 44–73.

Wong, K. K. (2015). Federal ESEA Waivers as Reform Leverage: Politics and Variation in State Implementation. *Publius: The Journal of Federalism*, *45*(3), 405–426. https://doi.org/10.1093/publius/pjv020

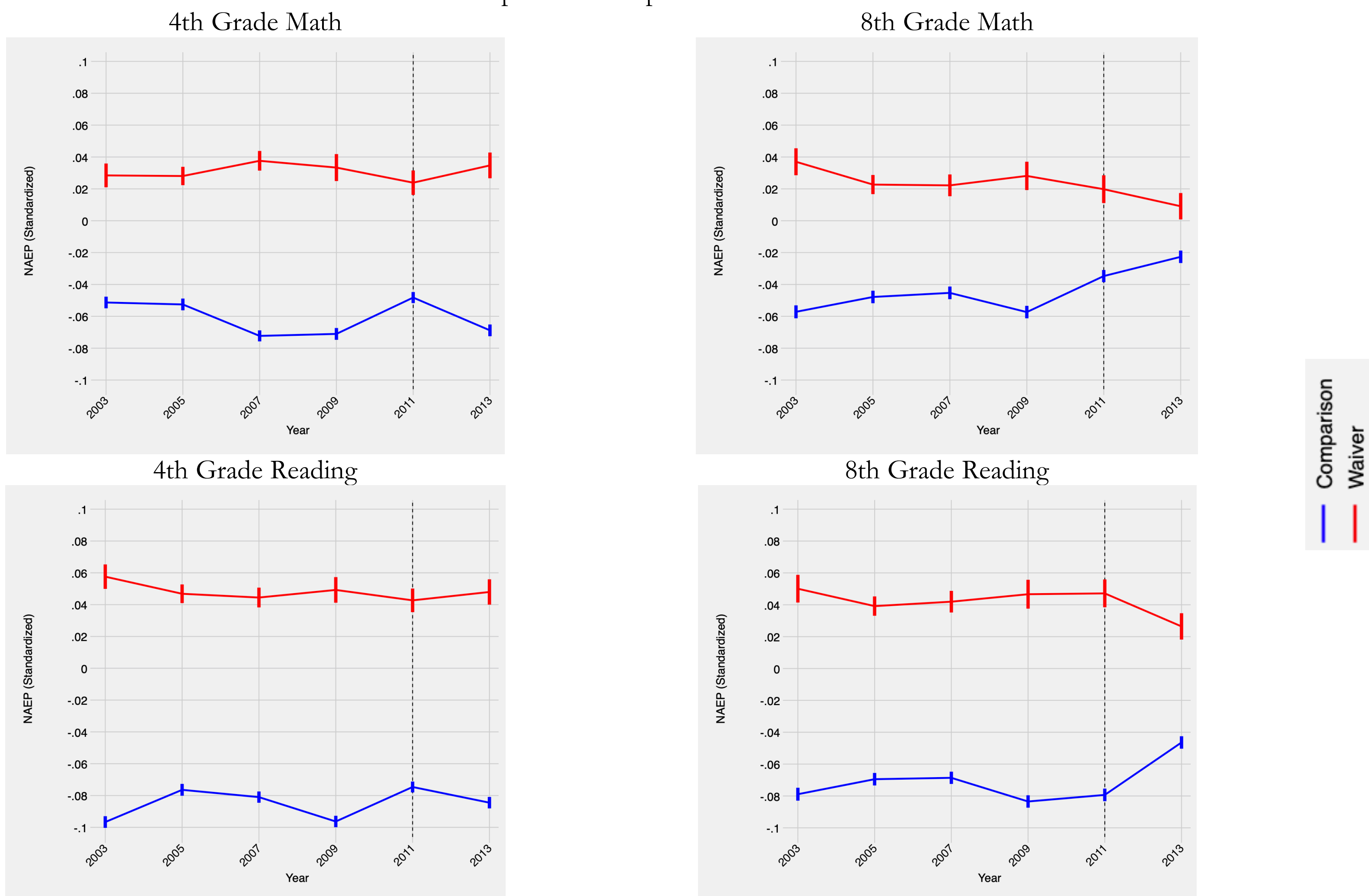
Wong, V. C., Wing, C., Martin, D., & Krishnamachari, A. (2018). Did states use implementation discretion to reduce the stringency of NCLB? Evidence from a database of state regulations. *Educational Researcher*, *47*(1), 9–33.

Figure 1. Waiver Receipt by State Prior to 2012-13 School Year



Note: Waiver receipt data collected from Center on Education Policy (2018), U.S. Education Department (2016), and Comprehensive School Performance Reports (U.S. Department of Education, 2018b).

Figure 2. NAEP Score Trends for Treatment and Comparison Groups



Note: The blue line is the comparison group and the red line is the treatment group. 2011 is the last NAEP year prior to waivers. Y axis is NAEP student outcomes standardized within subject/grade and year. Estimates adjusted using NAEP student level probability weights. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 1. Analytic Sample Characteristics by Grade, Subject, and Year

	2003	2005	2007	2009	2011	2013	Total
States (Treated)							
4th Grade Math	0	0	0	0	0	34	51
8th Grade Math	0	0	0	0	0	34	51
4th Grade Reading	0	0	0	0	0	34	51
8th Grade Reading	0	0	0	0	0	34	51
Districts							
4th Grade Math	2,770	3,270	2,800	3,260	2,820	2,760	17,680
8th Grade Math	2,750	3,100	3,050	3,080	3,100	2,790	17,870
4th Grade Reading	2,760	3,270	2,800	3,260	2,820	2,760	17,670
8th Grade Reading	2,760	3,100	3,050	3,090	3,090	2,790	17,880
Schools							
4th Grade Math	5,520	6,840	5,800	7,230	6,190	6,030	37,610
8th Grade Math	4,520	5,140	5,150	5,360	5,560	5,100	30,830
4th Grade Reading	5,520	6,840	5,790	7,230	6,190	6,020	37,590
8th Grade Reading	4,530	5,130	5,160	5,360	5,550	5,100	30,830
Students							
4th Grade Math	154,430	134,400	158,440	137,120	167,780	151,040	903,210
8th Grade Math	123,360	125,460	122,640	131,470	139,520	137,490	779,940
4th Grade Reading	150,200	129,020	152,810	144,840	170,990	154,040	901,900
8th Grade Reading	122,280	123,630	128,690	130,670	134,000	138,920	778,190

Note: See Appendix Table A1 for the states in the treatment and comparison groups by grade and subject. Sample size rounded for the number of states, districts, schools, and students in accordance with National Center for Education Statistics (NCES) nondisclosure rules. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 2. Descriptive Statistics for Analytic Sample

Characteristic	2003	2005	2007	2009	2011	2013
NAEP MG4	-0.0019	-0.0028	-0.0042	-0.0060	-0.0035	-0.0053
NAEP MG8	0.0011	-0.0040	-0.0036	-0.0050	-0.0016	-0.0032
NAEP RG4	-0.0010	-0.0008	-0.0033	-0.0058	-0.0017	-0.0032
NAEP RG8	0.0008	-0.0021	-0.0006	-0.0042	-0.0025	-0.0017
Female	0.4954	0.4959	0.4957	0.4938	0.4946	0.4906
IEP	0.1087	0.1046	0.1018	0.1083	0.1106	0.1222
LEP	0.0574	0.0643	0.0709	0.0594	0.0752	0.0715
FRPL	0.4255	0.4389	0.4444	0.4745	0.5220	0.5395
White	0.6315	0.6081	0.5940	0.5743	0.5437	0.5371
Black	0.1784	0.1707	0.1677	0.1793	0.1771	0.1743
Hispanic	0.1200	0.1473	0.1577	0.1642	0.1887	0.1945
Asia/PI	0.0422	0.0454	0.0470	0.0481	0.0500	0.0503
American Indian	0.0208	0.0207	0.0238	0.0223	0.0210	0.0201
Modal age for grade; At	0.5985	0.5956	0.5962	0.5945	0.5974	0.6019
Below	0.0028	0.0026	0.0026	0.0021	0.0021	0.0018
Above	0.3987	0.4017	0.4011	0.4035	0.4005	0.3962
School made AYP in 2003	0.5933	0.5884	0.5993	0.5864	0.5914	0.6008
N	581,420	573,130	612,780	567,340	614,650	572,610

Note: Student and school characteristics pooled across grade-subjects. Sample size rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP=National Assessment of Educational Progress test score standardized within grade-subject and year, IEP=Individualized Education Plan, LEP=Limited English Proficiency, FRPL=Free and Reduce Price Lunch, PI=Pacific Islander, AYP=Adequate Yearly Progress. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 3. NAEP Scores Regressed on Waiver Receipt

Outcome	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)
State FE	0.022	-0.042	0.004	-0.052*
	(0.027)	(0.023)	(0.025)	(0.022)
N	1,079,220	932,820	1,078,670	931,060
F	0.66	3.40	0.03	5.43
	(5)	(6)	(7)	(8)
State FE and Covariates	0.041*	-0.026	0.027	-0.006
	(0.018)	(0.014)	(0.016)	(0.015)
N	952,140	810,280	950,730	808,770
F	4600.16	5562.79	5083.33	5613.86
	(9)	(10)	(11)	(12)
District FE	0.041*	-0.033	0.021	-0.040*
	(0.018)	(0.018)	(0.018)	(0.018)
N	1,077,760	931,510	1,077,180	929,770
F	5.07	3.31	1.45	4.84
	(13)	(14)	(15)	(16)
District FE and Covariates	0.065***	-0.006	0.032*	0.007
	(0.015)	(0.013)	(0.013)	(0.014)
N	952,060	810,220	950,650	808,720
F	4663.04	5480.92	4787.06	5110.42
	(17)	(18)	(19)	(20)
School % in Improvement	0.001	0.001	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
N	952,140	810,280	950,730	808,770
F	4599.86	5562.04	5081.31	5619.46
	(21)	(22)	(23)	(24)
State FE with Matched Comparison Sample	0.019	-0.053	0.025	0.014
	(0.034)	(0.03)	(0.031)	(0.029)
N	546,920	475,100	548,070	474,180
F	0.31	3.21	0.63	0.24
	(25)	(26)	(27)	(28)
State FE with Covariates & Matched Comparison Sample	0.014	-0.059**	0.038	0.011
	(0.025)	(0.021)	(0.021)	(0.019)
N	491,610	424,120	492,130	423,360
F	3251.2	3761.36	3495.6	3452.4

Note: See Appendix Table A1 for the states in the treatment and comparison groups and the matched comparison sample. Covariates includes Female, Individualized Education Plan, Limited English Proficiency, race/ethnicity, modal age for grade, school AYP status in 2003. School % in Improvement is a continuous variable. In 2003 it equals zero because NCLB sanctions had not gone into effect. In 2005-2011 it equals the percent of schools in any stage of the improvement process for the comparison group. In 2013 it equals the percentage of priority and focus schools for treated states. Outcomes standardized within grade, subject, and year. Standard errors robust to clustering by school. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 4. NAEP Scores Regressed on Waiver Receipt: Event Study

Outcome	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)
Pre-Treatment 2003	-0.004	0.039	0.025	0.003
	(0.035)	(0.032)	(0.033)	(0.029)
Pre-Treatment 2005	-0.000	0.019	-0.004	-0.014
	(0.030)	(0.027)	(0.028)	(0.024)
Pre-Treatment 2007	0.031	0.007	0.002	-0.019
	(0.031)	(0.028)	(0.029)	(0.026)
Pre-Treatment 2009	0.027	0.025	0.024	-0.000
	(0.035)	(0.032)	(0.033)	(0.029)
Post-Treatment 2013	0.032	-0.024	0.014	-0.058*
	(0.035)	(0.030)	(0.032)	(0.028)
N	1,079,220	932,820	1,078,670	931,060
Adjusted R^2	0.031	0.034	0.028	0.027
F	0.41	0.68	0.26	0.80
State FE	X	X	X	X
	(5)	(6)	(7)	(8)
Pre-Treatment 2003	-0.016	0.051**	0.007	-0.004
	(0.021)	(0.019)	(0.020)	(0.020)
Pre-Treatment 2005	-0.009	0.029	-0.023	-0.017
	(0.018)	(0.017)	(0.017)	(0.018)
Pre-Treatment 2007	0.017	0.008	-0.021	-0.024
	(0.018)	(0.017)	(0.018)	(0.019)
Pre-Treatment 2009	0.008	0.018	0.006	-0.015
	(0.021)	(0.020)	(0.021)	(0.019)
Post-Treatment 2013	0.041	-0.005	0.021	-0.018
	(0.022)	(0.019)	(0.020)	(0.020)
N	952,140	810,280	950,730	808,770
Adjusted R^2	0.323	0.348	0.318	0.323
F	2821.74	3407.16	3106.24	3438.07
State FE	X	X	X	X
Covariate	X	X	X	X

Note: Reference category is the last wave prior to the NCLB waivers (2011). Standard errors are robust to clustering by school. See Appendix Tables A1 for composition of the treatment/comparison group and the matched comparison sample. Outcomes standardized within grade, subject, and year. See Table 3 for a full list of covariates. Standard errors robust to clustering by school. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect. *$p < 0.05$, **$p<0.01$, ***$p<0.001$. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 5. Differential Effects of Waivers for FRPL Eligible Students

Outcome	MG4	MG8	RG4	RG8	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Waiver Effect for FRPL Eligible	0.029	-0.036*	0.016	-0.039*	0.031	-0.037*	0.021	-0.012
	(0.020)	(0.018)	(0.018)	(0.018)	(0.018)	(0.015)	(0.016)	(0.016)
N	1,065,150	917,570	1,064,870	915,710	952,140	810,280	950,730	808,770
Adjusted R-squared	0.172	0.155	0.156	0.128	0.323	0.348	0.318	0.323
F	6415.05	5740.44	7140.11	5227.06	4348.40	5255.01	4815.64	5263.10
Covariates					X	X	X	X
Outcome	MG4	MG8	RG4	RG8	MG4	MG8	RG4	RG8
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Waiver Effect for FRPL Eligible	0.001	-0.042	0.014	0.017	0.004	-0.065**	0.038	0.006
	(0.026)	(0.023)	(0.024)	(0.023)	(0.025)	(0.021)	(0.022)	(0.020)
N	540,460	467,810	541,720	466,860	491,610	424,120	492,130	423,360
Adjusted R^2	0.164	0.146	0.147	0.118	0.313	0.332	0.309	0.302
F	4973.56	4043.88	5411.78	3505.53	3037.90	3512.61	3264.65	3221.67
Covariates					X	X	X	X
Matched Comparison Sample	X	X	X	X	X	X	X	X

Note: See Appendix Tables A1 for composition of the treatment/comparison group and the matched comparison sample. See Table 3 for a full list of covariates. Outcomes standardized within grade, subject, and year. Standard errors robust to clustering by school. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 6. Differential Effects of Waivers, Student Race/Ethnicity

Outcome	MG4	MG8	RG4	RG8	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Waiver Effect for White	0.020	-0.055***	0.009	-0.029	0.001	-0.084***	0.022	-0.007
	(0.018)	(0.015)	(0.016)	(0.015)	(0.026)	(0.021)	(0.022)	(0.020)
Waiver Effect for Black	0.042*	0.019	0.051**	0.016	0.002	0.002	0.065*	0.049
	(0.021)	(0.018)	(0.019)	(0.020)	(0.029)	(0.025)	(0.026)	(0.026)
Waiver Effect for Hispanic	0.116***	0.041*	0.065**	0.065***	0.069*	-0.015	0.068*	0.051*
	(0.021)	(0.018)	(0.020)	(0.019)	(0.028)	(0.025)	(0.027)	(0.024)
Waiver Effect for Asian/PI	0.044	0.030	0.053	0.020	0.031	-0.016	0.065	0.044
	(0.032)	(0.032)	(0.028)	(0.031)	(0.041)	(0.044)	(0.037)	(0.043)
N	952,140	810,280	950,730	808,770	491,610	424,120	492,130	423,360
Adjusted R^2	0.323	0.348	0.318	0.323	0.314	0.332	0.309	0.302
F	3419.68	4147.90	3780.48	4156.91	2395.86	2789.25	2579.84	2547.47
State FE	X	X	X	X	X	X	X	X
Covariates	X	X	X	X	X	X	X	X
Matched Comparison Sample					X	X	X	X

Note: See Appendix Tables A1 for composition of the treatment/comparison group and the matched comparison sample. See Table 3 for a full list of covariates. Outcomes standardized within grade, subject, and year. Standard errors robust to clustering by school. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect, PI=Pacific Islander. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 7. Intersectional Differential Effects of Waivers

Outcome	MG4	MG8	RG4	RG8	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Waiver Effect for White Advan	0.035	-0.041**	0.013	-0.020	0.018	-0.071**	0.025	0.001
	(0.019)	(0.015)	(0.016)	(0.016)	(0.027)	(0.022)	(0.022)	(0.021)
Waiver Effect for White FRPL Eligible	0.000	-0.071***	0.008	-0.042*	-0.026	-0.106***	0.019	-0.021
	(0.019)	(0.016)	(0.017)	(0.017)	(0.027)	(0.023)	(0.024)	(0.022)
Waiver Effect for Black Advan	0.065*	0.065*	0.116***	0.078**	0.001	0.035	0.115**	0.116**
	(0.029)	(0.025)	(0.026)	(0.030)	(0.039)	(0.034)	(0.035)	(0.042)
Waiver Effect for Black FRPL Eligible	0.031	0.035	0.111***	0.057	-0.043	0.001	0.109**	0.095*
	(0.031)	(0.028)	(0.028)	(0.032)	(0.041)	(0.036)	(0.039)	(0.045)
Waiver Effect for Hisp Advan	0.175***	0.086**	0.104***	0.076**	0.103**	-0.021	0.060	0.036
	(0.030)	(0.028)	(0.029)	(0.028)	(0.039)	(0.036)	(0.037)	(0.035)
Waiver Effect for Hisp FRPL Eligible	0.141***	0.057	0.098**	0.055	0.059	-0.056	0.054	0.015
	(0.031)	(0.030)	(0.031)	(0.029)	(0.040)	(0.038)	(0.039)	(0.037)
Waiver Effect for Asian Advan	0.061	0.049	0.083**	0.081*	0.025	0.018	0.107**	0.116*
	(0.039)	(0.039)	(0.030)	(0.035)	(0.049)	(0.055)	(0.039)	(0.050)
Waiver Effect for Asian FRPL Eligible	0.026	0.019	0.078*	0.059	-0.019	-0.017	0.101*	0.095
	(0.039)	(0.039)	(0.031)	(0.037)	(0.050)	(0.055)	(0.040)	(0.051)
N	952,140	810,280	950,730	808,770	491,610	424,120	492,130	423,360
Adjusted R^2	0.324	0.350	0.319	0.323	0.314	0.333	0.310	0.303
F	2297.31	2718.95	2512.76	2695.12	1538.04	1770.80	1640.70	1616.36
State FE	X	X	X	X	X	X	X	X
Covariates	X	X	X	X	X	X	X	X
Matched Comparison Sample					X	X	X	X

Note: See Appendix Tables A1 for composition of the treatment/comparison group and the matched comparison sample. See Table 3 for a full list of covariates. Outcomes standardized within grade, subject, and year. Standard errors robust to clustering by school. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect, Advan=Economically Advantaged (i.e., FRPL ineligible), Hisp=Hispanic. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 8. School Accountability System Characteristics Regressed on Waiver Receipt

Outcome	School Sanction Percent	Reward School Percent
	(1)	(2)
Waiver	-0.187***	0.054**
	(0.047)	(0.017)
N	357	357
Adjusted R^2	0.481	0.298
F	16.20	9.84
State FE	X	X
Year FE	X	X

Note: Data on schools sanctions and reward schools collected from Comprehensive School Performance Reports (U.S. Department of Education, 2018b). FE=Fixed Effect

Table 9. NAEP Scores Regressed on NCLB Waiver Sanctions

Outcome	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)
Focus School	-0.160**	-0.094*	-0.112*	-0.035
	(0.057)	(0.047)	(0.056)	(0.050)
Priority School	-0.671***	-0.633***	-0.538***	-0.502***
	(0.068)	(0.050)	(0.065)	(0.058)
N	137,460	183,440	138,170	181,590
Adjusted R^2	0.083	0.064	0.077	0.052
F	55.28	86.77	42.71	47.00
State FE	X	X	X	X
	(5)	(6)	(7)	(8)
Focus School	0.0001	0.017	0.062	0.031
	(0.040)	(0.036)	(0.042)	(0.037)
Priority School	-0.219***	-0.224***	-0.147**	-0.178**
	(0.052)	(0.048)	(0.054)	(0.059)
N	122,840	164,860	123,160	163,210
Adjusted R^2	0.316	0.357	0.325	0.333
F	622.63	1353.82	717.77	1194.35
State FE	X	X	X	X
Covariates	X	X	X	X

Note: Data on sanctions under NCLB and NCLB waivers (i.e., focus and priority) collected from Comprehensive School Performance Reports (U.S. Department of Education, 2018b). See Table 3 for a full list of covariates. Outcomes standardized within grade, subject, and year. Standard errors robust to clustering by school. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 10. NAEP Scores Regressed on Waiver Implementation Year

Outcome	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)
Waiver Effect 2013	0.020	-0.056*	-0.009	-0.069**
	(0.025)	(0.023)	(0.024)	(0.022)
Waiver Effect 2014	0.041	-0.080*	-0.033	-0.067*
	(0.039)	(0.034)	(0.037)	(0.033)
N	1,213,960	1,065,350	1,212,720	1,062,930
Adjusted R^2	0.030	0.033	0.027	0.025
F	0.60	3.72	0.42	5.19
State FE	X	X	X	X
	(5)	(6)	(7)	(8)
Waiver Effect 2013	0.037*	-0.039**	0.017	-0.024
	(0.016)	(0.013)	(0.014)	(0.013)
Waiver Effect 2014	0.072**	-0.046*	0.012	-0.020
	(0.025)	(0.022)	(0.023)	(0.022)
N	1,060,020	915,110	1,058,080	913,170
Adjusted R^2	0.324	0.349	0.324	0.325
F	4647.46	5745.01	5104.48	5935.04
State FE	X	X	X	X
Covariates	X	X	X	X

Note: Waiver Effect 2013 compares states that received waivers in 2013 relative to states that never received waivers. waiver Effect 2014 compares states that received waivers in 2014 relative to states that never received waivers. Standard errors are robust to clustering by school. See Appendix Tables A1 for composition of the treatment/comparison group and the matched comparison sample. Outcomes standardized within grade, subject, and year. See Table 3 for a full list of covariates. Standard errors robust to clustering by school. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2014-2015.

Table 11. Long-Term Outcomes of the Waiver

Outcome	MG4	MG8	RG4	RG8
Waiver 2013	0.008	-0.008	0.042	0.006
	(0.023)	(0.020)	(0.021)	(0.020)
N	1,149,100	1,004,740	1,146,920	1,002,210
Adjusted R^2	0.029	0.033	0.027	0.026
F	0.12	0.18	3.80	0.08
	(5)	(6)	(7)	(8)
Waiver 2013	-0.010	-0.013	0.025	0.016
	(0.016)	(0.013)	(0.014)	(0.014)
N	1,018,860	877,700	1,016,230	875,530
Adjusted R^2	0.320	0.347	0.319	0.322
F	5118.87	6063.15	5599.01	6196.98

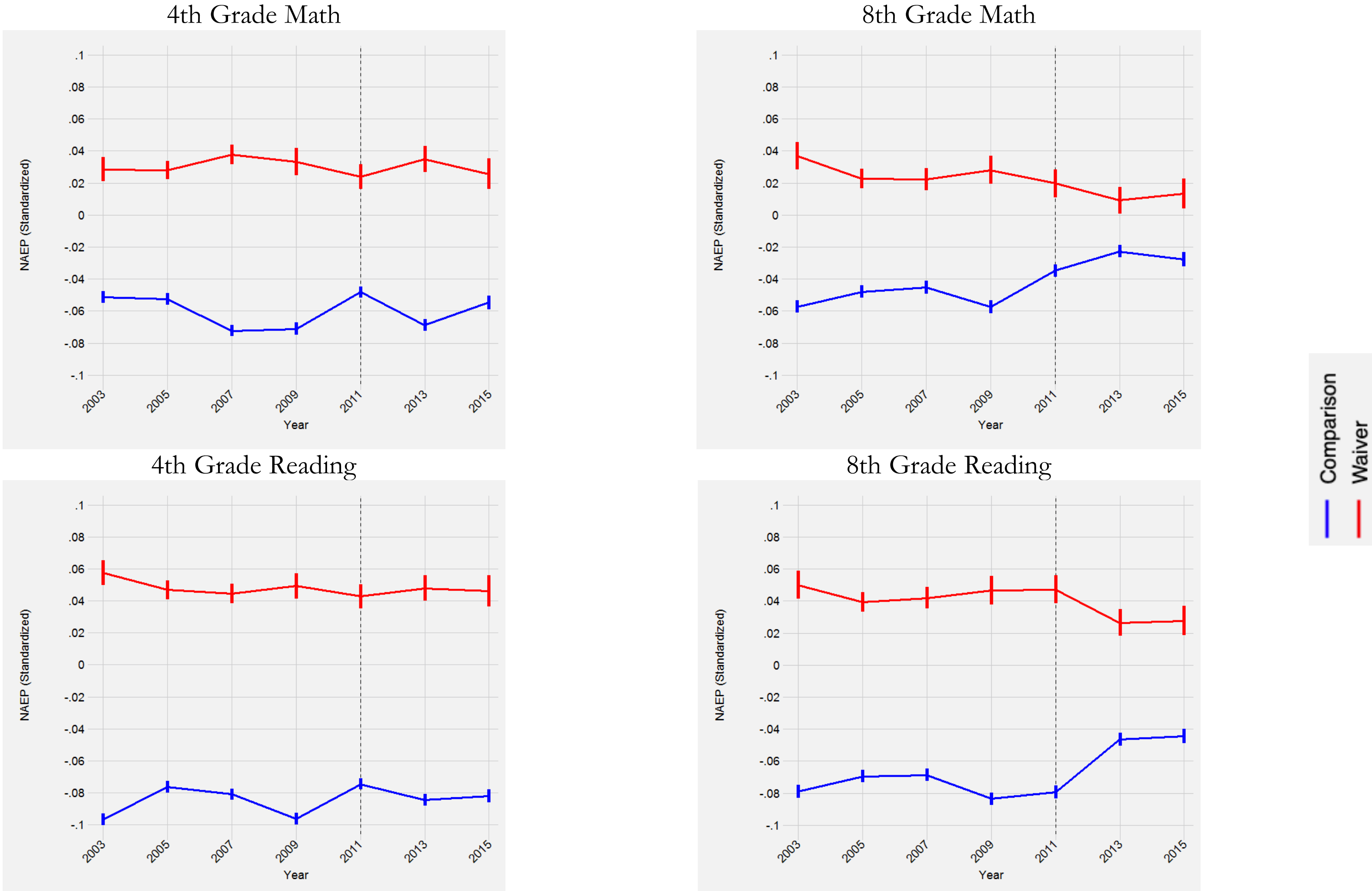
Note: Sample excludes states that never implemented waivers and includes NAEP data from 2015. waiver Effect 2013 compares states that received waivers in 2013 relative to states that received waivers in 2014. Standard errors are robust to clustering by school. See Appendix Tables A1 for composition of the treatment/comparison group and the matched comparison sample. Outcomes standardized within grade, subject, and year. See Table 3 for a full list of covariates. Standard errors robust to clustering by school. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2014-2015.

Table 12. Regressing NAEP Scores on Waivers Sanctions

Outcome	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)
Waiver Sanction	0.059***	0.013	0.150***	0.064**
	(0.018)	(0.018)	(0.018)	(0.021)
N	186,650	170,810	188,370	171,550
Adjusted R^2	0.001	0.000	0.005	0.001
F	11.13	0.52	66.79	9.26
	(5)	(6)	(7)	(8)
Waiver Sanction	0.033*	0.005	0.094***	0.038**
	(0.014)	(0.013)	(0.013)	(0.013)
N	155,770	142,760	157,140	143,430
Adjusted R^2	0.278	0.317	0.331	0.313
F	931.31	1270.80	1123.94	1248.81
Covariates	X	X	X	X
	(9)	(10)	(11)	(12)
Waiver Sanction Threat	0.160***	0.023	0.133***	0.026
	(0.020)	(0.017)	(0.017)	(0.017)
N	92,860	87,450	93,540	87,780
Adjusted R^2	0.290	0.313	0.327	0.300
F	667.51	825.29	743.60	723.34
Covariates	X	X	X	X
Matched Sample	X	X	X	X

Note: Sample restricted to include only schools in 2013 or 2015 where 15 to 85 percent of students were proficient on both math and reading exams. Sanction Threat is the estimated correlation from comparing student in schools would be sanctioned in the next year under NCLB but not the waivers to schools that would have been sanctioned under NCLB but not under the waivers. Standard errors are robust to clustering by school. See Appendix Tables A1 for composition of the treatment/comparison group and the matched comparison sample. Outcomes standardized within grade, subject, and year. See Table 3 for a full list of covariates. Standard errors robust to clustering by school. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, “Student and Teacher Survey,” 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2014-2015.

