WERE CHILDREN LEFT BEHIND? ESSAYS ON THE IMPACT OF NO CHILD LEFT BEHIND ON STATE POLICY AND SCHOOL CLOSURE

Elizabeth Kate Davidson

Submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy
under the Executive Committee
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

© 2016
ELIZABETH KATE DAVIDSON
ALL RIGHTS RESERVED

ABSTRACT

WERE CHILDREN LEFT BEHIND? ESSAYS ON THE IMPACT OF NO CHILD LEFT BEHIND ON STATE POLICY AND SCHOOL CLOSURE

Elizabeth Kate Davidson

Since 2002, the rules and regulations of the No Child Left Behind (NCLB) Act have dictated state and local education policy, influenced state and local reform efforts, and led to significant investments in building the capacity of state and local education agencies to meet its mandates. Using a nationally comprehensive data set on school- and student subgroup-level NCLB outcomes, these three studies are the first national studies exploring the ways in which state officials' interpretations of NCLB policy led to significant cross-state variation in school and subgroup outcomes across the country. I also investigate the extent to which NCLB accountability pressures and incentive structures led state and local officials to use school closure as a remedy for schools' persistence poor performance. I conduct the latter analysis for all U.S. public schools and separately for a subset of U.S. public schools, all U.S. charters schools, in order to account for the idiosyncrasies of charter school governance and oversight. I find that significant cross-state variation in the share of schools identified as "failing" according to NCLB rules can largely be explained by variation in states' NCLB implementation decisions, and that schools determined to have "failed" according to NCLB rules are more likely to close than schools that never "failed." For all public schools and for charter schools only, a school determined to have "failed" according to NCLB rules is significantly more likely to close than a school determined to have never "failed." Combined, these studies provide insight into the ways in which states' NCLB implementation decisions had significant and lasting impact on school outcomes and state and local reforms.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

ACKNOWLEDGEMENTS

Over the last eight years, I've been lucky to meet the most incredible people, have the most incredible opportunities, and live in the most incredible city. I'm relieved that this chapter of my life has come to a close, and I'm grateful to every person who encouraged me, challenged me, thought highly of me, thought not-so-highly of me, loved me, taught me, picked me up, or knocked me down. What follows is my meager attempt to thank some of those folks, but I can't possibly do them justice in so few words.

Throughout my life, I've always wanted to be involved in every activity, take every class, and make every friend. For this reason, I had a hard time deciding on a major at the University of Georgia. Luckily, I was assigned to an honors program faculty mentor, Professor Chris Cornwell, who sparked my early interest in Economics and Education and led me to major in Economics. In many ways, this journey started with him, and I'm grateful to him for steering me in this direction. Special thanks to Zach Wilcox who shared his immaculate notes with me for Professor David Mustard's Economics and Education class our senior year.

After graduating from UGA and spending two years in Philadelphia as a high school math teacher with Teach for America, I moved to NYC in 2008, a fresh-faced and wide-eyed 24-year-old. I'm less fresh-faced and wide-eyed now, but I have a lifetime of memories and friendships to show for it. Thanks to Katie Aholt, one of my first NYC friends and a key reason I moved to the city. Thanks to the "girls" for easing my transition, Natalie, Michaela, Haley, Anna-Kate, and Koty. Many thanks to my other "girls," Maurya, Ali, Audrey, Alexa, and Natasha. And thanks to all my friends in between, especially Anne-Marie, Kelly, Yvonne, Michelle, Felicia, and Laura.

When I first transitioned from teaching to this Ph.D. program, I missed interacting with students, serving the community, and making immediate and lasting change. In an effort to stay connected to those things, I took on several projects and opportunities throughout my program, which most definitely delayed my completion. One of those projects, ScriptEd, is now making a real difference in the lives of over 900 NYC students in over 37 NYC schools. Thank you to Maurya Couvares, my co-founder, for being such an awesome friend and strong partner—most especially, thanks for spending the past three years ensuring ScriptEd's continued growth. Even though it delayed my Ph.D., it was so worth it.

I'm grateful to those friends and mentors who I've worked alongside at Teach for America, North Star Academy College Preparatory High School, Teacher U/Relay Graduate School of Education, Columbia's Center for Public Research and Leadership, and especially, the New York City Department of Education and Ithaka S+R. I've learned so much from you all. Thank you especially to Martin Kurzweil, whose mentorship and support over the past (nearly) two years I deeply appreciate.

In the early years of my program, Randy Reback and Jonah Rockoff hired me as their research assistant when I was desperate for research experience. After the interview, I remember walking out of Randy's office having no idea whether I'd impressed them or not, and a few days later, I remember celebrating on the side of Interstate 80 after I pulled off the road during a road trip to read their email offering me a job. Thanks for believing in me and letting me join your project. I learned so much from you both. Thanks also to Heather Schwartz for her patience and willingness to advance my STATA coding skills in the early days of my assistantship. I think of you every time I write embedded loops with multiple locals!

I'm particularly indebted to Randy for his mentorship, guidance, and patience over the past seven years. He never gave up on me, even when I probably deserved it. He continually supported and encouraged me, even when I didn't believe in myself. He taught me how to write, how to design and execute a research project, how to think critically about complex questions, and how to overcome my anxieties about writing and finishing. I would never have finished if not for his steady hand, gently pushing me to keep moving forward. Thank you, Randy.

No one is more overjoyed that this journey is complete than my parents, Jim and Jamie Davidson. In the middle years of this process, they frequently asked about my progress and encouraged me to keep going. In the later years, they asked less frequently, but only because that's what I needed. Thank you for your emotional and financial support throughout the years. Thank you, mom, for graciously taking on wedding planning responsibilities when I was too stressed to deal. Thank you both for always believing in me.

Finally, I want to thank my fiancé (soon-to-be husband!!), Tim Pisacreta. You see the best parts of me and reflect them back to me when I need it most. Your love, guidance, friendship, and encouragement made this possible—not to mention your editing and pep talks. I'd do this all over again only because moving to NYC is the reason I get to spend the rest of my life with you.

Elizabeth Kate Davidson (Liz)
August 18, 2016

DEDICATION

This dissertation is dedicated to my maternal grandfather,
James (Jim) Harvey Alexander.
(b. March 29, 1921 – d. June 16, 2009)
Love and miss you, Papa.

INTRODUCTION

The No Child Left Behind (NCLB) Act of 2001 required that all states adopt an accountability system that rated its schools based on the performance of their students on annual state exams in reading and math. While lawmakers designed NCLB to allow states some flexibility in implementation, NCLB rules required that states evaluate each school against a set of escalating yearly student proficiency rate targets, the last of which, in 2014, is 100 percent of students in a school reaching proficiency. If a school's students did not meet the proficiency targets in a particular year, then the school "failed" in that year. For schools that failed, NCLB prescribed a set of consequences that were designed to provide students in these schools with additional support and provide strong incentives for states and districts to improve schools that failed.

There is some compelling evidence that NCLB has had a positive impact on student outcomes, especially for those students near the margin of meeting proficiency targets on state exams (Ahn & Vigdor, 2014; Springer, 2008; Dee & Jacob, 2011, Lauen & Gaddis, 2014; Reback, 2008; Rockoff & Turner, 2010). This evidence suggests that NCLB rules were at least mildly successful in creating the right incentives to impact educators' behavior. Yet there is little national evidence showing how state and local policies, derived by state and local officials from federal NCLB rules, created the incentive structures that influence outcomes. The three studies in this collection investigate how states' implementation of NCLB impacted school outcomes, how those school outcomes are related to state and local efforts to remedy poor school performance, and how the application of state and local reforms are influenced by state policy contexts, such as teachers' unions, state legislation, and other factors.

Political Origins of NCLB

The No Child Left Behind Act (NCLB), passed by Congress with significant bipartisan support in 2001, was the result of several decades of an expanding federal presence in matters of education policy, an expansion driven by several key events over the last half-century. In response to fears that U.S. schools were failing to adequately prepare students to compete with Russian students in math and science, Congress passed the National Defense of Education Act of 1958 that led to "new money [for schools] and new responsibilities for the federal government [in the education sphere]" (Chubb & Moe, 1990, p. 7). In the early 1960s, campaigns for racial equality and the issuance of the 1966 Coleman Report on Equality of Educational Opportunity highlighted the differences in educational outcomes between low income and affluent students. In an effort to "level the playing field" for disadvantaged students (Coleman, 1966), Congress passed the Elementary and Secondary Education Act (ESEA) of 1965 which doubled federal funding to schools through the Title I program—a program that, even today, provides funds and services to states' most disadvantaged students (McDonnell, 2005).

In the 1980s, a frail economy and a decades-long trend of declining student scores on the Scholastic Aptitude Test (SAT) led to a resurgence in the study of school effectiveness (Congressional Budget Office, 1986). The most notable report, *A Nation at Risk,* determined that "the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people" (National Commission on Excellence in Education, 1983). In response to these studies, states led a national reform movement to leverage school inputs to improve students' educational outcomes by

implementing a range of reforms. These reforms included augmenting course offerings and increasing graduation requirements, lengthening the school year and school day, strengthening teacher certification requirements and increasing teacher compensation, and implementing statewide compulsory tests of student performance (Firestone, et al., 1991). During this time, education spending became the largest share of state budgets in all fifty states, yet student outcomes remained largely stagnant (Stedman & Kaestle, 1991; Stedman, 1993; Linn, et al., 1990).

Extensive state reform efforts continued throughout the 1990s, but student results largely remained stagnant. One might conclude that state reforms were misguided, yet the economics literature provides limited insight on proven, scalable strategies to improve school quality. For example, the evidence is mixed that increasing school funding leads to better student outcomes (Card & Krueger, 1996; Coleman, 1966; Figlio, 1999; Hanushek, 1997; Hedges, et al., 1994; Pritchett & Filmer, 1997). Some compelling evidence exists that students feel safer and more supported and teachers feel more empowered and have more opportunities to collaborate in small learning communities and small schools, factors which lead to improvements in student outcomes (Bloom & Unterman, 2012; Abdulkadiroglu, et al., 2013; Barrow, et al., 2013; Kuziemko, 2006). Evidence also exists that consecutive years of small class sizes in early grades may improve students' long-term academic and social outcomes (Krueger & Whitmore, 2001; Schanzenbach, 2014). However, class size reductions large enough to affect student achievement carry large costs, are difficult to implement, and may inadvertently diminish teacher quality (Whitehurst & Chingos, 2011). Effective teachers, as measured by the value they add to student test scores, have a significant impact on students'

future earnings (Chetty, et al., 2012). Yet, recruiting, hiring, training, and retaining high quality teachers in low-performing schools remains a great challenge (Feng, 2010; Feng & Sass 2011).

When increased spending in schools produced little change in student outcomes, policy-makers believed that holding schools "accountable" for the performance of their students would improve school efficiency—schools would be incentived to perform better without corresponding increases in spending. The 1990s—or the "age of accountability"—marked an uptick in the number of states that were implementing accountability systems designed to punish low-performing schools with a series of defined consequences and reward high-performing schools and teachers with public recognition or financial incentives (Hanushek, 2005, p. 306).[1] Yet, accountability systems work "only to the extent that they motivate students, teachers, and schools to examine their performance and make changes to improve" (Hanushek & Raymond, 2002, p. 99). In addition to codifying remedies for persistent school failure, the school accountability era further cemented the idea that schools have the capacity to improve student outcomes—schools simply must be incentivized to do so.

In 1994, the passage of the Improving American Schools Act (IASA) and Goals 2000 by Congress highlighted the continued federal focus on equalizing students' school experiences in an effort to equalize student outcomes.[2] The IASA required that states increase funding to programs that help disadvantaged students meet rigorous learning standards by increasing

[1] Economic theory suggests that accountability systems "are a powerful mechanism for overcoming the principal:agent problem," where asymmetric information prohibits the "principal" from evaluating the performance of the "agent" and whether the "agent" is acting in the best interests of the "principal" (Figlio & Loeb, 2011, p.386). Accountability systems re-align divergent goals by evaluating the agent's performance on specific criteria and then imposing appropriate consequences and rewards. Accountability systems also resolve information asymmetry by publicly reporting "agent" performance on specific criteria so that "consumers" can use that information to make decisions or advocate for change.

[2] The IASA was the second reauthorization of the 1965 Elementary and Secondary Education Act (ESEA).

students' access to high-quality teachers, tutoring, and educational enrichment opportunities (USED, 1994a). Simultaneously, the USDOE developed Goals 2000, which required states to develop learning standards and implement assessments to test students' progress toward those standards (USED, 1994b). States were required to fully comply with the provisions of IASA 1994 to continue receiving federal Title I money, but compliance with Goals 2000 was optional and not tied to funding. As of 2000, however, 39 states had school-level accountability systems in place.

NCLB Overview

Persistent achievement gaps and expanding use of accountability systems at the state level set the foundation for a national effort to regulate accountability systems and augment accountability pressure. That national effort, NCLB, required states to construct accountability systems that used standardized tests to measure student and subgroup proficiency rates in math and reading by annually testing students in grades 3-8 and high school. States must report on the performance of all students in a school and on the performance of various student subgroups: low income students, students with disabilities, students with limited English proficiency, and students in specific racial/ethnic groups.

Using pre-NCLB assessment results, states set yearly proficiency rate targets against which they measured the average performance of each student subgroup and the school as a whole. Each year, the proficiency rate targets increased so that, by 2014, the proficiency rate target for every school in every state was set to 100 percent of students scoring proficient or above. In any given year, every subgroup must meet the prescribed target in each subject for

the school to "pass," or make Adequate Yearly Progress, or AYP.[3] If the average proficiency rate of any subgroup did not meet the target in either reading or math, then the subgroup, and therefore the school, failed to make AYP. Any school that failed to meet their AYP targets in a particular year faced defined consequences, and a school that failed across consecutive years faced consequences that become increasingly severe (USED, 2001).

NCLB rules prescribed the consequences states must implement when schools failed to make AYP. After two years of consecutive failure, schools were designated as "In Need of Improvement" and, among other things, required to offer students the opportunity to transfer to non-failing schools within the same district. After three years of consecutive failure, schools continued to be subject to the transfer rules and were also required to use Title I funds to provide students with after-school tutoring services. The consequences of failure became more severe once schools were subject to "Corrective Action"—the consequence phase that schools entered after four consecutive years of failing AYP. Once a school enters the Corrective Action phase, NCLB rules required that states and districts choose from a list of options—replacing some school staff, instituting a new curriculum, extending the school day or year, restructuring school organization—designed to drastically change the operations of the school.

After five consecutive years of failure, schools entered the final phase of NCLB reform, "Restructuring." In this phase, states and districts could (a) turn the management of the school over to the state, (b) turn the management of the school over to a private management organization, (c) replace all or most of the school's staff, or (d) an "other" option whereby the state or district proposed a significant reform with "substantial promise of enabling the school

[3] For a more complete description of NCLB rules, see Chapter I.

to make adequate yearly progress" (USED, 2001). The Center on Education policy reports that, in studied states, states and districts most often chose the "other" option (Center on Education Policy, 2008). In 2012, nearly 7,643 U.S. schools were either planning for or implementing reforms in the restructuring phase (USED, 2012)

Brief Dissertation Summary

NCLB's escalating annual proficiency rate benchmarks and prescribed consequences infused a new sense of urgency into school reform efforts. After the onset of NCLB, states and districts felt an increased pressure to "turnaround" schools at risk of failing AYP and raise student achievement rather than engage in sustained, lengthy efforts to improve (Ballou & Springer, 2009; Ladd & Lauen, 2010; Reback, 2008; Rouse, et al., 2007; Springer, 2008). States began to increasingly use school closure as a way to circumvent schools' going too far down the NCLB consequence timeline and requiring significant investments (Dardick & Ahmed-Ullah, 2012; Graham, 2012b; Phillips, 2011; Rich, 2012; Santos, 2011). At the same time, urban districts across the country were also experiencing declines in student enrollment due to residential shifts, demographic changes, and students leaving traditional public schools to enroll in public charter schools.[4] In response, many urban districts used student performance to

[4]From 2000-2010, charter school enrollment increased from 500,000 students to 1.7 million students nationwide. During the same period, Detroit lost 87,000 students from traditional public schools—a 57 percent decline in district enrollment. Similarly, Chicago and Pittsburgh experienced traditional school enrollment declines of around 20 percent. Kansas City only saw a 4 percent total enrollment decline, but when combined with the effect of students shifting from traditional public to charter public, traditional public schools lost 45 percent of their students over the last decade. Similarly, Philadelphia's decline in the school-age population was 11 percent, but when combined with student transition to public charters, 23 percent (PEW, 2011). Like Kansas City, Philadelphia district buildings were operating at only 67 percent capacity.

determine which schools to close (PEW, 2011; Giordano, 2012; Ozek at al., 2012; Carlson & Lavertu, 2015).[5]

This dissertation explores the impact of states NCLB' implementation decisions on school outcomes and describes the extent to which states and districts used school closure in response to those outcomes. Using several datasets, I compile and combine national data on school and student subgroup NCLB outcomes from 2003 to 2005, school district demographics from 2000, school characteristics from 2002 to 2011, indicators of whether schools closed between 2002 and 2011, provisions in state charter school laws as of 2006, and information about the presence of teachers' union and collective bargaining agreements in 2003. These data support my analysis across all three chapters, and the data on NCLB outcomes are publicly available for other researchers or policy-makers to use.