Appendix Figure A1. Long-Term NAEP Score Trends for Treatment and Comparison Groups



Note: The blue line is the comparison group and the red line is the treatment group. 2011 is the last NAEP year prior to waivers. Y axis is NAEP student outcomes standardized within subject/grade and year. Estimates adjusted using NAEP student level probability weights. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2014-2015.

Appendix Table A1. Receipt of Waivers by Month

State	Waiver Receipt	Treatment	Matched Sample	Notes
Alabama	Jun-13	0	0	
Alaska	May-13	0	0	
Arizona	Jul-12	1	1	
Arkansas	Jun-12	1	1	
California	X	0	0	Only CORE districts receive waiver
Colorado	Feb-12	1	1	
Connecticut	May-12	1	0	
Delaware	May-12	1	1	
District of Columbia	Jul-12	1	0	
Florida	Feb-12	1	1	
Georgia	Feb-12	1	0	
Hawaii	May-13	0	1	
Idaho	Oct-12	1	0	
Illinois	Apr-14	0	1	
Indiana	Feb-12	1	1	
Iowa	X	0	0	IA denied waiver
Kansas	Jul-12	1	1	
Kentucky	Feb-12	1	1	
Louisiana	May-12	1	1	
Maine	Aug-13	0	1	
Maryland	May-12	1	1	
Massachusetts	Feb-12	1	1	
Michigan	Jul-12	1	1	
Minnesota	Feb-12	1	0	
Mississippi	Jul-12	1	0	
Missouri	Jun-12	1	0	
Montana	X	0	0	MT withdraws waiver application
Nebraska	X	0	0	NE never formally applies
Nevada	Aug-12	1	0	
New Hampshire	Jun-13	0	0	
New Jersey	Feb-12	1	0	
New Mexico	Feb-12	1	1	
New York	May-12	1	1	
North Carolina	May-12	1	1	
North Dakota	X	0	0	ND withdraws waiver application
Ohio	May-12	1	1	
Oklahoma	Feb-12	1	1	
Oregon	Jul-12	1	0	
Pennsylvania	Aug-13	0	1	
Rhode Island	May-12	1	1	
South Carolina	Jul-12	1	0	
South Dakota	Jun-12	1	0	
Tennessee	Feb-12	1	1	
Texas	Sep-13	0	0	
Utah	Jun-12	1	0	
Vermont	X	0	0	VT withdraws waiver application
Virginia	Jun-12	1	0	
Washington	X	0	0	WA's waiver was revoked in 2014
West Virginia	May-13	0	1	
Wisconsin	Jul-12	1	1	
Wyoming	X	0	1	WY withdraws waiver application

Note: Waiver receipt data collected from Center on Education Policy (2018), U.S. Education Department (2016), and Comprehensive School Performance Reports (U.S. Department of Education, 2018b)

Appendix Table A2. Granger Test of Parallel Trends

Outcome	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)
Pre-Treatment 2003	-0.008	0.100**	0.041	-0.011
	(0.036)	(0.034)	(0.034)	(0.031)
Pre-Treatment 2005	-0.012	0.069*	-0.014	-0.027
	(0.033)	(0.031)	(0.032)	(0.029)
Pre-Treatment 2007	-0.010	0.028	-0.017	-0.039
	(0.033)	(0.032)	(0.031)	(0.029)
Pre-Treatment 2009	0.012	0.033	0.024	-0.016
	(0.034)	(0.033)	(0.033)	(0.031)
Post-Treatment 2013	0.026	0.008	0.063	-0.006
	(0.036)	(0.033)	(0.035)	(0.033)
N	903,210	779,940	901,880	778,190
Adjusted R^2	0.001	0.001	0.001	0.001
F	0.41	1.83	3.00	1.79

Note: See Appendix Tables A1 for composition of the treatment/comparison group and the matched comparison sample. Outcomes standardized within grade, subject, and year. Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table A3. Pre-Treatment Balance on Standards-Based Reforms

AMO MG4	0.584	Stand Consortium	0.673
	(0.467)		(0.479)
AMO MG8	0.390	Stand Adoption	0.289
	(0.429)		(0.224)
AMO RG4	0.412	Common Assessment	0.348
	(0.483)		(0.284)
AMO RG8	0.367	Sch Interventions	0.046
	(0.488)		(0.141)
Growth waiver	0.094	Measures Growth	0.263
	(0.147)		(0.161)
AMO Goal Structure	0.0001	Educ Eval Multiple Categories	0.380*
	(0.097)		(0.156)
Min N Reporting	0.016	Educ Eval Growth	0.450**
	(0.017)		(0.158)
Safe Harbor	0.099	Annual Educ Eval	0.147
	(0.185)		(0.135)
Prof Stand MG4	-0.002	Rewards Schools	0.131
	(0.005)		(0.150)
Prof Stand MG8	0.001	School Assistance	0.094
	(0.005)		(0.147)
Prof Stand RG4	0.004	School Sanctions	0.224
	(0.004)		(0.136)
Prof Stand RG8	0.008	School Acct before NCLB	.235***
	(0.005)		(.053)

Note: Results from bivariate models, where I regress whether a state ever received an NCLB waiver on a characteristic of their pre-treatment school accountability system. Models estimated at the state-level (N=51). AMO's and Proficiency Standards for each grade and subject are continuous. All other predictors are binary. Data on AMO's from Ed Data Express (2018a), Growth waiver receipt (Hoffer et al., 2011) NCLB accountability system features (Erpenbach et al., 2003; Erpenbach & Forte, 2005, 2007; Fast & Erpenbach, 2004; Forte & Erpenbach, 2006), proficiency standards (U.S. Department of Education, 2020), Bush/Obama era education policies (Howell & Magazinnik, 2017), pre-waiver school accountability policies (Dee & Jacob, 2009; Edweek, 2011). AMO=Annual Measurable Objective, Prof=Proficiency, Stand=Standard, Sch=School, Educ=Educator (Teacher or Principal), Eval=Evaluation, Acct=Accountability. *p < 0.05, **p<0.01, ***p<0.001.

Appendix Table A4. Baseline Outcome Differences by NCLB Waiver Characteristics

Outcome	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)
State Ever Receives Waiver	-0.158	-0.191	-0.058	-0.085
	(0.242)	(0.254)	(0.256)	(0.261)
N	50	50	50	50
	(5)	(6)	(7)	(8)
Placed on High Risk Status	0.363	0.293	-0.053	0.010
	(0.380)	(0.401)	(0.405)	(0.412)
N	50	50	50	50
	(9)	(10)	(11)	(12)
Months Until Waiver Receipt	-0.003	-0.006	-0.014	-0.011
	(0.019)	(0.020)	(0.020)	(0.020)
N	42	42	42	42
	(13)	(14)	(15)	(16)
Never Receive Waiver	0.470	0.633	0.377	0.513
	(0.306)	(0.317)	(0.327)	(0.329)
N	50	50	50	50

Note: Results from bivariate models, where I regress the state average NAEP test scores at baseline (2003) on a waiver characteristic. Models estimated at the state-level. Months Until waiver Receipt is a continuous variable equal to the months from February 2011 (when the waiver program was announced) to when the state received a waiver. All other variables are binary. waiver characteristics collected from Center on Education Policy (2018), U.S. Education Department (2016), and Comprehensive School Performance Reports (U.S. Department of Education, 2018b).

Appendix Table A5. Pre-Treatment Balance on Student Characteristics

Full Sample				
Characteristic	MG4	MG8	RG4	RG8
NAEP Score	0.0188*	0.0223*	0.0363*	0.0305*
Female	0.0013	0.0025	0.0005	0.0094*
IEP	0.043*	0.0087	0.0187*	-0.008
LEP	-0.3024*	-0.2489*	-0.3094*	-0.2742*
FRPL	-0.05*	-0.0442*	-0.0492*	-0.0419*
White	0.1248*	0.1094*	0.1193*	0.1137*
Black	0.1471*	0.122*	0.1349*	0.1322*
Hispanic	-0.3057*	-0.2749*	-0.292*	-0.2897*
Asian/PI	-0.1946*	-0.2135*	-0.2015*	-0.2186*
American Indian	0.1099*	0.1016*	0.0952*	0.0342
Modal age for grade; At	-0.0419*	-0.0382*	-0.0392*	-0.0362*
Below	0.0361	-0.012	-0.0248	-0.0111
Above	0.0416*	0.0385*	0.0396*	0.0365*
School made AYP in 2003	-0.0905*	-0.0831*	-0.0833*	-0.0805*
Matched Sample				
Characteristic	MG4	MG8	RG4	RG8
NAEP Score	0.003	-0.0011	0.0041	-0.0091*
Female	0.0012	-0.003	0.0061	0.0019
IEP	-0.002	-0.0158*	-0.0146	-0.0249*
LEP	0.0397*	0.065*	0.0555*	0.0703*
FRPL	0.0215	0.0228	0.0173	0.032*
White	-0.0294*	-0.0408*	-0.0314*	-0.0449*
Black	0.026	0.0424*	0.0216	0.0439*
Hispanic	0.0312	0.0454*	0.0407*	0.0526*
Asian/PI	-0.1025*	-0.1103*	-0.1052*	-0.1006*
American Indian	0.1578*	0.148*	0.1527*	0.1539*
Modal age for grade; At	0.0099*	0.0081	0.0067	0.0136*
Below	-0.0046	0.0069	-0.0136	0.0578
Above	-0.0098	-0.0082	-0.0066	-0.0144*
School made AYP in 2003	-0.0137	0.0514*	-0.0182	0.0562*

Note: Estimates from models where I regressed an indicator for whether a state received a waiver on each student or school characteristic in 2003. See Appendix Table A1 for the states in the treatment and comparison groups and the matched comparison sample. Standard errors are robust to clustering by school. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. IEP=Individualized Education Plan, LEP=Limited English Proficiency, FRPL=Free and Reduce Price Lunch, PI=Pacific Islander, AYP=Adequate Yearly Progress *$p < 0.05$.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003.

Appendix Table A6. Robustness to Time-Varying Policies

State Policy	MG4	MG8	MG4	MG4
Teacher Evaluation	0.061*	-0.023	0.026	-0.009
School Finance Reform	0.041*	-0.027	0.027	-0.007
Full Day Kindergarten	0.043*	-0.025	0.028	-0.006
Annual Teacher Evaluations	0.039*	-0.021	0.02	-0.012
Common Assessments	0.042*	-0.029*	0.026*	-0.001
Statewide Data System	0.041*	-0.025	0.027	-0.006
Data System with Identifiers	0.041*	-0.028	0.026	-0.005
Evaluation Firing	0.048*	-0.024	0.02	-0.01
Eval PD	0.035*	-0.028*	0.024*	-0.011
Eval Compensation	0.041*	-0.024	0.015	-0.012
Eval Responsibility	0.041*	-0.026	0.027	-0.006
Eval Grant Tenure	0.049*	-0.026	0.029	-0.013
Eval has Multiple Categories	0.047*	-0.021	0.026	-0.005
Evaluation Uses Student Growth	0.037*	-0.024	0.021	-0.009
Charter Authorizer	0.04*	-0.026	0.027	-0.006
Charter Building Funds	0.038*	-0.025	0.027	-0.006
Charter Cap	0.038*	-0.026	0.022	-0.007
School Turnaround	0.038*	-0.033*	0.026*	-0.006
Evaluation Growth Targets	0.037*	-0.027	0.027	-0.007
Alt Certification Pathways	0.041*	-0.026	0.027	-0.006
Alt Preparation Programs	0.04*	-0.021	0.026	-0.008
Vouchers	0.039*	-0.024	0.025	-0.004
High School Exit Exams	0.041*	-0.022	0.026	-0.005
Testing Grades 3-8	0.041*	-0.026	0.027	-0.006

Note: Reference category is the last year prior to the NCLB waivers (2011). Standard errors are robust to clustering by school. See Appendix Tables A1 for composition of the treatment/comparison group and the matched comparison sample. See Table 3 for a full list of covariates. Estimates adjusted using NAEP student-level probability weights. Eval=Evaluation, PD=Professional Development, Alt=Alternative. *$p < 0.05$. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table A7. Robustness to State Specific Linear Trends

Outcome	MG4	MG8	RG4	RG8
	(1)	(2)	(3)	(4)
Waiver Effect	0.004	-0.011	0.006	-0.022
	(0.013)	(0.012)	(0.012)	(0.012)
N	1,079,210	932,820	1,078,670	931,060
Adjusted R^2	0.033	0.035	0.029	0.028
F	0.08	0.80	0.24	3.48

Note: Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table B1. Main Regression Results with Covariates

	(1)	(2)	(3)	(4)
Waiver Receipt	0.041*	-0.026	0.027	-0.006
	(0.018)	(0.014)	(0.016)	(0.015)
Female	-0.121***	-0.108***	0.129***	0.223***
	(0.003)	(0.003)	(0.003)	(0.003)
IEP	-0.758***	-0.951***	-0.922***	-0.967***
	(0.005)	(0.006)	(0.006)	(0.006)
LEP	-0.532***	-0.709***	-0.657***	-0.842***
	(0.008)	(0.010)	(0.009)	(0.011)
FRPL Eligible	-0.439***	-0.379***	-0.442***	-0.362***
	(0.004)	(0.004)	(0.004)	(0.004)
Black	-0.657***	-0.652***	-0.510***	-0.534***
	(0.006)	(0.006)	(0.005)	(0.006)
Hispanic	-0.288***	-0.331***	-0.234***	-0.268***
	(0.007)	(0.007)	(0.006)	(0.007)
Asian/PI	0.329***	0.331***	0.187***	0.156***
	(0.014)	(0.012)	(0.011)	(0.010)
American Indian	-0.370***	-0.352***	-0.342***	-0.306***
	(0.014)	(0.014)	(0.015)	(0.015)
Other Race	-0.113***	-0.105***	-0.065***	-0.076***
	(0.012)	(0.014)	(0.012)	(0.014)
Below modal age	0.288***	0.313***	0.253***	0.205***
	(0.033)	(0.030)	(0.029)	(0.030)
Above modal age	-0.071***	-0.165***	-0.063***	-0.138***
	(0.003)	(0.003)	(0.003)	(0.003)
School Made AYP	0.168***	0.149***	0.155***	0.139***
	(0.007)	(0.007)	(0.007)	(0.007)
School had Safe Harbor	0.076**	0.037	0.097**	0.040
	(0.028)	(0.028)	(0.033)	(0.043)
Constant	0.468***	0.507***	0.335***	0.304***
	(0.007)	(0.006)	(0.007)	(0.006)
N	952,140	810,280	950,730	808,770
Adjusted R^2	0.323	0.348	0.318	0.323
F	4600.16	5562.79	5083.33	5613.86

Note: See Appendix Table A1 for the states in the treatment Sample size rounded in accordance with National Center for Education Statistics nondisclosure rules. Estimates adjusted using NAEP student level probability weights. FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix C. Observing School Sanctions 2003-2015

In most years I rely on the CSPR (U.S. Department of Education, 2018b). I use the CSPR to identify school sanctions in 2007, 2009, 2011, 2013, and 2015. In 2007 and 2009, some states reported no school identifiers (e.g., state, federal) and only names (e.g. school, district, state) were available. To recover NCES identifiers I used the STATA package reclink to "fuzzy name match" sanctioned schools to the Common Core of Data using school, district, and state names. Name matching allows me to observe NCLB waivers sanctions (i.e., focus, priority) but not NCLB sanctions (e.g., school improvement, corrective action, restructuring).

Appendix Table C1 describes the proportion of schools with NCES identifiers that were successfully merged into the NAEP. The percent of sanctioned schools with an identifier was quite high except in 2009 when only 73 percent of sanctioned schools were matched with an NCES identifier. Once merged into the NAEP data the differences between the number of sanctioned schools in the population is quite similar to the NAEP.

The trend in the percent of sanctioned schools is flat from 2005 to 2009, which reflects how states structured their AMO goals. NCLB allowed states to set their own targets if they were able to reach 100 percent proficiency by 2014. Many states deferred those goals until 2011 and later. The backloading of AMO targets accounts for the sharp increase in schools that were sanctioned. The proportion of sanctioned schools increases in 2013 because a very large proportion of schools in populous states (e.g., California, Washington, Illinois) were under sanction because they had not received waivers.

Appendix Table C: Sanctioned Schools

Year	Sanctioned Schools	Sanctioned with IDs	IDs Available
2003	NA	NA	NA
2005	6,093	5,394	88.5%
2007	10,266	9,820	95.7%
2009	10,812	10,389	96%
2011	20,700	19,820	95.7%
2013	20,338	20,264	99.6%
2015	16,126	15,986	99.1%
Year	**Schools under Sanction National**	**Schools under Sanction in NAEP**	**Difference**
2003	NA	NA	NA
2005	5.6%	3.9%	2%
2007	9.9%	14.8%	5%
2009	11.0%	17.2%	6%
2011	20.1%	21.2%	1%
2013	20.6%	18.0%	3%
2015	16.3%	18.8%	3%

Note: Sanctions schools determined based on the number of schools in CSPR reports overall. Sanctioned with IDs are the subset of schools in the CSPR report that IDs were either found or recovered. School under Sanction National is the number of sanctioned schools in CSPR report overall divided by the number of public schools. School under Sanction in the NAEP is the percent of schools that were under sanction in the NAEP sample. Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2014-2015.

CHAPTER 4

DOES THE COMMON CORE HAVE A COMMON EFFECT?: AN EXPLORATION OF EFFECTS ON ACADEMICALLY VULNERABLE STUDENTS

Introduction

In 2010 a national alliance of states moved rapidly to adopt the Common Core content standards (CC) due to concerns about low expectations for students. The popularity of the CC quickly declined because of insufficient support for implementation and the belief that reforms to content standards would harm students. About a quarter of the states that adopted the standards announced substantial revisions or revoked the adoption of the CC. Many modifications to the CC occurred before the standards were implemented in classrooms. But, state policymakers could not have judged whether the standards benefitted students prior to their implementation. Today, policymakers continue to debate whether or not to continue using the CC. The CC has received renewed attention because state laws mandate that states consider reform to content standards every 7-10 years. I provide new evidence about the effects of CC on student outcomes and achievement gaps that will inform decisions about future changes to the CC and content standards more broadly.

Questions about whether the standards have benefitted students remain unresolved (Polikoff, 2017). Studies have found the relationship between CC implementation and student outcomes is mixed (Loveless, 2014, 2015, 2016; Xu & Cepa, 2018; Gao & Lafortune, 2019; Song, Yang, & Garet, 2019). My study contributes to this growing literature by examining the influence of the CC throughout the country and on achievement.

I use the student level National Assessment of Educational Progress (NAEP) to estimate the initial causal effect of CC on student outcomes. I identify the effect of the CC by comparing student outcomes in states that were early implementors of the CC to late implementors of the CC. Within a difference-in-differences framework, I dynamically estimate the effect of preparing for the CC and implementation of the CC. In 2011, I estimate the effect of preparing for CC by comparing treated states that had begun preparation activities (e.g., professional development, content development) to comparison states that were still planning the implementation of the standards. In 2013, I estimate the effect of implementing the standards by comparing treated states that mandated alignment of instruction with the standards and comparison states where some had begun preparation activities. I restrict the sample to include the years from 2003 to 2013 to remove the endogeneity from changes to content standards after 2013. A tradeoff with my approach is that I capture the initial rather than long-term effect of the CC. The major barrier to identifying the causal effect of CC is that state capacity explains both the timing of implementation and changes to student outcomes. To mitigate concerns about this source of bias I demonstrate that early and late implementing states are quite similar across a broad range of capacities. I employ a critical quantitative approach to causal inference (Sablan, 2018) where I endeavor to disaggregate effects by race/ethnicity and to use quantitative intersectionality to test differences across diverse populations of students. In addition, when contextualizing my quantitative results I explore the role of racism and classism. My estimation strategy accounts for plausible confounders and identifies the initial causal effect of preparing for and implementation of the CC on student achievement overall and achievement gaps.

I find that robust to a variety of different estimators that CC increased NAEP scores in math, but not reading. The positive effect is larger among economically advantaged students than their peers who are eligible for Free and Reduced Priced Lunch (FRPL). Differences in state capacity for education reform and other policies adopted from 2003 to 2013 do not appear to explain the effects of CC on student outcomes.

The pattern of differential effects for academically vulnerable students is consistent with the hypothesis that the CC standards influenced student outcomes through raising expectations.

Students that struggle academically due to challenges that are a function of poverty (e.g., housing or food instability, lead exposure) will continue to struggle when state's raise expectations for their performance. But, when economically advantaged students face low expectations due to racist or classist beliefs about their ability to learn, raising expectations through changes to content standards could have a positive effect. States that have implemented the CC standards should refrain from making additional changes. But, without complementary policies meant to address student poverty the CC content standards will not lead to a closure of achievement gaps.

Common Core State Standards

The Common Core State Standards Initiative (CCSSI) was a joint project of the National Governors Association (NGA) and the Council of Chief State School Officers (CCSSO). The CCSSI pursued two standards-based reforms: development of new content standards, and development of new assessments. The CCSSI applied the CC brand to both projects, but there were key differences. The CC content standards were broadly supported by education reformers and stakeholders (e.g., AFT, NEA). Content standards are a list of learning goals that states define for teachers. States also set standards for curriculum and performance on summative assessments, but neither of these reforms were targeted by the CCSSI. CC is also used to described the Common Core testing consortia (i.e., Smarter Balanced, Partnership for Assessment of Readiness for College and Careers (PARCC)). The assessment consortia were groups of states that contracted with test writers to develop assessments that were aligned with the CC content standards. Additionally, companies have applied the CC brand to describe instructional materials (e.g., text books) that were aligned with either the content standards or assessments (Polikoff, 2015). However, instructional materials with CC branding are not explicitly part of the CCSSI. CC's public license allows businesses to use the CC brand for products that have educational purposes.

The CCSSI started writing the standards in 2008. Beginning in 2009 states began adopting the CC content standards in part due to incentives from Race to the Top and the Bill and Melinda Gates Foundation. By 2011, 45 states adopted the CC standards. By 2013, 15 states had implemented the standards in either math or reading. In 2014, the politics of CC soured, and Indiana became the first state to revoke the standards. The CC assessments first came online in 2015 after a year's long development process. I examine the effect of the content standards and use the term CC to refer solely to that reform. I narrow my focus because the overlapping implementation timelines present unique barriers to estimating the effects of each intervention.

State content standards prior to the CC varied widely in their rigor. A 2011 review gave a D or an F grade to 22 state English Language Arts standards and 15 state math standards (Carmichael et al., 2010). The rigor of state content standards has 3 main components: clarity/specificity, content and skills, and coverage (AFT, 2006). In some states content standards were described in a long narrative rather than an organized list. Not all states required the teaching of both content (e.g., literature, real-word examples) and skills (e.g., decoding, numeracy). Finally, in some states content standards did not cover every grade and subject. Prior to the CC, state expectations for students were quite low in many states.

Conceptual Framework

Theorized Benefits of Common Core

The CC content standards are more rigorous than previous content standards because they are specific and cover both content and skills for students in grades K through 12. Content standards may improve student achievement by clarifying and therefore changing what teachers ought to teach in the classroom. Ravitch explains that content standards, "define what teachers and schools should be trying to accomplish. They can raise the quality of education by establishing clear expectations about what students must learn if they are to succeed. If the goals of teaching and learning are spelled out, students understand that their teachers are trying to help them meet externally defined standards and parents know what is expected of their children in school" (2011, pp. 25–26). Content standards change the state's expectations for what is taught, which in turn changes what students learn.

The CC also could close achievement gaps by raising expectations for academically vulnerable students. A rich tradition of research has focused on Pygmalion effects or the ways that teacher expectations matter for student achievement (Rosenthal, 1987). Teachers have lower expectations for students who are Black and from low-income families (Ferguson, 2003; Gershenson et al., 2016). If the CC raised and equalized teacher expectations for academically vulnerable students to the same level as advantaged students then it could in turn close achievement gaps (Gamoran, 2008).

The CC may also improve student outcomes via other education policies linked to content standards. Contents standards serve as one of three key components in standards-based reform, along with assessments and accountability (Smith & O'Day, 1990). Content, "standards are the foundation upon which almost everything else rests" (Carmichael et al., 2010). Content standards like the CC determine the skills measured on assessments, which states use to determine which schools receive sanctions under accountability systems. Similarly, content standards influence other school activities (e.g., professional development, teacher evaluation, curricula). The CC could improve student outcomes via its influence on these other school policies.

CC could also improve the effectiveness of education technologies. Variation in standards across states creates barriers to the sharing of educational materials (Bleiberg & West, 2014). For example, if every state had different standards then a website designed for sharing lesson plans would have less value then if every state had the same standards. Universal adoption of standards produce network effects (Swann, 2000). As the number of system users (i.e., teachers) increases the size of the benefit for every network participant also increases. Standards also make it easier for firms to develop new products by decreasing development costs. Standardization creates a larger market and necessitates the development of fewer specialized products. The larger number of users and fewer product skews allows firms to increase their investment in developing new education technologies.

Theorized Tradeoffs of Common Core

There are several reasons to remain skeptical that the CC would have a positive effect on students. The committee that developed the CC did not represent the full range of grades and subjects. Although, many educators participated in writing the standards, teachers with expertise in early childhood grades were excluded (Ravitch, 2014). A lack of teacher input may have led to standards that were not developmentally appropriate. For example, critics of the CC argue the standards focused too much on skills and underemphasized imaginative play.

CC critics have also argued that the reading standards are also criticized for being overly prescriptive (Stotsky, 2013). There is general agreement that content standards ought to set goals for student learning while remaining agnostic to how educators achieve those goals. The CC reading standards specify that in elementary school teachers ought to use 50 percent informational texts and 50 percent fiction texts. Stotsky (2013) and other opponents of the CC have argued that requiring the use of informational texts violates the norm that teachers choose instructional materials in their classrooms. They further claim that the CC removes teacher autonomy and negatively influences the quality of instruction.

Educators did not receive sufficient supports to implement the CC (Xu & Cepa, 2018). Superintendents reported challenges related to finding adequate staff and financial resources to support all of the necessary implementation activities (Rentner, 2013). States were also implementing the CC when the Great Recession was causing funding cuts. States were raising their expectations for students but with fewer resources. A further complicating factor is that the schools serving large academically vulnerable populations have less capacity to implement the CC, which could end up disadvantaging the students the policy was intended to help. Staff from high-poverty districts reported less confidence in their capacity to implement the CC (A. B. Brown & Clift, 2010; Finnan & Domenech, 2014). The CC could have led to a decline in student performance as teachers and schools adjusted to the increased demands of the CC (Schmidt & Houang, 2012).

Teacher support is a critical component of any education reform, but it is particularly important for the CC. Today, equal numbers of teachers support and oppose the CC (Cheng et al., 2018). The lack of confidence in the CC is a particularly salient issue because changes to content standards will only have effects if teachers change their expectations for students. If teachers believe that the standards are not appropriate for their students then they will not make any changes to their instruction.

Content Standards on Student Outcomes

States began to pursue standards-based reform in the 1990s. These efforts also included implementing more rigorous content standards like the Principles and Standards for School Mathematics. However, there are no studies from the pre-CC period that isolate the effect of content standards on student outcomes. Two comprehensive literature reviews on the effects of standards-based reforms on students found no studies that estimated the effect of reforms to content standards on students (Hamilton et al., 2009; Lauer et al., 2005). Few studies were conducted because of the inherent complexity in examining standards-based reform. State changes to content standards virtually always coincided with reforms to assessments, accountability systems, or curricula. Contemporaneous standards-based reforms make it difficult to identify the effect of the content standards on student achievement. The interconnectedness of standards-based reform led

Dutro to conclude that, "We may never be able to directly answer the question 'What impact are state content standards having on student learning?'" (2002, p. 6). Fortunately, the CC differs from previous standards-based reform efforts because changes to assessments and accountability lagged behind changes to content-standards.

There are several studies that have examined the effect of the CC on student achievement. Loveless (2014, 2015, 2016) examines whether the similarity of a state's standards to CC is correlated with NAEP outcomes. He finds relatively small positive effect sizes ranging from 0.01 to 0.04 SDs. Overall, the descriptive differences between states that strongly implemented CC to states that did not adopt the standards appears to be small and insignificant. Xu and Cepa (2018) examine the effect of CC on ACT scores in Kentucky. They exploit the variation in exposure to CC across three cohorts. Students in the second two cohorts that received the CC had significantly higher ACT scores (0.03–0.04 SDs) compared to students in the first cohort.

Gao and Lafortune (2019) examine CC implementation in California and its effect on student outcomes. Using a statewide survey they collected information about districts' implementation processes. They exploit the variation in the timing of local adoption – as measured by the year in which a district adopted a CC aligned textbook – to examine the impact of CC standards adoption on student outcomes. In elementary and middle schools, the CC is associated with improvements in ELA achievement. In high schools, adoption districts saw their advanced placement passing rate increase by 1.3 percentage points.

Song, Yang, and Garet (2019) estimate the effect of adopting the College and Career Ready (CCR) content standards on NAEP state average test scores. CCR content standards includes three categories of states: CC implementing states, states that made substantive revisions to the CC, and states that never adopted the CC (i.e., developed their own standards). Content standards for states that made substantive revisions (Korn et al., 2016) and states that never adopted the CC have important differences with states that implemented the CC (Norton et al., 2017). Song, Yang, and Garet (2019) find moderately sized and significant negative effects of CCR on 4th grade average state NAEP scores (0.06 to 0.10 SDs). The analysis suggests that CCR had a significant negative effect on Black and Hispanic students in 4th grade reading and for students with disabilities in 8th grade math. In 4th grade math and 8th grade reading, they find statistically insignificant effects.

I isolate the effect of implementing the CC rather than other changes to content standards more broadly (CCR content standards). The CC is one example of CCR standards. States (e.g., Indiana, Texas) also wrote their own CCR standards. CCR standards were substantively different (23-27 percent) from the CC (Norton et al., 2017). The non-trivial differences in treatment motivate my focus on the effects of CC.

A weakness of Song, Yang, and Garet's (2019) approach is that the CCR treatment is endogenous. They assign states that had high rigor standards prior to the CCR to the comparison group and low rigor standards to the treatment group. States (Indiana, Oklahoma) that adopted the CC, revoked CC, and implemented their own standards, were assigned to their comparison group. They also include states (New York, North Carolina, and Pennsylvania) in their treatment group that implemented the CC, but also made major revisions. After the adoption of CC, but before either revising or revoking the content standards, student outcomes in states that revise the CC decline prior to implementation of the new CCR standards. It is likely that educators are reacting to the whiplash of multiple changes to content standards prior to the start of treatment. My study avoids confounding teacher reactions by using data from before states made changes to the CC.

Several qualitative studies have examined how the implementation of rigorous content standards can change instruction. Collaborating with other teachers improved the confidence of teachers that were developing CC aligned content materials (Herman et al., 2016). Teachers that do not feel they have authority over the implementation of content standards were less likely to make changes to their instruction (Edgerton & Desimone, 2019). Teacher collaboration and autonomy may mediate the effect of the CC on student outcomes via changes to instruction.

Contribution

I develop a measure of CC content standard implementation for each state in specific grade-subjects (4th grade math, 8th grade math, 4th grade reading, 8th grade reading). State definitions of "full" content standards implementation varied considerably. Some states only considered the standards implemented if the CC standards and assessments were in place. Other states only considered the standards fully implemented when they were required for all grades and subjects. In addition, many states staggered the implementation of the CC across grades and subjects. Using a measure that is specific to states, grades, and subjects allows me to more precisely estimate the effect of CC.

I am able to isolate the effect of the CC and related preparation activities (e.g., PD, curricula). I estimate effects by comparing early implementors of the CC to late implementors of the CC in the period before states began making endogenous changes to their content standards. During the period of study, virtually every state makes a change to their content standards, which makes it challenging to identify a defensible comparison group. My solution is to exploit variation in the implementation of the CC over time.

Finally, I am also able to estimate the intersectional effects of CC. Previous studies have used the state-level NAEP to examine the effects of the CC. State-level datasets can test for changes in outcomes between two groups of students (i.e., Black and white). But, a unique advantage of the student level data is that I can estimate effects of CC for students that belong to multiple academically vulnerable groups (i.e., Black and FRPL). The intersectional effects of CC allow greater insight into how the benefits of CC were distributed across diverse groups of students.

Research Questions

Specifically, I ask the following questions:

1. To what extent did Common Core affect student achievement?
2. To what extent did Common Core close or exacerbate achievement gaps?

Data, Measures, and Sample

I use data from four subject/grade NAEP datasets (4th grade math, 8th grade math, 4th grade reading, and 8th grade reading) over six waves (2003, 2005, 2007, 2009, 2011, and 2013). The NAEP study uses a complex three-stage sampling design to allow for valid inferences about student achievement outcomes for the nation as a whole, each state, and certain school districts (Rogers et al., 2014). Two strengths of the NAEP are that the assessment items rarely changed across waves, and that the sample includes students from diverse backgrounds (including students with Individualized Education Plans and those with Limited English Proficiency) (Rogers et al., 2014). The NAEP assesses a broader set of skills than the average state summative assessment. The broadness of the state frameworks makes the NAEP particularly useful for examining the CC, which expands the scope of what states expect teachers to learn. Another strength is the low-stakes nature of the NAEP assessment for students and teachers. Accountability pressures on students and teachers could induce measurement error in tests that states use to evaluate schools (D. Koretz, 2017; D. Koretz & Hamilton, 2006; D. M. Koretz & Barron, 1998). NAEP's purpose is to inform policy and practice, mitigating the incentive for cheating or gaming.

I merged into the NAEP, data on pre-CC content standards from the American Federation of Teachers (AFT, 2006) and the Fordham Institute (Carmichael et al., 2010; Finn Jr et al., 2006; Klein et al., 2005). I categorize pre-CC standards as either low or high-rigor. Low-rigor standards are "clearly inferior" to the CC according to Carmichael and colleagues (2010). Standards in the other group were either "indistinguishable from the CC" or were "superior to the CC".[12] I also merged in Adequate Yearly Progress data from 2003 as a measure of baseline school achievement data (Reback et al., 2013). Finally I merge in data on education reforms adopted during the period of study including teacher evaluation (Bleiberg & Harbatkin, 2018), ESEA Waivers (Center on Education Policy, 2018), high-school exit exams, and alternative pathways to teaching (Howell & Magazinnik, 2017).

Dependent Variable

To construct my outcomes of interest, I rely on test score information from six waves of the NAEP. The NAEP is a matrix-based assessment in which each student completes a sample of test items. The NAEP provides plausible values that are created through an Item Response Theory (IRT) procedure. NAEP then transforms the plausible values into scale scores. I then standardized the scale scores within grade, subject, and year to have a mean of 0 and a standard deviation of 1. I use the first plausible value as my dependent variable.[13]

Treatment Indicator

To measure changes to state content standards, I conducted a document analysis (Bowen, 2009) (See Appendix C). I collected 123 documents from state education agencies (e.g., reports, websites, grant and waiver applications, implementation timelines), surveys, interviews, media reports. All documents were collected from online sources. I made extensive use of the Internet Archive to obtain documents that were taken offline. I define standards implementation as the state

---

[12] The two measures of pre-CC standards rigor are strongly correlated. I use the Fordham measure because it is available in multiple years. The AFT variable identifies fewer states with low-rigor standards, which restricts the power in my preferred specification.
[13] The means of the NAEP test scores are different than zero in the analytic sample due to listwise deletion. The results are insensitive to other approaches that use the plausible values. See section on *Multiple Plausible Values* for more details.

mandating the alignment of instruction and curricula with a set of standards for a specific grade-subject (i.e., 4th grade math, 8th grade math, 4th grade reading, and 8th grade reading). My measure of Common Core excludes states that implemented CCR standards that were not CC or made substantive changes to CC (e.g., major revisions, rebranding, revoking the standards) through 2015.

The differences in definitions of standards implementation motivate my use of document analysis, which is particularly valuable for studying dynamic historical events like state policy implementation (Bowen, 2009). Document analysis is also useful tool when implementation timelines are not congruent across sources. Whenever possible, I triangulate sources and discuss divergent cases. Ideally, multiple sources of different types (i.e., government documents, interview data, media reports) describe the same implementation date. For all states I use multiple sources to corroborate the implementation date of the CC standards. I measure when the CC standards were adopted, when implementation was planned, when implementation occurred if at all, and when an alternative set of standards was implemented. Analyzing state specific documents across time increases my confidence that I have observed when implementation occurred. For example, if the documents show that a state adopted the CC standards in May 2010, one month later describes plans to implement in 2013, and then reports in December 2014 that implementation occurred in 2013, then my assertion that implementation occurred in 2013 is valid. I find that states adopted the CC standards from February 2010 to June 2012 and implemented the standards from the 2012 school year to the 2015 school year. Two states implemented CC in 2012 and ten more followed in 2013 (See Appendix Table A1).

I define two CC treatment indicators. *CC 2011* measures preparation for CC for early implementing states compared to late implementing states. Schools were engaged in a variety of activities to prepare for the CC prior to formal implementation of the standards (e.g., professional development, curriculum). The crux of the CC intervention is raising expectations for student learning. The formal change in state content standards is observable for a precise school year. However, there is also an informal change where educators adjust their own expectations. CC implementation (*CC 2013*) is the effect of state mandated alignment of instruction with the CC for early implementing states compared to late implementing states.

Surveys of state and school leaders support the notion that CC preparation activities were underway prior to formal implementation. States required districts to engage in CC preparation activities. Among a sample of 36 states that had adopted the CC in 2010, 13 states required districts to provide professional development for teachers and principals to support implementation of the CC, and 22 reported that districts were expected to do so (Kober & Rentner, 2011a). Among CC adopters, 11 states required districts to align teacher evaluation systems with CC, and 10 required the alignment of new curriculum materials and/or instructional practices with CC. Thirty-seven states reported providing, guiding or funding professional development on the CC in the 2011 school year (Webber et al., 2014). Sixty-six percent of school districts in states that had adopted the CC reported intentions to develop a comprehensive plan and timeline for implementing the CCSS in either 2011 or 2012 (Kober & Rentner, 2011b).

The NAEP teacher survey shows a jump in the emphasis of professional development on content standards in 2011 when compared to 2009. About 1 percent more teachers in 2011 reported that the extent to which they learned about content standards during professional development was large compared to 2009. A national survey (Markow et al., 2013) found that 46 percent of principals and 62 percent of teachers reported that a great deal of teachers in their school were using the CC in the 2012 school year when only 3 states (Nevada, Kentucky, and the District of Columbia) were requiring full implementation the standards. Teachers began aligning their instruction to the CC before state mandates. If there is an effect of CC, I ought to be able to detect it in 2011 and would expect its size to increase in 2013.

Covariates

The NAEP student survey contains a robust set of student characteristics. I control for exogenous student characteristics including gender, whether the student has an Individualized Education Plan (IEP), Limited English Proficiency (LEP) status, eligibility for Free or Reduced-Price Lunch (FRPL), and race/ethnicity. I also add measures for whether the student is at, above, or below the modal age for their grade level. These exogenous student characteristics control for observable differences between the students in states that were early and late implementors of the CC that are correlated with student outcomes. I also include a baseline measure of school achievement (AYP status in 2003) and lagged state average NAEP scores.[14] Baseline AYP status controls for pre-treatment differences in student outcomes.

Sample

Table 1 describes the states I assign to the treatment and comparison groups (See Appendix Table A2). I observe 8 states to implement the CC early in 4th grade math, 7 states to implement the standards early in 8th grade math, and 10 states to implement early in reading (4th and 8th grade). Figure 1 visually displays which states implemented the standards early by grade and subject. Early implementing states are spread out through the nation and appear to be diverse politically and demographically (LaVenia et al., 2015). For each grade and subject there are about 24 comparison states that implemented the standards late (2014 or 2015).

States were excluded from the analytic sample for three reasons. First only, states that had low rigor content standards prior to the CC were included (Carmichael et al., 2010). Ideally, I would use states that had no content standards as a control group, but every state had content standards

---

[14] To create the lagged state average for 2003 I use scores from 2002 for reading and 2000 for math.

prior to the CC. States that made major revisions to their content standards during the year 2014 and 2015 were also excluded. Substantive revisions made to the CC standards would likely confound the true effect of the CC. Finally, I exclude states that never adopted the CC, but did reform their content standards (i.e., Alaska, Texas). These states adopted standards that are substantively different from the CC. Each grade-subject includes about 2,000 school districts and about 4,000 schools. In total there are about half a million students for each grade subject.

Table 2 includes descriptive statistics for the analytic sample. The first column contains mean student, school, and locale characteristics for the pre-treatment period (2003-2009). The second and third columns describe means for the CC preparation (2011) and implementation year (2013). Most observable characteristics change very little across time. There was an increase in the number of students eligible for FRPL, likely due to the Great Recession. NAEP scores decline slightly in the pre-treatment period compared to the treatment period, except for 4th grade math.

Estimation Strategy

I estimate the causal effect of the CC on student achievement in a difference-in-differences framework. I compare states that were early implementors of the CC (2011 to 2013) to late implementors (2014 to 2015). I begin by estimating a series of models that assume the following general form:

$$(3)\quad y_{icst} = \beta_1 CC_s \times 2011_t + \beta_2 CC_s \times 2013_t + \rho \boldsymbol{F'}_{it} + \tau \boldsymbol{G'}_{ct} + \alpha_s + \pi_t + e_{icst}$$

Where *y* is a NAEP test score (standardized by subject/grade and year) for student *i*, school *c*, state *s*, and year *t*. *CC* ×*2011* is a binary variable equal to 1 if a state is preparing to implement for the CC in 2011. *CC* ×*2013* is a binary variable equal to 1 if a state has mandated alignment of instruction with the CC. β_1 is the effect of preparing for CC on NAEP scores within states that were early implementors of the CC (2013). β_2 is the coefficient of interest, the effect of implementing the CC on student outcomes within states that were early implementors of the CC

(2013). $\boldsymbol{F}$ and $\boldsymbol{G'}$ are vectors of time-varying student and school covariates. α_s is a vector of either state or school district fixed effects. π_t is a year fixed effect and *e* is an idiosyncratic error term clustered by school.[15] I estimate each model 4 times using each of the NAEP datasets (4th grade math, 8th grade math, 4th grade reading, 8th grade reading).

I then estimate a non-parametric event-study specification, which models pre- and post-treatment effects in a fully flexible way:

$$(2)\quad y_{icst} = \tau_1 CC \times 2003_{st} + \tau_2 CC \times 2005_{st} + \tau_3 CC \times 2007_{st} + \tau_4 CC \times 2011_{st} + \tau_5 CC \times 2013_{st} + \rho \boldsymbol{F}'_{it} + \tau \boldsymbol{G}'_{ct} + \alpha_s + \mu_{icst}$$

The coefficients in the event study estimate effects relative to outcomes in 2009, the last year prior to CC. For the pre-treatment years τ_1 , τ_2, and τ_3 model anticipatory effects of CC relative to 2009. In the two post-treatment years τ_4 and τ_5 estimate the effect of CC relative to 2009. Equation 2 includes state or district fixed effects and the full set of covariates in equation 1.

To answer the second research question I add interactions between the treatment indicators, membership in race/ethnic groups, and eligibility for FRPL. Here I employ a critical quantitative approach (Sablan, 2018). I leverage the detailed information about student race/ethnicity by not aggregating racial subgroups. For example, I test for effects within groups of Hispanic/Latinx student (e.g., Mexican, Cuban, Puerto Rican). I also test whether the effect of CC differed for race/ethnic groups across levels of socio-economic status (i.e., FRPL eligibility).

Threats to Causal Inference

The key assumption required for estimating a causal effect is that outcomes for students in treated states (early CC implementors) would have followed the same trajectory as students in

[15] Following Abadie, Athey, Imbens, & Wooldridge (2017) I cluster my standard errors at the school level. I cluster the standard errors by school because the errors of students in schools are correlated due to the IRT procedure employed by NAEP. In addition, clustering at the school level is appropriate because there are schools in the population that I do not observe in the sample.

comparison states (late CC implementors) in the absence of treatment. If the treatment and comparison groups had systematically different pre-treatment trends then the assumption of parallel trends is likely violated. Figure 2 shows a flat pre-treatment trend for both treatment and comparison states prior to the implementation of CC. For 4th grade math and 8th grade math the mean outcome differs by less than 1.5 percent of a standard deviation. Visually the pre-treatment trends in math appear flat for both the treatment and comparison groups. For 4th and 8th grade reading there is visual evidence that the assumption of parallel trends is violated. The pre-treatment trends for the treatment and comparison groups in reading cross, which implies their trajectory post-treatment may be attributable to something other than implementing the CC. The differing pre-treatment trends invalidates the differences-in-differences estimate of CC's effect on reading outcomes.

A salient issue when estimating the effect of CC are changes that states made to standards after the adoption of CC. Starting in 2014, several states made major revisions to the CC and some revoked them entirely. In 2014 and later, teachers will react to announced changes and revisions, which will change how the CC influences student outcomes. I avoid potentially endogenous teacher reactions by restricting the period of study from 2003 to 2013. The sample restriction also avoids conflating the effect of the CC standards with the CC assessments which were first used in 2015. A remaining issue is the possibility of unobserved state reforms that occurred contemporaneously with the implementation of CC and influence student outcomes. For example, if states implemented teacher evaluation at the same time as CC, then teacher evaluation would bias the effect of CC . The fixed effects control for any time-invariant state or district policy that would bias the effect of CC. Additionally, I find that the results in math are robust to controlling for time-varying education policies (See Appendix Tables A5 and A6).

A final concern is systematic differences between the treated states that chose to implement the CC early and the comparison states that chose to implement the CC late. For example, if the states implemented the standards early because they knew they had high levels of capacity then the high levels of capacity could explain any positive effects. It is also possible that late implementing states waited because they thought they lacked the capacity to implement the CC. The lack of capacity could also explain changes in student outcomes. State capacity could vary based on experience with implementing rigorous content standards. To account for the threat from pre-treatment differences in state capacity, I restrict the sample in my preferred model to include only states that had low-rigor standards prior to CC. States that had high rigor standards prior to the CC may have also implemented other standards-based reforms that could bias the estimate of interest. The sample restriction also improves the contrast between the treatment and comparison groups. There are no significant differences between treatment and comparison states on observable measures of state capacity (i.e., educational resources, political capacity, standards-based reforms, prior content-standards rigor) for education reform (See Appendix Figures A1, A2, A3, A4). Additionally, the state and district fixed effects will also account for any state- and district-level selection bias, respectively, that is not accounted for by the covariates.

Results

Figure 2 depicts the trends in outcomes for the treatment and comparison groups. Each panel describes the trend for a NAEP grade-subject (4th grade math, 8th grade math, 4th grade reading, 8th grade reading). The X axis is the NAEP year and the Y axis is NAEP student outcomes standardized within subject, grade, and year. CC (red line) describes average outcomes for students in states that were early implementors of the CC. Comparison (blue line) describes the average outcomes for students in states that were late implementors of the CC. 2009 is the last wave prior to the start of preparation for CC in 2011 and the implementation of standards in 2013. Average 4th

and 8th grade math outcomes for comparison states are about flat from 2003 to 2013. In 4th grade math, average outcomes increase for states that were preparing for CC and had implemented CC. In 4th grade math, average NAEP scores were about 3 percent of a standard deviation (SDs) higher in 2013 compared to 2009 and about 2 percent of a SD higher in 2013 compared to 2011. In 8th grade math, the outcomes for treatment states increase in 2011 before dipping in 2013. The pattern of results for reading do not suggest any change in scores after the implementation of CC.

Regressions

Table 3 includes the descriptive regression results from models without any sample restrictions. Models 1 through 4 include math results and models 5 through 8 include reading results. Columns 1 and 3 include state fixed effects and columns 2 and 4 add covariates. After adding covariates to the model the effect of fully implementing the CC on 4th grade math is about 6 percent of a standard deviation. There is no detectable effect of the CC on 8th grade math outcomes or reading outcomes in either 2011 or 2013. The estimates in Table 3 likely underestimate the effect of CC because of poor contrast between the treatment and comparison groups.

Table 4 describes the effect of the CC on math outcomes in the analytic sample with district and state fixed effects. Columns 1 and 2 include state fixed effects and columns 3 and 4 include district fixed effects. In columns 2 and 4, I include covariates. The estimates in Table 4 compare students in early implementing states to students in late implementing states. Implementing the CC appears to have a positive and significant effect on math scores in Table 4. In column 2, model 1, the effect of preparing for the CC on 4th grade math scores is about 4 percent of a SD and the effect of implementing the CC is about 10 percent of an SD. For 4th grade the effect of implementing the CC is about twice as large as the effect of preparing for the CC. The effects are larger than in Table 4 than in Table 3 due to the sharper contrast from excluding states that had rigorous content standards and removing states that made endogenous changes to content standards.

Table 5 shows the effect of the CC on 4th and 8th grade reading outcomes in the analytic sample with district and state fixed effects. The non-parallel pre-treatment trends for reading (Figure 2) violates the assumption required to identify causal effects in a difference-in-differences framework. The pre-treatment trends cross multiple times and imply the direction of the bias could be either negative or positive. The results in Table 5 are consistent with the descriptive results. Mean reading outcomes for treated states in 2011 and 2013 are insignificantly different from 2009. The standard errors are not sufficiently precise to rule out even small effect sizes (0.02 to 0.03 SD).[16]

Event Study

Table 6 includes the results from the event study. The 4 columns describe results from each of the NAEP datasets (4th grade math, 8th grade math, 4th grade reading, 8th grade reading). The models in Table 6 include district fixed effects and covariates. The pre-treatment coefficients test for the presence of anticipatory effects. The pre-treatment estimates are both individually and jointly indistinguishable from zero. In the event study there is no evidence of trends in student performance prior to the CC, after the inclusion of controls. For 4th grade math the effect of implementing the CC is about twice as large as the effect of preparing for the standards (9.5 percent of a SD).[17] In 8th grade math the effect of CC is about 4 percent of an SD in both post-treatment years. Consistent with the previous models, the effects of CC are significant in math but not reading.

Differential Effects for Academically Vulnerable Students

Table 7 adds interactions between membership in academically vulnerable populations and implementation of the CC. The first row describes the main effect of implementing the CC, which here is interpretable as the effect of implementing the CC for white economically advantaged

16 See Appendix Table B1 for regression results that include covariate estimates.
17 See Appendix Table B2 for regression results that include covariate estimates.

students (FRPL ineligible). The subsequent rows compare outcomes for academically vulnerable students relative to white economically advantaged students. Appendix Figure A5 visualizes the effect of CC for race/ethnic groups. Across race/ethnic groups CC contributes to the closure of achievement gaps. The white-Black achievement gap is about 5 percent an SD smaller after the CC in 4th grade math and about 6 percent of an SD smaller in 4th grade reading. In 4th and 8th grade math, CC shrinks the white-Hispanic achievement gap by about 16 percent of an SD. 4th grade math outcomes for FRPL eligible students decline by 6 percent of an SD after CC. In math, the benefits of CC were shared across race/ethnic groups, but not across socio-economic status. Similar to main the results, I cannot detect the effect of CC on academically vulnerable populations on reading outcomes.

Intersectional Effects

Figure 3 describes the effect of implementing the CC for economically advantaged and disadvantaged students from different race/ethnic groups. To produce the estimates in Figure 3, I add interactions between implementing CC, membership in a race/ethnic group, and a measure of economic advantage (FRPL eligibility). The effect of CC on 4th and 8th grade math is about 5 percent of an SD larger for economically advantaged white students when compared to economically disadvantaged white students. The effect of CC on 4th grade math is about 15 percent of an SD larger for economically advantaged Black students when compared to economically disadvantaged Black students. In 8th grade math there is no significant difference between CC's effect on economically advantaged and disadvantaged Black students. The positive effect of CC was larger for economically advantaged white and Black students than the effect of CC for economically disadvantaged white and Black students in math. The CC had a larger positive effect (0.1 SD) for economically disadvantaged Hispanic students than for economically advantaged Hispanic students in 4th grade math.[18] A possible explanation is that expectations of economically disadvantaged Hispanic student's English comprehension was systematically lower than economically advantaged Hispanic students, but there were no differences in English comprehension. If true, raising expectations for Hispanic students would disproportionately benefit economically disadvantaged Hispanic students.

Mechanisms

I hypothesized that the CC would change teacher instructional strategies and school administration. I argue that the clearest channel through which CC could influence student outcomes is by changing teacher expectations for students. I do not observe teacher expectations, but I observe teacher reports of several instructional strategies that are conceptually related. For example, differentiated instruction involves tailoring teaching to the needs of individual students. A common differentiated instruction strategy is to develop student specific achievement standards. Teachers set different goals for students based on their expectations. The CC should decrease differentiated instructional overall and more specifically decrease the use of differentiated achievement standards because the CC ought to raise and equalize expectations for all students. The CC's influence on expectations could also change other instructional activities. Under the direction to raise expectations, teachers may increase their emphasis on core subjects, discuss student's current performance more frequently, set more explicit learning goals, and adjust teaching strategies to help students succeed. If the CC influenced any of these instructional strategies then it could mediate the effect of the content standards on student outcomes.

Teachers need to have the proper resources and supports in place to implement the CC. The level of resources are a salient issue for understanding the CC's effects because teachers in schools

---

[18] The estimates here aggregate Hispanic students into a single group for the sake of parsimony. The effects for Mexican, Puerto Rican, Cuban, and other Hispanic students are qualitatively similar.

that serve academically vulnerable students report not having the resources they need to implement the CC (A. B. Brown & Clift, 2010; Finnan & Domenech, 2014). For CC to work teachers also need to have a clear sense of the standards and understand how they should change their instruction. Professional development for content standards should increase if there is any effect on CC. Edgerton and Desimone (2019) find that teachers need to have discretion over content standards for them to make changes to their instruction. A related concern is that the CC would change how schools used instructional materials. Teachers might face pressure to use scripted lesson plans or teacher tools. Standards could also make technologies more efficient for teachers (Bleiberg & West, 2014). If CC increased the expectations that teachers had for students or if teacher's used technology in the classroom more frequently then it could improve student outcomes.

Using the NAEP teacher and school surveys, I test each of my hypotheses in a mediation framework. Data on hypothesized mediators was only available in three years (2009, 2011, 2013), which allows for only one year of pre-treatment data. The Likert scale items have either 4 or 5 possible responses. To simplify the interpretation of the estimates I standardize each item within grade, subject, and year. In Tables 9 and 10, I use a single independent variable indicating either preparation for or implementation of the CC. Each of the models include district fixed effects and student covariates.

Tables 9 and 10 describe correlations between CC and dimensions of teacher's instruction in math and reading. The preponderance of the teacher[19] and school constructs[20] I test were not

[19] I constructed the Differentiated Instruction factor using 5 items, asking to what extent do teachers: Set different achievement standards for some students, Supplement the regular course curriculum with additional material for some students, Have some students engage in different classroom activities, Use a different set of methods in teaching some students, Pace my teaching differently for some students.

[20] School mediators measured the extent a school's program was structured according to: curriculum standards or frameworks, District curriculum standards or curriculum guides, Results from state/district assessment, In-school curriculum frameworks and standards for learning, Results from school assessments, Recommendations from school reading/language arts department, Discretion of individual teachers, and Commercially designed programs.

significantly correlated with CC and did not mediate the effect of CC on student outcomes. CC is also associated with less frequent usage of different achievement standards for students. CC is associated with a 0.03 SD decline in the use of different achievement standards in 4th grade math, 4th grade reading, and 8th grade reading. The negative correlation between CC and achievement standards was twice as large in 8th grade math (0.07 SD). The negative relationship between CC and differentiated instruction is consistent with the idea that the CC content standards equalized teacher expectations for students. CC's relationship with other dimensions of teacher instruction is inconsistent across grades and subjects. CC is associated with a 0.03 SD increase in subject emphasis, except for 8th grade where the relationship is negative (-0.03 SD). CC is correlated with decreases in subject emphasis, setting goals, and determining instructional adjustment (0.02-0.03 SD) except for 4th grade math where each correlation is insignificant.

I hypothesized that teachers need to have autonomy over the standards and have instructional resources and supports to implement the CC. I am not able to detect a significant relationship between the CC and teacher discretion or usage of commercially designed products. I argued that the CC could increase teacher's use of technology. But, the available evidence suggests that CC was associated with less frequent use of computers for instruction (0.01-0.02 SD) in 8th grade math, 4th grade reading, and 8th grade reading. CC is associated with a significant increase in teacher participation in professional development on content standards (0.01-0.02 SD) in each subject except 4th grade reading. The CC is correlated with teachers reporting significantly more instructional resources in math (0.018-0.019 SD). In reading, the association between CC and reporting sufficient instructional resources is insignificant and precise enough to reject very small correlations (0.01 SD). Teachers appear to have some but not all of the resources and supports they needed to implement the CC.

Long-Term Outcomes

In Table 10, I estimate the effect of CC after 2013. I supplement the event study approach I use in Table 6 with data from 2015 and 2017. As expected the pre- and post-treatment estimates are quite similar for the years 2003 through 2013. The estimated effect of CC declines slightly (0.004 SD) from 2013 to 2015 and then 0.02 SD from 2015 to 2017 in 4th grade math. In 8th grade math the estimated effect of CC is about 0.045 SD in 2015 and 0.07 SD in 2017. In 4th grade reading and 8th grade reading there are no detectable effects of the CC for either 2015 or 2017. The positive effect of the CC primarily occurs in 2011 and 2013 during the preparation and implementation phase.

The effect of CC in 2015 and 2017 is biased for two reasons. First, teachers are reacting to the decline in support for the CC among education reformers and state education officials. Starting in 2014, many states made changes to their standards or announced they were considering changes. In a tumultuous policy environment, the effect of standards-based reforms like CC is decreased because of teacher reactions to policy churn (Hess, 1998; Hess & McShane, 2014; O'Day & Smith, 2019). Teachers in states that implemented the standards before 2015 will be less sensitive because they received treatment when subsequent changes to the CC were exceedingly unlikely. But, for teachers in states that were late implementors they could reasonably assume the content standards would not be strictly enforced or revoked quickly. If true then the CC treatment would have no effect on late implementors and early implementing states would see few additional changes relative to 2011 and 2013. The flat long-term (2015 and 2107) results are consistent with the hypothesis that teachers are reacting to changes in the CC treatment.