My first chapter,[6] coauthored with Randall Reback, Jonah Rockoff, and Heather Schwartz, explores the effects of variation in state implementation of the NCLB on school and student subgroup outcomes. We use detailed information about states' NCLB-mandated accountability systems to isolate those factors that explain cross-state differences in school AYP failure rates, which, in 2003, ranged from one to 80 percent. We find that, contrary to widely held beliefs, variation in states' school AYP failure rates is only loosely related to students'

[5]The following districts considered school performance and/or the impact of the school closure on neighborhoods (for each district, I list the total number of schools in the district that were closed during the indicated time period): Chicago (44 total schools closed, between 2001-2009), Detroit (59 total schools closed, in 2009 and 2010), Washington, D.C. (23 total schools closed, in 2008), Pittsburgh (22 total schools closed, in 2006), Kansas City (29 total schools closed, in 2009 and 2010), Milwaukee (20 total schools closed, between 2005-2010), Philadelphia (planning to close 65 schools from 2013-2017), and New York (140 closed or phasing out, between 2002-2012) (PEW, 2011).

[6] Davidson, E., Reback, R., Rockoff, J., & Schwartz, H.L. (2015). Fifty Ways to Leave a Child Behind: Idiosyncrasies and Discrepancies in States' Implementation of NCLB. Educational Researcher, 44(6), 347-358.

performance on state exams; rather, subtle differences in states' NCLB policies combined with states' school and student characteristics led to extreme differences in NCLB outcomes.

My second and third chapters investigate the extent to which states and districts considered schools' NCLB performance in the early years of NCLB in subsequent school closure decisions. In my second chapter, I use a probit estimation with state fixed effects to explore the relationship between NCLB performance and school closure for all U.S. public schools. I use data on school characteristics from the National Center for Education's Common Core of Data (CCD) to determine, using federally-assigned school identifiers, which schools closed and never re-opened between 2004 and 2011.[7] To understand if state policy contexts affect schools that failed and schools that never failed differently, I interact schools' NCLB performance with a measure of the diffusion of teachers' unions in the state.

In my third chapter, I focus on a subset of all U.S. public schools, public charter schools.[8] My analytic approach in chapter three is similar to my approach in chapter two, except I also include and compare indices that measures the "strictness" of the accountability provisions in each state's charter school law. Strict charter laws include more explicit and rigid rules that govern the monitoring, renewal, and termination of schools' charter contracts. This analysis also compares my results for charter schools to non-charter public schools to explore whether

[7] In this paper, I use a broad definition of school closure: I identify as "closed" any school whose federally-assigned identifier disappears from the data and never reappears. Therefore, I do not distinguish between schools that a) close and the building that housed the school does not house a new school, b) close but a new, distinct school opens in the same building with or without many of the same students, c) are closed and re-opened as two or more smaller schools or academies with or without the same students, or d) are converted to charter schools or re-opened under new governance with or without many of the same students. It is also marginally possible, but unlikely, that I designate a school as "closed" due to a data error where the federally-assigned identifier is changed for some reason other than a significant reorganization or a legitimate change in governance.

[8] Public charter schools are public schools that receive public financing but are typically governed independently from the local school district and, to varying degrees, are exempt from state and district rules related to spending, staffing, and other operations.

effects of AYP failure on school closure differ across the two groups.

The three studies in this collection contribute to the body of evidence on NCLB outcomes and especially, the more recent discussion on the use of school closure as a remedy for persistent low-performance. To my knowledge there are no national studies that document the complex effects of states' NCLB implementation decisions, the scope of school closures, or the impact of accountability pressures and other policy factors on the use of school closure as a mechanism for reform. The findings of these studies provide important context for policy-makers about the trade-offs and implications of large-scale, complex accountability systems.

CHAPTER ONE – FIFTY WAYS TO LEAVE A CHILD BEHIND: IDIOSYNCRASIES AND DISCREPANCIES IN STATES' IMPLEMENTATION OF NCLB

With Randall Reback of Columbia University, Jonah Rockoff of Columbia University, and Heather Schwartz of RAND Corporation

Introduction

The American public education system has had a history of strong local community control of public schools. U.S. public schools are predominantly funded through a combination of state and local tax revenues. Since the Elementary and Secondary Education Act of 1965, the federal government has been supplementing these revenues by awarding funds to states for allocation to public schools serving students from low-income families. These federally-funded revenues are known as "Title I funding." Cohen and Moffit (2009) describe how the first 35 years of Title I funding included several rounds of debates concerning schools' flexibility in using their Title I funds and whether the impacts of Title I funds on student achievement should be evaluated. Interest in preserving America's history of strong local control of schools sometimes clashed with desires to attach strings to Title I funding to increase its efficacy as a poverty reduction program. Local control generally won the day during those first 35 years. The relative size of the Title I program did not grow (3.6% of public school revenues in 1969-1970 and only 2.6% in 1999-2000[9]), nor did a national system emerge to evaluate whether Title I funds were improving student achievement. States and school districts arguably had greater incentives to monitor the fiscal compliance of their Title I funds than to assess whether these funds were

[9] Calculated based on statistics reported by the U.S. Department of Education (1994, 2001, 2012).

going to the most productive outlets (Gordon and Reber, 2015). Several federal administrations during that time encouraged the adoption of national standards and the development of state accountability systems for schools, but these were voluntary (Manna, 2006).

In 2001, the re-authorization of the Elementary and Secondary Education Act marked the single greatest expansion of the federal role in education policy since the original 1965 Act (Manna, 2010). This re-authorization, known as the *No Child Left Behind (NCLB) Act*, broke new ground by mandating schools be held accountable for their students' achievement as a condition of states' receipt of Title I funds. NCLB requires states to construct school accountability systems using standardized tests to measure student proficiency rates in math and English Language Arts (ELA). A school fails to make Adequate Yearly Progress (AYP) if proficiency rates fall short of that year's targets. This AYP determination was based not only on the proficiency rates of schools' general student populations but also on the proficiency rates of various ethnic and categorical subgroups of students, such as students from low-income families.

NCLB changed education policy by leveraging Title I funds to compel states to develop standardized testing systems for assessing student proficiency levels. NCLB increased the size of the Title I program—from roughly $8.5 billion appropriated in 2000-2001 to $13.6 billion appropriated in 2005-2006 (U.S. Department of Education, 2001, 2006). Title I funding remained equivalent to only about 3 percent of total public school operating expenditures, though this percentage remains much higher for some school districts than others. NCLB did not establish an evaluation system for the impact of Title funds, or funding in general, on student achievement. One of the few direct changes to the use of Title I funds was to allow

students from low-income families to purchase after-school tutoring services (called supplemental education services), by re-directing Title I funds away from schools that had failed to make AYP. Rather than holding states or schools accountable for the use of Title I funds, NCLB forced states to hold schools accountable for their students' proficiency rates.

From NCLB's inception, federal policymakers avoided a "one size fits all" policy and encouraged states to adapt NCLB guidelines to meet the demands of their particular contexts. For example, states could choose their own exams and set definitions of proficiency on those exams. Many states already had their own testing and accountability systems prior to NCLB, and so the impact of NCLB could depend on whether students were already being tested under similar accountability systems (Dee and Jacob, 2011; Dee, Jacob, & Schwartz, 2013).

The early years of NCLB thus provide an important example of how variation in state policy implementation can cause a federal law to have very different consequences across the country. While previous studies have examined states' and schools' implementation of NCLB (Manna, 2006, 2010; Srikantaiah, 2009; Hamilton et al, 2007), these studies each examine a limited number of states or localities. No prior study has used national data to examine the link between states' initial NCLB implementation decisions and their schools' ratings. In so doing, our work provides a concrete example of the effects of the expanding federal role within education (Cohen & Moffit, 2009; Henig, 2013; Manna, 2006, 2010).

Using a newly-assembled national data set, we investigate the following questions:

(1) Which types of schools failed during the early years of NCLB? How are student demographics, school grade levels, and schools' urbanicity related to failure rates?

(2) Which performance targets did schools fail to meet? Did schools frequently fail due to the performance of one student subgroup alone?

(3) What explains cross-state differences in school failure rates? Are these differences associated with student demographics or with specific state policy implementation decisions?

We find that wide cross-state differences in failure rates were largely the result of subtle differences in states' own NCLB rules. A common misconception regarding wide variation in AYP failure rates across states is that this variation was driven by more obvious state policy differences, such as the difficulty of the exam questions and the proficiency standards. In fact, school failure rates are only weakly related to student proficiency rates. A better understanding of how subtle policy differences influenced schools' ratings during the early years of NCLB may inform current efforts to reform NCLB and other school accountability programs. Even if states are given wide flexibility in the design of their accountability and testing systems, policy-makers may wish to remove loopholes that create disparate standards for schools via haphazard differences in rules and calculation methods. Flexibility need not come at the cost of transparency.

Our paper proceeds as follows. Section II provides an overview of NCLB policies, and Section III describes our data. Section IV describes which types of schools most frequently failed and which performance targets they failed to meet. Section V describes cross-state variation in school failure rates, Section VI explores reasons for this variation, and Section VII briefly discusses the implications for current policy decisions.

NCLB Overview

A school's performance rating under NCLB is based on student proficiency rates on statewide tests, student participation rates on those tests, and an additional state-selected indicator of student performance.[10] Both the campus as a whole and various student subgroups— racial/ethnic subgroups, students eligible for free/reduced priced lunch, students with limited English language proficiency, and students with disabilities—must meet all of the performance targets for the school to make AYP.[11]

The three core mandatory elements of NCLB pertain to annual testing of virtually all public school students in certain grade levels and subjects, an increasing bar for the fraction of students demonstrating proficiency on these tests, and annual determinations of school performance with consequences for schools that fail to make AYP. NCLB required states to administer baseline student exams in the spring of 2002 and to adopt school accountability systems for the school year 2002-2003. States selected their own exams and defined proficiency on those exams. States then determined a schedule for the percentage of students who must meet proficiency each year, with targets increasing annually up to a mandated 100% target for 2014. States could set different benchmarks by grade level and by subject area, but

[10] We provide a brief overview of NCLB in this section and refer the reader to the U.S. Department of Education's *Desktop Reference* (2002) and to Manna's *Collision Course* book (2010) for more details on NCLB policies. Manna also provides revealing anecdotes concerning the challenges faced by states and schools in implementing these policies.

[11] Students are counted in all subgroups to which they belong. For example, a Hispanic student who is limited English proficient and eligible for free lunches will contribute to eight different proficiency rates—the campus-wide group, the Hispanic subgroup, the limited English proficient subgroup, and the free/reduced priced lunch subgroup proficiency rates in math and ELA. Subgroup proficiency rates only influence the school's AYP rating if there are sufficient numbers of students enrolled at the school (and meeting the "continuous enrollment" definition described elsewhere in the paper).

not by student subgroup. To prevent schools from strategically exempting low-performing students from taking exams, NCLB dictates that student subgroups are required to meet a 95% participation rate on both math and ELA exams. The final category of school performance is the state-selected "other" academic indicator. NCLB rules allowed for flexibility in states' selection of elementary and middle schools' other indicators, and most states used attendance rates. NCLB rules required that states use graduation rates for high schools' other indicator.[12]

In addition to the stigma of failing to make AYP, there are additional consequences for schools serving low-income populations that receive funding under the federal Title I program. Students at failing Title I schools have the opportunity to transfer to non-failing schools within the same district. After consecutive years of AYP failure, these schools' students from low-income families are entitled to use school funds to purchase private tutoring services (called "supplemental education services"). If these schools fail to make AYP for several years, then they are subject to closure or restructuring.

Beyond these core requirements, there are three key areas where states have latitude in calculating AYP. We summarize them here and provide further detail in the sections that follow. The first area relates to acceptable adjustments to student proficiency rates under the law. Even if a subgroup's or school's performance falls below the proficiency target for the given school year, the school may still make AYP because NCLB allows states to employ various

[12] Initially, NCLB permitted states to use their own formulae for calculating graduation rates. In December 2008, the U.S. Department of Education announced that all states must use a standardized four-year graduation rate formula. The U.S. DOE requested states implement the new formula as soon as possible but required states to comply by 2010-2011 (U.S. DOE, 2008).

statistical techniques and contingencies to adjust proficiency rates.[13] Two types of adjustments permitted under NCLB are the application of confidence intervals and the use of "safe harbor." Confidence intervals provide leniency around proficiency rate targets to account for small numbers of tested students. They lower a student group's effective proficiency targets based on the number of tested students in that group at that school—the smaller the group, the larger the confidence interval.[14] "Safe harbor" rules offered leniency to schools that missed proficiency targets but had students make large gains in proficiency rates from the previous year. To make AYP under the safe harbor rule, states typically require a 10% reduction in the fraction of students failing to reach proficiency.

The second area where states have latitude is determining which students count towards the accountability system. In the initial years of implementation, not all states applied consistent definitions of special needs categories exempted from the general standardized test. However, the U.S. Department of Education (U.S. DOE) later issued exemption rules to close loopholes related to testing of students with disabilities. But several other discrepancies

[13] Beyond the formal NCLB rules, states also allowed school districts and schools to submit appeals of schools' AYP ratings. Acceptable grounds for appeal varied by state. For example, in Colorado, schools could successfully appeal AYP failure if the sole reason for failure was the performance of the subgroup of students with disabilities and if this subgroup did meet its targets in another year. In several states, (e.g., Iowa and Michigan), schools could appeal by retroactively exempting students from contributing to participation rates if the students had experienced significant medical emergencies.

[14] The confidence interval adjustment lowers the target from p to $p - [\sqrt{\frac{p(1-p)}{n}} * C]$, where p is the unadjusted proficiency rate target in decimal form, n is the number of students contributing to the proficiency rate, and C is the critical value for the specified confidence interval, such as 1.96 for a 95% two-sided confidence interval. For example, in Alaska, the 2003 ELA proficiency target was 64% and the state used a 99% confidence interval adjustment. An Alaskan student subgroup with 20 students would only have to reach 36% proficiency that year to make AYP, because .36= $[.64 - [\sqrt{\frac{.64(1-.64)}{20}} * 2.575]$, where 2.575 is the critical value for the 99% confidence interval.

remain. Not all states hold the same racial and ethnic subgroups of students separately accountable for meeting proficiency rate targets; for example, Asian American students might be a separate category in one state but not in another. In addition, states determine how long students must be enrolled in the same school for their test performance to contribute to schools' AYP determinations. These "continuously enrolled students" comprise the denominator of the participation rate calculation. A state with a very strict definition of continuous enrollment only counts students enrolled at their schools for one calendar year prior to testing. More commonly, states count students who were tested in the spring and had been enrolled at their schools since late September or October. Schools could also exempt students from contributing to participation rates if the students experienced significant medical emergencies. To protect student anonymity and avoid using unreliable measures of subgroup performance, the proficiency rate of a student subgroup only affected its school's AYP determination if the number of students in that subgroup exceeded a specific threshold. States had flexibility in choosing that minimum subgroup size threshold. Most states chose a minimum subgroup size between 30 to 40 students, but the range extended from 5 students to 100 students. In some states, minimum group size was a function related to school population. For example, California's subgroups were held accountable if they either had 100 tested students or at least 50 tested students that composed at least 15% of the schools' total tested population.

A third, often-overlooked area of flexibility is which grade levels of students were tested and the methods of aggregating performance across grade levels. Although tested grade levels

became more standard as of 2005-2006,[15] the aggregation of scores across tested grade levels within a school was not. For schools that served multiple tested grade levels, states could decide whether to aggregate statistics across all of the tested grade levels or to consider the student proficiency levels of each grade separately. For example, in a state like Washington that considered each tested grade's proficiency level separately, both 4th graders and 7th graders in a hypothetical school would each need to exceed proficiency targets, making it more likely the school could fail AYP. However, other state AYP criteria pertaining to minimum subgroup size and confidence intervals could offset that challenge. Specifically, Washington counted the number of tested students in each grade separately to determine the size of the confidence interval to apply to that grade level's proficiency rate. This means the respective confidence intervals for 4th grade and for 7th grade proficiency rates were more generous than a confidence interval applied to a proficiency rate that pooled 4th and 7th graders. It is also more likely that the number of 4th graders or 7th graders, when considered separately, would fall below Washington's minimum subgroup size threshold, rendering 4th or 7th grade proficiency rates inapplicable to a school's AYP determination.

NCLB Data

NCLB has greatly expanded the amount of student performance data available to researchers and the public, though dissemination of data has been uneven across states. To promote studies of NCLB, we approached each of the 50 states individually in an attempt to

[15] As of 2005-2006, states were required to test students in grades 3 through 8 and in one high school grade. Before this, states were required to test in at least one elementary grade, at least one middle school grade, and at least one high school grade. Consequently, tested grade levels varied across states during the first few years of NCLB. On the one extreme, states like Maryland tested in all grades 3 through 8 for AYP determinations. On the other extreme, states like New Jersey only tested grades 4, 8, and 11 up until 2004-2005.

form the most complete school-level data set concerning the early years of NCLB. We used a combination of methods to obtain the most comprehensive and accurate data possible—primarily requesting data directly from state education departments and downloading data from state websites.

The resulting school-level data set includes school-level AYP determinations and the subcomponents for these determinations. Our variables include indicators of whether the school as a whole and each individual student subgroup made AYP, school-and subgroup-level average student proficiency rates on state assessments, and the number of students tested in the school and in each student subgroup. For the school years 2002-2003 and 2003-2004, we filled in otherwise missing data with information provided by the American Institutes for Research (2005) and the Council of Chief State School Officers (2005). The resulting data and our state-by-state documentation of sources are publicly available.[16] For 2004-2005, we use school and subgroup proficiency target data from the American Institutes for Research (2005).

Descriptive Evidence on Failing Schools

Looking nationwide from 2003 to 2005, there were clear observable differences between AYP failing and non-failing schools (Table 1). AYP failing schools were more likely to have higher total student enrollments, to have larger enrollments of poor and minority students, and to be designated as Title I schools. On average, schools that failed all three years had nearly double the percentage of students eligible for free and reduced-priced lunch as schools that made AYP all three years. Failing schools also had fewer teachers per student and

[16] Data for the first two years of NCLB are currently accessible from our "No Data Left Behind" website at www.gsb. columbia.edu/nclb (Authors, 2011).

were disproportionately located in urban school districts. Middle schools and high schools failed far more frequently than elementary schools.

Most schools failed to make AYP due to proficiency rate requirements as opposed to participation rates. The majority of failing schools had groups of students not meeting proficiency rate targets in both subjects. In 2005, 52% of failing schools missed proficiency rate targets in both subjects, 24% of failing schools missed ELA proficiency rate targets only, 20% of failing schools missed Math proficiency rate targets only, and the remaining 4% of failing schools satisfied all of their proficiency rate targets but not their participation rate targets. The number of schools failing due to participation alone was substantially lower in 2005 than in the prior two years, suggesting that schools took action to ensure that sufficient numbers of students were tested.[17]

While schools were potentially accountable for many student subgroups, the rate at which different subgroups caused schools to fail AYP varied widely. Such differences could simply have been due to whether a subgroup was large enough to be held accountable. Figure 1 shows the percentage of schools where various subgroups counted toward AYP in 2004, as well as the rates at which these subgroups failed to make AYP. The total height of each bar illustrates the fraction of schools where that subgroup's proficiency rate counted towards the AYP determination, while the shaded areas of the bars represent the fraction of schools where that subgroup failed to make AYP. White and economically disadvantaged subgroups were held

[17] Participation data are not available for as many states in 2003 and 2004 as in 2005. When we restrict the sample to the 31 states with data available for all three years, then we observe a downward trend in the fraction of schools failing only due to participation: from 17% in 2003 to 14% in 2004 to 5% in 2005.

accountable in about 43% and 37% of schools, respectively, while fewer than 4% of schools had a Native American subgroup held accountable.

However, conditional on being accountable, subgroup failure rates varied considerably. Figure 1 reveals that White and Asian subgroups rarely failed, while more than half of all accountable Native American subgroups and students-with-disabilities subgroups failed to meet proficiency targets. The students-with-disabilities subgroup was also the most likely to be the only subgroup failing their schools' proficiency targets: 57% of accountable students-with-disabilities subgroups were the only group to fail to meet targets at their schools.

Cross-State Differences in Failure Rates

Figure 2 illustrates the wide variation in states' AYP failure during the first three years of NCLB. Figure 2 is a density plot, the continuous version of a histogram, so the area under the curve represents the proportion of states falling in various ranges of values for the fraction of their schools failing to make AYP. For example, in the first year of AYP designations (2003), approximately 40% of states had AYP failure rates between 20 and 40 percent.[18] That same year, 32% of the nation's schools failed AYP, but failure rates ranged from 1% in Iowa to 82% in Florida. The national failure rate declined to 26% by 2005, but failure rates ranged from 2% in Oklahoma to 66% in Hawaii.

[18] This is found by calculating the area under the blue-colored curve, which equates to approximately 2 units on the y-axis multiplied by 0.2 (=.4-.2) units on the x-axis.

Failure rates changed substantially over time in some states. Alabama's failure rate jumped from 4% in 2003 to 68% in 2004.[19] Tennessee's failure rate declined from 47% in 2003 to 7.6% in 2005.

Failure rates by school level also varied substantially within some states. For example, only 11% of Georgia's elementary schools failed to meet AYP in 2003, yet 72% of its high schools failed. Similarly, only 20% of West Virginia's elementary schools failed in 2003, yet more than 80% of its high schools failed.

A common misconception is that this wide variation in failure rates resulted from cross-state differences in the proportion of students identified as proficient. In reality, states' school failure rates were not strongly related to their students' performance. Figure 3 illustrates the lack of a strong relationship between school failure rates and student proficiency rates, showing student performance on states' math exams for the spring of 2004 against their states' school failure rates. Based on corresponding linear regression, a one percentage point increase in state math proficiency rates is associated with only a statistically insignificant 0.05 percentage point decline in the fraction of a state's schools making AYP.[20] This weak relationship arises because states determined NCLB proficiency targets based on their own pre-NCLB student proficiency rates. In essence, states were grading their schools on a curve, with state-specific curves based on the starting points and trajectories for proficiency targets. For example, Iowa set 2003

[19] In 2002-2003, Alabama had an interim accountability system that used students' grade-level, not subgroup-level, norm-referenced scores to determine school-level AYP status. By 2003-2004, Alabama transitioned to a NCLB-compliant accountability system.

[20] The relationship with state ELA proficiency rates is also statistically insignificant and small, only a 0.16 percentage point decline in the fraction of schools making AYP. If we regress states' school AYP failure rates on quadratic terms for their states' proficiency rates in each subject (i.e., four independent variables total), the R-squared is .07 but the adjusted R-squared is only .02. The joint significance level of these estimated coefficients is 0.56.

proficiency targets at 64% in math and 65% in ELA, while Missouri chose 8.3% and 18.4%, respectively.

Even states with similar starting points had dramatically different rates of schools failing AYP. For example, proficiency targets in Louisiana and Florida differed by less than 7 percentage points, but their 2003 school failure rates differed by more than 75 percentage points. Reback, Rockoff, & Schwartz (forthcoming) document how a sizable fraction of schools that did not make AYP in their own states would have very likely made AYP in many other states.

Explaining Cross-State Variation in Failure Rates

Various dimensions of NCLB implementation contributed to the wide variation in school AYP failure rates.[21] No individual state policy decision appears to have been the primary culprit. Instead, failure rates appear to have been influenced by interactions among several decisions and states' school characteristics (e.g., enrollment size, grade spans, and ethnic diversity of students). Given that we only have a sample of 50 states and a host of potentially important explanatory variables, there are insufficient degrees of freedom to tease out the relative importance of state policy variables via regression analysis. To examine the nature of these complex interactions, we instead describe five categories of policy decisions that we have identified as having had substantial impacts on some states' school failure rates. We provide examples of states where failure rates were strongly influenced by these decisions. The first of

[21] To determine each state's confidence intervals, safe harbor policies, and other AYP formulae choices, we referred to their approved state accountability workbooks. We obtained the workbooks from http://www2.ed.gov/admins/lead/account/stateplans03/index.html in January of 2007. Where possible, we selected criteria that applied to the 2003-2004 school year. However, as the workbooks were updated sometimes annually and often overwrote prior versions, we are not always able to determine when states adopted their criteria. For example, many states began to apply a 75% confidence interval to safe harbor determinations in 2005-2006.

these categories covers implementation errors that were rectified within the first couple of years of NCLB, but the remaining categories encompass policy decisions that continue to affect school failure rates. We focus on examples below, and Table 2 provides some relevant policy information for all fifty states. The states in Table 2 are sorted in ascending order by the percent of schools failing to make AYP in 2004.

- *A few states initially deviated from NCLB rules.*
    a. *Calculations.* Iowa continued to develop its AYP formula and data collection processes throughout the initial two years of NCLB. Using proficiency rate and participation rate data we retrieved from Iowa's Department of Education website, we applied Iowa's AYP formula and found higher failure rates than the state's official published rates.[22] In 2003 and 2004, respectively, 20% and 3% of Iowa's schools made AYP even though they had at least one accountable subgroup missing the 95% participation target.[23] Iowa did have an appeals process by which schools can petition to have up to 1% of students excused from participation due to illness, but the reported participation rates were often too low to have warranted a successful appeal. Data disaggregated by grade level is unavailable for Iowa, but we can examine proficiency rates for the 90% of Iowa's schools that served only one tested grade level.[24] Among these schools in 2004,

[22] During the summer of 2004–the months when state officials typically make AYP determinations – the state official responsible for AYP determinations suffered an injury that required a leave of absence (Deeter, personal communication, 3/5/13). This disruption and subsequent understaffing may have led to inconsistencies in Iowa's AYP determinations and may partially explain why Iowa's failure rates were extraordinarily low: less than 1% in 2003 and less than 5% in 2004.
[23] In 2004, Iowa used a uniform averaging procedure for both its proficiency and participation rates. If either the 2004 proficiency (participation) rates or the average of the 2003 and 2004 proficiency (participation) rates were greater than or equal to the proficiency target (95%), the subgroup met the proficiency (participation) target.
[24] In 2003 and 2004, Iowa tested students in grade 4, 8, and 11.

27% of schools that Iowa labeled as making AYP should not have made AYP by our calculations due to either: (a) a subgroup with a participation rate below 95%, or (b) a subgroup with a proficiency rate too low to meet the required targets, even after considering safe harbor and the most generous possible confidence interval adjustment.[25]

b. *Alternative Assessments.* Because the students-with-disabilities subgroups' performances were often the only reason for a school failing to make AYP, states' policies toward these subgroups have substantial ramifications. NCLB requires states to incorporate nearly all special education students' scores on regular, grade-level assessments in AYP determinations. Student scores on alternative assessments can account for no more than 1% of a school's total scores. Texas state officials petitioned to "phase-in" the 1% rule over time, but the U.S. DOE denied their request. In 2003, the Texas State Education Agency ignored the U.S. DOE's ruling and approved the appeals of 1,718 schools whose special education subgroup failed due to NCLB's 1% rule. These approvals prevented the failure of 22% of Texas schools (Hoff, 2005). In 2004, the U.S. DOE issued new guidance allowing states to petition to raise the 1% limit; in 2007, the U.S. DOE raised this limit from 1% to 2% (U.S. DOE, 2007).

[25] This 27% estimate is actually conservative because we lack data on the size of Iowa's student subgroups. We apply the confidence interval formula by setting the subgroup size to 30, the minimum size for holding a subgroup accountable in Iowa. The actual, larger N's would yield smaller confidence intervals, so we may be overstating the number of subgroups that should have made AYP.

c. *Applying a large confidence interval to safe harbor calculations.* NCLB gives states the option of applying these safe harbor calculations, as well as a further option to apply a 75% confidence interval to safe harbor calculations. Fourteen states incorporated this safe harbor confidence interval as allowed. Louisiana and Massachusetts, however, applied confidence intervals that were more generous than allowed – Louisiana employed a 99% confidence interval and Massachusetts employed a 95% confidence interval. In Louisiana, this added increment helped more than 62% of otherwise failing economically disadvantaged subgroups, 79% of otherwise failing Black subgroups, and 90% of otherwise failing students-with-disabilities subgroups avoid failing status.[26] Applying such a wide confidence interval adjustment to a safe harbor rule even allows some subgroups to make AYP when their proficiency rates *fell* instead of rose from the prior year. For example, the 31 fourth graders at McDonogh Elementary School #7 in Orleans Parish, LA, had a proficiency rate of 20% in ELA on state exams in 2002, which fell to 16.1% for the fourth graders in the same school in 2003. This 2003 performance failed to meet both the AYP ELA target of 36.9% and the lower target established by the confidence interval adjustment. To qualify for safe harbor without a confidence interval adjustment, the fourth grade group would need a 28% proficiency rate in 2003, representing a 10% reduction in the prior year's 80% failure rate. Louisiana's 99% confidence interval

[26] Reported figures are for math performance in 2003. The analogous figures for ELA performance are 49%, 57%, and 90%, respectively.

applied to this 28% target, however, set the safe harbor target rate at 7%, meaning the fourth grade 2003 proficiency rate could have met Louisiana's safe harbor criteria even if its proficiency rate was as low of 7%. The extremely generous confidence intervals applied to the safe harbor rule allowed McDonogh to make AYP even though its proficiency rate had actually declined by 4 percentage points.

- *States use more and less generous confidence interval adjustments*. States varied in the generosity of the confidence interval rules they adopted—ranging from no confidence intervals to 90, 95, or even 99%. States can reduce school failure rates by using larger confidence interval adjustments. As shown in Table 2, twenty-three states opted to use the maximum 99% confidence intervals. This typically meant that they used a 2.33 critical value, meaning a subgroup would still make AYP if their proficiency rate was within 2 times the standard deviation of the target proficiency rate. Yet failure rates in states with 99% confidence intervals were not substantially different from those in the fourteen states using 95% confidence intervals; in fact, the average state failure rate across 2004 and 2005 was slightly higher for the states using 99% confidence intervals (24% versus 21%).[27] The interaction of the other AYP decisions about continuous enrollment, minimum subgroup size, tested grade levels, and baseline proficiency rates helps to explain this counterintuitive result.

[27] For these calculations, we only include states that used standard confidence interval adjustments applied to both student subgroups and the overall student population.