Robustness Checks

Balance on State Capacity

I construct the treatment and comparison groups based on when states choose to implement the CC. If there were systematic differences in the capacity of early implementing and late implementing states then it would bias the effect of CC. Capacity for state education reform is a multifaceted concept (Manna, 2006). To test whether there were differences between treatment and comparison sates, I collected measures of education resources, political capacity, standards-based reforms, and content standards rigor. Using a state-level (N=51) dataset I ran bivariate models, where I regressed an indicator for whether states were early or late implementors of the CC on state capacity characteristics that were measured prior to CC. The results from the models are visualized in Appendix Figures A1 through A4. There are no observable differences in state capacity between early and late implementing states. The document analysis suggests that the availability of the CC assessments influenced when states chose to implement the CC content standards. Test writers were developing and piloting the assessments from 2010 to 2014 and they were first administered in 2015. Forty-seven percent of states that were early implementors of the CC standards chose to use the CC assessments in 2015, whereas 73 percent of states that were late implementors of the standards chose to implement the standards and assessments in the same year (2015). Unobservable differences in capacity to implement content standards may have biased the results. But, the document analysis suggests that states were influenced by the availability of the assessments, which is exogenous prior to 2015.

Balance on Observable Characteristics

The characteristics of students in treatment and comparison group states could have also motivated when states chose to implement the CC. For example, if early implementors of the CC had more students that were academically vulnerable then they may have pursued other changes that could explain improvements in student outcomes. Appendix Table A3 describes results from bivariate models, where I regressed an indicator for whether states were early or late implementors of the CC on student characteristics measured in 2003. All of the differences are either statistically insignificant, quite small (less than 0.05 standard deviations), or fall within the range (0.05 SDs to 0.25 SDs) where covariate adjustment is an appropriate solution (Institute of Education Sciences, 2017). The balance between treatment and comparison states is consistent with LaVenia, Cohen-Vogel, and Lang (2015) who investigated the innovation and diffusion of CC. They find that student characteristics (i.e., internal determinants) did not influence the adoption of CC.

Endogenous Time-Varying State Policies

Another barrier to obtaining unbiased estimates of the effect of CC is endogenous state policies. State education policies would bias the estimated effect of CC if they were time-varying, implemented at the same time as CC, and correlated with student outcomes. To test whether the positive effects of CC are robust to controlling for other policies I constructed a database of 23 state education policies that were adopted during the period of study (Howell & Magazinnik, 2017; Jordan & Grossmann, 2018). These state policies cover a wide variety of education reforms and include many of the most popular policies adopted by states from 2003 to 2013. Appendix Tables A5 and A6 contain the results from the district fixed effects model with covariates.[21] If adding a state education reform measure as a control attenuates the effect of CC, then it suggests that policy may have accounted for the results. I estimate three versions of each model. The first using the adoption date of a state policy, the second lagging adoption 1 year, and the third lagging adoption 2 years. Lagging adoption simulates a plausible implementation year for these policies which could also

[21] The results for the reading remain insignificant and about the same size.

confound the effect of CC. Each row contains a specified state education reform that is added as a control. The first two columns contain the results for implementing CC and the year a policy is adopted. The third and fourth columns include the results from lagging the results 1 year and the fifth and sixth columns lagging 2 years. The effects of implementing CC are robust to controlling for state education policies. The sign and size of the effect remain virtually unchanged in each of the models.

State Specific Linear Trends

I follow the robustness check recommended by Angrist and Pischke (2008) for difference-in-differences by adding state-specific linear time trends to the model. The state-specific linear trends model allows each state that implements CC to have a different trend. If the results are robust to the inclusion of the state trends then it mitigates my concern that unobserved confounding variables remain. Appendix Table A4 describe the results with district fixed effects, covariates, and state specific linear trends. The effects of CC remain significant for CC implementation in 2013 for 4th grade math but not 8th grade math. The effect of CC for math is about 7 percent of a standard deviation. The results from the models with state specific linear trends increase my confidence that the CC had an effect on 4th grade math outcomes.

Assessment Alignment

A potential concern is that the CC changed the alignment (i.e., tighter, weaker) between content standards and the NAEP, which could explain the effect of the CC. The most salient issue is that "gaming" (i.e., teaching to the test) of state accountability systems would begin to influence NAEP scores contemporaneous with treatment due to alignment with CC (Figlio & Loeb, 2011). Porter and colleagues (2011) find that the NAEP frameworks have significantly higher alignment with the CC than previous state assessments. Alignment between the CC and the NAEP ELA framework was an explicitly stated goal (Common Core State Standards Initiative, 2010). However, Porter et al. (2011) explains that alignment between the NAEP and CC is inflated, because the NAEP assesses content at and below grade level. In addition, the size of CC effects by subject and grade are not correlated with the change in alignment between content standards and the NAEP framework (see Appendix Table A7). The effect of CC is largest for 4th grade math and insignificantly different from zero for reading subjects. Despite the alignment between the NAEP ELA framework and CC there do not appear to be effects on NAEP reading scores. The effects of CC on math but not reading suggest the increased alignment between NAEP and CC does not explain the results here.

Multiple Plausible Values

The NAEP uses a matrix-based assessment where a portion of the full test is administered to each student. An IRT procedure is then used to estimate plausible values of that student's true outcome. Here the dependent variable is the first plausible value. Another approach is to use the first 5 plausible values in a multiple imputation framework (Little & Rubin, 1989). The multiple imputation strategy accounts for the variance in the estimates of student learning. Both strategies for using the plausible values yield similar results (Jerrim et al., 2017). The results are robust to these approaches because variation in each plausible value is approximately the same. Appendix Table A8 includes the results using the multiple imputation procedure, which are qualitatively similar to the main results in Tables 4 and 5.

Discussion

CC had a small positive effect on math scores (0.04-0.1 SD) and no detectable effect on reading scores. The benefits of CC were clearest in 4th grade math. Critically, the effect of CC varies across academically vulnerable students. The CC had a large positive effect on Black economically advantaged students across grades and subjects. Academically vulnerable students whose families equipped them with the benefits of high Socio-Economic Status (SES) in the form of economic capital benefitted when the CC raised expectations. However, for students from economically disadvantaged families that faced other barriers to academic success the CC backfired. Raising expectations without addressing the structural issues burdening economically disadvantaged students will at best maintain the status quo. Higher expectations provide the greatest benefit to students when students also have the resources needed to succeed.

A nascent consensus in the literature on the effects of CC on student outcomes is emerging. Consistent with the analysis herein, Loveless (2014, 2015, 2016), Xu and Cepa (2018), Gao and Lafortune (2019) all find small positive correlations between the CC and student outcomes. Song, Yang, and Garet (2019) find largely negative effects of the CCR standards, which includes CC implementing states, states that made major revisions, and states that never adopted the CC. The results of both studies are consistent if the negative effects are isolated amongst states that revised or revoked their standards after adopting the CC.

There is scant evidence that replacing the CC standards will benefit students. I find no evidence that student outcomes declined due to the implementation of CC. Making multiple substantive changes to content standards sends a confusing signal to teachers and schools. I find the CC increased math outcomes in states that chose to implement the content standards before switching to a new assessment. States should consider focusing on implementing one standards-based reform at a time. Conversely, pursuing changes to content standards and assessments at the same time may put too must strain on schools.

The positive effect on math outcomes and null results on reading outcomes is consistent with previous research on content standards. The effects of school interventions are frequently larger in math than in reading. Factors like home environment, other coursework, and extracurricular activities have a greater influence on reading relative to math outcomes (Early et al., 2014). In addition, the Principles and Standards for School Mathematics written by the National Council of Teachers of Mathematics (NCTM) may have prepared math teachers for more rigorous content standards. The NCTM standards described principles for learning core mathematics concepts and were akin to a first draft of the math CC standards. The experience from using the NCTM could have enhanced the clarity, specificity, and coverage of the math standards. There is no analogue to the NCTM standards for reading. The developers of the CC reading standards faced the challenge of writing the first national reading standards and the appropriate role of informational texts, which remains unresolved (Porter-Magee, 2012). Between the experience with the NCTM and the challenges of writing reading standards it is possible that the CC math standards were relatively better than the CC reading standards.

I argued that the CC standards influenced student outcomes through raising teacher expectations for their students. I have a paucity of data to measure teacher expectations. However, results do suggest that the implementation of CC is associated with a decrease in differentiated instruction overall and the specific practice of setting different learning goals for students. The effect of CC on differentiated instruction is consistent with the idea that CC causes teachers to raise and equalize their expectations for student learning.

My analyses have a few salient limitations. I am unable to estimate unbiased long-term effects of the CC on student achievement. It is proper to characterize the main results as the initial effects of CC. The generally positive pattern of results persists in 2015 and 2017 for early implementing states. However, the flat outcomes for late implementors suggests that some reaction to treatment biases the estimates. In addition, the effects I find are attributable to the CC standards and all associated preparation activities (e.g., professional development, coaching, curriculum). I am unable to isolate the effect of just changing the content standards. Finally, it is not possible to rule out that unobservable differences in state capacity to implement content standards account for the results I describe.

I find that the benefits of CC were isolated amongst economically advantaged students as measured by eligibility for FRPL, which is a noisy measure of Socio-Economic Status. In future research I hope to better understand which forms of economic, social, or cultural capital explain differential effects. Another potential line of research would examine how the CC changed teacher instruction via collaboration and autonomy. The CC does not work equally well for all students across schooling contexts. Understanding what causes those differences is key to improving the next generation of content standards.

Chapter 4 References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. (2017). *When Should You Adjust Standard Errors for Clustering?* (Working Paper No. 24003). National Bureau of Economic Research. https://doi.org/10.3386/w24003

AFT. (2006). *Smart Testing: Let's Get It Right How assessment-savvy have states become since NCLB?* https://files.eric.ed.gov/fulltext/ED497883.pdf

American National Election Studies. (2013). *2012 Time Series Study*. https://electionstudies.org/data-center/2012-time-series-study/

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Baker, B. D., Srikanth, A., & Weber, M. (2020). *School Funding Fairness Data System*. Rutgers Graduate School of Education/Education Law Center: http://www.schoolfundingfairness.org/

Bleiberg, J., & Harbatkin, E. (2018). Innovation and Diffusion of Teacher Evaluation Reform: A Convergence of Federal and Local Forces. *Educational Policy*.

Bleiberg, J., & West, D. (2014). *In defense of the Common Core standards*. Brookings Institution.

Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, *9*(2), 27.

Brown, A. B., & Clift, J. W. (2010). The unequal effect of adequate yearly progress: Evidence from school visits. *American Educational Research Journal*, *47*(4), 774–798.

Brown, C., Hess, F. M., Lautzenheiser, D. K., & Owen, I. (2011). State Education Agencies as Agents of Change: What It Will Take for the States to Step Up on Education Reform. *American Enterprise Institute for Public Policy Research*.

Carmichael, S. B., Martino, G., Porter-Magee, K., & Wilson, W. S. (2010). The State of State Standards–and the Common Core–in 2010. *Thomas B. Fordham Institute*.

Center on Education Policy. (2018). *NCLB/ESEA Waivers*. https://www.cep-dc.org/displayTopics.cfm?DocumentSubTopicID=48

Cheng, A., Henderson, M. B., West, M., & Peterson, P. (2018). *Public Support Climbs for Teacher Pay, School Expenditures, Charter Schools, and Universal Vouchers*. https://www.educationnext.org/2018-ednext-poll-interactive/

Common Core State Standards Initiative. (2010). *Http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf*. http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Dutro, E. (2002). Do State Content Standards Make a Difference? An Illustration of the Difficulties of Addressing That Pressing Question. *Mid-Western Educational Researcher*, *15*(4), 2–6.

Early, D. M., Rogge, R. D., & Deci, E. L. (2014). Engagement, Alignment, and Rigor as Vital Signs of High-Quality Instruction: A Classroom Visit Protocol for Instructional Improvement and Research. *The High School Journal*, *97*(4), 219–239. https://doi.org/10.1353/hsj.2014.0008

Edgerton, A. K., & Desimone, L. M. (2019). Mind the gaps: Differences in how teachers, principals, and districts experience college-and career-readiness policies. *American Journal of Education*, *125*(4), 593–619.

Erpenbach, W. J. (2008). *Statewide Educational Accountability Systems Under the NCLB Act—A Report on 2008 Amendments to State Plans*.

Erpenbach, W. J. (2011). Statewide Educational Accountability Systems under the NCLB Act: A Report on 2009 and 2010 Amendments to State Plans. *Council of Chief State School Officers*.

Erpenbach, W. J., & Forte, E. (2005). Statewide Educational Accountability Under the No Child Left Behind Act—A Report on 2005 Amendments to State Plans. *Council of Chief State School Officers*.

Erpenbach, W. J., & Forte, E. (2007). Statewide Educational Accountability Systems Under the NCLB Act—A Report on 2007 Amendments to State Plans. *Council of Chief State School Officers*.

Erpenbach, W. J., Forte-Fast, E., & Potts, A. (2003). *Statewide Educational Accountability under NCLB. Central Issues Arising from An Examination of State Accountability Workbooks and US Department of Education Reviews under the No Child Left Behind Act of 2001*.

Fast, E. F., & Erpenbach, W. J. (2004). Revisiting Statewide Educational Accountability Under NCLB: A Summary of State Requests in 2003-2004 for Amendments to State Accountability Plans. *Council of Chief State School Officers*.

Ferguson, R. F. (2003). Teachers' perceptions and expectations and the Black-White test score gap. *Urban Education*, *38*(4), 460–507.

Figlio, D., & Loeb, S. (2011). School accountability. In *Handbook of the Economics of Education* (Vol. 3, pp. 383–421). Elsevier.

Finn Jr, C. E., Julian, L., & Petrilli, M. J. (2006). The State of State Standards, 2006. *Thomas B. Fordham Foundation & Institute*.

Finnan, L. A., & Domenech, D. A. (2014). *Common Core and Other State Standards: Superintendents Feel Optimism, Concern and Lack*. AASA.

Forte, E., & Erpenbach, W. J. (2006). Statewide Educational Accountability Under the No Child Left Behind Act: A Report on 2006 Amendments to State Plans. A Summary of State Requests in 2005-06 for Amendments to Their Educational Accountability Systems Under NCLB. *Council of Chief State School Officers*.

Gamoran, A. (2008). *Standards-Based Reform and the Poverty Gap: Lessons for" No Child Left Behind"*. Brookings Institution Press.

Gao, N., & Lafortune, J. (2019). *Common Core State Standards in California: Evaluating Local Implementation and Student Outcomes*. Public Policy Institute of California. https://www.ppic.org/publication/common-core-state-standards-in-california-evaluating-local-implementation-and-student-outcomes/

Gershenson, S., Holt, S. B., & Papageorge, N. W. (2016). Who believes in me? The effect of student–teacher demographic match on teacher expectations. *Economics of Education Review*, *52*, 209–224.

Hamilton, L. S., Stecher, B. M., & Yuan, K. (2009). Standards-based reform in the United States: History, research, and future directions. *Santa Monica, CA: RAND Corporation, RP-1384. As of June*.

Herman, J., Epstein, S., & Leon, S. (2016). Supporting Common Core Instruction With Literacy Design Collaborative: A Tale of Two Studies. *AERA Open*, *2*(3), 2332858416655782. https://doi.org/10.1177/2332858416655782

Hess, F. M. (1998). *Spinning wheels: The politics of urban school reform*. Brookings Institution Press.

Hess, F. M., & McShane, M. Q. (2014). *Common core meets education reform: What it all means for politics, policy, and the future of schooling*. Teachers College Press.

Hoffer, T. B., Hedberg, E. C., Brown, K. L., Halverson, M. L., Reid-Brossard, P., Ho, A. D., & Furgol, K. (2011). Final Report on the Evaluation of the Growth Model Pilot Project. *US Department of Education*.

Howell, W. G., & Magazinnik, A. (2017). Presidential Prescriptions for State Policy: Obama's Race to the Top Initiative. *Journal of Policy Analysis and Management*, *36*(3), 502–531.

Institute of Education Sciences. (2017). *What works clearinghouse: Standards handbook (Version 4.0).*

Jerrim, J., Lopez-Agudo, L. A., Marcenaro-Gutierrez, O. D., & Shure, N. (2017). What happens when econometrics and psychometrics collide? An example using the PISA data. *Economics of Education Review*, *61*, 51–58.

Jordan, M. P., & Grossmann, M. (2018). *Correlates of US State Public Policies*.

Klein, D., Braams, B. J., Parker, T., Quirk, W., Schmid, W., & Wilson, W. S. (2005). The State of State MATH Standards, 2005. *Thomas B Fordham Foundation and Institute*.

Kober, N., & Rentner, D. (2011a). *States' Progress and Challenges in Implementing Common Core State Standards*. Center on Education Policy. https://eric.ed.gov/?id=ED514598

Kober, N., & Rentner, D. S. (2011b). *Common Core State Standards: Progress and Challenges in School Districts' Implementation*. Center on Education Policy.

Koretz, D. (2017). *The Testing Charade: Pretending to Make Schools Better*. University of Chicago Press.

Koretz, D., & Hamilton, L. S. (2006). *Testing for accountability in K-12*. https://www.rand.org/pubs/external_publications/EP20060030.html

Koretz, D. M., & Barron, S. I. (1998). *The Validity of Gains in Scores on the Kentucky Instructional Results Information System (KIRIS)*.

Korn, S., Gamboa, M., & Polikoff, M. (2016, November 3). *Just How Common are the Common Core States?* https://www.c-sail.org/resources/blog/just-how-common-are-standards-common-core-states

Lauer, P. A., Snow, D., Martin-Glenn, M., Van Buhler, R. J., Stoutemyer, K., & Snow-Renner, R. (2005). The Influence of Standards on K-12 Teaching and Student Learning: A Research Synthesis. *Mid-Continent Research for Education and Learning (McREL)*.

LaVenia, M., Cohen-Vogel, Lora, & Lang, L. B. (2015). The Common Core State Standards Initiative: An Event History Analysis of State Adoption. *American Journal of Education*, *121*(2), 145–182. https://doi.org/10.1086/679389

Little, R. J., & Rubin, D. B. (1989). The analysis of social science data with missing values. *Sociological Methods & Research*, *18*(2–3), 292–326.

Loveless, T. (2014). *2014 Brown Center report on American education: How well are American students learning? Part III: A progress report on the Common Core*. Brookings Institution. https://www.brookings.edu/research/2016-brown-center-report-on-american-education-how-well-are-american-students-learning/

Loveless, T. (2015). *2015 Brown Center Report on American Education: How Well Are American Students Learning? Part II: Measuring effects of the Common Core*. Brookings Institution. https://www.brookings.edu/research/2016-brown-center-report-on-american-education-how-well-are-american-students-learning/

Loveless, T. (2016). *2016 Brown Center Report on American Education: How Well Are American Students Learning? Part I: Reading and math in the Common Core*. Brookings Institution. https://www.brookings.edu/research/2016-brown-center-report-on-american-education-how-well-are-american-students-learning/

Manna, P. (2006). *School's in: Federalism and the national education agenda*. Georgetown University Press.

Markow, D., Macia, L., & Lee, H. (2013). *The MetLife survey of the American teacher: Challenges for school leadership*. Metropolitan Life Insurance Company.

NCSL. (2020). *Adopting Agency: State Agency or Actor Who Adopted the Common Core*. https://www.ccrslegislation.info/ccr-state-policy-resources/adopting-agency/

Norton, J., Ash, J., & Ballinger, S. (2017). *Common Core revisions: What are states really changing? EdTech Times*. Abt Associates. https://www.abtassociates.com/insights/perspectives-blog/common-core-revisions-what-are-states-really-changing

O'Day, J. A., & Smith, M. S. (2019). *Opportunity for All: A Framework for Quality and Equality in Education*. Harvard Education Press.

Polikoff, M. S. (2015). How well aligned are textbooks to the common core standards in mathematics? *American Educational Research Journal*, *52*(6), 1185–1211.

Polikoff, M. S. (2017). Is Common Core "Working"? And Where Does Common Core Research Go From Here? *AERA Open*, *3*(1), 233285841769174. https://doi.org/10.1177/2332858417691749

Porter, A., McMaken, J., Hwang, J., & Yang, R. (2011). Common core standards: The new US intended curriculum. *Educational Researcher*, *40*(3), 103–116.

Porter-Magee, K. (2012, August 15). Common Core Opens The Second Front In The Reading Wars. *Shanker Institute*. https://www.shankerinstitute.org/blog/common-core-opens-second-front-reading-wars

Ravitch, D. (2011). *National standards in American education: A citizen's guide*. Brookings Institution Press.

Ravitch, D. (2014, January 18). Everything you need to know about Common Core. *Washington Post*. https://www.washingtonpost.com/news/answer-sheet/wp/2014/01/18/everything-you-need-to-know-about-common-core-ravitch/

Reback, R., Rockoff, J., Schwartz, H. L., & Davidson, E. (2013). *Barnard No Child Left Behind Database, 2002-2003 and 2003-2004*. Barnard Columbia NCLB Data Project. http://www.gsb.columbia.edu/nclb

Rentner, D. S. (2013). Year 3 of Implementing the Common Core State Standards: An Overview of States' Progress and Challenges. *Center on Education Policy*.

Rogers, A., Tarsitano, C., & Sikali, E. (2014). *National Assessment of Educational Profress (NAEP) 2013 Mathematics and Reading Grades 4 and 8 Assessments Restricted-Use Data Files Data Companion*. National Center for Education Statistics.

Rosenthal, R. (1987). Pygmalion effects: Existence, magnitude, and social importance. *Educational Researcher*, *16*(9), 37–40.

Sablan, J. R. (2018). Can You Really Measure That? Combining Critical Race Theory and Quantitative Methods: *American Educational Research Journal*. https://doi.org/10.3102/0002831218798325

Salazar, T. (2014). 50 Ways to Test: A Look at State Summative Assessments in 2014-15. *Education Commission of the States*.

Schmidt, W. H., & Houang, R. T. (2012). Curricular coherence and the common core state standards for mathematics. *Educational Researcher*, *41*(8), 294–308.

Song, M., Yang, R., & Garet, M. (2019). *Effects of States' Implementation of College- and Career-Ready Standards on Student Achievement*. Meeting of the American Education Research Association. https://www.c-sail.org/sites/default/files/Effects%20of%20CCR%20standards%20on%20stu%20achievement_4-2019_AERA%20DRFAT.pdf

Stotsky, S. (2013, February). What's Wrong with Common Core ELA Standards? *The Center for Education Reform*. https://edreform.com/2013/02/whats-wrong-with-common-core-ela-standards/

Swann, G. P. (2000). The economics of standardization. *University of Manchester, Manchester, UK*.

U.S. Department of Education. (2017). *ED Data Express*. https://eddataexpress.ed.gov/

U.S. Department of Education. (2020). *NAEP State Mapping*. https://nces.ed.gov/nationsreportcard/studies/statemapping/

Webber, A., Troppe, P., Milanowski, A., Gutmann, B., Reisner, E., & Goertz, M. (2014). *State Implementation of Reforms Promoted under the Recovery Act. A Report from Charting the Progress of*

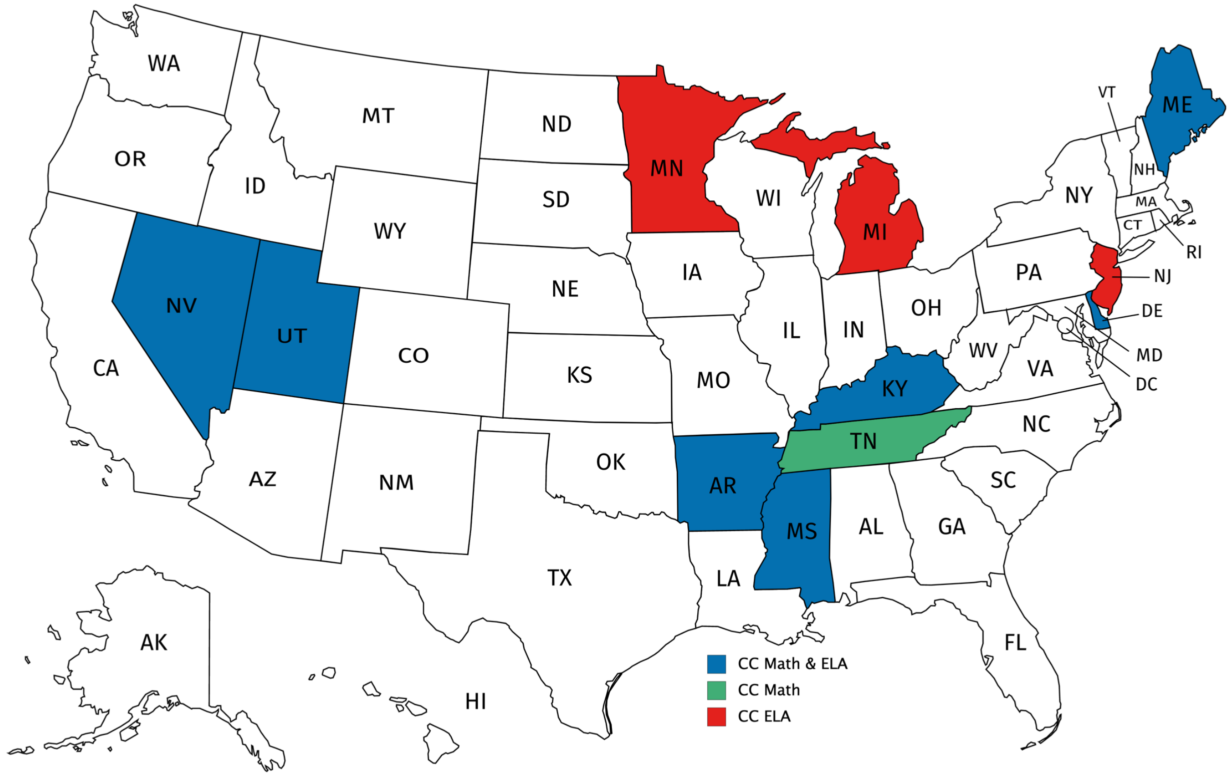
Education Reform: An Evaluation of the Recovery Act's Role. NCEE 2014-4011. National Center for Education Evaluation and Regional Assistance. https://eric.ed.gov/?id=ED544746

Woods, J. (2015). *State Summative Assessments: 2015-16 school year*. Education Commission of the States. https://www.ecs.org/state-summative-assessments-2015-16-school-year/

Woods, J. (2018, April). *Math and English language arts assessments and vendors for grades 3-8 (2017-18)*. http://ecs.force.com/mbdata/mbquestrt?rep=SUM1801

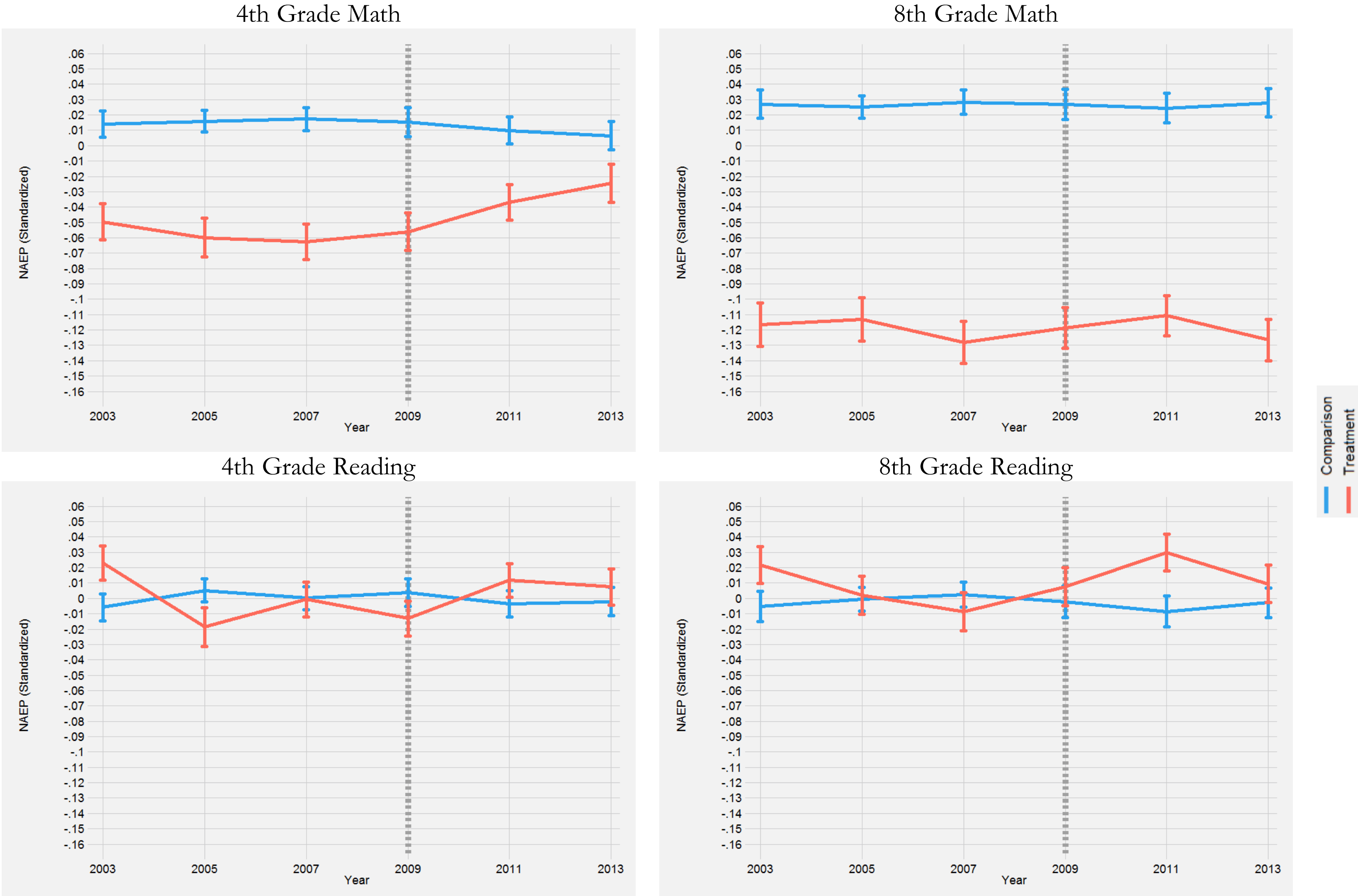
Xu, Z., & Cepa, K. (2018). Getting College-Ready during State Transition toward the Common Core State Standards. *Teachers College Record*, *120*(6), n6.