At the other end of the spectrum, four states did not employ any confidence interval adjustment at all—Florida, Ohio, South Carolina, and Virginia—and this dramatically increased their school failure rates as a result. The average failure rate in these states was 57% in 2003 and 44% in 2004. Florida identified over 80% of its schools as failing AYP in 2003. If Florida had instead applied even a 95% confidence interval that year, we estimate that 14% of its schools failing to meet proficiency targets would have instead made AYP.[28] Michigan applied 99% confidence interval adjustments but only for schools with very small campus-wide enrollments. If Michigan had instead applied 99% adjustments to all of its schools in 2004, we estimate that the percent of its schools failing to meet at least one proficiency target would have declined from 19% to 5%.

Some states altered their school failure rates by adjusting confidence interval policies over time. During the first two years of NCLB, South Carolina did not employ confidence interval adjustments on either absolute subgroup proficiency rates or safe harbor calculations. In 2005, South Carolina amended its accountability system to include a one standard error band adjustment (i.e., a 68% confidence interval adjustment), and the proportion of schools failing to make AYP in South Carolina promptly fell by ten percentage points from the prior year.

[28] Florida also had low cutoffs for minimum subgroup size. Their subgroups for limited English proficient students, students with disabilities, and Black students had relatively low proficiency rates and were frequently held accountable: in 2003, these groups were accountable for math performance in 27%, 80%, and 68% of schools respectively. Florida's schools thus failed frequently and only 11% of them had at least one subgroup pass via safe harbor.

Confidence intervals applied to safe harbor were another important source of cross-state variation in failure rates. By 2004, all but one state – Alabama[29] - employed safe harbor calculations. Yet sixteen states applied confidence intervals to their safe harbor calculations[30], and the other states did not. As discussed above, Louisiana and Massachusetts applied improperly large confidence intervals to safe harbor calculations, whereas fourteen other states applied the permitted 75% confidence interval to safe harbor calculations. Polikoff & Wrabel (2013) describe how the number of schools making AYP due to safe harbor has increased over time in California, one of the states applying a 75% confidence interval to its safe harbor calculations.

- *Some states adopt homogenous targets across grade levels whereas others do not.* As mentioned earlier, states were allowed to set grade-specific, subject-specific proficiency rate targets or could set uniform targets across grade levels and subjects. In most states, high school student proficiency rates were lower than those in younger grade levels. Because proficiency targets were based on pre-NCLB performance levels, states setting uniform targets may have thus been setting up relatively easy targets for elementary and middle schools to reach—particularly if high school students' proficiency rates lagged far behind. Twenty-three states employed this policy.[31] Of these, Texas and Pennsylvania provide examples of states with lagging high school proficiency rates. In 2002, the

[29] Alabama employed safe harbor adjustments in 2005.
[30] The postal abbreviations for these sixteen states are: AK, CA, CT, DE, KS, LA, MA, ME, MO, NJ, NV, PA, SD, UT, WI, and WY.
[31] The postal abbreviations for the twenty-three states with homogenous targets across grade levels are: AK, CA, CT, DE, FL, HI, ID, IL, IN, LA, MA, MO, MT, NH, NY, OK, OR, PA, SC, TN, TX, VA, and WI.

proficiency rates in both Texas and Pennsylvania were at least 7 percentage points greater in elementary schools than in high schools for both ELA and math. These states' decision to use uniform targets across grade levels led to low failure rates among elementary schools. For Texas in 2004, only 1% of elementary schools failed to make AYP, 17% of high schools failed, and the overall failure rate was 6% of schools. Similarly, for Pennsylvania, only 7% of elementary schools failed to make AYP, 27% of high schools failed, and the overall failure rate was 15% of schools.

Setting a more easily obtained proficiency rate target for elementary and middle schools relative to high schools can lower states' school failure rates for both computational and meaningful reasons. On the purely computational side, high schools are larger and less numerous than elementary schools, so a relatively low elementary school failure rate means a low proportion of *schools* failing AYP even though the proportion of *students in schools* failing AYP may be much higher. But on a more substantive note, given the safe harbor policy, having fewer schools close to the margin for meeting their student proficiency rate targets can decrease school failure rates. Schools that expect to perform close to their proficiency rate targets do not benefit from a safe harbor policy—if their proficiency rates improve from the prior year then they would already be meeting their proficiency targets without using safe harbor. Safe harbor is more likely to enable schools to make AYP if schools' proficiency rates are nowhere near the targets to begin with. So, all else equal, states will have lower school failure rates if they have more elementary and middle schools that will easily meet their proficiency targets even if they also have more high schools that are nowhere near

these targets, since some of these high schools might still meet AYP via safe harbor.

South Carolina was operating an interim accountability system in the initial year of NCLB that provides a counterexample to Texas and Pennsylvania. South Carolina applied pre-NCLB proficiency rates of students in grades 3 to 8 to elementary, middle, and high schools, because South Carolina had not yet calculated high school proficiency rates for a sufficient number of prior years. Fewer students scored proficient or above in high schools than in elementary or middle schools, so applying the grades 3-8 proficiency rate as a baseline caused 97% of South Carolina's high schools to fail AYP in 2003. When separate targets were established for high schools in 2004, the high school failure rate decreased to 52%.

- *States established different minimum subgroup sizes and held a different number of subgroups accountable.* The all or nothing nature of the AYP designations increases the risk of failure for schools with greater numbers of accountable student subgroups (Kane and Staiger, 2002, 2003; Simms, 2013). Within states, schools with a greater number of accountable subgroups were indeed more likely to fail AYP. Across states, there is a mild correlation between schools' average number of accountable student groups and their failure rates. Figure 4 displays this comparison for 2004. If we regress failure rates on the number of accountable student groups and this variable squared, then this produces an R-squared of less than .07 and the joint significance is .23.

  But Figure 4 also reveals that this relationship would have been stronger if not for a few outliers—the low failure rates in Louisiana, Montana, and Texas. With these

three outlier states omitted, the R-squared from the quadratic term regression jumps to .14, with a joint significance of .05.[32] The other policy implementation decisions described above created exceptionally low failure rates in these three states. Louisiana had low cutoffs for minimum subgroup size and thus had a larger number of accountable subgroups per school, but Louisiana used wide confidence intervals that, in combination with small subgroup sizes, made the effective proficiency target quite low. Texas used a uniform proficiency target across grade levels, resulting in extremely low failure rates among its elementary and middle schools. Montana did not use any minimum subgroup size, so subgroups would technically be held accountable even if there was only one student in that group. However, Montana's small schools and 95% confidence interval policy meant that subgroups were so small that they would make AYP even with few students passing.

Because the performance of the students-with-disabilities subgroup was often the only reason for a school failing to make AYP, one might expect states' policies toward this subgroup to influence their schools' failure rates. The fraction of schools with accountable subgroups will depend not only on states' minimum subgroup size rules but also on how they allocated students with disabilities across schools. School failure rates were initially higher in states with larger fractions of schools with accountable students-with-disabilities subgroups. If we regress state failure rates on a quadratic for the fraction of schools where these subgroups were accountable for math performance in 2003, then the R-squared is .13, with joint significance of .09 and

[32] The adjusted R-squared increases from .02 to .10 when these three states are omitted.

adjusted R-squared of .08. But this relationship disappeared by 2004: the R-squared declined to .02, the joint significance was .70, and the adjusted R-squared was negative. States with higher fractions of accountable students-with-disabilities subgroups tended to mitigate this effect by having more generous confidence interval adjustments. In 2004, five of the eight states with the highest fractions of schools holding these subgroups accountable for math performance used 99% confidence interval adjustments.

- *States defined continuous enrollment differently.* Five states—Hawaii, Illinois, Iowa, New Jersey, and Wisconsin—used starting dates for continuously enrolled students that precede September of the school year of the testing.[33] In these states, students who have transferred schools prior to the first day of the school year will not affect their schools' AYP determinations. Two of these states, Hawaii and Wisconsin, chose early enrollment cutoff dates because they test students during fall months. If mobile students tended to be relatively low achieving or if school districts tended to strategically wait to enroll students at particular schools (Jennings and Crosta, 2011), then these long required enrollment windows would make it easier for schools to make AYP. Hawaii already had a high failure rate in 2003 in spite of their early enrollment cutoff date, due to low participation rates and low proficiency rates in the subgroups of students with disabilities and students with limited English proficiency. The other four states may have had much

[33] We thank Jennifer Jennings and Heeju Sohn for providing information on states' rules for continuous enrollment and testing dates, collected from state government websites.

higher failure rates if they had used post-September enrollment cutoffs, since the fraction of students excluded from the accountable pool was sometimes quite high. In Wisconsin, for example, 14% of 4th grade students, 10% of 8th grade students, and 8% of 10th grade students were enrolled during test administration in November of 2003 but did not contribute to their schools' proficiency rate calculations, because they had not been enrolled in the same school since late September of 2002.

Discussion

The early years of NCLB provide an important example of how variation in state policy implementation can cause a federal law to have very different consequences across the country. Discrepancies in states' AYP formulae teach us that details have important ramifications. Complex and off the radar of all but the most embedded policymakers and researchers, esoteric differences in rules had substantive impacts on schools due to the escalating sanctions under NCLB. Purposefully or not, some states took advantage of loopholes that made it much easier for schools to meet targets. Variation in these rules has only increased in recent years, as some states have received waivers allowing their schools to avoid failure designations even if their students do not reach 100% proficiency by 2014 (Riddle & Kober, 2012; U.S. DOE, 2012). These waivers are idiosyncratic to each state, so that cross-state variation in the minutia of accountability policy rules is as complicated and important as ever (Polikoff et al., 2014).

While flexibility may be a positive aspect of NCLB or other school accountability systems, many of the discrepancies in states' NCLB rules reflect arbitrary differences in statistical formulae rather than substantive policy disagreements. When states and districts

design test-based accountability policies, schools may be best served by a consistent set of directions about acceptable statistical practices and common definitions. The federal government could convene a panel of experts or commission a professional association such as the American Statistical Association to provide guidance on sound statistical practices related to confidence interval setting, safe harbor exceptions, and minimum subgroup sizes. Formulae for these procedures, if used, could then be standardized. These formulas themselves attempt to adjust evaluation to treat schools in a fair and just manner. Standardizing rules for exceptions and adjustments does not eliminate this quest for fairness. Rather, using uniform accounting practices might promote transparency and better insulate state accountability systems from the political whims of governors and state legislatures. While our own analysis does not investigate whether arbitrary differences across states were harmful, we are hard-pressed to think of a compelling reason why citizens should prefer these arbitrary differences in accounting.

Even after statistical definitions are standardized, school accountability policies could still provide states and districts with discretion in their *substantive* choices of how to measure school effectiveness and which sanctions or rewards to attach to performance outcomes. Ideally, consequences for schools in an accountability system should be linked to student learning rather than the idiosyncrasies of state rules. This ideal might be better served if the federal government offered states a selection from a menu of accountability systems, while maintaining precise definitions and formulae within each of these systems.

CHAPTER TWO – CLOSING UP SHOP: SCHOOL CLOSURE IN THE CONTEXT OF NO CHILD LEFT BEHIND

Introduction

Throughout the U.S., low-income and minority students fall short of national proficiency standards in math and reading at high rates. Yet targeted efforts at raising proficiency rates among minority students and eliminating achievement gaps between socioeconomic and racial/ethnic groups seem ineffective (Hanushek, 1986; Ladd, 1999; Hanushek & Raymond, 2005; Reardon, 2011). In response, in 2002, Congress passed the No Child Left Behind (NCLB) Act, which required that states and districts conduct annual assessments of their public school students in order to determine whether schools were making sufficient progress in improving the performance of their students, particularly low-income and minority students. For those schools that failed to make progress each year, NCLB rules prescribed a set of consequences which became increasingly severe each year schools' low-performance persisted. In addition, states only had until 2014 to drastically improve their lowest-performing schools and bring them up to NCLB standards. I posit that the threat and implementation of these consequences as well as the abbreviated timeline for improvement may have encouraged districts and states to use school closure to address schools' persistent low-performance. In this paper, I investigate the extent to which schools' NCLB performance ratings predict their likelihood of closure.

Prior to the onset of NCLB, districts rarely considered student performance as a motivation for closing a school; instead, districts closed schools due to declining enrollment, decrepit facilities, or budget constraints (Boyd, 1979; Burlingame, 1979; Colton & Frelich, 1979;

Valencia, 1984). Recently, researchers have uncovered an increase in school closure decisions based on schools' chronic underperformance (De La Torre & Gwynne, 2009; Kirshner, Gaertner & Pozzoboni, 2009; Lipman & Haines, 2007; Steiner, 2009; Young, 2009; Brummet, 2014; Carlson & Lavertu, 2015; Kemple, 2016). Additionally, some urban districts—faced with closure decisions due to declining enrollments—designated student performance as the primary closure criterion. Districts like Chicago, Pittsburgh, and Cleveland met demand by preemptively closing their lowest performing schools (Enberg et al., 2011; PEW, 2011; Carlson & Lavertu, 2015). This evidence suggests that some states are likely considering schools' NCLB performance in closure decisions, but these reports have only a narrow view on national trends and do not pay specific attention to schools' NCLB rating.

The implementation of NCLB provided states and districts with new information about school performance that state and district leaders could use to inform their overall school improvement strategy, including school closure. Schools that met predetermined performance targets were designated as making Adequate Yearly Progress (AYP). Schools that did not meet their targets failed to meet AYP. Multiple years of AYP failure meant that states and districts must intervene to improve performance, and NCLB rules prescribed these interventions, or consequences. In response to these consequences or in an effort to avoid their costly implementation, states and districts may have used school closure as an intervention for schools that persistently under-performed.

State policy environments may also impact a school's likelihood of closure. In states where their political presence is strong, teachers' union labor contracts and political tactics influence many state and district decisions related to school operations, staffing, and financing.

Instead, teachers' union leaders advocate for improvement through reductions in class size, increases in teacher salaries, and improvements in school conditions (Gadlin, 2013; Santos, 2011; UFT, 2016; NEA, 2008). Teachers' union political influence, therefore, may have limited the ability of state and district leaders to use school closure as a mechanism for school reform. Therefore, in addition to schools' AYP performance, I also explore the impact of the presence of teachers' unions on schools' likelihood of closure.

Many factors may impact the relationship between schools' NCLB performance and their likelihood of closure—the complexity of NCLB rules and states' implementation of those rules may give inconsistent signals about school quality, unrelated or adverse school or district conditions may necessitate closure regardless of student performance, or states' own policy environments may mitigate or amplify state reform efforts. To explore these factors, I use national datasets that contain information on whether schools "passed" or "failed" AYP from 2003-2005 as well as school characteristics from 2002-2011. I find that schools' performance in the early years of NCLB is highly predictive of school closure and that each additional year of NCLB "failure" increased the likelihood a school is closed. Yet states' individual policy environments matter—the effect of a school's NCLB "failure" on school closure appears to be mitigated by the strong presence of teachers' unions in that state. All else equal, "failing" schools in states with a strong union presence were less likely to close than "failing" schools in states' with a weaker union presence. Finally, controlling for schools' relative performance in the state does reduce the size of the effect of failing AYP on schools' likelihood of closure, but the effect remains substantive and significant.

To my knowledge, this is the first national study that catalogues school closure in the era of NCLB. Further, no national study exists that examines the relationship between schools' AYP failure and subsequent closure. Several important points underscore any rigorous analysis of NCLB ratings' impact on the likelihood of school closure. States differed both in the ways they implemented NCLB and in their dealings with low-performing schools prior to NCLB. State implementation of NCLB's regulations led to extreme cross-state variation in the difficulty schools faced in meeting AYP performance targets. A school that failed AYP in one state may have passed AYP in another. Thus, the strength of the signal of schools' AYP status varies by state. Further, many states had existing accountability systems that pre-dated NCLB and often held schools accountable to the rules of these systems and NCLB simultaneously. Existing accountability systems may have incentivized districts to close low-performing schools prior to the onset of NCLB regulations. Perhaps districts would have closed these same schools in the absence of NCLB. This obfuscation makes it difficult to identify the exact mechanism through which districts made school closure decisions. My findings suggest, however, that the visibility and stigma of schools' AYP status may have provided districts with the political leverage needed to close schools that persistently failed AYP.

This paper proceeds as follows: In Section II, I provide a brief overview and background of NCLB and school closure. Section III provides descriptions of each dataset and the methods I use to construct the variables of interest. Section IV and V report descriptive statistics and the methodology I implement, respectively. My findings are reported in Section VI, and Section VII concludes.

Background

On the whole, there is some convincing evidence that the accountability pressure established by NCLB led to meaningful improvements in student outcomes (Ahn & Vigdor, 2014; Springer, 2008; Dee & Jacob, 2011, Lauen & Gaddis, 2014; Reback, 2008; Rockoff & Turner, 2010). Dee and Jacob (2011) find that the onset of NCLB in 2003 led to significant improvements in student performance on the math portion between the 1992 and 2007 National Assessment of Education Progress (NAEP), and that the math gains were greatest in those states that did not have formal accountability systems prior to NCLB. These gains were mostly concentrated among low-income students and students in the lower half of the achievement distribution. Lauen and Gaddis (2014) find compelling evidence that basing AYP determinations on subgroup-level performance, rather than whole-school performance alone, led to test score gains for racial minority students and students eligible for free- and reduced-priced lunch in North Carolina. These gains were greater for schools near the bottom of achievement distribution than for schools closer to the margin of making AYP.

As typically seen with most complex policy design and implementation, some unintended negative consequences accompanied any realized improvements in student outcomes. NCLB implementation varied considerably across states which led to extreme variation in states' school AYP failure rates. In the early years of implementation, there is some evidence that states gamed the system to reduce the number of schools that failed AYP (Davidson, 2013). While Reback, Rockoff, and Schwartz (2011) find neutral or positive effects of NCLB on student achievement, they find that teachers[34] in schools at-risk of failing AYP

[34] The findings were especially pronounced for teachers of math and reading.

experience greater worry about their job security and expect to leave the teach profession sooner than teachers in schools with little risk of failing AYP. There is compelling evidence that schools' focused more on tested grade-levels and subject-areas to the detriment of non-testing grades and subjects, like art and social studies (Reback, et al., 2011). Similarly, some evidence suggests that schools may target interventions to students on the margin of scoring proficient while neglecting students at the low and high end of the test score distribution (Neal & Schanzenbach, 2010; Booher-Jennings, 2005). Finally, when poor performance leads to severe consequences, some schools and districts have resorted to cheating to avoid consequences and their associated stigma (Aviv, 2014).

Many policy-makers argued that, even though NCLB may have led to small improvements in student outcomes, some NCLB components were problematic and refinements to these components would likely to lead to even greater improvements. First, NCLB's principal goal of all schools reaching 100 percent proficient by 2014 was unrealistic (Linn, 2003; Manna, 2010; Hess & Finn, 2007; Darling-Hammond, 2006). In addition, researchers and policy-makers argued that, to measure school improvements, NCLB methodology should shift from static proficiency rates, which are based on the performance of different groups of students each year, to measures of student growth, which track the performance of the same group of students over time (Ladd & Lauen, 2010; Hanushek & Raymond, 2005; Ladd & Walsh 2002; Toch & Harris, 2008). Similarly, some practitioners argued that NCLB's measures were too narrowly defined which led educators to focus on a narrow set of activities or outcomes (Dee et al., 2012). Instead, some practitioners and policy-makers felt that accountability systems should expand the measures that are included in schools'

performance ratings. Many policy-makers agreed that binary performance ratings (i.e., passing AYP or failing AYP) lacked the nuance necessary to appropriately target interventions and prioritize and tailor supports (USDOE, 2008).[35]

In response to these concerns and Congressional inaction on the reauthorization of ESEA, the USDOE allowed states to submit waivers to circumvent or refine specific components of NCLB. More recently, the Obama administration introduced the Race to the Top (RTT) initiative as part of the American Recovery and Reinvestment Act (ARRA) of 2009. RTT represented a concerted effort to shift the education reform discourse from NCLB's top-down federal control and forced compliance to incentives for improvement through competitive grants. In theory, RTT was designed to avoid the implementation problems of NCLB by "relying on incentives instead of sanctions to drive state reform" (McGuinn, 2012, p. 138). Appendix A includes a fuller discussion of these federal policy shifts, both of which explicitly included school closure as a consequence or intervention for schools whose students are persistently low-performing.

School Closure in the U.S.

While there is some evidence that NCLB has positively impacted student outcomes, there are no national studies that document the scope of school closures or analyze the impact of school closure on students. This paper attempts the former, but for the latter, the complexity of identification and data collection make such an analysis difficult. Selection issues make

[35] In Chapter I, my coauthors and I argue that additional complexity in performance determinations would likely lead to greater variation in implementation and diminish the comparability of ratings across districts and states (Davidson, 2013).

investigations into the effects of school closure on student performance difficult because students in schools that close due to low student performance are not randomly assigned. Even after controlling for student background characteristics, students in schools that persistently failed AYP may be different in immeasurable ways from similar students who attend schools that never failed AYP. To date, much of the research in this area has focused on particularly cities or regions, and some preliminary evidence has emerged that suggests that, in some cases, school closure can be an effective strategy to improve future student performance for students in closed schools with persistently low-performing students. Although this paper does not focus on the impact of closure on future student achievement, the size and direction of these effects are important context when exploring the scale and reasons for closure nationally.

The limited evidence on the effects of school closure on student performance is mixed. On average, school closure appears to have either a neutral or negative impact on student achievement in the year of the closure announcement as well as up to three years after the school closes (De la Torre & Gwynne, 2009; Enberg, et al., 2011; Ozek et al., 2012; Kirshner et al., 2009; Larsen, 2014). Ozek et al., (2012) find significant declines of 0.10 standard deviations in reading and between 0.12 – 0.20 standard deviations in math in the years following school closure. These estimates suggest that school closure leads to more than a 20 to 30 percent loss of a year's learning. De la Torre and Gwynne (2009) find that student performance effects are "neither negative nor positive" for 1, 2, and 3 years after transfer. Yet, there is some compelling evidence that student test scores decline either in the year preceding school closure or the months following the closure announcement (Kirshner, at al., 2009; De La Torre & Gwynne, 2009; Ozek, et al., 2012). De La Torre and Gwynne (2009) find that, compared to their matched

control group peers, student reading scores in imminently closing schools are 1.5 months below their expected level. Similarly, Ozek et al., also find significant declines of 0.19 standard deviations in student math scores and 0.11 standard deviations in reading scores immediately following the closure announcement ($p < 0.01$).[36]

While the average effects are neutral or positive, students in closed schools who, after closure, attend schools with higher-performing students experience significant and persistent increases in outcomes. Transferring to a school with higher-performing students has significant and lasting positive effects on student performance (Kirshner, et al., 2009; De la Torre & Gwynne, 2009; Kemple, 2016). In contrast, students who transferred to schools whose average student proficiency rates are in the bottom quartile experienced a losses equivalent to over a month in reading and half a month in math (De la Torre & Gwynne, 2009). Yet, the evidence suggests that there is no guarantee that students from closed schools will attend schools with higher-performing students, and assigning students to receiving schools does not ensure that they will attend them (Enberg, et al., 2011; Kirshner, et al., 2009, Ozek, et al., 2012). In response to 18 elementary and middle school closures in Chicago, around 40 percent of the 5,445 affected students transferred to schools on academic probation and around 42 percent of

[36] The extent of students' displacement after school closure may impact the difficulty of their transition and, perhaps, their future performance. Nearly all districts in this review transferred their students to new schools instead of re-opening new schools in the existing buildings. In Chicago, 17 of the 38 schools that closed were not re-opened as new schools by 2009 (De La Torre & Gwynne, 2009). Even in cases where those buildings were reused, districts usually delayed at least one school year before opening a new school in the building. Many districts continue to have vacant school properties due to enrollment declines and school reconfiguration. Some districts are leasing buildings to charters or selling them. As of 2011, Milwaukee and Kansas City had 27 and 39 vacant buildings, respectively, while Pittsburg had 18. In 2011, Detroit listed 92 school buildings for sale and reported generating $5 million in revenue from sales and leases of buildings (PEW, 2011).

transferred to schools with average student proficiency rates in the bottom quartile of the district (De la Torre & Gwynne, 2009).

School closure also effects various other student outcomes—e.g., student mobility, attendance, and graduation rates. Students whose schools close are more likely than their peers to leave their new school during and after the next school year even though they had lower rates of mobility prior to closure (De la Torre & Gwynne, 2009; Ozek, et al., 2012). On average, high student mobility negatively impacts student outcomes (Rumberger, 2003; Kerbow, 1996; Temple & Reynolds, 1999; Hanushek, et al., 2004; Xu et al., 2009). There is some evidence that school closure has significant negative effects on graduation, dropout, and attendance rates, but these findings are weak and may suffer from selection effects (Kirshner, at al., 2009; Enberg, et al., 2011).

Data

To explore the relationship between AYP failure and school closure, I use data on the characteristics of all U.S. public schools in school years 2001-2002 to 2010-2011[37] from the Common Core of Data (CCD), a 50-state dataset compiled from each state by the National Center for Educational Statistics (NCES). I limit my sample to those 85,619 U.S. schools designated as "regular" by the CCD across the included years.[38] The data include annual information on school demographics, total student enrollment, and other characteristics, such as grades served, geographic location, student/teacher ratio, and federal Title I eligibility. Due to the complexity of NCES' data collection process, a limited amount of CCD school

[37] Throughout this paper, I refer to school years using the year of the spring term.
[38] The CCD characterizes schools as "regular," "other/alternative," "special education," and "vocational."

demographics data is missing or inaccurate in particular years. See Appendix B for a description of the adjustments I made to account for these gaps and errors.

The core of my analysis relies on identifying which schools closed between school years 2003 and 2011. For each school, the CCD data includes federally assigned unique school identifiers (IDs). The IDs are never duplicated—new schools are assigned unique ids, and closed schools' IDs disappear from the reported data. I use these unique IDs to link each year's CCD data file to the next which allows me to flag those school IDs that disappear from the data between years and never reappear in subsequent years.[39] If a school ID disappears between years, I infer that the school closed. To be included in my sample, a school must be "open" or appear in the 2002 CCD data file. To account for errors in federal data reporting (i.e., a state mistakenly fails to report a school in one year only to add them to the reported data in the following year), I only categorize schools as "closed" if the school ID does not appear in any subsequent year. This approach also excludes schools that temporarily close and re-open in a future year. I do not categorize schools as closed that have "restructured" or changed their staff or governance but still operate under the same federally assigned ID. Conversely, it may be the case that a school ID disappears from the data because the school was closed to facilitate a restructuring process or a change in governance, but a "new" school immediately re-opens in the next school year in the same building. Under my current approach, I am unable to distinguish this "closure" from a more permanent or distinct school closure.

The NCES also provides extensive school district-level data through two databases. First, I use 2000 district-level Census data from the publicly available NCES' School District

[39] For New York State, I use state assigned school IDs instead of federally assigned NCES school IDs.

Demographic System (SDDS) to describe the demographic characteristics of school districts' residents. I include measures of residents' education-level, median income, and racial composition. Second, NCES's 2003 restricted-use Schools and Staffing Survey (SASS) reports an indicator of whether districts' teachers are covered by either labor unions or collective bargaining agreements. For each state, I use this district-level indicator and the SASS sampling weights to estimate the percentage of states' districts where any form of collective bargaining occurs.

School- and subgroup-level NCLB performance data are available nationally for school years 2003, 2004, and 2005 (Reback, et al., 2011). For 2003 and 2004, I use indicators of whether schools passed or failed Adequate Yearly Progress (AYP) from state departments of education. When state departments were unable to provide AYP determinations, I supplemented the national dataset with AYP performance indicators from two incomplete public datasets—the American Institutes for Research (AIR) National AYP and Identification Database (NAYPI) and the Council of Chief State School Officers' School Data Direct (SDD). For 2005, AYP performance indicators come exclusively from the national dataset compiled by AIR. For many states in 2003 and most states in 2004, I also use school- and subgroup-level proficiency rates on states' standardized exams which were also provided by state departments of education.[40] Compared to AYP determinations, school average proficiency rates are a more straightforward measure of the performance of schools' students and may be an equally important signal to state and district officials in making closure decisions.

[40] State proficiency rates, like states' school failure rates, depend on state-specific NCLB implementation. States vary in their definition of proficiency standards and the difficulty of their state exams.

Descriptive Trends

Table 3 lists the descriptive statistics of the 85,619 schools included in the sample. All data in Table 3 is from school year 2002 unless otherwise indicated. Nationally, 42 percent of all public schools failed AYP at least once, and 12 percent failed AYP all three years for which I have national data. The rate of school closure between 2004 and 2011 represents a substantial proportion of U.S. schools. Nearly nine percent of US schools (n=6,458 schools) closed between 2004 and 2011. In this paper, I typically examine the effect of AYP failure between 2003 and 2005 on schools' likelihood of closing immediately after 2005, or between 2006 and 2011. Six percent of U.S. schools closed during this period and, perhaps unsurprisingly, the rate of school closure varies fairly significantly across states. Figure 5 displays the 2004-2011 school closure rate for individual states. Washington D.C. had the highest closure rate during this period, closing more than 25 percent of its schools. While New York City experienced many school closures during this period, New York State only closed just over 5 percent of its schools. The state with the smallest share of closures was Hawaii—less than 1 percent of Hawaii's schools closed.

My "sample" includes nearly all public schools in the U.S. In 2002, U.S. public schools had, on average, nearly equal shares of Black and Hispanic students—17 percent and 15 percent, respectively. Forty-one percent of U.S. students were economically disadvantaged, defined by their eligibility for free- or reduced-priced lunch. Unsurprisingly, the majority (61 percent) of U.S. schools serve elementary students. Only 17 percent and 18 percent of schools serve high school and middle school students, respectively. The average enrollment at U.S. schools was 569 students and between 2002 and 2006, U.S. schools maintained steady levels of

enrollment, on average. Each state uniquely determines student proficiency on state exams, but nationally, about 50 percent of students were designated by their states as "proficient" in math and reading. Finally, labor unions and collective bargaining contracts are in place in the majority of U.S. school districts: on average, 70 percent of states' districts are covered by collective bargaining or labor unions, and 33 states have more than 75 percent of their districts covered by collective bargaining or labor unions.

Table 4 lists descriptive characteristics across schools that closed and schools that remained open. AYP failure rates were higher for closed schools than for schools that remained open—closed schools were significantly more likely to have failed AYP at least once, at least twice, and all three years between 2003 and 2005. Nearly 50 percent of closed schools failed AYP at least once compared to 41 percent of schools that remained open. Closed schools were significantly more likely to have experienced declines in enrollment between 2002 and 2006 than schools that remained open. Closed schools' enrollments declined by more than seven percent during this period. In contrast, schools that remained open essentially maintained their enrollments over this time period. Interestingly, schools that closed were significantly more likely to be in states with larger shares of districts covered by collective bargaining agreements. Seventy-three percent of closed schools were located in states with more than 75 percent of their districts covered by collective bargaining agreements while only 66 percent of schools that remained open were located in these states.

Compared to schools that remained open, closed schools enrolled significantly higher shares of Black students and students eligible for free or reduced-priced lunch but significantly lower shares of white students. Twenty-eight percent of students in closed schools were Black

compared to only 15 percent in schools that remained open; fifty-two percent of students in closed schools were eligible for free- and reduced-priced lunch compared to only 40 percent in schools that remained open. In contrast, white students account for 56 percent of students in closed schools compared to 65 percent in schools that remained open. Closed schools were much more likely than schools that remained open to be located in cities and much less likely to be in suburbs. Thirty-six percent of closed schools were located in cities compared to only 24 percent of schools that remained open. On the other hand, 27 percent of closed schools were located in a suburb compared to 36 percent of schools that remained open. Closed schools are only slightly more likely to be located in rural areas than schools that remained open, yet closed schools and schools that remained open are equally as likely to be located in towns.