Figure 1. Implementation of Common Core in English Language Arts and Math by State in 2013



Note: See Appendix Table A2 for the states in the treatment and comparison groups by grade and subject. New Jersey implemented the CC by 2013 in 4th grade math, 4th grade reading, and 8th grade reading, but not 8th grade math.

Figure 2. NAEP Score Trends



Note: The blue line is the comparison group and the red line is the treatment group. Treatment centered on 2009, the last wave prior to adoption of Common Core. Y axis is NAEP student outcomes standardized within subject/grade and year. Estimates adjusted using NAEP student-level probability weights.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Figure 3. Common Core Effects by Race/Ethnicity and Economic Disadvantage

Note: Differential effects estimated using the regression model from Table 6 that includes the full set of covariates and district fixed effects. Economically disadvantage defined as student eligible for Free and Reduced-Price Lunch.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 1. Analytic Sample Characteristics by Year

	2003	**2005**	**2007**	**2009**	**2011**	**2013**	**Total**
States (Treated)							
4th Grade Math	0	0	0	0	8	8	33
8th Grade Math	0	0	0	0	7	7	33
4th Grade Reading	0	0	0	0	10	10	33
8th Grade Reading	0	0	0	0	10	10	33
Districts							
4th Grade Math	1,820	2,480	2,100	2,440	2,090	1,990	12,920
8th Grade Math	1,610	2,200	2,150	2,190	2,130	1,880	12,160
4th Grade Reading	2,200	2,650	2,260	2,650	2,260	2,130	14,150
8th Grade Reading	2,090	2,360	2,360	2,360	2,310	2,040	13,520
Schools							
4th Grade Math	3,360	4,800	4,000	4,910	4,040	3,850	24,960
8th Grade Math	2,510	3,320	3,330	3,410	3,320	2,970	18,860
4th Grade Reading	3,950	4,890	4,090	5,110	4,170	3,920	26,130
8th Grade Reading	3,030	3,410	3,430	3,540	3,470	3,080	19,960
Students							
4th Grade Math	83,040	82,100	95,750	82,330	94,700	84,530	522,450
8th Grade Math	64,720	76,960	74,190	78,640	77,640	78,140	450,290
4th Grade Reading	94,900	81,200	96,040	90,990	100,740	89,030	552,900
8th Grade Reading	78,190	77,860	80,480	81,300	77,410	81,100	476,340

Note: See Appendix Table A2 for the states in the treatment and comparison groups by grade and subject. Appendix Table A2 also describes which states were excluded from the analytic sample. Sample size rounded for the number of districts, schools, and students in accordance with National Center for Education Statistics nondisclosure rules.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 2. Descriptive Statistics

Characteristic	**2003-2009**	**2011**	**2013**
NAEP 4th Grade Math	-0.021	-0.004	0.005
NAEP 8th Grade Math	0.010	0.013	-0.001
NAEP 4th Grade Reading	0.069	0.041	0.040
NAEP 8th Grade Reading	0.094	0.094	0.072
Female	0.495	0.492	0.490
IEP	0.104	0.111	0.122
LEP	0.036	0.045	0.046
FRPL	0.391	0.458	0.477
White	0.617	0.595	0.584
Black	0.148	0.136	0.135
Mexican	0.086	0.100	0.106
Puerto Rican	0.024	0.025	0.024
Cuban	0.012	0.010	0.010
American Indian	0.010	0.010	0.010
Other Race	0.006	0.018	0.023
Modal age for grade; At	0.593	0.593	0.597
Below	0.002	0.002	0.002
Above	0.404	0.406	0.402
School made AYP in 2003	0.632	0.636	0.646
N	1,379,850	355,940	338,080

Note: Sample size rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. NAEP= National Assessment of Educational Progress test score standardized within grade-subject and year, IEP=Individualized Education Plan, LEP=Limited English Proficiency, FRPL=Free and Reduce Price Lunch, PI=Pacific Islander, AYP=Adequate Yearly Progress.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 3. NAEP Scores Regressed on Common Core, Full Sample

NAEP Outcome	MG4	MG4	MG8	MG8
	(1)	(2)	(3)	(4)
CC 2011	0.025	0.026	0.012	0.008
	(0.022)	(0.014)	(0.021)	(0.014)
CC 2013	0.043	0.060***	-0.009	-0.005
	(0.022)	(0.016)	(0.021)	(0.014)
Covariates		X		X
State FE	X	X	X	X
N	1,043,790	891,460	902,140	770,680
Adj R^2	0.032	0.318	0.035	0.351
F	2.10	3287.10	0.33	3984.86

NAEP Outcome	RG4	RG4	RG8	RG8
	(5)	(6)	(7)	(8)
CC 2011	0.021	0.012	0.034	0.021
	(0.020)	(0.013)	(0.018)	(0.013)
CC 2013	0.016	0.005	0.007	0.006
	(0.021)	(0.013)	(0.018)	(0.013)
Covariates		X		X
State FE	X	X	X	X
N	1,042,660	917,430	900,490	793,900
Adj R^2	0.029	0.314	0.027	0.324
F	0.74	3776.67	1.84	4160.81

Note: See Appendix Table A2 for the states in the treatment and comparison groups by grade and subject. Covariates includes Female, Individual Education Plan, Limited English Proficiency, race/ethnicity, modal age for grade, school AYP status in 2003, and lagged average state scores. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 4. Effect of Common Core on NAEP Scores, Math

4th Grade Math	(1)	(2)	(3)	(4)
CC 2011	0.031	0.043*	0.033	0.050**
	(0.024)	(0.017)	(0.019)	(0.016)
CC 2013	0.088***	0.106***	0.078***	0.104***
	(0.025)	(0.018)	(0.020)	(0.017)
Covariates		X		X
State FE	X	X		
District FE			X	X
N	592,270	522,500	592,170	522,450
Adj R^2	0.030	0.306	0.172	0.350
F	6.45	2194.78	7.55	2026.47

8th Grade Math	(5)	(6)	(7)	(8)
CC 2011	0.023	0.031	0.037*	0.045**
	(0.021)	(0.016)	(0.017)	(0.015)
CC 2013	0.049*	0.035*	0.060***	0.044**
	(0.021)	(0.016)	(0.018)	(0.015)
Covariates		X		X
State FE	X	X		
District FE			X	X
N	519,860	450,320	519,810	450,280
Adj R^2	0.030	0.329	0.157	0.368
F	2.95	2610.81	6.69	2548.42

Note: Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. *$p < 0.05$, **$p<0.01$, ***$p<0.001$.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, “Student and Teacher Survey,” 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 5. Effect of Common Core on NAEP Scores, Reading

4th Grade Reading	(1)	(2)	(3)	(4)
CC 2011	0.027	0.012	0.015	0.003
	(0.021)	(0.014)	(0.016)	(0.013)
CC 2013	0.008	0.014	0.003	0.013
	(0.023)	(0.016)	(0.017)	(0.015)
Covariates		X		X
State FE	X	X		
District FE			X	X
N	615,880	552,960	615,770	552,890
Adj R^2	0.018	0.299	0.135	0.332
F	0.86	2599.18	0.46	2447.76

8th Grade Reading	(5)	(6)	(7)	(8)
CC 2011	0.034	0.014	0.026	0.017
	(0.020)	(0.015)	(0.015)	(0.014)
CC 2013	0.029	0.008	0.029	0.016
	(0.021)	(0.015)	(0.016)	(0.014)
Covariates		X		X
State FE	X	X		
District FE			X	X
N	538,550	476,370	538,480	476,330
Adj R^2	0.018	0.299	0.125	0.333
F	1.97	2794.63	2.56	2675.90

Note: Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. *$p < 0.05$, **$p<0.01$, ***$p<0.001$.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 6. Effect of Common Core on NAEP Scores, Event Study

NAEP Subject/Grade	Math 4	Math 8	Read 4	Read 8
	(1)	(2)	(3)	(4)
Pre-Treatment 2003	-0.035	-0.032	0.007	0.026
	(0.021)	(0.018)	(0.018)	(0.019)
Pre-Treatment 2005	-0.006	0.003	0.020	0.005
	(0.020)	(0.017)	(0.017)	(0.018)
Pre-Treatment 2007	0.001	0.008	0.023	-0.024
	(0.018)	(0.017)	(0.017)	(0.017)
Post-Treatment 2011	0.041*	0.041*	0.015	0.018
	(0.019)	(0.018)	(0.016)	(0.017)
Post-Treatment 2013	0.095***	0.040*	0.025	0.018
	(0.020)	(0.018)	(0.018)	(0.018)
Covariates	X	X	X	X
District FE	X	X	X	X
N	522,450	450,280	552,890	476,330
Adj R^2	0.350	0.368	0.332	0.333
F	1753.72	2204.83	2114.62	2311.99

Note: Reference category is the last wave prior to adoption (2009). Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 7. Differential Effects of Common Core for Academically Vulnerable Students

NAEP Subject/Grade	Math 4	Math 8	Read 4	Read 8
	(1)	(2)	(3)	(4)
CC 2013	0.106***	0.050*	-0.011	0.003
	(0.022)	(0.020)	(0.017)	(0.015)
CC 2013 x Black	0.051*	0.046	0.062**	0.046
	(0.026)	(0.028)	(0.022)	(0.025)
CC 2013 x Hispanic	0.157***	0.164***	0.031	0.084**
	(0.032)	(0.031)	(0.028)	(0.029)
CC 2013 x Asian	0.084	0.188**	0.061	0.037
	(0.047)	(0.057)	(0.037)	(0.041)
CC 2013 x American Indian	0.005	0.326***	-0.051	-0.080
	(0.126)	(0.086)	(0.118)	(0.111)
CC 2013 x FRPL	-0.059**	-0.076***	0.013	-0.017
	(0.020)	(0.019)	(0.016)	(0.016)
Covariates	X	X	X	X
District FE	X	X	X	X
N	522,450	450,280	552,890	476,330
Adj R^2	0.354	0.365	0.333	0.328
F	1782.63	2177.24	2154.13	2275.67

Note: Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. $^{*}p < 0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 8. Effects of CC on Teaching Constructs, Math

4th Grade Math

	(1)	(2)	(3)	(4)	(5)
Outcome	Computer Usage	Subject Emphasis	Differentiated Instruction	Instructional Time	PD Content Standards
CC	-0.0021	0.0352*	-0.033*	0.0199*	0.006*
	(0.0036)	(0.0027)	(0.0036)	(0.0029)	(0.0027)
	(6)	(7)	(8)	(9)	(10)
Outcome	Instructional Resources	Discuss current performance	Set goals	Determine adjustments	Achievement Standards
CC	0.0179*	-0.0023	-0.0056	-0.0069	-0.0334*
	(0.0038)	(0.0036)	(0.0036)	(0.0037)	(0.0037)

8th Grade Math

	(11)	(12)	(13)	(14)	(15)
Outcome	Computer Usage	Subject Emphasis	Differentiated Instruction	Instructional Time	PD Content Standards
CC	-0.0254*	-0.0311*	-0.0692*	-0.0159*	0.0153*
	(0.0042)	(0.0044)	(0.0042)	(0.0037)	(0.0031)
	(16)	(17)	(18)	(19)	(20)
Outcome	Instructional Resources	Discuss current performance	Set goals	Determine adjustments	Achievement Standards
CC	0.0197*	-0.0177*	-0.027*	-0.029*	-0.0715*
	(0.0044)	(0.0042)	(0.0044)	(0.0043)	(0.0043)

Note: CC is estimated effect pooled across 2011 and 2013. Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Models 1 through 10 use the 4th grade math sample and Models 11 through 20 use the 8th grade math sample. Standard errors are robust to clustering by school. All regressions include district fixed effects and covariates. Computer Usage, Subject Emphasis, and Differentiated Instruction are factors constructed from several survey questions. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. *$p < 0.05$, **$p<0.01$, ***$p<0.001$.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Table 9. Effects of CC on School Constructs, Reading

4th Grade Reading

	(1)	(2)	(3)	(4)	(5)
Outcome	Computer Usage	Subject Emphasis	Differentiated Instruction	Instructional Time	PD Content Standards
CC	-0.0113*	0.0265*	-0.0285*	0.0119*	0.0008
	(0.0032)	(0.0031)	(0.0031)	(0.0025)	(0.0032)
	(6)	(7)	(8)	(9)	(10)
Outcome	Instructional Resources	Discuss current performance	Set goals	Determine adjustments	Achievement Standards
CC	0.0032	-0.0132*	-0.0133*	-0.0111*	-0.0256*
	(0.0033)	(0.0039)	(0.004)	(0.0038)	(0.0033)

8th Grade Reading

	(11)	(12)	(13)	(14)	(15)
Outcome	Computer Usage	Subject Emphasis	Differentiated Instruction	Instructional Time	PD Content Standards
CC	-0.0268*	0.036*	-0.0285*	-0.0135*	0.0169*
	(0.0036)	(0.0038)	(0.0036)	(0.003)	(0.0037)
	(16)	(17)	(18)	(19)	(20)
Outcome	Instructional Resources	Discuss current performance	Set goals	Determine adjustments	Achievement Standards
CC	-0.0002	-0.0175*	-0.0213*	-0.0157*	-0.0243*
	(0.004)	(0.0046)	(0.0045)	(0.0045)	(0.0037)

Note: CC is estimated effect pooled across 2011 and 2013. Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Models 1 through 10 use the 4th grade reading sample and Models 11 through 20 use the 8th grade reading sample. Standard errors are robust to clustering by school. All regressions include district fixed effects and covariates. Computer Usage, Subject Emphasis, and Differentiated Instruction are factors constructed from several survey questions. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. $^{*}p < 0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

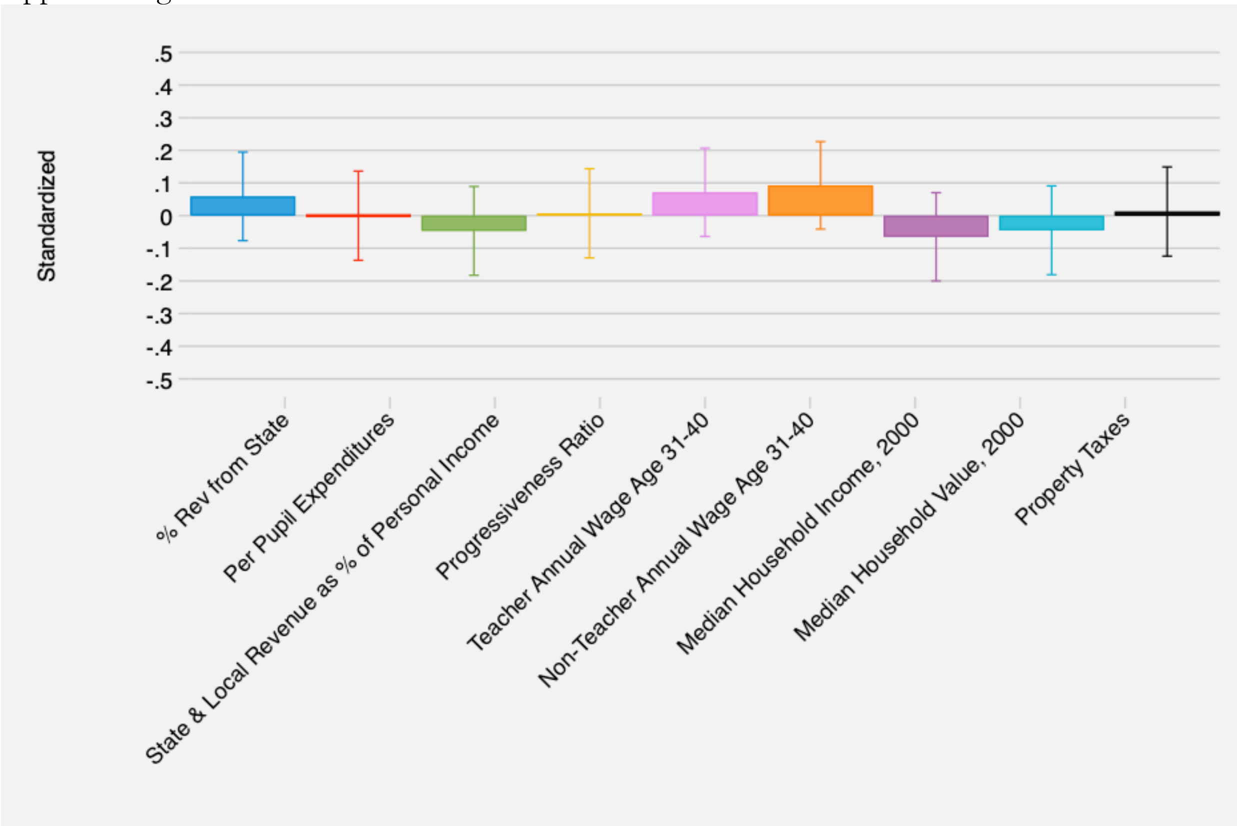
Table 10. Effect of Common Core on NAEP Scores, Long-Term Outcomes

NAEP Subject/Grade	Math 4	Math 8	Reading 4	Reading 8
	(1)	(2)	(3)	(4)
2003	-0.035	-0.026	0.013	0.029
	(0.021)	(0.018)	(0.018)	(0.018)
2005	-0.005	0.005	0.023	0.007
	(0.019)	(0.017)	(0.017)	(0.017)
2007	0.001	0.009	0.020	-0.018
	(0.018)	(0.017)	(0.017)	(0.017)
2011	0.042*	0.041*	0.014	0.025
	(0.019)	(0.017)	(0.016)	(0.017)
2013	0.097***	0.036*	0.024	0.014
	(0.020)	(0.018)	(0.017)	(0.018)
2015	0.093***	0.045*	-0.003	0.010
	(0.021)	(0.020)	(0.018)	(0.018)
2017	0.070**	0.069***	0.025	0.012
	(0.022)	(0.019)	(0.018)	(0.018)
Covariates	X	X	X	X
District FE	X	X	X	X
N	656,240	584,150	676,840	599,330
Adj R^2	0.351	0.371	0.340	0.338
F	1816.10	2175.84	2449.70	2659.26

Note: Reference category is the last wave prior to adoption (2009). Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards. See Appendix Table A2 for detailed exclusion criteria. Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. $^{*}p < 0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

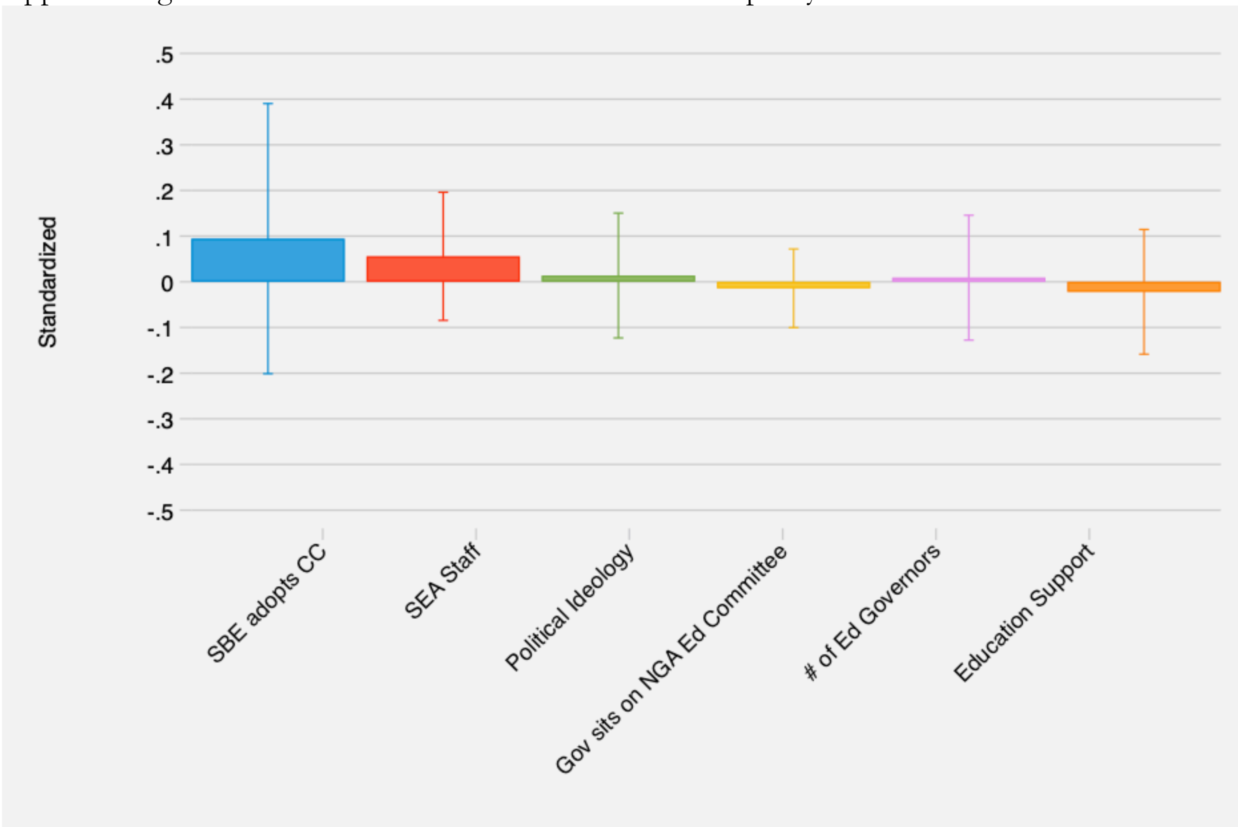
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, “Student and Teacher Survey,” 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013, 2014-2015, 2016-2017.

Appendix Figure A1. Pre-Treatment Balance on Educational Resources



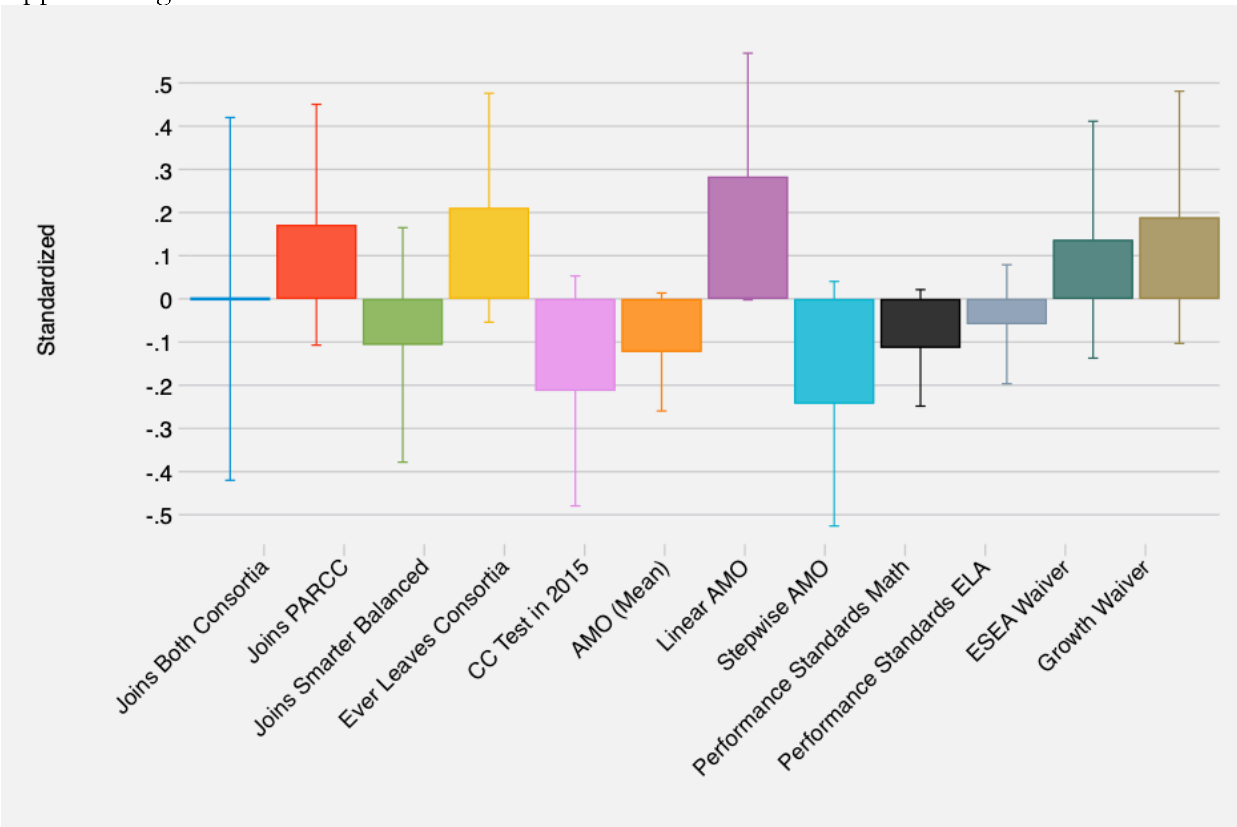
Note: Estimates are from state-level models (N=51) where I regress an indicator for whether a state implements CC by 2013 on each state characteristic. Each characteristic is a state average from 2009 except for Median Household Income and Value, which were measured in 2000. Education resource data from School Funding Fairness Data System (Baker et al., 2020).

Appendix Figure A2. Pre-Treatment Balance on Political Capacity



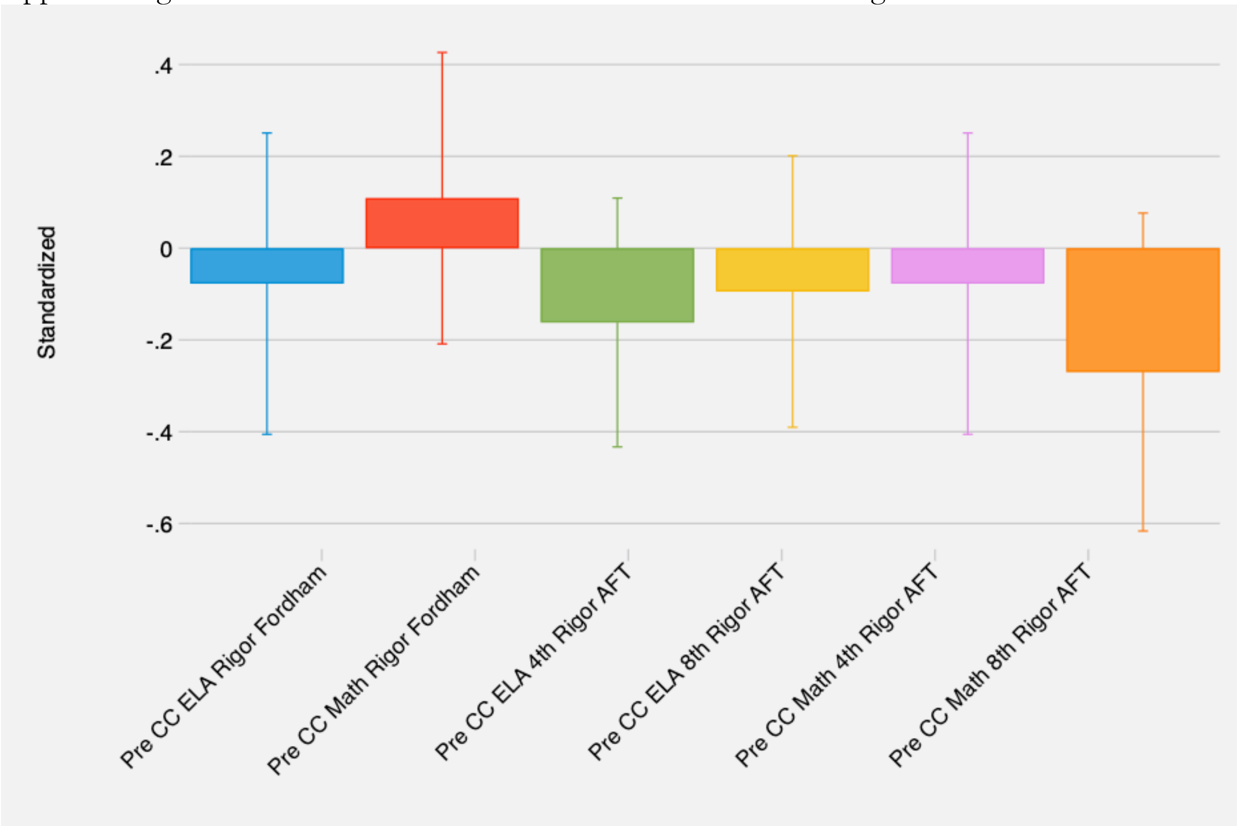
Note: Estimates are from state-level models (N=51) where I regress an indicator for whether a state implements CC by 2013 on each state characteristic. Political capacity data were collected from several sources: CC adopting institution (NCSL, 2020), State Education Agency staff in 2011 (C. Brown et al., 2011), support for education spending (American National Election Studies, 2013).