Schools that served high school grades were more likely to have remained opened—19 percent of schools that remained open were high schools whereas 12 percent of closed schools were high schools. Schools that closed were more likely to be elementary and middle schools. Schools that served elementary grades comprise 66 percent of closed schools and 63 percent of schools that remained open. The larger share of elementary schools among schools that closed may partly explain why closed schools have fewer students, on average, compared to schools that remained open. Significant differences also exist for middle schools—24 percent of closed schools were middle schools while 20 percent of schools that remained open were middle schools.

Schools that closed and schools that remained open also differ in terms of average student enrollment and district-level characteristics. Schools that closed enrolled 229 fewer students, on average, than schools that remained open. Between 2002 and 2006, enrollment in

closed schools declined by 7.3 percent. Schools that remained open experienced an increase in enrollment of 0.2 percent during the same period. Census data reveal differences in the demographics of school districts in which schools closed and school districts in which schools remained open. Schools that closed are in districts where, on average, the median income is 13 percent lower than the median income of districts of schools that remained open. There are also differences between the two groups in the educational attainment of districts' residents. Districts of closed schools have smaller shares of adults with high school diplomas and bachelor's degrees. On average, districts of closed schools have 14 percent fewer residents with bachelor's degrees, but only about 2 percent fewer residents with high school diplomas. Compared to districts of schools that remained open, districts of closed schools have smaller shares of white and Hispanic residents and larger shares of Black residents, on average.

Table 5 displays the characteristics of school by their frequency of AYP failure between 2003 and 2005. Schools that failed AYP at least once during this period were significantly different from schools that never failed AYP. Failing schools have larger average student enrollments, serve higher rates of poor and Black and Hispanic students, and are more likely to be located in urban areas. On average, the share of black students in schools that failed AYP all three years is nearly three times the share of black students in schools that never failed AYP—32 percent in schools that failed all three years and 11 percent in schools that never failed. The share of Hispanic students in schools that failed AYP all three years is twice the share of Hispanic students in schools that never failed AYP. Interestingly, even though the closure rate is higher as schools failed more often, average school enrollment is also higher as schools failed

more often—this trend is in contrast to my earlier report that closed schools, on average, enroll significantly fewer students than schools that remained open.

As schools fail more frequently during the period, differences between failing schools and non-failing schools grow larger. Failing schools are, in all cases, nearly twice as likely to serve middle and high school grades as never failing schools. Twenty-nine percent of schools that failed all three years were high schools while only 15 percent of schools that never failed AYP were high schools. Similarly, 72 percent of schools that never failed AYP were primary schools while only 36 percent of schools that failed AYP all three years were primary schools. Across schools that failed and never failed AYP, there are large differences in the share of students who are eligible for free- and reduced-priced lunch, yet there are fairly small differences in the percentage of these schools that are eligible for Title I. District characteristics of failing schools are also different than district characteristics of never failing schools. Failing schools' districts have lower median incomes, larger shares of Black and Hispanic residents, smaller shares of white residents, and smaller shares of residents with High School or college diplomas.

Between 2006 and 2011, closure rates for schools that failed AYP at least once exceeded closure rates for schools that never failed AYP. Nearly five percent of the 37,841 never-failing schools closed between 2006 and 2011 while nearly eight percent of the 8,132 schools that failed AYP all three years closed between 2006 and 2011. Figure 6 shows the annual closure rates for schools that never failed AYP and schools that failed at least once for each individual year between 2004 and 2011. The 2004 closure rate for schools that never failed AYP was slightly higher than for schools that failed AYP at least once. But between 2005 and 2009, the

gap between schools that never failed and schools that failed at least once widened, with the widest gap between the two groups in 2006. In 2009, the closure rate for schools that failed AYP at least once peaked at more than 1.8 percent of all U.S. schools. In the next year, the closure rate for schools that failed at least once sharply declined, dropping below the rate for schools that never failed. Figure 7 shows the cumulative closure rate for schools that never failed AYP and schools that failed AYP at least once. Across all years, schools that failed AYP at least once had a higher cumulative rate of closure than schools that never failed AYP.

Methods

NCLB legislation required that states and districts implement specific reforms, or "consequences," once schools fail AYP. After multiple years of consecutive AYP failure, states and districts must implement substantive reforms like changes in governance or school restructuring. States and districts may decide to close schools in anticipation of continued failure and in an effort to avoid implementing costly consequences. Thus, schools' AYP status is likely related to their probability of closing. School closure rates vary across schools and districts both within-states and across states. So, I include a comprehensive set of school- and district-level characteristics that allow me to disentangle the effects of AYP failure from other factors associated with low student performance. Since the school closure literature indicates that school closure disproportionately impacts poor, Black, and Hispanic students (De La Torre & Gwynne, 2009; Kirshner, Gaertner & Pozzoboni, 2009; Enberg et al., 2011; PEW, 2011; Steiner, 2009), I control for the share of a school's students who are Black, Hispanic, Asian, and Native-American, as well as the share of students eligible for free- or reduced-priced lunch. I also

control for school district characteristics like racial composition, median income, and residents' education level.

School geographic location and grade-levels served may impact district closure decisions. Students in geographically isolated schools will likely face greater disruption due to closure than students in schools with easily accessible alternatives. I include, for each school, geographic indicators of school locale since schools' likelihood of closure may also depend on the relative size of the district.[41] Finally, districts may implement closure unequally across school levels, so I also control for the level of the school—high school, middle school, primary school and "other" grade configurations.

The school closure literature also suggests that states and districts close schools due to changing residential patterns that result in declines in enrollment. It is also possible that continued poor performance results in families leaving a school in order to enroll their students' in schools with higher-performing students. Therefore, I include a measure of the percent change in schools' enrollment between 2002 and 2006 to capture the effects of enrollment decline on schools' likelihood of closure. To minimize the effects of random fluctuations and data reporting errors in schools' annual enrollment data, I compare the average of schools' 2002, 2003, and 2004, total enrollment with schools' 2006 total enrollment to generate the percent change in enrollment measure.

Some significant differences may exist between state policies and demographics that could potentially confound the results of my analysis—e.g., states' own accountability systems, cross-state differences in the difficulty of making AYP, the extent of states' racial and economic

[41] I include indicators for whether schools are located in towns, cities, rural areas, or suburban areas.

segregation, state rules governing the existence and operation of charter schools, or states' allowances for collective bargaining. The inclusion of state fixed effects controls for any correlation between these state differences and schools' AYP failure. Even with the inclusion of these controls, however, the methods described herein only allow me to provide descriptive, not causal, evidence of a relationship between AYP failure and school closure. To make causal claims would require a quasi-experimental analysis that accounts for unobservable school and district characteristics that may influence school closure decisions.

Two conditions restrict the years of school closure consider in my analysis. First, states implemented NCLB during the 2002-2003 school year, so I can only attribute districts' closure decisions made after 2002-2003 (*t=2003)* to school AYP failure.[42] Second, districts typically released schools' AYP status during the late summer or even after the subsequent school year had begun. So, closure decisions based on schools' AYP determinations would lag one school year. I posit that schools' AYP status in year *t-1* predicts school closure during the summer of year *t.* Take, for instance, a school that closes immediately after the 2004 school year (summer 2004). I attribute this closure to the school's AYP performance in school year 2003 since it's likely the school was not designated as "failing" AYP until it was too late to decide to close the school and the 2004 school year had already begun. If a district closed a school in the middle of the 2004 school year in response to its 2003 AYP failure, then the designation in my data would be the same as if the district closed at the end of the 2004 school year. Federal data rules would

[42] For the remainder of this paper, I refer to school years by the Spring/Summer Year. 2002-2003 is 2003.

require the district report data for that school in the 2004 school year, so the 2004 CCD data would include the school's unique identifier but the 2005 CCD data would not.

A. Cumulative Effects

Each subsequent year of school failure inevitably moves schools further down NCLB's consequence timeline. Therefore, the cumulative effect of multiple years of AYP failure likely leads to districts facing greater incentives to remedy school failure. School closure is likely a remedy districts use. Here, I define $Close_{jq,t1\text{-}t2}$ as a school *j* in state *q* that was operational at least in 2002 and closed between t_1 and t_2. My baseline analysis estimates school-level probit regressions examining the effects of one, two, or three years of AYP failure between 2003 and 2005 on schools' likelihood of closing between 2006 and 2011.

$$(1.1) \quad Pr(Close_{jq,2006\text{-}2011} = 1) = \alpha + AYP_{jq,2003\text{-}2005}\,\lambda + \beta_1(Enroll_{jq2006} - Enroll_{jq2002})/Enroll_{jq2002} + X_{jq2002}\beta_3 + Z_{jq2000}\beta_4 + \delta_q + \zeta_{jqt}$$

Here, the cumulative effects of AYP failure are measured by three mutually exclusive indicators that predict the probability of closure: the school failed exactly once between 2003 and 2005, the school failed exactly twice between 2003 and 2005, or the school failed exactly three times between 2003 and 2005. ($Enroll_{jq2006} - Enroll_{jq2002})/Enroll_{jq2002}$ is the percent change in enrollment between 2002 and 2006, X_{jq2002} is a vector of baseline school characteristics in 2002, and Z_{jq2000} is a vector of district-level characteristics from 2000. Finally, δ_q is a vector of state fixed effects, and ζ_{jqt} is a normally distributed error term.

States' collective bargaining environments and political context are captured by state fixed effects, but being in a state that is more heavily unionized may impact schools that failed AYP and schools that never failed AYP differently. An additional model (1.2) includes an

interaction term to test if the relationship between schools' AYP failure and school closure is suppressed in states' with higher rates of districts covered by collective bargaining. Specifically, I include a dummy variable—$Union_{jq2003}$—that equals one if states' percent of districts covered by collective bargaining exceeds 75 percent. While a 75 percent rate of coverage is below the median in states' coverage, it represents a natural cutoff in the distribution.[43]

$$(1.2) \quad Pr(Close_{jq,2006\text{-}2011} = 1) = \alpha + AYP_{jq,2003\text{-}2005}\lambda + (AYP_{jq,2003\text{-}2005} \times Union_{jq2003})\vartheta + \beta_1(Enroll_{jq2006} - Enroll_{jq2002})/Enroll_{jq2002} + X_{jq2002}\beta_3 + Z_{jq2000}\beta_4 + \delta_q + \zeta_{jqt}$$

In other versions of equations 1.1 and 1.2 above, I add schools' within-state proficiency ranking as an additional control variable. Schools' within-state proficiency rate ranking controls for where a school falls, within its own state, in the distribution of the percentage of schools' students who score proficient or above on state exams. Schools with smaller shares of students meeting proficiency have lower rankings than schools with larger shares of students meeting proficiency. Because each state administers its own exam, a within-state proficiency rate ranking allows me to compare schools' relative performance across states. To account for a non-linear relationship between schools' relative performance and closure, I run separate versions of the model with squared and cubic relative proficiency rate terms.

Here, a drop in the size and significance of the estimated coefficient on the school AYP failure indicator may mean that the relationship between AYP failure and school closure in

[43] The median collective bargaining/union coverage rate is 92%. Using the median as the cutoff leads to an increase both the size and significance of the interaction terms' estimated coefficients. Fourteen states have collective bargaining/union coverage rates below 50%.

previous models is largely driven by schools' within-state performance ranking. The time period of this study precedes the onset of NCLB waivers which required intensive interventions for schools in the bottom five percent of performance in the state, but states may have utilized this approach prior to waiver submissions. If many of the lowest-performing schools failed AYP, then excluding schools' relative performance may misconstrue the relationship between AYP failure and school closure. Replacing schools' AYP indicators with schools' within-state percentile proficiency rate might similarly predict school closure. Yet, exogenous variation in AYP ratings may still exist even after controlling for schools' relative proficiency rates.[44] Holding student proficiency rates constant, the difficulty of making AYP may vary due to a number of factors—factors like school demographics, subgroup size, or dispersion of schools' average subgroup proficiency rates (Davidson et al., 2013). As a reminder, the identification in this study is not causal, but descriptive. A small and insignificant relationship between schools' relative performance in the state and school closure would not provide causal evidence of a link between AYP failure and school closure. Rather, including relative proficiency rate rankings allows me to explore whether AYP failure is uniquely associated with school closure, or whether schools that close simply enroll students who, on average, perform the worst on state exams.

B. Individual Years

Since schools that closed prior to 2006 are necessarily excluded from the previous analysis, selection bias potentially impacts the validity of the cumulative AYP failure effects. If the AYP failure rate of these closed schools is relatively high, the estimates above may be

[44] In states with available data, I averaged 2003 and 2004 overall school student proficiency rates. In states without 2003 data, I use 2004 proficiency rates only. Proficiency rates are available in 47 states. Proficiency data is missing in Alabama, Nebraska, and New Hampshire. Including proficiency rates reduces my sample from 75,753 schools to 68,911 schools.

muted. There may also be heterogeneous differences in closure rates across the included years of schools' NCLB outcomes. Individual years of AYP failure may predict closure better than others if districts preemptively closed schools to avoid implementing harsher consequences. In the baseline, I estimate three separate models for each year 2004, 2005, and 2006. I estimate the effects of a particular year of AYP failure on the probability a school closed between the subsequent year and the end of the sample, 2011 (*t to 2011*).

$$(2.1) \quad \Pr(Close_{jq,t\text{-}2011}=1) = \alpha + \lambda\, AYP_{jqt\text{-}1} + \varphi_1(Enroll_{jq2006} - Enroll_{jq2002})/Enroll_{jq2002} + X_{jq2002}$$
$$\varphi_3 + Z_{jq2000}\, \varphi_4 + \delta_q + \zeta_{jqt}$$

Results

Cumulative Effects

The "baseline" column of Table 6 presents the main probit regression results.[45] I report the average estimated marginal effects as well as the probit coefficient estimates and standard errors. Across both models, each indicator of AYP failure significantly predicts an increase in the likelihood of closure. Further, the more frequently a school fails AYP (once, twice, or three times), the more likely a school will close. In my baseline specification (column 1), the effects of failing AYP once or twice are similar in magnitude. Failing AYP once predicts a 0.5 percentage point increase in the likelihood of closure ($p < 0.01$), and failing AYP twice predicts a 1.1 percentage point increase in the likelihood of closure ($p < 0.01$). The effect of failing AYP three years is significantly larger than failing AYP once or twice ($p < 0.01$). Failing AYP all three years

[45] See Appendix C for a version of Table 6 that reports the same results but only for non-charter public schools (i.e., the analysis excludes charter schools).

predicts a 2.9 percentage point increase in the likelihood of closure ($p < 0.01$). Given a nine percent closure rate across years, the estimated effect of failing AYP all three years translates to an additional 186 closed schools.

Schools' 2002 total student enrollment and the percent change in enrollment across years are significantly associated with schools' closure probability. Smaller schools have higher relative closure probabilities than larger schools. All else equal, a school with an initial enrollment of 1,000 students has a 0.9 percentage point higher probability of closure than a school with 1,100 students ($p < 0.01$). Similarly, schools that experience enrollment declines between 2002 and 2006 are more likely to close. On average, a 10 percentage point decline in enrollment from 2002 to 2006 predicts a 6.5 percentage point increase in the likelihood of failure ($p < 0.01$). In other words, a school with a 20 percentage point drop in enrollment between 2002 and 2006 is 6.5 percentage points more likely to close than a school with a 10 percentage point drop in enrollment during the same period.

Compared to schools located in towns, schools located in cities and rural areas face significantly different probabilities of closure. Schools in cities are 1.8 percentage points more likely to close ($p < 0.01$) while schools in rural areas are 1.1 percentage points less likely to close ($p < 0.01$). Small but significant differences exist in the closure probabilities of towns and suburbs—suburban schools are 0.5 percentage points more likely to close than schools in towns ($p < 0.05$).

No measureable differences exist in the closure probabilities of high schools and the comparison group, elementary schools. There are large and significant differences, however, between elementary schools and middle schools and between elementary schools and schools

designated as "other."[46] Middle schools are 1.9 percentage points more likely to close than elementary schools ($p < 0.01$) and "other" schools are 2.5 percentage points more likely to close ($p < 0.01$).