Appendix Figure A3. Pre-Treatment Balance on Standards-Based Reforms



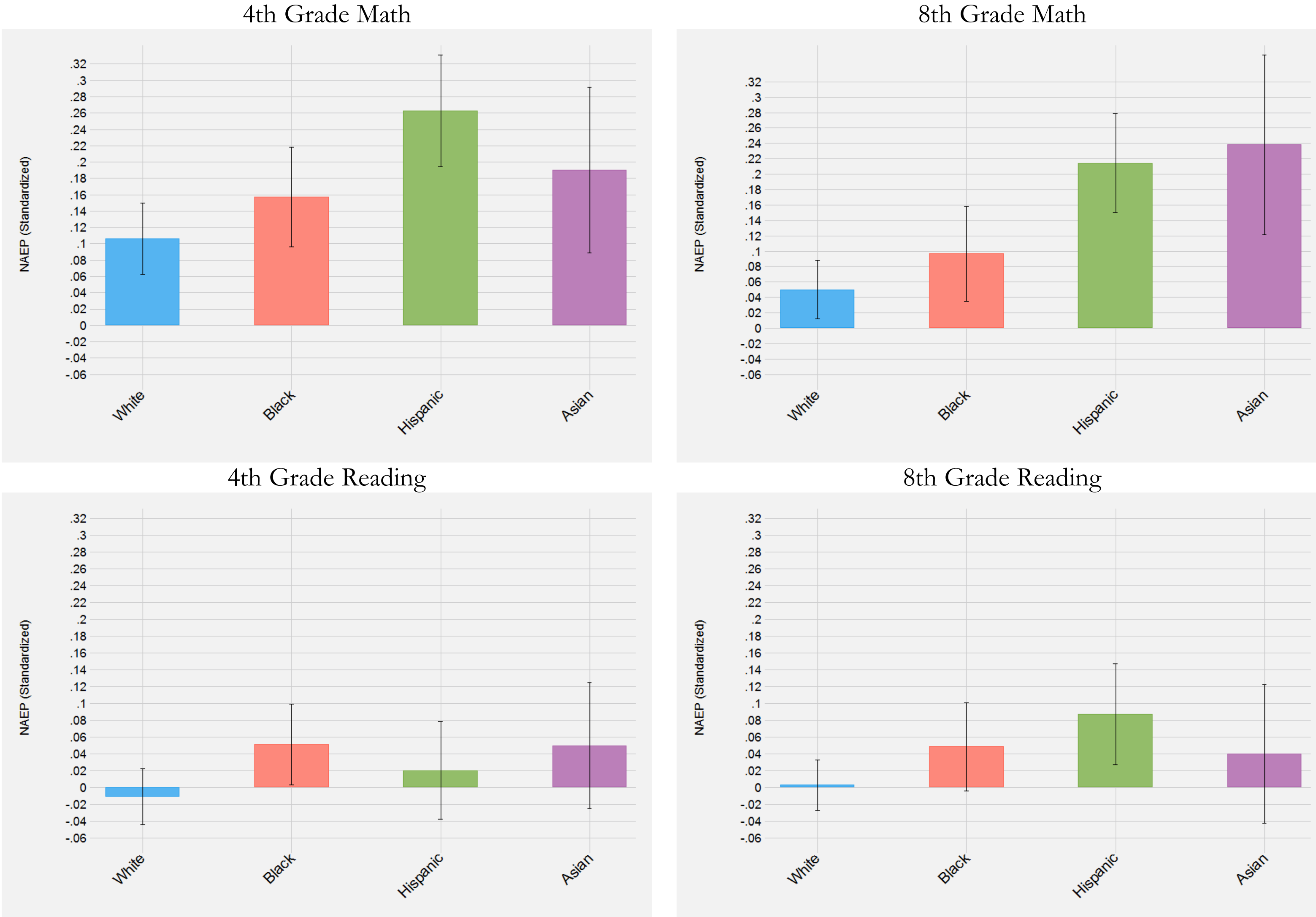
Note: Coefficients are from state-level models (N=51). I regressed an indicator for whether a state implements CC by 2013 on each state characteristic. I collected data on CC consortia and assessments collective from state reports on summative assessments (Salazar, 2014; Woods, 2015, 2018) and data on the state accountability systems CSSO and Education Department reports (Erpenbach, 2008, 2008, 2011; Erpenbach et al., 2003; Erpenbach & Forte, 2005, 2007; Fast & Erpenbach, 2004; Forte & Erpenbach, 2006; Hoffer et al., 2011; U.S. Department of Education, 2017, 2020).

Appendix Figure A4. Pre-Treatment Balance on Content Standards Rigor



Note: Coefficients are from state-level models (N=51). I regressed an indicator for whether a state implements CC by 2013 on each state characteristic. Data on content standards rigor were collective from studies of content standards (AFT, 2006; Carmichael et al., 2010).

Appendix Figure A5. Common Core Effects on Achievement Gaps



Note: Differential effects estimated using the regression model from Table 7 that includes the full set of covariates and district fixed effects. Economically disadvantage defined as student eligible for Free and Reduced-Price Lunch.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table A1. Standards Rigor, Adoption, and Implementation 2003-2017

State	Pre-CC Rigor Math	Pre-CC Rigor ELA	Adoption	CC Math	CC ELA	Withdrawal/Revise
Alabama	B+	B	Nov-10	2013	2014	
Alaska	D	F	Never	Never	Never	
Arizona	B	B	Jun-10	2014	2013	Dec-16
Arkansas	C	D	Jul-10	2013	2013	Apr-16
California	A	A	Aug-10	2015	2015	
Colorado	C	B+	Dec-10	2014	2014	Aug-14
Connecticut	D	D	Jul-10	2014	2014	
Delaware	B	F	Aug-10	2013	2013	
DC	A	A	Jul-10	2013	2012	
Florida	A	B	Jul-10	2015	2015	Jan-19
Georgia	A-	B+	Jul-10	2013	2013	
Hawaii	C	C	Jun-10	2014	2014	
Idaho	B	C	Jan-11	2014	2014	
Illinois	D	D	Jun-10	2014	2014	
Indiana	A	A	Aug-10	Never	Never	Mar-14
Iowa	C	F	Jul-10	2015	2015	
Kansas	F	C	Oct-10	2014	2014	
Kentucky	D	D	Feb-10	2012	2012	
Louisiana	C	B+	Jul-10	2014	2014	Mar-16
Maine	C	C	Apr-11	2013	2013	
Maryland	D	C	Jun-10	2014	2014	
Massachusetts	B+	A-	Jul-10	2014	2014	
Michigan	A-	D	Jun-10	2013	2013	
Minnesota	B	C	Sep-10	Never	2013	
Mississippi	C	D	Jul-10	2013	2013	
Missouri	D	D	Jun-10	2015	2015	Apr-16
Montana	F	F	Nov-11	2014	2014	

Nebraska	C	F	Never	Never	Never	
Nevada	C	C	Jun-10	2012	2012	
New Hampshire	D	C	Jul-10	2014	2014	
New Jersey	C	C	Jun-10	4th-2013;8th-2014	2013	May-16
New Mexico	C	C	Oct-10	2014	2014	
New York	B	C	Jul-10	2013	2013	Dec-15
North Carolina	D	D	Jun-10	2013	2013	Jul-14
North Dakota	C	D	Jun-10	2014	2014	May-16
Ohio	C	C	Jun-10	2014	2014	
Oklahoma	B+	B+	Jun-10	Never	Never	Jun-14
Oregon	B+	C	Oct-10	2015	2015	
Pennsylvania	F	D	Jul-10	2014	2014	Sep-14
Rhode Island	D	D	Jul-10	2014	2014	
South Carolina	C	D	Jul-10	2015	2015	May-14
South Dakota	C	C	Nov-10	2015	2015	Mar-18
Tennessee	C	A-	Jul-10	2013	2014	May-15
Texas	C	A-	Never	Never	Never	
Utah	A-	C	Aug-10	2013	2013	
Vermont	F	D	Aug-10	2014	2014	
Virginia	C	B+	Never	Never	Never	
Washington	A	C	Jun-12	2015	2015	
West Virginia	B	D	May-10	2015	2015	Dec-15
Wisconsin	F	D	Jun-10	2015	2015	
Wyoming	F	D	Jun-12	2015	2015	

Note: Pre-CC Rigor Math/Pre-CC Rigor ELA describes the rigor or state content standards in Math and ELA prior to the adoption of Common Core in 2010 (Carmichael et al., 2010). Adoption is the month and year a state adopted the CC. CC Math/CC ELA is the Spring from the school year that states required teachers to align instruction with the CC in either Math or ELA. Withdrawal/Revise is the date that a state either withdrew from or made major revisions to the CC.

Appendix Table A2. Treatment, Comparison, and Excluded States by State, Grade, and Subject

State	Math 4th Grade	Math 8th Grade	Read 4th Grade	Read 8th Grade
Alabama	Excluded †	Excluded †	Excluded †	Excluded †
Alaska	Excluded ±	Excluded ±	Excluded ±	Excluded ±
Arizona	Comparison	Comparison	Excluded †	Excluded †
Arkansas	Treatment	Treatment	Treatment	Treatment
California	Excluded †	Excluded †	Excluded †	Excluded †
Colorado	Excluded ‡	Excluded ‡	Excluded †/‡	Excluded †/‡
Connecticut	Comparison	Comparison	Comparison	Comparison
Delaware	Treatment	Treatment	Treatment	Treatment
DC	Excluded †	Excluded †	Excluded †	Excluded †
Florida	Excluded †	Excluded †	Excluded †	Excluded †
Georgia	Excluded †	Excluded †	Excluded †	Excluded †
Hawaii	Comparison	Comparison	Comparison	Comparison
Idaho	Comparison	Comparison	Comparison	Comparison
Illinois	Comparison	Comparison	Comparison	Comparison
Indiana	Excluded †	Excluded †	Excluded †	Excluded †
Iowa	Comparison	Comparison	Comparison	Comparison
Kansas	Comparison	Comparison	Comparison	Comparison
Kentucky	Treatment	Treatment	Treatment	Treatment
Louisiana	Comparison	Comparison	Excluded †	Excluded †
Maine	Treatment	Treatment	Treatment	Treatment
Maryland	Comparison	Comparison	Comparison	Comparison
Massachusetts	Excluded †	Excluded †	Excluded †	Excluded †
Michigan	Excluded †	Excluded †	Treatment	Treatment
Minnesota	Comparison	Comparison	Treatment	Treatment
Mississippi	Treatment	Treatment	Treatment	Treatment
Missouri	Comparison	Comparison	Comparison	Comparison
Montana	Comparison	Comparison	Comparison	Comparison
Nebraska	Comparison	Comparison	Comparison	Comparison
Nevada	Treatment	Treatment	Treatment	Treatment
New Hampshire	Comparison	Comparison	Comparison	Comparison

New Jersey	Treatment	Comparison	Treatment	Treatment
New Mexico	Comparison	Comparison	Comparison	Comparison
New York	Excluded ‡	Excluded ‡	Excluded ‡	Excluded ‡
North Carolina	Excluded ‡	Excluded ‡	Excluded ‡	Excluded ‡
North Dakota	Comparison	Comparison	Comparison	Comparison
Ohio	Comparison	Comparison	Comparison	Comparison
Oklahoma	Excluded †	Excluded †	Excluded †	Excluded †
Oregon	Excluded †	Excluded †	Comparison	Comparison
Pennsylvania	Excluded ‡	Excluded ‡	Excluded ‡	Excluded ‡
Rhode Island	Comparison	Comparison	Comparison	Comparison
South Carolina	Comparison	Comparison	Comparison	Comparison
South Dakota	Comparison	Comparison	Comparison	Comparison
Tennessee	Treatment	Treatment	Excluded †	Excluded †
Texas	Excluded ±	Excluded ±	Excluded ±	Excluded ±
Utah	Excluded †	Excluded †	Treatment	Treatment
Vermont	Comparison	Comparison	Comparison	Comparison
Virginia	Comparison	Comparison	Excluded †	Excluded †
Washington	Excluded †	Excluded †	Comparison	Comparison
West Virginia	Comparison	Comparison	Comparison	Comparison
Wisconsin	Comparison	Comparison	Comparison	Comparison
Wyoming	Comparison	Comparison	Comparison	Comparison

Note: Treatment indicates that a state implemented CC for a specified subject and grade. Treatment states implemented in 2012 or 2013. Implementing states required teachers to align their instruction with the CC in a specified grade and subject. The specific implementation years are available in Appendix Table A1. The comparison group is all states that implement the treatment after 2013 , did not make major revisions to their standards from 2010-2015, and had low rigor standards. Excluded †=Pre-CC standards high indicates that a state was excluded from either the treatment or comparison group because pre-treatment standards rigor was too high (Carmichael et al., 2010). Excluded ‡=Major Reviser indicates that a state was excluded from either the treatment or comparison group because the state made a major revision the standards (2010-2015). Excluded ±=Alternate CCR indicates that a state was excluded from either the treatment or comparison group because the states implemented another set of College and Career Ready standards that differed substantively from the CC.

Appendix Table A3. Pre-Treatment Balance on Student Characteristics

Characteristic	**4th Grade Math**	**8th Grade Math**	**4th Grade Reading**	**8th Grade Reading**
Female	0.0015	0.0012	0.0022	-0.005
IEP	-0.0222*	-0.012	-0.0497*	-0.0239*
LEP	-0.0857*	-0.0743*	0.013	0.0079
FRPL	0.0538*	0.0573*	0.0058	0.0205
White	0.0167	0.0238*	-0.0091	-0.0257
Black	0.0568*	0.0425*	0.037	0.0583*
Mexican/Chicano	-0.0389*	-0.0646*	-0.037*	-0.0542*
Asian/PI	-0.0689*	-0.0605*	0.0286	0.0581
Puerto Rican	0.0523*	0.0312	0.0399*	0.0498
Cuban	-0.1459*	-0.1491*	-0.1157*	-0.0192
American Indian	-0.0663*	-0.0709*	-0.142*	-0.1759*
Modal age for grade; At	-0.0183*	-0.0188*	0.0251*	0.0302*
Below	-0.0567*	0.0178	-0.0149	0.1008*
Above	0.0189*	0.0186*	-0.025*	-0.0315*
School made AYP in 2003	-0.0766*	-0.0282	-0.0422	0.0078

Note: Estimates from models where I regressed an indicator for whether a state implements CC by 2013 on each student characteristic or school characteristics in 2003. Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. *$p < 0.05$

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table A4. Robustness to State Specific Linear Trends

NAEP Subject/Grade	Math 4	Math 8	Read 4	Read 8
CC 2011	0.029	0.021	-0.011	0.048*
	(0.022)	(0.020)	(0.019)	(0.019)
CC 2013	0.069*	0.011	-0.006	0.065**
	(0.027)	(0.025)	(0.024)	(0.024)
Covariates	X	X	X	X
District FE	X	X	X	X
N	522,450	450,280	552,890	476,330
Adj R2	0.351	0.368	0.333	0.334
F	2018.96	2550.42	2437.22	2684.23

Note: Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. $^{*}p < 0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, “Student and Teacher Survey,” 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table A5. Robustness to State Policies, 4th Grade Math

	Adoption		Lagged 1 Year		Lagged 2 Year	
State Policy	**CC 2011**	**CC 2013**	**CC 2011**	**CC 2013**	**CC 2011**	**CC 2013**
Annual Teacher Evaluations	0.0424*	0.0955*	0.0401*	0.0995*	0.0505*	0.0931*
Common Assessments	0.0491*	0.1027*	0.0491*	0.1031*	0.0518*	0.1031*
Statewide Data System	0.0499*	0.1041*	0.0486*	0.1027*	0.0572*	0.1054*
Data System with Identifiers	0.0421*	0.0926*	0.0552*	0.1016*	0.0643*	0.1013*
Evaluation Firing	0.0485*	0.1036*	0.0509*	0.0867*	0.0509*	0.0771*
Eval PD	0.0486*	0.1052*	0.0774*	0.1082*	0.0531*	0.1018*
Eval compensation	0.0465*	0.1017*	0.0321*	0.0991*	0.0516*	0.0989*
Eval Responsibility	0.0495*	0.1024*	0.047*	0.1023*	0.0497*	0.1024*
Eval Grant Tenure	0.0552*	0.1085*	0.0556*	0.1057*	0.0492*	0.1114*
Eval has Multiple Categories	0.0432*	0.1039*	0.0717*	0.1138*	0.0549*	0.0905*
Evaluation Uses Student Growth	0.0499*	0.1059*	0.0511*	0.1135*	0.0521*	0.1038*
Charter Authorizer	0.0497*	0.1149*	0.0483*	0.0976*	0.047*	0.1017*
Charter Building Funds	0.049*	0.1027*	0.0485*	0.1026*	0.0485*	0.1026*
Charter Cap	0.0489*	0.1033*	0.0539*	0.0964*	0.0499*	0.105*
School Turnaround	0.0558*	0.1057*	0.0508*	0.105*	0.0519*	0.0988*
Evaluation Growth Targets	0.0542*	0.1062*	0.052*	0.1126*	0.0505*	0.1085*
Alt Certification Pathways	0.036*	0.0904*	0.0503*	0.0894*	0.0503*	0.0894*
Alt Preparation Programs	0.0506*	0.1046*	0.0474*	0.1016*	0.0474*	0.1016*
Vouchers	0.0498*	0.0972*	0.0499*	0.1015*	0.0497*	0.1039*
High School Exit Exams	0.0476*	0.1017*	0.0484*	0.1025*	0.0369*	0.0901*
Teacher Evaluation	0.05*	0.1058*	NA	NA	NA	NA
School Finance Reform	0.0484*	0.1027*	NA	NA	NA	NA
Full Day Kindergarten	0.0495*	0.104*	NA	NA	NA	NA

Note: Estimates are the effect of CC after a control for a time variant state policy is added as a covariate. NA indicates that a policy was adopted in 2012 or later. Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. *$p < 0.05$.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table A6. Robustness to State Policies, 8th Grade Math

	Adoption		Lagged 1 Year		Lagged 2 Year	
State Policy	**CC 2011**	**CC 2013**	**CC 2011**	**CC 2013**	**CC 2011**	**CC 2013**
Annual Teacher Evaluations	0.0504*	0.0509*	0.0506*	0.0458*	0.0447*	0.0473*
Common Assessments	0.045*	0.0446*	0.0445*	0.0437*	0.0451*	0.0438*
Statewide Data System	0.0434*	0.0424*	0.0444*	0.0437*	0.0471*	0.0446*
Data System with Identifiers	0.0442*	0.0428*	0.044*	0.0447*	0.0437*	0.0445*
Evaluation Firing	0.0421*	0.043*	0.0452*	0.0411*	0.0455*	0.0197
Eval PD	0.0425*	0.0446*	0.058*	0.0463*	0.0479*	0.0395*
Eval compensation	0.0405*	0.0406*	0.0384*	0.0421*	0.0454*	0.0393*
Eval Responsibility	0.0458*	0.0466*	0.0492*	0.0468*	0.0459*	0.0468*
Eval Grant Tenure	0.0365*	0.0367*	0.0497*	0.0451*	0.0457*	0.0404*
Eval has Multiple Categories	0.0401*	0.0436*	0.0485*	0.046*	0.0476*	0.0351*
Evaluation Uses Student Growth	0.0445*	0.0465*	0.0462*	0.048*	0.0474*	0.0439*
Charter Authorizer	0.045*	0.0449*	0.0447*	0.0422*	0.0431*	0.0427*
Charter Building Funds	0.0457*	0.0457*	0.0456*	0.045*	0.0456*	0.045*
Charter Cap	0.0486*	0.0473*	0.0463*	0.0415*	0.0447*	0.0449*
School Turnaround	0.0523*	0.0466*	0.0494*	0.0491*	0.044*	0.0468*
Evaluation Growth Targets	0.0513*	0.0464*	0.0461*	0.0475*	0.0451*	0.0491*
Alt Certification Pathways	0.0503*	0.0492*	0.045*	0.046*	0.045*	0.046*
Alt Preparation Programs	0.0493*	0.0467*	0.0535*	0.0527*	0.0535*	0.0527*
Vouchers	0.0442*	0.032*	0.0447*	0.04*	0.0439*	0.0434*
High School Exit Exams	0.05*	0.0496*	0.0486*	0.0482*	0.0474*	0.0474*
Teacher Evaluation	0.0451*	0.0436*	NA	NA	NA	NA
School Finance Reform	0.0466*	0.0464*	NA	NA	NA	NA
Full Day Kindergarten	0.0427*	0.042*	NA	NA	NA	NA

Note: Estimates are the effect of CC after a control for a time variant state policy is added as a covariate. NA indicates that a policy was adopted in 2012 or later. Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. *$p < 0.05$

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table A7. CC Effects by Grade and Subject

NAEP Subject/Grade	Math 4	Math 8	Read 4	Read 8
	(1)	(3)	(5)	(7)
CC 2011	0.050**	0.045**	0.003	0.017
	(0.016)	(0.015)	(0.013)	(0.014)
CC 2013	0.104***	0.044**	0.013	0.016
	(0.017)	(0.015)	(0.015)	(0.014)
Covariates	X	X	X	X
District FE	X	X	X	X
N	522,450	450,280	552,890	476,330
F	0.350	0.368	0.332	0.333
Adjusted R^2	2026.47	2548.42	2447.76	2675.90
State & CC Alignment	0.20	0.20	0.17	0.17
NAEP & CC Alignment	0.28	0.21	0.25	0.24
Alignment Change	0.08	0.10	0.08	0.07

Note: Alignment based on Table 7 from Porter et al. (2011). These cells include an index measuring the alignment between a specified test and CC. Estimates are the effects of CC from Tables 4 and 5. Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table A8. CC Effects with Multiply Imputed Plausible Values

NAEP Outcome	Math 4	Math 4	Math 8	Math 8
4th Grade Math	(1)	(2)	(3)	(4)
CC 2011	0.043*	0.049**	0.036*	0.051**
	(0.018)	(0.017)	(0.017)	(0.016)
CC 2013	0.104***	0.103***	0.034*	0.045**
	(0.019)	(0.018)	(0.017)	(0.016)
Covariates	X	X	X	X
State FE	X		X	
District FE		X		X
N	522,500	522,450	450,320	450,280
NAEP Outcome	Reading 4	Reading 4	Reading 8	Reading 8
8th Grade Math	(5)	(6)	(7)	(8)
CC 2011	0.007	-0.003	0.009	0.015
	(0.017)	(0.017)	(0.016)	(0.015)
CC 2013	0.010	0.012	0.005	0.013
	(0.016)	(0.016)	(0.017)	(0.017)
Covariates	X	X	X	X
State FE	X		X	
District FE		X		X
N	552,960	552,890	476,370	476,330

Note: Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. *p < 0.05, **p<0.01, ***p<0.001.

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table B1. NAEP Scores Regressed on Common Core, Regressions with Covariates

NAEP Subject/Grade	Math 4	Math 8	Read 4	Read 8
	(1)	(2)	(3)	(4)
CC 2011	0.050**	0.045**	0.003	0.017
	(0.016)	(0.015)	(0.013)	(0.014)
CC 2013	0.104***	0.044**	0.013	0.016
	(0.017)	(0.015)	(0.015)	(0.014)
Female	-0.125***	-0.111***	0.130***	0.225***
	(0.003)	(0.003)	(0.003)	(0.003)
IEP	-0.748***	-0.957***	-0.969***	-1.003***
	(0.006)	(0.006)	(0.007)	(0.007)
LEP	-0.494***	-0.621***	-0.658***	-0.749***
	(0.011)	(0.013)	(0.011)	(0.015)
FRPL	-0.378***	-0.318***	-0.372***	-0.295***
	(0.004)	(0.005)	(0.004)	(0.004)
Black	-0.563***	-0.598***	-0.432***	-0.478***
	(0.007)	(0.007)	(0.007)	(0.007)
Asian	0.188***	0.193***	0.074***	0.071***
	(0.012)	(0.015)	(0.010)	(0.011)
American Indian	-0.326***	-0.332***	-0.302***	-0.244***
	(0.018)	(0.018)	(0.018)	(0.019)
Other Race	-0.115***	-0.150***	-0.048**	-0.024
	(0.017)	(0.020)	(0.017)	(0.019)
Mexican	-0.312***	-0.354***	-0.271***	-0.321***
	(0.007)	(0.008)	(0.006)	(0.008)
Puerto Rican	-0.501***	-0.456***	-0.385***	-0.373***
	(0.011)	(0.013)	(0.011)	(0.013)
Cuban	-0.626***	-0.586***	-0.528***	-0.578***
	(0.014)	(0.022)	(0.014)	(0.024)
Other Hispanic	-0.206***	-0.343***	-0.157***	-0.310***
	(0.007)	(0.008)	(0.006)	(0.008)
Modal age for grade; Below	0.302***	0.363***	0.253***	0.195***
	(0.038)	(0.035)	(0.034)	(0.035)
Modal age for grade; Above	-0.051***	-0.142***	-0.027***	-0.091***
	(0.003)	(0.004)	(0.003)	(0.004)
School Made AYP	0.156***	0.150***	0.139***	0.174***
	(0.010)	(0.012)	(0.010)	(0.013)
Safe Harbor	0.095**	0.080*	0.084**	0.010
	(0.035)	(0.032)	(0.026)	(0.042)
Lagged State Score	0.370***	0.203***	0.273***	0.178***
	(0.043)	(0.043)	(0.040)	(0.042)
Covariates	X	X	X	X
District FE	X	X	X	X
N	522,450	450,280	552,890	476,330
Adj R^2	0.350	0.368	0.332	0.333
F	2026.47	2548.42	2447.76	2675.90

Note: Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, Adj=Adjusted FE=Fixed Effect. $^{*}p < 0.05$, $^{**}p<0.01$, $^{***}p<0.001$.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table B2. Event Study Estimate of CC with Covariates

NAEP Subject/Grade	Math 4	Math 8	Read 4	Read 8
	(1)	(2)	(3)	(4)
Pre-Treatment 2003	-0.035	-0.032	0.007	0.026
	(0.021)	(0.018)	(0.018)	(0.019)
Pre-Treatment 2005	-0.006	0.003	0.020	0.005
	(0.020)	(0.017)	(0.017)	(0.018)
Pre-Treatment 2007	0.001	0.008	0.023	-0.024
	(0.018)	(0.017)	(0.017)	(0.017)
Post-Treatment 2011	0.041*	0.041*	0.015	0.018
	(0.019)	(0.018)	(0.016)	(0.017)
Post-Treatment 2013	0.095***	0.040*	0.025	0.018
	(0.020)	(0.018)	(0.018)	(0.018)
Female	-0.125***	-0.111***	0.130***	0.225***
	(0.003)	(0.003)	(0.003)	(0.003)
IEP	-0.748***	-0.957***	-0.969***	-1.003***
	(0.006)	(0.006)	(0.007)	(0.007)
LEP	-0.494***	-0.621***	-0.658***	-0.749***
	(0.011)	(0.013)	(0.011)	(0.015)
FRPL	-0.378***	-0.318***	-0.372***	-0.295***
	(0.004)	(0.005)	(0.004)	(0.004)
Black	-0.563***	-0.598***	-0.432***	-0.478***
	(0.007)	(0.007)	(0.007)	(0.007)
Asian	0.188***	0.193***	0.074***	0.071***
	(0.012)	(0.015)	(0.010)	(0.011)
American Indian	-0.326***	-0.332***	-0.303***	-0.244***
	(0.018)	(0.018)	(0.018)	(0.019)
Other Race	-0.115***	-0.150***	-0.048**	-0.025
	(0.017)	(0.020)	(0.017)	(0.019)
Mexican	-0.312***	-0.354***	-0.271***	-0.321***
	(0.007)	(0.008)	(0.006)	(0.008)
Puerto Rican	-0.501***	-0.456***	-0.385***	-0.373***
	(0.011)	(0.013)	(0.011)	(0.013)
Cuban	-0.626***	-0.586***	-0.528***	-0.578***
	(0.014)	(0.022)	(0.014)	(0.024)
Other Hispanic	-0.206***	-0.343***	-0.157***	-0.310***
	(0.007)	(0.008)	(0.006)	(0.008)
Modal age for grade; Below	0.301***	0.363***	0.253***	0.194***
	(0.038)	(0.035)	(0.034)	(0.035)
Modal age for grade; Above	-0.051***	-0.142***	-0.027***	-0.091***

	(0.003)	(0.004)	(0.003)	(0.004)
School Made AYP	0.156***	0.150***	0.139***	0.174***
	(0.010)	(0.012)	(0.010)	(0.013)
Safe Harbor	0.095**	0.080*	0.085**	0.011
	(0.035)	(0.032)	(0.026)	(0.042)
Lagged State Score	0.364***	0.194***	0.272***	0.176***
	(0.044)	(0.043)	(0.041)	(0.042)
Covariates	X	X	X	X
District FE	X	X	X	X
N	522,450	450,280	552,890	476,330
Adjusted R-squared	0.350	0.368	0.332	0.333
F	1753.72	2204.83	2114.62	2311.99

Note: Reference category is the last wave prior to adoption (2009). Sample excludes states with high rigor pre-CC standards and states that made substantive changes to their standards (See Appendix Table A2 for detailed exclusion criteria). Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, FE=Fixed Effect. NAEP= National Assessment of Educational Progress test score standardized within grade-subject and year, IEP=Individualized Education Plan, LEP=Limited English Proficiency, FRPL=Free and Reduce Price Lunch, AYP=Adequate Yearly Progress *p < 0.05, **p<0.01, ***p<0.001

Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix Table B3. Differential Effects by Race/Ethnicity and Economic Disadvantage

NAEP Subject/Grade	Math 4	Math 8	Read 4	Read 8
	(1)	(2)	(3)	(4)
CC 2013	0.107***	0.055**	-0.017	0.005
	(0.022)	(0.020)	(0.017)	(0.015)
CC 2013 x FRPL	-0.057**	-0.072***	0.015	-0.015
	(0.022)	(0.021)	(0.018)	(0.017)
CC 2013 x Black	0.125*	0.011	0.160***	0.092*
	(0.056)	(0.049)	(0.037)	(0.039)
FRPL X Black	0.016	0.052***	0.009	-0.000
	(0.012)	(0.012)	(0.012)	(0.012)
CC 2013 x Black x FRPL	-0.089	0.028	-0.116**	-0.065
	(0.059)	(0.052)	(0.043)	(0.051)
CC 2013 x Hispanic	0.005	0.140*	0.012	0.004
	(0.051)	(0.060)	(0.047)	(0.049)
FRPL X Hispanic	0.137***	0.145***	0.110***	0.077***
	(0.013)	(0.013)	(0.013)	(0.014)
CC 2013 x Hispanic x FRPL	0.157*	0.000	0.020	0.097
	(0.063)	(0.061)	(0.053)	(0.055)
CC 2013 x Asian	0.069	0.172*	0.035	0.054
	(0.060)	(0.076)	(0.041)	(0.045)
FRPL X Asian	-0.124***	-0.097***	-0.115***	-0.087***
	(0.020)	(0.024)	(0.017)	(0.019)
CC 2013 x Asian x FRPL	0.058	0.043	0.097	-0.042
	(0.091)	(0.114)	(0.085)	(0.097)
Covariates	X	X	X	X
District FE	X	X	X	X
N	531,120	452,900	560,080	478,840
Adj R^2	0.354	0.366	0.334	0.329
F	1528.93	1870.58	1840.23	1930.37

Note: Sample excludes states with rigorous pre-CC standards and states that implemented the CC, but made major revisions. Standard errors are robust to clustering by school. See Table 3 for a full list of covariates. Sample sizes rounded in accordance with NCES nondisclosure rules. Estimates adjusted using NAEP student-level probability weights. CC=Common Core, FE=Fixed Effect. FRPL=Free and Reduced-Price Lunch. *$p < 0.05$, **$p<0.01$, ***$p<0.001$.
Source: U.S. Department of Education, National Center for Education, Statistics, NAEP, "Student and Teacher Survey," 2002-2003, 2004-2005, 2006-2007, 2008-2009, 2010-2011, 2012-2013.