The 0.5 percentage point effect of a school's title I status on its closure probability is statistically significant ($p < 0.01$). More than 50 percent of sampled schools are designated as Title I schools so a 0.5 percentage point increase in the probability of closure translates to a 212 school increase in the number of closed schools. The effect of a school's percentage of poor students is similarly substantive. A 10 percentage point increase in the proportion of poor students predicts a 0.8 percentage point increase in a schools' likelihood of closure ($p < 0.01$). The most substantive of all effects—schools' percent of Black students—suggests that a 10 percentage point increase in the population of Black students is associated with a 4.5 percentage point increase in the probability of school closure ($p < 0.01$). For schools' share of Hispanic students, the effect of 1.2 percentage points is less substantive but still significant ($p < 0.10$). The negative effects on closure of schools' Asian student population are substantive and significant ($p < 0.05$). Similarly, the share of a school's Native American population is negatively related to the probability of school closure. A 10 percentage point increase in the population of Native American students is associated with a 3.2 percentage point decrease in the probability of school closure ($p < 0.01$). These results suggest that Black, Hispanic, and poor students are disproportionately impacted by school closure.

[46] "Other" schools serve grade spans misaligned with the normal primary, middle, and high school classifications. For example, an "other" school may serve kindergarten through 12th grade or 4th through 10th grade.

District-level controls are also highly predictive of school closure. A 10 percentage point increase in the percent of residents with bachelor's degrees predicts a 5.9 percentage point decline in a school's likelihood of closure ($p < 0.01$). Similarly, a 10 percentage point increase in the percentage of Hispanic residents predicts a 6.7 percentage point decrease in the likelihood of school closure ($p < 0.01$). These results suggest that large shares of Hispanic students in a school and large shares of Hispanic residents in a neighborhood act as opposing forces in the model. Schools with relatively large shares of Hispanic students are more likely to close while schools in neighborhoods with relatively large shares of Hispanic residents are less likely to close. On the other hand, even though the share of schools' Black students has a substantive and significant effect on schools' likelihood of closure, districts' percentage of Black residents seems to have a trivial effect on schools' likelihood of closure.

The column labeled "Baseline + Interaction" in Table 6 reports the results from my 2nd specification, which includes the percentage of a state's districts covered by collective bargaining contracts and labor unions as an interaction term. In this specification, the estimated coefficients on the interaction term indicate that the impact of failure on school closure diminishes as the proportion of states' districts covered by collective bargaining contracts increases. The main effect now represents the relationship between AYP failure and school closure in states with fewer than 75 percent of districts covered by collective bargaining or unions.[47] For these states, the effects of failing AYP once are nearly double from the baseline model. Here, failing AYP once leads to a 1.3 percentage point increase in the

[47] The closure rate for schools in states with fewer than 75 percent of districts covered by unions is 4.9 percent. The closure rate for schools in states with more than 75 percent of their districts covered by unions is 6.7 percent. This difference in closure rates is significant at the 0.01 level.

probability of closure ($p < 0.01$). Failing AYP twice leads to a 1.6 percentage point increase ($p < 0.01$). Failing AYP three times leads to a 3.1 percentage point increase in the likelihood of closure ($p < 0.01$). The interaction coefficient reveals that the effect of failing AYP once is suppressed by 1.0 percentage point in states with more than 75 percent of districts covered by collective bargaining ($p < 0.01$). Thus, the relationship between failing once and school closure is reduced to 0.3 percentage points. Similarly, a greater saturation of collective bargaining contracts is associated with a 0.6 percentage point decline in the probability of closure for schools that failed AYP exactly twice ($p = 0.17$). Failing AYP twice in states with greater than 75 percent of districts covered by collective bargaining is reduced from a 1.6 percentage point to 1.0 percentage point increase in the likelihood of closure. The interaction coefficient for failing AYP three times is smaller and insignificant ($p = 0.741$).

Inclusion of Proficiency Rates as Additional Control

The column "Baseline + Prof Rank" in Table 6 presents the results when I add schools' within-state proficiency rate ranking as an additional control in the baseline model. AYP determinations are complex, rely on statistical formula and calculations, and account for various measures of school performance. In addition to schools' current year student test performance, schools' demographic composition, prior school year performance, and participation rates may contribute to schools' likelihood of AYP failure. If, after the inclusion of schools' relative performance, a relationship between AYP failure and school closure still exists, it strengthens the descriptive evidence that states and districts consider the AYP label, and not just student performance, when making closure decisions. It would suggest that there is

something unique about the AYP failure label that influences closure decisions irrespective of student performance.

The inclusion of the relative performance of a school's students eliminates the effects of failing once and twice on school closure. Yet, failing AYP three times is still associated with a 1.4 percentage point increase in the likelihood of closure for all sampled schools ($p < 0.01$). As expected, the relative performance of a school's students is negatively and significantly associated with school closure—a ten-percentile increase in the relative performance of a school's students predicts a 0.4 percentage point decrease in the likelihood of closure ($p < 0.01$). Thus, moving from the bottom quartile of performance (25^{th} percentile) to the upper quartile of performance (75^{th} percentile) is associated with a two percentage point decline in the likelihood of school closure.[48]

Compared to the other models, adding in schools' within-state proficiency rate ranking has a fairly substantial impact on the effects of school and district demographics on their likelihood of closure. The inclusion of the within-state proficiency rate ranking reduces the size of the coefficient for the share of Black students. In this model, a 10 percentage point increase in the share of Black students is associated with a 3.1 percentage point increase in the likelihood of closure rather than a 4.5 percentage point increase. On the other hand, the

[48] Table 7 presents the results for the same model but also includes squared and cubic terms of schools' proficiency rate ranking. The inclusion of these non-linear measures leads to a very slight reduction in the size and precision of the majority of the main estimates, but for a few of the main estimates, an increase in the size and precision. Most importantly, the inclusion of non-linear proficiency rate ranking terms leads to a sizeable increase in the coefficient on the linear proficiency rate ranking term. In the original model, a 10 percentile increase in a school's proficiency rate ranking is associated with a 0.4 percentage point decrease in the likelihood of closure ($p < 0.01$). In the non-linear model, a 10 percentile increase in a school's proficiency rate ranking is associated with a 2.5 percentage point decrease in the likelihood of closure. On the whole, including non-linear terms has little impact on the estimates of AYP failure but has a substantive impact on the estimates for relative proficiency rate rankings.

inclusion of the within-state proficiency rate ranking augments the size of the coefficient for the share of Native American students, and to a lesser degree, the share of Asian students. In this model, a 10 percentage point increase in the share of Native American students is associated with a 4 percentage point decline in the likelihood of closure rather than a 3.2 percentage point decline. In addition, the inclusion of the within-state proficiency rate ranking eliminates the effect of the share of students eligible for free- and reduced-priced lunch.

The effects of district demographics respond differently to the inclusion of the within-state proficiency rate ranking than the effects of school demographics. With the inclusion of the within-state proficiency rate ranking, the coefficient on the share of Black residents is larger and more precisely estimated. Yet, the coefficients on the share of district residents who are Hispanic and have bachelor's degrees are diminished. A 10 percentage point increase in the share of a Hispanic residents is now associated with a 5.5 percentage point decline in the likelihood of school closure rather than a 6.7 percentage point decline as in previous models. Similarly, a 10 percentage point increase in the share of Bachelor's degree recipients is associated with a 4.3 percentage point decrease in the likelihood of school closure compared to 5.9 percentage point decline in model versions without schools' proficiency ranking.

Interestingly, for schools in states with a lower saturation of districts with collective bargaining, failing AYP once, twice, and three times significantly increased schools' probability of closure. The column "Baseline + Interaction + Prof Rank" in Table 6 reports the results for this specification. For these schools, failing AYP once and twice predicts a 0.7 percentage point increase in closure probability ($p < 0.05$ & $p < 0.10$, respectively). Failing AYP three times predicts a 1.6 percentage point increase in the likelihood of closure ($p < 0.01$). Here, the

interaction terms for failing AYP once is also significant. Being in a state with higher rates of districts covered by collective bargaining reduces once-failing schools' likelihood of closing by 0.8 percentage points ($p < 0.01$). The coefficients on the interaction terms for schools that failed AYP twice and three times are small and insignificant.

Individual Years

Table 8 displays the probit regression results of individual years of AYP failure. For all schools, failing AYP in 2004 has the largest effect on a schools' likelihood of closure—failing in 2004 predicts a 2.2 percentage point increase in the likelihood of closure. The smallest effect comes in 2003. Failing AYP in 2003 predicts a 1.5 percentage point increase in schools' likelihood of closure ($p < 0.01$). In 2005, failing AYP predicts a 1.6 percentage point increase in school closure ($p < 0.01$).

None of the estimated coefficients on the interaction terms are substantive or significant. In 2003 and 2005, however, my main effects grow larger once I include the interaction term. With the inclusion of the interaction term, the main effect includes only those schools in states with fewer than 75 percent of its districts covered by collective bargaining. For these schools, failing AYP in 2003 predicts a 1.8 percentage point increase in the likelihood of closure after 2004 compared to a 1.5 percentage point increase for all schools. For schools in states with less than 75 percent of its districts covered by collective bargaining, failing AYP in 2005 predicts a 1.9 percentage point increase in the likelihood of closure after 2006 compared to a 1.6 percentage point increase for all schools. The estimated effects of AYP failure in both the 2004 baseline and interaction specifications are fairly similar—failing AYP in 2004 predicts a 2.2 percentage point increase in the probability of school closure.

Discussion

The rise of federal accountability ushered in a new era of school reform marked by an increased urgency for states and districts to remedy students' low-performance, greater incentives for states and districts to intervene faster, and increased costs to states and districts for failing to improve. This paper provides compelling evidence that states responded to these pressures by closing schools that failed AYP, especially in states that have fewer districts with union contracts or collective bargaining agreements. For all schools, although the relationship between school AYP failure and school closure is somewhat mitigated by the inclusion of schools' within-state proficiency rate rank, a strong, significant effect remains for schools that failed AYP three times.

For schools that fail AYP only once, my findings also indicate that teachers' unions and collective bargaining contracts may reduce the likelihood of states and districts using school closure as a consequence to AYP failure. There is some evidence, however, that as a school fails AYP more often, the relationship between collective bargaining agreements and school closure is suppressed. Failing AYP twice or three times, however, appears to increase the demand for intervention sufficiently to overcome any negative pressure on closure that results from a greater saturation of collective bargaining agreements. Further, for schools in states with fewer districts covered by collective bargaining, the effect of failing AYP once, twice, and three times remains positive and significant even with the inclusion of states' within-state proficiency rate rank. This finding suggests that, in states with fewer districts covered by collective bargaining, a school's AYP failure (once, twice, or three times) was more important in school closure decisions than its proficiency rate ranking. Perhaps the political contexts in these states gives

districts more flexibility in responding to AYP failure and avoiding the resulting consequences. In these contexts, districts may face limited opposition to school reform efforts in general, but especially contentious reforms like school closure. In states with more unionized districts, however, a school's proficiency rate ranking was at least as important as failing AYP once or twice. This finding suggests that the political context in these states limits districts' ability to respond to AYP failure. Yet, the effect of failing AYP three times is substantive and significant whether states are heavily unionized or not, and for schools in heavily unionized states and less-heavily unionized states, failing AYP three times was likely more important in closure decisions than proficiency rate rankings. State political contexts appear to have little influence on district responses once schools have demonstrated persistent AYP failure.

This paper also demonstrates that Black students disproportionately experience school closure and attend failing schools. Closed schools enroll Black students at nearly twice the rate as schools that remained open, and schools that failed AYP all three years enroll Black students at nearly three times the rate of schools that never failed AYP. Further, I find that, all else equal, a 10 percentage point increase in the share of Black students is associated with a 4.5 percentage point increase in a schools' likelihood of closure—one of the largest effects in my model. The inclusion of schools' proficiency rate rank eliminates the effect of Hispanic enrollment on school closure which suggests that there are fundamental differences in the experiences of Hispanic students and Black students with respect to school closure. In addition, these results reinforce recent research on the positive benefits to Black students of school integration and suggest that focused efforts to further integrate schools, through means other

than school closure, may alleviate the burdens of closure disproportionately faced by Black students and lead to more sustained improvements in Black students' performance.

Unlike NCLB, more recent accountability efforts require states to intervene in schools in the bottom 5 percent of performance, as measured by student proficiency rates. This approach is a shift from NCLB, where schools received binary labels rather than further-differentiated ratings that may allow districts and states prioritize interventions or tailor support. Instead, states had to intervene with prescribed interventions for every school labeled as failing, even though schools labeled as failing AYP may not have been the schools with the lowest student proficiency rates. Further, in states with high AYP failure rates, implementing NCLB consequences at so many schools may have diverted resources and support away from the neediest schools. This paper provides some evidence that, under NCLB, states did prioritize interventions for those schools at the bottom of the proficiency rate distribution, but that as a school failed AYP more often, its AYP status was likely still an important factor in closure decisions.

This paper provides some evidence that the accountability pressure generated by NCLB's prescribed consequences and defined timeline led states and districts to respond to AYP failure with school closure. This paper stops short, however, of exploring whether school closure, or other NCLB interventions, had a positive impact on student performance. While the nascent school closure literature suggests student outcomes improve when a student from a closed school enrolls in a high-performing school, there is no evidence that students from closed schools will systematically enroll in higher-performing schools or that, on average, students from close schools even have access to higher-performing school options.

In this paper, I find that NCLB rules and consequences led to increased rates of school closure for schools that failed AYP; yet, school closure was an implied, *but not explicit*, consequence of AYP failure. More recent accountability efforts explicitly refer to closure as a consequence which suggests closure rates may continue to rise. With so little known about the long-term impact of school closure on student outcomes, state and district officials should proceed with caution.

CHAPTER THREE – THE STRONG ARM OF THE LAW: HOW STATE LAW INFLUENCES CLOSURE DECISIONS FOR PERSISTENTLY FAILING CHARTER SCHOOLS

Introduction

Charter schools, first introduced in Minnesota in 1991, are public schools that operate under distinct governance arrangements from traditional public schools. Originally, charter schools were conceived to promote experimentation and innovation in the provision of education (Budde, 1988; Shanker, 1988). Eventually, policymakers and other stakeholders recognized the potential for charters schools to introduce market competition in education in the form of school choice so as to promote greater efficiencies in the provision of education (Chubb & Moe, 1990; Henig, 1994; Nathan, 1996; Belfield & Levin, 2005; Kolderie, 1990). Unlike traditional public schools, charters, in theory, can incubate innovate practices because they are free from bureaucratic constraints related to staffing, financing, and curriculum that impede improvements traditional public schools (Hassel, 1999; Finn, et al., 2000; Kolderie, 1990). Between 2000 and 2015, enrollment in charter schools grew from around 350,000 to nearly 2.7 million students, which represents approximately five percent of the total enrollment in U.S. public schools. As an increasing number of states introduced charter school legislation, the number of charter schools has grown exponentially—from around 1,500 in 2000 to around 6,600 in 2015 (NAPCS, 2016).

Charter schools, unlike traditional public schools, are governed, in large part, by state charter school laws which dictate where charter schools can operate, how they can operate, how state and local education funds are allocated to them, what entity or entities are responsible for their oversight, how they are evaluated, and in some cases, how many of them

can exist in the state (Holyoke, et al., 2009). State charter laws underpin the provisions that can be included in charter school contracts, or "charters," which include the specific agreements that dictate charter school operations. Yet, charter schools, as public schools, are also subject to the rules and requirements of federal accountability policies, which until recently, included the rules of states' accountability systems under the No Child Left Behind Act (NCLB) of 2001.

The impact of charter schools on student performance is decidedly mixed (Hanushek et al. 2007; CREDO 2009; Zimmer et al. 2009; Betts & Tang, 2011; Wohlstetter, et al., 2013). There is compelling evidence that, on average, students in charter schools in some cities and states outperform their peers in traditional public schools (Hoxby & Rockoff, 2004; Abdulkadiroglu, et al., 2011; Hoxby et al., 2009, CREDO, 2015; Angrist, et al., 2011; Betts & Tang, 2011), but in other areas, students in charter schools, on average, do as well as or decidedly worse than their peers in traditional public schools (Bifulco & Ladd, 2006; Buddin & Zimmer, 2005; Betts & Tang, 2011; Gleason, Clark, Tuttle, & Dwoyer, 2010).