Appendix C. Common Core Adoption, Implementation, Revision, & Withdrawal

Appendix C describes states changes (i.e., adoption, implementation, revision, and withdrawal) to content standards (hereinafter standards) and summative assessments. I collected all documents from March 2017 to 2019. All years refer to the spring of the school year.

Alabama
Alabama adopted the CC standards in November 2010 (CCSSI, 2013). The state joined both PARCC and Smarter Balanced testing consortia in 2010 (Salazar, 2014). Alabama reported in January 2012 that the full implementation of the Math standards will occur in 2013 for Math and 2014 for ELA (Anderson et al., 2012). A local advocacy group reported that Alabama "begins implementing the College and Career Ready Standards…in grades K-12" in August 2012 for ELA and August 2013 for Math (A+ Education Partnership, 2014). The advocacy group report corroborates the interview data from Achieve (2013). Alabama dropped out of both consortia entirely and used ACT Aspire as its summative assessment in 2015 (Woods, 2015).

Alaska
Alaska never adopts the CC standards (CCSSI, 2013; Certica Solutions, 2017; Ujifusa, 2016). The state also never participated in the CC consortia or used their assessments (Salazar, 2014; Woods, 2015). Alaska adopts their College and Career Ready Standards in June 2012 (WestEd, 2018) with full implementation by 2015 (Achieve, 2013).

Arizona
Arizona adopted the CC standards in June 2010 (CCSSI, 2013; Certica Solutions, 2017). Arizona initially joined PARCC in 2010 (Woods, 2015). The state's Round II Race to the Top application submitted in June 2010 (Arizona Governor's Office of Economic Recovery, 2010) describes an incremental approach to implementation that finishes in 2014. An October 2013 state document describes the timeline targeting CC standards implementation for 4th grade ELA and full implementation or 8th grade ELA in 2013 (AZ DOE, 2013). The state defines targeted implementation as, "instructional shifts, specific content emphasis by strand, and an intentional increase of rigor in the classroom" and full implementation as "complete transition to standards with fidelity" (AZ DOE, 2013). Arizona fully implemented Math in 2014 (AZ DOE, 2013). Data from Achieve (2013) and Certica (2017) corroborate these dates. In 2014 Arizona left both testing consortia and used an assessment developed by AIR (Creno, 2014). Arizona voted to rebrand the CC standards in October 2015 and then replace the standards in December 2016. The rebranded standards remain in place through the 2017 school year (National Council of State Legislators, 2017).

Arkansas
Arkansas adopted the CC standards in July 2010 (CCSSI, 2013; Certica Solutions, 2017). They joined PARCC in 2010 (Salazar, 2014). Later the state dropped out of PARCC and used ACT Aspire as its summative assessment in 2015 (Woods, 2015). State documents from April 2011 describe the implementation of CC standards in 4th/8th grade and Math/ELA in 2013 (Arkansas Department of Education, 2011). The reported implementation timeline is consistent with the state's ESEA Waiver applications from February 2012 (Arkansas Department of Education, 2012). Subsequently Arkansas implements the CC standards in 2013 (Achieve, 2013; Certica Solutions, 2017). The Arkansas Board of Education revokes the CC standards in April 2016 and created new standards that were implemented in 2018 (C. Howell, 2016).

California

California adopted the CC standards in August 2010 (CCSSI, 2013; Certica Solutions, 2017). California joined the Smarter Balanced consortia in 2010 (Salazar, 2014). The initial plan was for full adoption of the standards in 2014 (Best & Cohen, 2013; California Department of Education, 2012). The state delayed implementation of the standards until 2014 (California Department of Education, 2014; Griffith, 2012). By 2015 most but not all California school districts had implemented the CC standards (Harrington, 2017). The process of implementation is California is unique in part due to the CORE districts. The CORE districts received an ESEA waiver in August 2013 and these districts implemented the CC standards from 2013 to 2015 (Knudson & Garibaldi, 2015). California began using the Smarter Balanced test as their summative assessment in 2015 (Woods, 2015).

Colorado

Colorado joined the PARCC consortia in 2010 (Salazar, 2014). Colorado technically adopts the CC standards in August 2010 (CCSSI, 2013; Certica Solutions, 2017). The state engaged in it's CC adoption, "with the expectation that the Colorado Department of Education would honor the work and values of the Colorado Academic Standards previously written by Colorado educators and adopted by the board to create the best mathematics and reading, writing, and communicating standards for the State of Colorado" (Colorado Department of Education, 2019). In December 2010, Colorado adopts a set of College and Career Ready standards that melds elements of the CC standards and Colorado Academic Standards. Official state documents describe the Colorado Academic Standards rather than the CC standards as implemented as of 2014 (Colsman, 2017). An independent analysis of state standards describes the Colorado Academic Standards as a "major" modification of the CC standards (Korn et al., 2016). The difference is substantive enough that classify Colorado as "major reviser" and do not consider them to have implemented the CC standards. They began using the PARCC assessment in 2015 (Woods, 2015) and developed a new test in 2017 (Garcia, 2017).

Connecticut

Connecticut adopted the CC standards in July 2010 (CCSSI, 2013; Certica Solutions, 2017). The state joined the Smarter Balanced consortia in 2010 (Salazar, 2014). 2012 documents from the Connecticut Department of Education describe plans for full implementation in 2014 (Connecticut State Department of Education, 2013). The state met that implementation timeline and has kept the standards (Achieve, 2013; AFT Connecticut, 2019). They began using the Smarter Balanced assessment in 2015 (Woods, 2015).

Delaware

Delaware adopts the CC standards in August 2010 (CCSSI, 2013; Certica Solutions, 2017). Delaware initially joined PARCC using their assessment in 2015 and then switches to joining Smarter Balanced (Salazar, 2014; Woods, 2015). Delaware's ESEA Waiver from February 2012 describes plans for full implementation of the standards in 2013 (Delaware Department of Education, 2012). A survey of state education officials (Achieve, 2013) and a news article (Albright, 2014) corroborate full implementation in 2013. Delaware began using the Smarter Balanced assessment in 2015 (Woods, 2015).

District of Columbia

The District of Columbia adopted the CC standards in July 2010 (CCSSI, 2013; Certica Solutions, 2017). They initially joined Smarter Balanced and then switched to PARCC (Salazar, 2014; Woods, 2015). According to their May 2010 Race to the Top Proposal the District of Columbia planned to implement the ELA standards in 2012 and the Math standards in 2013 (Government of the District of Columbia, 2010). The District of Columbia implemented the ELA standards in 2012 and the Math standards in 2013 (Achieve, 2013; Certica Solutions, 2017). They began using the PARCC assessment in 2015 (Woods, 2015).

Florida
Florida adopted the CC standards in July 2010 (Certica Solutions, 2017). A January 2012 document review indicates the state plans to implement the standards for grades 4 and 8 in 2015 (Anderson et al., 2012). Initially Florida joined PARCC, but in September 2013 they leave the consortia (Hatter, 2013). The state never uses a CC developed assessment instead using an assessment from AIR (Woods, 2015). State documents confirm the 2015 implementation of the CC standards (Certica Solutions, 2017; Florida Department of Education, 2014). Florida revokes the CC standards in January 2019 (Gore, 2019).

Georgia
Georgia adopted the CC standards in July 2010 (CCSSI, 2013; Certica Solutions, 2017). Georgia initially joins PARCC but leaves the consortia (Salazar, 2014). The state never implements a CC test (Woods, 2015). A January 2012 document review indicates the state plans to implement the standards in 2013 (Anderson et al., 2012) which is consistent with a report from the Council of Chief State School officers (Griffith, 2012). 2013 implementation date corroborated by a presentation from state superintendent (Barge, 2014) and interview data (Achieve, 2013).

Hawaii
Hawaii adopted the CC standards in June 2010 (CCSSI, 2013; Certica Solutions, 2017). California joined the Smarter Balanced consortia in 2010 (Salazar, 2014). Hawaii planned to implement the standards by 2014 (Best & Cohen, 2013; Hawaii Department of Education, 2019). The state website corroborates that implementation of standards for 4th and 8th grade occurred in 2014 (Hawaii Department of Education, 2019), which is consistent with another document analysis (EdGate Correlation Services, 2019a). Hawaii began using the Smarter Balanced assessment in 2015 (Woods, 2015).

Idaho
Idaho adopted the CC standards in January 2011 (CCSSI, 2013; Certica Solutions, 2017). Idaho joined the Smarter Balanced consortia in 2010 (Salazar, 2014). State documents from 2012 describe plans for implementation in 2014 (Best & Cohen, 2013; Idaho State Department of Education, 2012). Implementation did occur in 2014 (Boise State Public Radio, 2014; Certica Solutions, 2017). They began using the Smarter Balanced assessment in 2015 (Woods, 2015).

Illinois
Illinois adopted the CC standards in June 2010 (CCSSI, 2013; Certica Solutions, 2017). Illinois joined the PARCC consortia in 2010 (Salazar, 2014). The state's ESEA Waiver request from February 2012 describes planned implementation of the standards in 2014 (Illinois State Board of Education, 2012). State documents from 2015 are consistent with implementation in 2014 (Illinois State Board of Education, 2015). They began using the PARCC assessment in 2015 (Woods, 2015).

Indiana

Indiana adopts the standards in August 2010 (Certica Solutions, 2017). Indiana initially joined PARCC, but dropped out in July 2013 (Nelson, 2013; Salazar, 2014). The state never used a CC assessment. Indiana planned to implement the standards in 2014, but "paused" implementation prior to the start of the 2014 school year (Salazar & Christie, 2014). In March 2014, the state legislature passes a law to repeal the standards (Elliott, 2014).

Iowa
Iowa adopts the standards in July 2010 (CCSSI, 2013; Certica Solutions, 2017). Iowa joined Smarter Balanced (Salazar, 2014; Woods, 2015). A January 2011 report from the Iowa Department of Education describes the planned implementation of the CC standards in 2015 (Iowa Department of Education, 2011). A subsequent report from January 2015 corroborates the implementation of the standards in 2015 (Iowa Department of Education, 2015). In August 2014 the state left Smarter Balanced and uses a test from the University of Iowa (Hart, 2014; Woods, 2015).

Kansas
Kansas adopts the standards in July 2010 (CCSSI, 2013; Certica Solutions, 2017). Kansas joined Smarter Balanced in 2010 (Salazar, 2014; Woods, 2015). The state's July 2012 ESEA flexibility request describes the state's plan to implement the CC standards in 2014 (Kansas Department of Education, 2012). An advocacy group blog post corroborates 2014 (Get It Right, 2015) as the year of implementation which is consistent with other sources (Achieve, 2013; Certica Solutions, 2017). In December 2013, Kansas leaves PARCC and announces plans to develop its own assessment (Gewertz, 2013).

Kentucky
Kentucky adopted the CC standards in February 2010 (CCSSI, 2013; Certica Solutions, 2017). They joined both consortia in 2010, but were never a governing member and develop their own test (Salazar, 2014; Woods, 2015). The state's May 2010 Race to the Top Application describes the plans to implement the standards in 2012 (Kentucky Department of Education, 2010). Kentucky's ESEA Waiver application from August 2014 confirms full implementation in 2012 (Kentucky Department of Education, 2014), which is corroborated by interview data from Achieve (2013).

Louisiana
Louisiana adopted the CC standards in July 2010 (CCSSI, 2013; Certica Solutions, 2017). The state initially joined PARCC in 2010 and uses that assessment in 2015 (Woods, 2015). The state's May 2010 Race to the Top Phase 2 application states that the state will fully "roll out" the CC standards by 2014 (Louisiana Department of Education, 2010). A Louisiana Department of Education press release in March 2014 describes the active implementation of the CC standards in that year (Louisiana Department of Education, 2014). In March 2016, Louisiana "technically" revokes the CC standards and new standards were used in the 2017 school year (Guidry, 2016). Independent analyses are not in agreement about whether the new standards were substantively different from the CC (Korn et al., 2016; Ujifusa, 2016). In 2016, Louisiana used a modified PARCC assessment (Schaffhauser, 2015).

Maine
Maine adopted the CC standards in April 2011 (CCSSI, 2013; Certica Solutions, 2017). The state joined Smarter Balanced and uses their test through 2015 (Salazar, 2014). In June 2015, Maine leaves the consortia and adopts a new assessment (Ujifusa, 2015). The state's May 2010 Race to the Top application describes plans for a 2013 implementation date for the CC standards (Maine Department

of Education, 2010). The state describes actively implementing the standards in a September 2012 ESEA Waiver Request (Maine Department of Education, 2012).

Maryland

Maryland adopted the CC standards in June 2010 (CCSSI, 2013; Certica Solutions, 2017). Maryland joined the PARCC consortia in 2010 (Salazar, 2014) and began using its assessment in 2015 (Woods, 2015). Maryland's June 2010 Race to the Top application describes a plan to implement the standards in 2014 (Maryland Department of Education, 2010). The state Department of Education website explains that the standards were implemented in 2014 (Maryland Department of Education, 2019), which is corroborated by interview data from Achieve (2013).

Massachusetts

Massachusetts adopted the CC standards in July 2010 (CCSSI, 2013; Certica Solutions, 2017). The state initially joined PARCC and used that assessment through 2015 (Salazar, 2014; Woods, 2015). In November 2015 they left PARCC and began developing their own test (Zernike, 2015). In 2012, a state document review finds that the state plans to implement the CC standards in 2014 (Best & Cohen, 2013). In June 2015, the state's ESEA Waiver application corroborates that the state implemented the CC standards in 2014 (Massachusetts Elementary and Secondary Education, 2015).

Michigan

Michigan adopted the CC standards in June 2010 (CCSSI, 2013; Certica Solutions, 2017). The state initially joined Smarter Balanced, but never used their assessment (Salazar, 2014; Woods, 2015). A document from the state department of education from August 2010 describes the states plan to implement the CC standards in 2013. Michigan's ESEA flexibility request from July 2014 describes the state implementing the CC standards in 2013 (Michigan Department of Education, 2014), which is corroborated by interview data from Achieve (2013).

Minnesota

In September 2010, Minnesota adopts the CC English Language Arts standards, but the state does not implement the CC standards in Math (CCSSI, 2013; Certica Solutions, 2017). Minnesota never joined either consortia and developed their own assessment (Salazar, 2014; Woods, 2015). In their November 2011 ESEA Waiver Application, Minnesota's Department of Education describes its plans for implementing the CC standards in ELA, but not Math in 2013 (Minnesota Department of Education, 2011). The Minnesota Department of Education says that the CC English Language Arts standards were implemented in 2013 (Minnesota Department of Education, 2019b) and the state also implement their College and Career Ready Math Standards in 2013 (Minnesota Department of Education, 2019a).

Mississippi

Mississippi adopted the CC standards in July 2010 (CCSSI, 2013). The state belonged to the PARCC consortia in 2010 (Salazar, 2014). Mississippi planned for the first year of teaching students in grades 3 through 8 with the CC standards to be 2013 according to interviews with state education officials from January 2012 (Anderson et al., 2012), which is corroborated a review by CCSSO (Griffith, 2012). State documents from May 2013 confirm that the standards were used in the prior school year (Mississippi Department of Education, 2013), which is corroborated by other sources (Achieve, 2013; Certica Solutions, 2017). Mississippi used the PARCC test in 2015, but in January of that year chooses to use a new test in 2016 (Le Coz, 2015).

Missouri
Missouri adopted the CC standards in June 2010 (CCSSI, 2013). The state belonged to the Smarter Balanced consortia in 2010 (Salazar, 2014). A review of state records from 2012 (Griffith, 2012) and a blog post from September 2013 (Reischman, 2013) both indicate the state planned to implement the standards in 2015. They implement the CC standards in 2015 according to state records (Missouri Department of Elementary and Secondary Education, 2015). Missouri used the Smarter Balanced assessment for some but not all grades in 2015 and then use a new test the next year (Salazar, 2014; Woods, 2015). Missouri officially replaced the CC standards in April 2016 (Ballentine, 2016). Schools could use the new standards in 2017 on a voluntary basis and are required to use the new standards in 2018 (Ballentine, 2016).

Montana
Montana adopted the CC standards in November 2011 (CCSSI, 2013). The state belonged to the Smarter Balanced consortia in 2010 and still uses their assessment (Salazar, 2014; Woods, 2015). A state document from November 2011 describes the planned CC standards implementation date as 2014 (Montana Office of Public Instruction, 2011). The implementation of the CC standards occurred in 2014 according to multiple sources (ABC Montana, 2014; Achieve, 2013; Certica Solutions, 2017).

Nebraska
Nebraska never adopts the CC standards (CCSSI, 2013; Korn et al., 2016; Ujifusa, 2015). The state also never participates in the CC consortia and never uses a CC branded assessment (Salazar, 2014; Woods, 2015). The state implemented their College and Career Ready standards for ELA in 2014 and Math in 2015 (Achieve, 2013).

Nevada
Nevada adopted the CC standards in June 2010 (CCSSI, 2013). The state belonged to the Smarter Balanced consortia in 2010 and still uses their assessment (Salazar, 2014; Woods, 2015). State documents from March 2011 describe implementation of the standards for grades 3 through 8 as 2012 for ELA and Math (Nevada Department of Education, 2011). Multiple sources report that the state implemented the CC standards in 2012 (Achieve, 2013; Bennett, 2015; Certica Solutions, 2017).

New Hampshire
New Hampshire adopted the CC standards in July 2010 (CCSSI, 2013). The state belonged to the Smarter Balanced consortia in 2010 and continues to use their assessment through 2017 (NH Department of Education, 2019; Salazar, 2014; Woods, 2015). The New Hampshire Department of Education ESEA Waiver request from September 2012 describes a planned implementation date of 2014 (New Hampshire Department of Education, 2012, p. 3). The state completed the instructional transition to CC standards for all grades/subject in 2014 (New Hampshire Department of Education, 2015).

New Jersey
New Jersey adopted the CC standards in June 2010 (CCSSI, 2013). The state belonged to the PARCC consortia in 2010 and continues to use their assessment (Salazar, 2014; Woods, 2015). New Jersey's ESEA waiver application from 2011 describes a staggered implementation process were full implementation will occur no later than 2014 (New Jersey Department of Education, 2011). In August 2014, state documents show that the CC standards were implemented in 2013 for grades K-12 for ELA and some grades for math (grades 3-5; 9-12) (New Jersey Department of Education,

2014). The CC standards were implemented for Math grades 6-8 in 2014 (New Jersey Department of Education, 2014). New Jersey makes a major revision (Ujifusa, 2016) to their standards in May 2016 which goes into place in 2018 (Clark, 2016).

New Mexico

New Mexico adopted the CC standards in November 2010 (CCSSI, 2013). The state belonged to the PARCC consortia in 2010 and still uses their assessment (Salazar, 2014; Woods, 2015). State documents from March 2012 describe plans for CC standards implementation by 2014 for ELA and Math (New Mexico Public Education Department, 2012). Interviews with state officials indicate that grades 4 through 12 implemented the standards in 2014 (Achieve, 2013; EdGate Correlation Services, 2019b).

New York

New York adopted the CC standards in July 2010 (CCSSI, 2013). The state belonged to the PARCC consortia in 2010 (Salazar, 2014). New York planned to align instruction for Math and ELA in 2013 according to state documents from July 2011 (Engage NY, 2011). In 2013, instruction in grades K-8 is aligned with the CC standards (Engage NY, 2019). In December 2015 (Darville et al., 2015), as commission appointed by Governor Cuomo recommends a major revision to the CC standards in 2016 (DiSare, 2016; Ujifusa, 2016). New York remained an advisory board member of PARCC from 2010 through 2015, but never used the consortia's assessment (Salazar, 2014; Woods, 2015).

North Carolina

North Carolina adopted the CC standards in June 2010 (CCSSI, 2013). The state was initially an advisory board member in the PARCC consortia, but left the consortia and never used their assessment (Salazar, 2014; Woods, 2015). Documents (dated July 2011) from the North Carolina Department of Instruction describe plans for full implementation of the CC standards in 2013 (North Carolina Department of Public Instruction, 2011). A July 2014 law directs the state to rewrite the CC standards (Salazar & Christie, 2014). The new standards do not go into place until after 2017 (WestEd, 2018).

North Dakota

North Dakota adopted the CC standards in June 2011 (CCSSI, 2013). The state originally belonged to both CC consortia, but left PARCC and stayed in Smarter Balanced (Salazar, 2014). State documents from February 2012 describe plans for full implementation of the CC standards by 2014 (North Dakota Department of Instruction, 2012). The first year of implementation was 2014 according to interview data with state education officials (Achieve, 2013). The state used the Smarter Balanced test in 2015, but then left Smarter Balanced and switched to a non CC assessment (Burnette II, 2016). The state announced a major revision to the standards in May 2016 that takes effect in 2018 (Nowatzki, 2016).

Ohio

Ohio adopted the CC standards in June 2010 (CCSSI, 2013) The state originally belonged to the PARCC consortia (Salazar, 2014). As of 2012, state plans were to implement the standards in 2014 according to their ESEA waiver request (Ohio Department of Education, 2012). The state implemented the standards in 2014 for grades K-12 (Achieve, 2013; Ohio Department of Education, 2015). Ohio uses the PARCC assessment in 2015 (Woods, 2015) but then switches to an AIR assessment for 2016 (O'Donnell, 2015).

Oklahoma
Oklahoma adopted the CC standards in June 2010 (CCSSI, 2013). The state originally belonged to the PARCC consortia as an advisory board member (Salazar, 2014). The state had planned to implement the standards in 2015 (Griffith, 2012). But in June 2014, Oklahoma became to second state to revoke the standards (Oklahoma Governor's Office, 2014).

Oregon
Oregon adopted the CC standards in October 2010 (CCSSI, 2013). The state originally belonged to the Smarter Balanced consortia and uses their assessment from 2015 to 2017 (Salazar, 2014; Woods, 2015, 2018). The state planned for full implementation of the standards by 2015 as of their 2012 ESEA waiver request (Oregon Department of Education, 2011). They implement the standards in 2015 according to multiple sources (Achieve, 2013; Oregon Department of Education, 2015).

Pennsylvania
Pennsylvania adopted the CC standards in July 2010 (CCSSI, 2013). Pennsylvania initially belonged to both testing consortia, but left both prior to 2015 and never used a CC assessment (Salazar, 2014; Woods, 2015, 2018). The state planned to implement the CC standards in 2014 (Griffith, 2012) and did use the standards for that one year (Achieve, 2013). The State Board of Education replaced the CC standards in March 2014 (Kraft, 2014) with standards that were substantially different (Achieve, 2017; Korn et al., 2016).

Rhode Island
Rhode Island adopted the CC standards in July 2010 (CCSSI, 2013). The state planned to implement the standards in 2014 according to their waiver application from May 2012 (RIDE, 2012) and met that timeline according to their July 2015 waiver renewal application (RIDE, 2015). Rhode Island used the PARCC assessment from 2015 through 2017 (Salazar, 2014; Woods, 2015, 2018).

South Carolina
South Carolina adopted the CC standards in July 2010 (CCSSI, 2013). They joined both consortia as an advisory board member (Salazar, 2014; Woods, 2015). South Carolina planned to implement the CC standards in 2015 (Griffith, 2012) and used the CC standards in 2015 before the legislature voted in May 2014 to create new standards for use in 2016 (Salazar & Christie, 2014). The state left both consortia and never used their respective assessments (Salazar, 2014; Woods, 2015, 2018).

South Dakota
South Dakota adopted the CC standards in November 2010 (CCSSI, 2013). The state originally belonged to the Smarter Balanced consortia and used their assessment from 2015 to 2017 (Salazar, 2014; Woods, 2015, 2018). The standards were fully implemented in 2015 (CSSO, 2016). In March 2018 the state board replaced the CC with substantially different standards (Raposa, 2018).

Tennessee
Tennessee adopted the CC standards in July 2010 (CCSSI, 2013). The state originally belonged to the PARCC consortia (Salazar, 2014; Woods, 2015, 2018). Tennessee planned to implement the Math standards in 2013 and ELA in 2014 (Pepper et al., 2013; TN Core, 2012). In April 2014, the state legislature voted to delay the use of the PARCC tests (Zubrycki, 2014) and ultimately never uses a CC branded assessment (Salazar, 2014; Woods, 2015, 2018). In May 2015, Governor Haslem signed a law requiring the state to implement new standards by 2018 (Tatter, 2015).

Texas

Texas never adopts the CC standards (CCSSI, 2013). They also never join a CC consortia or use a CC branded assessment (Salazar, 2014; Woods, 2015, 2018). Texas' College and Career Readiness Standards were implemented in 2012 (Achieve, 2013).

Utah

Utah adopted the CC standards in August 2010 (CCSSI, 2013). Utah was originally a member of the Smarter Balanced consortia, but left and never used a CC branded assessment (Salazar, 2014; Woods, 2015, 2018). In their May 2010, Race to the Top application the state describe their plan to implement standards by 2013 (Utah State Office of Education, 2010). Utah chose a staggered implementation approach. Their 2015 ESEA flexibility document explains that by 2013 all school districts had aligned curricula and instruction with the CC standards (Utah State Office of Education, 2015).

Vermont

Vermont adopted the CC standards in August 2010 (CCSSI, 2013). The state originally belonged to the Smarter Balanced consortia and uses their assessment from 2015 to 2017 (Salazar, 2014; Woods, 2015, 2018). Vermont planned to implement the standards according to 2012 survey data (Griffith, 2012) and implemented the CC standards in 2014 (Achieve, 2013; Certica Solutions, 2017; EdGate Correlation Services, 2019c).

Virginia

Virginia never adopts the CC standards (CCSSI, 2013). They also never join a CC consortia or use a CC branded assessment (Salazar, 2014; Woods, 2015, 2018). Virginia implements their College and Career Ready standards for Math in 2012 and ELA in 2013 (Achieve, 2013).

Washington

Washington adopted the CC standards in July 2011 (CCSSI, 2013). The state originally belonged to the Smarter Balanced consortia and uses their assessment from 2015 to 2017 (Salazar, 2014; Woods, 2015, 2018). The state planned to implement the CC standards by 2015 according to a state document from January 2012 (OSPI, 2012) and does implement the standards in that year according to multiple interviews and document reviews (Achieve, 2013; Certica Solutions, 2017; EdGate Correlation Services, 2019d).

West Virginia

West Virginia adopted the CC standards in June 2010 (CCSSI, 2013). The state initially joined Smarter Balanced and uses their assessment through 2017 (Salazar, 2014; Woods, 2015, 2018). West Virginia planned to use the CC standards in 2015 (Achieve, 2013; Griffith, 2012). The state used the standards in 2015 according to their ESEA waiver application (West Virginia Department of Education, 2015), but in December 2015 the West Virginia Board of Education announced that new standards would be used in 2017 (Associated Press, 2015). In February 2017 they decided to leave the consortia an use a different assessment in the next year (West Virginia Board of Education, 2017).

Wisconsin

Wisconsin adopted the CC standards in June 2010 (CCSSI, 2013). The state initially joined Smarter Balanced (Salazar, 2014). Wisconsin planned to implement the CC standards in phases with full implementation in 2015 (Achieve, 2013). The state's ESEA Waiver from July 2015 (Wisconsin

Department of Public Instruction, 2015a) and news articles detailing Governor Walker's opposition to the standards in April 2015 and 2019 corroborate the survey data (Beck, 2015; Zettel, 2019). The state uses the Smarter Balanced assessment in 2015 and then switches to a new assessment for 2016 (Wisconsin Department of Public Instruction, 2015b).

Wyoming

Wyoming adopted the CC standards in June 2012 (CCSSI, 2013). The state originally belonged to both test consortia, but left PARCC and stayed in Smarter Balanced (Salazar, 2014). They planned to implement the standards in 2015 (Achieve, 2013). Official state documents show that the state used the Common Core standards starting in 2015 and kept them through 2017 when a regular standards review cycle began (Wyoming Department of Education, 2015, 2018). The state used the Smarter Balanced test in 2015, but then left Smarter Balanced and switched to a non CC assessment (Burnette II, 2016).

Appendix C References

A+ Education Partnership. (2014, February). *Academic Standards in Alabama*. Retrieved from https://eric.ed.gov/?id=ED560224

ABC Montana. (2014, November 12). Missoula County Public Schools Implementing Common Core. Retrieved November 3, 2019, from ABC Fox Montana website: https://www.abcfoxmontana.com/news/missoula-county-public-schools-implementing-common-core/article_0b31fd54-3137-5256-a7d6-ad87ab41d4ce.html

Achieve. (2013). *Closing the expectations gap: 2013 annual report on the alignment of state K–12 policies and practice with the demands of college and careers*. Achieve, Inc Washington, DC.

Achieve. (2017). *Strong Standards: A Review of Changes to State Standards Since the Common Core*. Retrieved from https://www.achieve.org/strong-standards

AFT Connecticut. (2019). Common Core State Standards. Retrieved October 10, 2019, from http://aftct.org/common-core

Albright, M. (2014, November 21). Scores to plunge on new standardized test. *Delaware Online*. Retrieved from https://www.delawareonline.com/story/news/local/2014/11/21/scores-plunge-new-standardized-test/19348591/

Anderson, K., Harrison, T., & Lewis, K. (2012). *Plans to Adopt and Implement Common Core State Standards in the Southeast Region States. Issues & Answers*. Retrieved from Regional Educational Laboratory Southeast at SERVE Center website: https://eric.ed.gov/?id=ED528960

Arizona Governor's Office of Economic Recovery. (2010, May 28). *Arizona RTTT Round II Application*. Retrieved from https://web.archive.org/web/20170703223100/https://www2.ed.gov/programs/racetothetop/phase2-applications/arizona.pdf

Arkansas Department of Education. (2011, April 7). *Common Core State Standards Implementation Timeline for Arkansas Public Schools*. Retrieved from http://www.arkansased.gov/public/userfiles/Learning_Services/Curriculum and Instruction/CCSS/timeline_040711.pdf

Arkansas Department of Education. (2012, February 27). *Arkansas ESEA Flexibility Request*. Retrieved from https://www2.ed.gov/policy/eseaflex/ar.pdf

Associated Press. (2015, December 17). *W.Va. Board of Education repeals Common Core standards*. Retrieved from https://www.heraldmailmedia.com/news/tri_state/west_virginia/w-va-board-of-education-repeals-common-core-standards/article_c29e9a42-a53f-11e5-81e0-a744009949b0.html

AZ DOE. (2013, October 10). *Arizona's College and Career Ready Standards Statewide Implementation Plan*. Retrieved from https://web.archive.org/web/20141212082321/http://www.azed.gov/azccrs/files/2013/10/azccrs-statewide-implementation-plan_10102013.pdf

Ballentine, S. (2016, April 20). Missouri education officials replace Common Core standards. *St. Louis Post-Dispatch*. Retrieved from https://www.stltoday.com/news/local/education/missouri-education-officials-replace-common-core-standards/article_050fbd0a-5dce-54f8-a502-c655ab409fe7.html

Barge, J. (2014, June). *Historical Overview of Georgia's Standards*. Retrieved from https://www.georgiascienceteacher.org/Resources/Documents/Legislative/Common%20Core%20Study%20Committee%207.30.14.pdf

Beck, M. (2015, April 23). Scott Walker says his budget repeals Common Core, but it only reiterates existing law. *Wisconsin's State Journal*. Retrieved from https://madison.com/news/local/education/blog/scott-walker-says-his-budget-repeals-common-core-but-it/article_f84ad85b-b4c6-55bd-a2d9-538916dfb336.html

Bennett, B. (2015). *Policies and practices of parental involvement and parent-teacher relations in Irish primary education: A critical discourse analysis*. University College Dublin.

Best, J., & Cohen, C. (2013). *The Common Core: Are State Implementation Plans Enough?* Retrieved from Mid-continent Research for Education and Learning website: https://eric.ed.gov/?id=ED544604

Boise State Public Radio. (2014). *Your Ultimate Guide To Common Core In Idaho*. Retrieved from https://www.boisestatepublicradio.org/topic/your-ultimate-guide-common-core-idaho

Bowen, G. A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal*, *9*(2), 27.

Burnette II, D. (2016, May 3). North Dakota, Wyoming Move Away From Smarter Balanced Tests. *Education Week - State EdWatch*. Retrieved from http://blogs.edweek.org/edweek/state_edwatch/2016/05/north_dakota_and_wyoming_move_away_from_common_core_smarter_balanced_tests.html?cmp=SOC-SHR-FB

California Department of Education. (2012). *Common Core State Standards Systems Implementation Plan for California*. Retrieved from http://www.cde.ca.gov/re/cc/

California Department of Education. (2014, April). *Common Core State Standards Systems Implementation Plan for California*. Retrieved from https://www.cde.ca.gov/re/cc/documents/ccsssimplementationplan.pdf

CCSSI. (2013). Standards in Your State. Retrieved October 24, 2019, from Standards in Your State website: http://www.corestandards.org/standards-in-your-state/

Certica Solutions. (2017). *Common Core State Standards Adoption Map*. Retrieved from http://statestandards.certicasolutions.com/common-core-state-adoption-map/

Clark, A. (2016, May 5). N.J. revises, renames Common Core academic standards. Retrieved November 3, 2019, from NJ.COM website: https://www.nj.com/education/2016/05/nj_common_core_standards_christie.html

Colorado Department of Education. (2019). *Colorado Academic Standards: History and Development*. Retrieved from https://www.cde.state.co.us/standardsandinstruction/cas-historyanddevelopment

Colsman, M. (2017, January 25). *Upcoming Standards Review and Revision Process*. Retrieved from https://leg.colorado.gov/sites/default/files/jec_standards_review_and_revision_presentation_1-25-17.pdf

Connecticut State Department of Education. (2013, September 7). Common Core State Standards in Connecticut. Retrieved October 10, 2019, from https://web.archive.org/web/20130907091204/http://www.sde.ct.gov/sde/cwp/view.asp?a=2618&q=322592

Creno, C. (2014, May 30). Arizona withdraws from PARCC. *Azcentral*. Retrieved from https://www.azcentral.com/story/news/arizona/politics/2014/05/30/arizona-withdraws-parcc-testing-group/9773249/

CSSO. (2016). *A Path of Progress: State and District Stories of High Standards Implementation*. Retrieved from https://ccsso.org/resource-library/path-progress-state-and-district-stories-high-standards-implementation

Darville, S., DiSare, M., & Wall, P. (2015, December 10). Gov. Cuomo's Common Core task force calls for evaluation freeze, test changes. Retrieved November 15, 2019, from Chalkbeat website: https://www.chalkbeat.org/posts/ny/2015/12/10/gov-cuomos-common-core-task-force-calls-for-evaluation-freeze-test-changes/

Delaware Department of Education. (2012, February 28). *ESEA Flexibility Request*. Retrieved from https://www2.ed.gov/policy/eseaflex/de.pdf

DiSare, M. (2016, September 21). New York state recommends changes to over half the Common Core learning standards. Retrieved November 3, 2019, from Chalkbeat website: https://www.chalkbeat.org/posts/ny/2016/09/21/new-york-state-recommends-changes-to-over-half-the-common-core-learning-standards/

EdGate Correlation Services. (2019a). Standards—Hawaii. Retrieved November 3, 2019, from https://correlation.edgate.com/standards/cc/d-k/standard-hi.html

EdGate Correlation Services. (2019b). Standards—New Mexico. Retrieved November 3, 2019, from https://correlation.edgate.com/standards/cc/n-o/standard-nm.html

EdGate Correlation Services. (2019c). Standards—Vermont. Retrieved November 3, 2019, from https://correlation.edgate.com/standards/cc/p-w/standard-vt.html

EdGate Correlation Services. (2019d). Standards—Washington. Retrieved November 3, 2019, from https://correlation.edgate.com/standards/cc/p-w/standard-wa.html

Elliott, S. (2014, March 12). Common Core bill passed; heads to Pence. Retrieved October 11, 2019, from Chalkbeat website: https://www.chalkbeat.org/posts/in/2014/03/12/common-core-bill-passed-heads-to-pence/

Engage NY. (2011, July 1). Changes to New York State Standards, Curricula, and Assessments. Retrieved November 3, 2019, from https://web.archive.org/web/20120522224322/http://engageny.org/wp-content/uploads/2011/07/ccsstimeline.pdf

Engage NY. (2019). New York State P-12 Common Core Learning Standards. Retrieved November 3, 2019, from https://www.engageny.org/resource/new-york-state-p-12-common-core-learning-standards

Florida Department of Education. (2014, February 2). Common Core State Standards. Retrieved October 10, 2019, from https://web.archive.org/web/20140202223403/http://www.fldoe.org/schools/ccc.asp

Garcia, N. (2017, June 14). Colorado backing away from PARCC English and math tests, forging its own path. Retrieved November 2, 2019, from Chalkbeat website: https://www.chalkbeat.org/posts/co/2017/06/14/colorado-will-no-longer-give-parcc-english-and-math-tests-forging-its-own-path/

Get It Right. (2015, October 5). Kansas Administrators Show How Common Core Shifts Professional Development. Retrieved October 11, 2019, from https://learningfirst.org/blog/kansas-administrators-show-how-common-core-shifts-professional-development

Gewertz, C. (2013, December 12). Consortium Watch: Kansas Drops Out of Smarter Balanced Testing Group. *Education Week - Curriculum Matters*. Retrieved from http://blogs.edweek.org/edweek/curriculum/2013/12/consortium_watch_kansas_drops_.html?cmp=SOC-SHR-FB

Gore, L. (2019, January 31). Florida eliminating Common Core. *Al*. Retrieved from https://www.al.com/news/2019/01/florida-eliminating-common-core.html

Government of the District of Columbia. (2010, May). *Race to the Top Application*. Retrieved from https://web.archive.org/web/20170703225652/https://www2.ed.gov/programs/racetothetop/phase2-applications/district-of-columbia.pdf

Griffith, D. (2012, November 6). Moving the Common Core State Standards from Adoption to Implementation to Sustainability. Retrieved from ASCD Public website: http://inservice.ascd.org/moving-the-common-core-state-standards-from-adoption-to-implementation-to-sustainability/

Guidry, L. (2016, April 1). Are the new standards really different from Common Core? *Shreveport Times*. Retrieved from https://www.shreveporttimes.com/story/news/education/2016/04/01/new-standards-really-different-common-core/82477808/

Harrington, T. (2017, August 25). Understanding the Common Core State Standards in California: A quick guide. *EdSource*. Retrieved from https://edsource.org/2017/understanding-the-common-core-state-standards-in-california-a-quick-guide/585006

Hart, S. (2014, August 7). Iowa Withdraws From Common Core Assessment Consortia. Retrieved November 1, 2019, from Caffeinated Thoughts website: https://caffeinatedthoughts.com/2014/08/iowa-withdraws-common-core-assessment-consortia/

Hatter, L. (2013, September 23). *Common Core "PARCC" Tests Face Uncertain Future In Florida After Governor's Executive Order*. Retrieved from https://news.wfsu.org/post/common-core-parcc-tests-face-uncertain-future-florida-after-governors-executive-order

Hawaii Department of Education. (2019). Hawaii Common Core Standards. Retrieved October 10, 2019, from http://www.hawaiipublicschools.org/TeachingAndLearning/StudentLearning/CommonCoreStateStandards/Pages/home.aspx

Howell, C. (2016, July 22). Schools to start using new standards. *Arkansas Online*. Retrieved from www.nwaonline.com/news/2016/jul/22/schools-to-start-using-revised-standard/

Idaho State Department of Education. (2012). Common Core State Standards. Retrieved October 10, 2019, from SDE website: http://www.sde.idaho.gov/site/common

Illinois State Board of Education. (2012, February 23). *ESEA Flexibility Request*. Retrieved from https://www2.ed.gov/policy/eseaflex/il.pdf

Illinois State Board of Education. (2015). Illinois Learning Standards. Retrieved October 10, 2019, from https://www.isbe.net

Iowa Department of Education. (2011, January). *Iowa Core Annual Report*. Retrieved from https://www.legis.iowa.gov/docs/publications/SD/20899.pdf

Iowa Department of Education. (2015, January). *Iowa Core Annual Report*. Retrieved from https://www.legis.iowa.gov/docs/publications/DF/662489.pdf

Kansas Department of Education. (2012, July 11). *Kansas ESEA Flexibility Request*. Retrieved from https://www2.ed.gov/policy/eseaflex/approved-requests/ks.pdf

Kentucky Department of Education. (2010, May). *Kentucky Phase II Race to the Top Application*. Retrieved from https://web.archive.org/web/20170703232709/https:/www2.ed.gov/programs/racetothetop/phase2-applications/kentucky.pdf

Kentucky Department of Education. (2014, August 14). *Kentucky ESEA Flexibility Request Revised Submission With Amendments To Principles 1, 2 And 3 August 14, 2014*. Retrieved from https://www2.ed.gov/policy/eseaflex/approved-requests/ky2reqamend814.pdf

Knudson, J., & Garibaldi, M. (2015). None of us are as good as all of us: Early lessons from the CORE districts. *San Mateo, CA: American Institutes for Research*.

Korn, S., Gamboa, M., & Polikoff, M. (2016, November 3). Just How Common are the Common Core States? Retrieved from https://www.c-sail.org/resources/blog/just-how-common-are-standards-common-core-states

Kraft, R. (2014, September 19). Quarrel over Common Core: A Pennsylvania Primer. *WFMZ.Com*. Retrieved from https://www.wfmz.com/news/quarrel-over-common-core-a-pennsylvania-primer/article_c3a035d1-6ff9-51eb-8fc4-a2f1a1e31254.html

Le Coz, E. (2015, January 16). Miss. Withdraws from Common Core testing. *The Clarion Ledger*. Retrieved from https://www.clarionledger.com/story/news/2015/01/16/mississippi-withdraw-parcc/21859553/

Louisiana Department of Education. (2010, May). Our Children Can't wait Louisiana's Blue Print for Education Reform. Retrieved October 14, 2019, from https://web.archive.org/web/20170703233321/https:/www2.ed.gov/programs/racetothetop/phase2-applications/louisiana.pdf

Louisiana Department of Education. (2014, March 5). Department Releases 2014-2015 Curriculum Package. Retrieved October 14, 2019, from https://www.louisianabelieves.com/newsroom/news-releases/2014/03/05/department-releases-2014-2015-curriculum-package

Maine Department of Education. (2010, May). *Race to the Top Phase 2 Application: Maine*. Retrieved from https://web.archive.org/web/20170212160529/https://www2.ed.gov/programs/racetothetop/phase2-applications/maine.pdf

Maine Department of Education. (2012, September). *ESEA Waiver Request*. Retrieved from https://www2.ed.gov/policy/eseaflex/me.pdf

Maryland Department of Education. (2010, June). *Maryland Race to the Top Phase 2 Application*. Retrieved from https://web.archive.org/web/20170703233953/https:/www2.ed.gov/programs/racetothetop/phase2-applications/maryland.pdf

Maryland Department of Education. (2019). MD College and Career-Ready Standards. Retrieved October 14, 2019, from https://mdk12.msde.maryland.gov/instruction/commoncore/Pages/index.aspx

Massachusetts Elementary and Secondary Education. (2015, June 15). *ESEA Flexibility Request: Massachusetts*. Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/marenewalreq2015.pdf

Michigan Department of Education. (2014, July). *Michigan ESEA Waiver Amended Document*. Retrieved from https://www2.ed.gov/policy/eseaflex/approved-requests/miamendreq822.pdf

Minnesota Department of Education. (2011, November 14). *ESEA Flexibility Request: Minnesota*. Retrieved from https://www2.ed.gov/policy/eseaflex/mn.pdf

Minnesota Department of Education. (2019a). Academic Standards (K-12). Retrieved October 14, 2019, from https://education.mn.gov/MDE/fam/stds/

Minnesota Department of Education. (2019b). English Language Arts. Retrieved October 14, 2019, from https://education.mn.gov/MDE/dse/stds/ela/

Mississippi Department of Education. (2013, May 17). State Accountability and Assessment Transitional Timeline. Retrieved November 3, 2019, from https://web.archive.org/web/20140113233543/http://www.mde.k12.ms.us/docs/communications-library/transitional-timeline.pdf?sfvrsn=2

Missouri Department of Elementary and Secondary Education. (2015). *Missouri Learning Standards & the Common Core State Standards Information Packet for Legislators*. Retrieved from https://dese.mo.gov/sites/default/files/ccss-legislators.pdf

Montana Office of Public Instruction. (2011, November 1). *Montana Common Core Standards Timeline*. Retrieved from https://web.archive.org/web/20170303220522/http://www.opi.mt.gov/pdf/Assessment/MCPresents/MCCC/11NovTimeline.pdf

National Council of State Legislators. (2017). *Common Core Status Map*. Retrieved from https://www.ccrslegislation.info/ccr-state-policy-resources/common-core-status-map/

Nelson, L. (2013, July 30). Another state drops out of PARCC - Indiana changed school grading system for donor—Penn State's Graham Spanier in court—Race to the Top risk-free. *POLITICO*. Retrieved from https://www.politico.com/morningeducation/0713/morningeducation11278.html

Nevada Department of Education. (2011, March 2). *Common Core State Standards Nevada Transition Plan*. Retrieved from https://web.archive.org/web/20130724115353/http://www.doe.nv.gov/NDE_Offices/APAC/Nevada_Academic_Standards/Common_Core_Standards/Resources/CCSS_NV_Transition_Plan_Overview/

New Hampshire Department of Education. (2012). *New Hampshire ESEA Flexibility Request for Window 3* (p. 24). Retrieved from https://www2.ed.gov/policy/eseaflex/nh.pdf

New Hampshire Department of Education. (2015). *Approved ESEA Flexibility Request*. Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/nhrenewalreq2015.pdf

New Jersey Department of Education. (2011, November 14). *ESEA Flexibility Request from New Jersey*. Retrieved from https://www2.ed.gov/policy/eseaflex/nj.pdf

New Jersey Department of Education. (2014, August 12). Core Curriculum COntent Standards Timeline. Retrieved November 3, 2019, from https://web.archive.org/web/20140812174508/http://www.nj.gov/education/cccs/timeline.htm

New Mexico Public Education Department. (2012, March 1). New Mexico Common Core State Standards: Transition Timeline. Retrieved November 3, 2019, from https://web.archive.org/web/20180507195051/https://newmexicocommoncore.org/pages/view/22/transition-timeline/11/

NH Department of Education. (2019). Smarter Balanced Assessment Consortium. Retrieved November 2, 2019, from https://www.education.nh.gov/instruction/assessment/sbac/index.htm

North Carolina Department of Public Instruction. (2011, July 27). Calendar for Roll-Out of New North Carolina Standards and Assessments. Retrieved November 3, 2019, from https://web.archive.org/web/20140730092751/http://www.ncpublicschools.org/docs/acre/timeline/timeline.pdf

North Dakota Department of Instruction. (2012, February 15). Implementation Schedule North Dakota Common Core Standards. Retrieved November 3, 2019, from https://web.archive.org/web/20121107163614/http://www.dpi.state.nd.us/standard/schedule.pdf

Nowatzki, M. (2016, May 3). Baesler: ND replacing Common Core with new standards for math, English. Retrieved November 3, 2019, from The Dickinson Press website: /news/4023889-baesler-nd-replacing-common-core-new-standards-math-english

O'Donnell, P. (2015, July 1). Ohio dumps the PARCC Common Core tests after woeful first year. *Cleveland Plaindealer*. Retrieved from https://www.cleveland.com/metro/2015/06/ohio_dumps_the_parcc_common_core_tests_after_woeful_first_year.html

Ohio Department of Education. (2012, February). *ESEA Flexibility Request (Original Submission)*. Retrieved from https://www2.ed.gov/policy/eseaflex/oh.pdf

Ohio Department of Education. (2015, March). *ESEA Flexibility Renewal*. Retrieved from https://www2.ed.gov/policy/eseaflex/oh.pdf

Oklahoma Governor's Office. (2014, June 5). Gov. Fallin Signs HB 3399 to Repeal and Replace Common Core Standards. Retrieved November 3, 2019, from https://www.ok.gov/triton/modules/newsroom/newsroom_article.php?id=223&article_id=14279

Oregon Department of Education. (2011, May). *Oregon Common Core State Standards Fact Sheet*. Retrieved from https://www.oregon.gov/ode/educator-resources/standards/mathematics/Documents/commoncorefactsheet.pdf

Oregon Department of Education. (2015, July 23). *Oregon Approved ESEA Flexibility Request*. Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/orrenewalreq2015.pdf

OSPI. (2012, January). *Implementing the Common Core State Standards in Washington State*. Retrieved from https://web.archive.org/web/20140729143419/http://www.k12.wa.us/CoreStandards/pubdocs/CCSSTimeline.pdf

Pepper, M. T., Burns, S. K., Kelly, T., & Warach, K. (2013). *Tennessee Teachers' Perceptions of Common Core State Standards*. Retrieved from TN Consortia website: https://news.vanderbilt.edu/files/RESULTS-Tennessee_Teachers_Perceptions_of_Common_Core_State_Standards.pdf

Raposa, M. (2018, March 24). South Dakota replaced Common Core, but did it really? *Argus Leader*. Retrieved from https://www.argusleader.com/story/news/education/2018/03/24/south-dakota-replaced-common-core-but-did-really/451123002/

Reischman, C. (2013, September 16). Common Core: Missouri's journey to implementation. Retrieved November 3, 2019, from The Missouri Times website: https://themissouritimes.com/6736/common-core-missouris-journey-implementation/

RIDE. (2012, May 29). *Approved ESEA Flexibility Request*. Retrieved from https://www2.ed.gov/policy/eseaflex/approved-requests/ri.pdf

RIDE. (2015, July). *ESEA Flexibility Renewal*. Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/rirenewalreq2015.pdf

Salazar, T. (2014). 50 Ways to Test: A Look at State Summative Assessments in 2014-15. *Education Commission of the States*.

Salazar, T., & Christie, K. (2014). States and the (Not So) New Standards–Where Are They Now? State Academic Standards: Activity around the Common Core. *Education Commission of the States*.

Schaffhauser, D. (2015, December 1). Louisiana To Try Blend of PARCC and State-Developed Assessments -. *The Journal*. Retrieved from https://thejournal.com/articles/2015/12/01/louisiana-to-try-blend-of-parcc-and-state-developed-assessments.aspx

Tatter, G. (2015, May 12). Haslam signs Common Core bill into law. *Chalkbeat*. Retrieved from https://chalkbeat.org/posts/tn/2015/05/12/haslam-signs-tennessee-common-core-bill-into-law/

TN Core. (2012, March). *The Common Core State Standards: Tennessee's Transition Plan*. Retrieved from https://web.archive.org/web/20160712191509/http://www.tncore.org/sites/www/Uploads/files/Common_Core_Plan.pptx

Ujifusa, A. (2015, June 22). Maine Leaves Common-Core Test Consortium. Retrieved November 1, 2019, from Education Week—State EdWatch website: http://blogs.edweek.org/edweek/state_edwatch/2015/06/maine_leaves_common-core_test_consortium.html?cmp=SOC-SHR-FB

Ujifusa, A. (2016, November 2). Map: Tracking the Common Core State Standards - Education Week. *Education Week*. Retrieved from https://www.edweek.org/ew/section/multimedia/map-states-academic-standards-common-core-or.html

Utah State Office of Education. (2010, May). *Phase 2 Race to the Top Application*. Retrieved from https://web.archive.org/web/20170704012308/https://www2.ed.gov/programs/racetothetop/phase2-applications/utah.pdf

Utah State Office of Education. (2015, July 23). *Approved ESEA Flexibility Request (Amended)*. Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/utrenewalreq2015.pdf

West Virginia Board of Education. (2017, February). West Virginia Board of Education Votes to Reduce Testing in Schools and Move Away from Smarter Balanced. Retrieved November 2, 2019, from https://wvde.state.wv.us/news/3357/

West Virginia Department of Education. (2015, March 3). *ESEA Flexibility Renewal*. Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/wvrenewalreq2015.pdf

WestEd. (2018). *State of the States*. Retrieved from The Center on Standards & Assessment Implementation website: https://www.csai-online.org/sos

Wisconsin Department of Public Instruction. (2015a, July 27). *Wisconsin ESEA Flexibility Request*. Retrieved from https://www2.ed.gov/policy/elsec/guid/esea-flexibility/flex-renewal/wirenewalreq15.pdf

Wisconsin Department of Public Instruction. (2015b, September 9). Wisconsin Forward Exam. Retrieved November 2, 2019, from https://dpi.wi.gov/assessment/forward

Woods, J. (2015). *State Summative Assessments: 2015-16 school year*. Retrieved from Education Commission of the States website: https://www.ecs.org/state-summative-assessments-2015-16-school-year/

Woods, J. (2018, April). Math and English language arts assessments and vendors for grades 3-8 (2017-18). Retrieved October 24, 2019, from http://ecs.force.com/mbdata/mbquestrt?rep=SUM1801

Wyoming Department of Education. (2015, July). *Proposed Upcoming Standards Reviews—5 year cycle*. Retrieved from https://web.archive.org/web/20150707002325/http://edu.wyoming.gov/downloads/standards/2015/Standards-Timeline-5-year-cycle-for-website.pdf

Wyoming Department of Education. (2018, October 19). *Wyoming State Content Standards Implementation Timeline*. Retrieved from https://edu.wyoming.gov/downloads/standards/2018/StandardsImplementationTimeline.pdf

Zernike, K. (2015, November 21). Massachusetts's Rejection of Common Core Test Signals Shift in U.S. *The New York Times*. Retrieved from https://www.nytimes.com/2015/11/22/us/rejecting-test-massachusetts-shifts-its-model.html

Zettel, J. (2019, November 12). Wis. Gov. Walker sets sights on Common Core, vouchers. *USA TODAY*. Retrieved from https://www.usatoday.com/story/news/politics/2014/11/12/wis-gov-walker-sets-sights-on-common-core-vouchers/18940681/

Zubrycki, J. (2014, April 17). State legislature votes to delay Common Core-aligned assessments. Retrieved November 2, 2019, from Chalkbeat website:

https://www.chalkbeat.org/posts/tn/2014/04/17/state-legislature-votes-to-delay-common-core-aligned-assessments/

Appendix. NAEP Plausible Values and Conditioning Model Bias

The IRT procedure used to produce the plausible values in NAEP includes controls for a variety of student characteristics. The plausible values are therefore a function of student characteristics. The Mislevy (1991) guidance is not to use covariates in the conditioning model when there is shared variance with variables of interest that are not in the conditioning models. He cautions that if covariates are included in could attenuate the estimated effect of the policy. However, Mislevy also argues that adding more student controls and their nested interactions to the conditioning model drastically reduces the size of the bias, which led to changes in the IRT procedure for the main NAEP (used here). Additionally, Jacob and Rothstein (2018) find no evidence for this source of bias.

References

Jacob, B., & Rothstein, J. (2016). The measurement of student ability in modern assessment systems. Journal of Economic Perspectives, 30(3), 85–108.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56(2), 177–196.