Every anecdote in the news or academic literature about a charter school whose students are beating the odds is matched by an anecdote about charter school financial mismanagement, high rates of teacher turnover, questionable discipline policies, and poor student outcomes (CREDO, 2012; Tuttle, et al., 2015; Deruy, 2016; CREDO, 2014, Losen, et al., 2016). Many opponents to charter schools argue that states have failed to take sufficient action to increase charter school oversight and intervene when objectionable practices or persistent poor outcomes occur (SRI International, 2000; Bulkley, 2001).

In this paper, I investigate the extent to which charter school's NCLB performance and the accountability provisions of state charter laws, in combination, impact a charter school's

likelihood of closure. To determine which charter schools closed over the period from 2002 to 2011, I use national datasets on student and school characteristics, and to determine charter schools' NCLB performance, I use a national dataset on AYP outcomes from 2003-2005. In addition, I constructed a state-specific index that measures the "strictness" of states' charter school laws as of 2006. Strict laws include rigid and specific provisions for charter school approval, renewal, evaluation, and termination. Mild laws include flexible and general provisions for the same. My index is based on those provisions of charter school law that I believe are most likely to influence state and district interventions in failing charter schools. However, to test those assumptions, I also include a related index from the prior literature.

My identification strategy compares charter schools that failed AYP at least once, at least twice, or at least three times, to charter schools that never failed AYP. I use a probit regression to estimate the likelihood that a charter school that failed AYP between 2003 and 2005 closed anytime between 2006 and 2011. To account for state-level policies that may correlate with states' AYP failure rates or charter school closure rates, I include state fixed effects. I compare the results for charter schools to similar results for all U.S. *non-charter* public schools to highlight any differences in outcomes across sectors and to connect my Chapter III estimation to my Chapter II results. Finally, I interact school AYP performance with each of the three included indices to determine if strict accountability provisions in state charter law enhance the effects of AYP failure on charter school closure.

My results show that, all else equal, failing AYP is associated with a large and significant increase in the likelihood a charter school closed. A charter school that fails AYP at least once is nearly six percentage points more likely to close than a charter school that never failed AYP.

The main coefficient for failing AYP at least once, at least twice, and all three years is much larger for charter schools than non-charter public schools; however, due to the small number charter schools that failed AYP twice or three times, only the differences in the coefficients for failing AYP at least once are statistically significant at the 0.10 level. There is also compelling evidence that strict provisions in state charter law increase the likelihood that a failing charter school closed, although there is some evidence that mild charter law provisions have a similar effect.

By exploring the relationship between charter school governance and AYP outcomes, this is the first national study to catalogue and explain charter school closure in the context of NCLB. Further, my efforts to categorize state charter school laws by their accountability provisions expands on and refines early efforts by Shober, Manna, and Witte (2006). By examining the impact on closure of multiple measures of state charter school law, my study also contributes to the literature on the ways in which state regulatory and policy environments impact the charter school landscape across states.

This paper proceeds as follows: In Section II, I provide an overview of the literature on charter school legislation and governance. Section III provides a brief description of the included data, and Section VI reviews the data's descriptive trends. Section V and VI provide an overview of the methodology and findings, respectively. Section VII concludes.

Background

Even though charter schools were conceptualized to improve public education and promote innovation, more often than not, charter school legislation is introduced or refined due to political pressures from interest groups and politicians "seeking to make their mark in

education reform" rather than attempting to solve education crises (Holyoke, et al., 2009, p. 49; Bulman & Kirp, 1999; Bulkley, 2005; Henig, 1994). These political pressures, combined with spill-over from successful policies in neighboring states, often led policy-makers to substantively change the provisions of states' charter school legislation over time (Center for Education Reform, 2015, 2012, 2006; Holyoke, et al., 2009; Mintrom & Vergari, 1998; Renzulli & Roscigno, 2005; Shober, et al., 2006). Often, characteristics of charter laws are categorized in terms of their "flexibility" (sometimes referred to as "autonomy")—the extent to which charter schools are given the operational latitude to respond to specific stakeholder needs—or "accountability"—the extent to which charter school performance is assessed against a clear set of standards with resulting rewards and consequences (Wohlstetter, Wenning, & Briggs, 1995; Shober et al., 2006).

Typically, legislators and policy-makers must delicately balance the principles of flexibility and accountability in the design of state charter laws. Relatively high levels of accountability may result in an inability to respond to specific stakeholder needs, often the needs of students' families. Relatively high levels of flexibility may mean some stakeholders' needs are adversely prioritized over others (Shober, 2006). Shober et al., (2006) find that recalibrations of state law have tended to favor flexibility over accountability. While this recalibration may be in the best of interest of charter school operators, there is little evidence that diminishing accountability pressures are in the best interest of students. There is some evidence, however, that specific accountability- or flexibility-related provisions, rather than these types of provisions as a whole, are related to improvements in charter school student performance. For instance, caps on the number of charter schools, an accountability-related

provision, are associated with lower-than-expected student growth, and allowing multiple entities to serve as authorizers, a flexibility-related provision, is associated with lower-than-expected student growth (CREDO, 2009).

The onset of NCLB may have inadvertently overridden these state-level recalibrations toward flexibility since charter schools, as public schools, are held to NCLB's rules and consequences. Unlike non-charter public schools, charter schools are monitored and evaluated by a wide variety of entities or "authorizers," not always the local school district. State charter law dictates what types of entities can serve as charter school authorizers. In some states, only local education agencies can authorize charter schools. In other states, state education agencies or not-for-profit organizations, including higher-education institutions, can serve as authorizers. Authorizers are responsible for evaluating operator applications and granting charters, or contracts, to qualified candidates. Charter contracts are influenced by state charter law, but often include performance expectations and other provisions not specifically covered by law. Authorizers are responsible for ensuring that charters comply with state and local regulations and must review charter performance in order to make renewal and termination determinations (Vergari, 2001).[49]

Authorizers are largely responsible for implementing NCLB consequences for persistently failing charter schools, but authorizers are rarely subject to external monitoring or oversight from the state education agency (Bulkley, 2001; Wohlstetter, et al., 2013; Ed Trust,

[49] There is some evidence that the type and number of authorizers in a state impacts the number and performance of charter schools in that state. For instance, the USDOE reports that local education agencies tend to grant fewer charters, on average, than state education agencies or not-for-profits (USDOE, 2002). In Ohio, students in charter schools with not-for-profit authorizers underperformed relative to their peers in charter schools with state, regional, or district authorizers.

2016).[50] As such, there are real questions about authorizers' capacity to intervene in low-performing charter schools. As of 2002, most authorizers had extremely limited resources to provide technical support or oversight (USDOE, 2002, Fordham, 2006, 2008). Due to community support, political pressures, or perverse incentives, many authorizers find closing low-performing charter schools extremely difficult (Bulkley, 2001; SRI International, 2000). Even though a large number of low-performing charter schools continue to operate, there is some evidence that closures of low-performing charter schools are leading to an improvement in the average performance of students in the charter sector (CREDO, 2015).[51]

The primary reason that charter schools close is due to financial mismanagement rather than low-performance. The Center for Education Reform (CER) (2011) estimates that, since 1992, over 80 percent of charter school closures are due to financial distress, mismanagement[52], issues with facilities, or district obstacles. Of the 19 percent of charters that close for academic reasons, however, CER has little to no information on how or why authorizers made non-renewal decisions. The authors suggest that states with laws that allow for multiple, independent authorizers are better able to enforce performance standards, but these findings should be interpreted with caution as the evidence that supports this conclusion is not causal and CER has a well-documented pro-charter bias.

Over the last decade, there have been numerous attempts to categorize state charter laws based on their included provisions to determine how those provisions impact the number

[50] For a fuller discussion of NCLB policy, see Chapter I. For a fuller discussion of the prior literature of NCLB and school closure, see Chapter II.
[51] The authors find little to no improvement in the average performance of students in individual charter schools.
[52] According to CER (2011), mismanagement is defined as "administrator or sponsor misbehavior" (p. 9).

and type of charter schools, the demographic distribution of students across charter schools, the effectiveness of charter school authorizers, and the performance of charter school students (Shober et al., 2006; Bierlein Palmer and Gau, 2003; AFT, 1996; Miron, 2005). Some analyses use these categorizations of charter law to understand how charter law provisions change over time in response to political pressures and legislative action (Holyoke, et al., 2009).

One of the most widely-recognized ratings is produced by the Center for Education Reform (CER). Since 1999, CER has produced annual report cards that assign each state a score, grade, and ranking that indicates whether the state law is "strong" or "weak" according to 10 criteria (see Appendix D for a full description of CER criteria) Strong laws, as measured by CER's criteria, are typically those that limit the barriers to open a charter school, impose fewer restrictions on charter school operations, and ensure equitable funding for charter schools as compared to non-charter public schools. Since 2010, the National Alliance for Public Charter Schools (NAPCS) has also produced a widely-recognized annual ranking of charter school laws based on 20 criteria (see Appendix D for a full description of the NACPS criteria). The NAPCS ranking, like CER, also rates the "strength" of charter school law provisions related to barriers to entry, operations, and funding. The NAPCS rating, however, also includes ratings on factors related to charter school performance and accountability.

Shober, Manna, and Witte (2006) reviewed states' charter school laws as of 2002 to construct ratings that measure criteria related to flexibility and accountability (see Appendix D for a full description of the Shober, et al., criteria). Much like the CER rating and the NACPS ranking, Shober, et al., rated states based on the extent to which state law limits the barriers to opening and operating a charter school. A more "flexible" law has fewer barriers; a less

"flexible" law includes more barriers. Shober, et al., also rated states based on accountability-related provisions--whether or not the charter school must use state standards, the extent of the control that charter schools have over the students they admit, and whether or not states were required to report information about student performance to determine state accountability rankings.

Data

Like Chapter II, I include school-level demographic information from the Common Core of Data (CCD), and school-level NCLB outcomes available on the Barnard Columbia NCLB Data Project website. For more information on the methods for identifying which schools closed between school years 2002-2003 and 2010-2011, review the Data Section in Chapter II. Because Chapter III focuses only on the impact of NCLB performance on charter schools, my analysis includes all charter schools that appear in the CCD as "operational" and "regular" in the school year 2001-2002. For some analyses, I compare charter school demographics and outcomes to the demographics and outcomes of all U.S. non-charter public schools. Non-charter public schools are defined as schools that appear in the CCD as "operational" and "regular" in the school year 2001-2002 but are not identified as charter schools.

To quantify the impact of state charter school law on the likelihood of closure for charter schools that fail AYP, I need to account for variation in state laws regarding the ways that charter school authorizers considered school and student performance in granting and revoking charter agreements. Because my study is focused on whether or not schools closed during a specific time period, 2006 through 2011, I also need to account for state charter law provisions in effect at the beginning of that period. While the NACPS state charter law ranking

includes measures related to school performance and accountability, the earliest available ranking is from 2010, which is near the end of the examined period. The CER index, on the other hand, extends as far back as 1999 but does not include any accountability-related measures.

The Shober, et al., accountability index (referred to throughout as "Shober Accountability Index") is based on the contents of state law as of 2002, before the onset of NCLB. States may have amended their charter laws to accommodate the rules of NCLB, or the criteria used to construct the index may no longer be applicable due to the requirements of NCLB. In addition, the Shober Accountability Index is based on a limited number of criteria—curricular standards, annual performance reporting requirements, and student admission policies—some of which are more relevant to this study than others. For these reasons, the Shober accountability index is a useful but imperfect measure of the state policy context faced by charter schools during the NCLB era.

To create a richer and more relevant measure, I reviewed state charter school laws as of 2006 and constructed a new accountability index based on four dimensions of state law that I posit contribute to a charter schools' likelihood of closure – 1) whether the criteria for charter renewal included meeting specific academic goals set forth in the charter, 2) whether schools were required to annually report academic performance to their authorizer, 3) whether authorizers could revoke the charter at any time due to schools' failure to satisfy academic

progress set forth in the charter[53], and 4) the maximum length of a charter contract, in years.[54] I use these four dimensions, or criteria, to construct a four-point accountability index.

My index measures the "strictness" of each states' charter school law as of 2006. I score each criterion based on its restrictiveness; a score of one indicates the least restrictiveness and a score of four indicates the most restrictiveness. For example, a state charter law with no termination clause would receive a score of one for the termination clause criterion because there are no restrictions in place. A state charter that limits a charter agreement to a maximum of three years would receive a score of four because three years is maximally restrictive—a relatively short maximum charter length implies that charter schools have to apply for reauthorization and undergo a review more often. For each criterion, Table 9 describes the evidence that is associated with each score. For each state, I average the scores (equally weighted) across the four criteria to generate an average state-level index score. A summary of the index scores for each state is in Appendix E. To complete my analysis, I separate states into three approximately equal groups – states with the lowest scores have mild levels of accountability, states with mid-range scores have a moderate level of accountability, and states with the highest scores have the strictest levels of accountability. Ten states with no charter laws as of January 2006 receive an index score of "missing" and are subsequently excluded from the sample. Figure 9 displays the distribution of schools across the index score range. Table 9 outlines the categorical ratings of mild, moderate, and strict for each state.

[53] In contract law, this clause is typically referred to as a "termination clause."
[54] I refer to this index as the "EKD Index" throughout.

In 2006, states' charter school laws varied in their requirements for charter school accountability. Across states, the maximum length of a charter contract allowable by state law ranged from three years to an unlimited number of years. The average maximum length was six years. In Maryland, for instance, state law did not specify the maximum length of initial charter contracts but rather gave local school boards the discretion to set the length of the charter contract. Because this approach gave authorizers the most flexibility and least oversight, I assign Maryland's maximum charter contract length criterion a score of one, the mildest score possible. Minnesota law, on the other hand, restricted the initial charter contract length to a maximum of three years—the shortest specified length of any state—so I scored Minnesota's maximum charter contract length criterion a four, the strictest score possible.

In 2006, few states required schools' performance goals be outlined in the charter contract. North Carolina was the exception. North Carolina law stated, "The State Board of Education, or a chartering entity subject to the approval of the State Board of Education, may terminate or not renew a charter …[for] failure to meet the requirements for student performance contained in the charter" (North Carolina State Statutes, 2006). Even when state law specified that student performance goals be specified in the charter contract, charter authorizers often had ultimate discretion in granting charter contract renewals—authorizers in North Carolina could choose to renew the contract even when confronted with convincing evidence of poor performance. Because North Carolina law included authorizer discretion, I assign North Carolina's renewal criterion a score of three. Louisiana law, on the other hand, referred ambiguously to requirements for "improvement," but required that authorizers non-renew contracts when students perform poorly: "No charter shall be renewed unless the

charter renewal applicant can demonstrate, using standardized test scores, improvement in the academic performance of pupils over the term of the charter school's existence" (Louisiana State Statutes, 2006). Because Louisiana law mandated that authorizers non-renew contracts when students underperform, I assign Louisiana's renewal criterion a score of four.

There was little variation across states in 2006 regarding whether or not their laws specified termination clauses for poor performance, so the score range for this criterion is only 1-3 rather than 1-4. Most states (n=28) allowed authorizers to terminate charter contracts *at any time* for poor performance; I gave these states a score of three, the strictest score possible. Some states (n=12) required authorizers to hold a public hearing, invoke a probationary period, or hold a vote of key stakeholders before they could terminate a charter contract. For these states, I gave a score of two, a moderate score. The laws of all but one state, Maryland, included some form of a termination clause. I score Maryland's law a one, the mildest possible score.

Finally, state laws varied in the extent to which they required schools submit annual performance reports or undergo annual evaluations to measure students' progress from year-to-year. States with the mildest laws only required that charter schools submit the same reports as traditional non-charter public schools. These states, like Rhode Island, Wisconsin, and New Mexico, all receive a score of one on the annual performance reports criterion. On the other hand, states with the strictest laws required an annual in-depth evaluation that included a site visit to closely evaluate students' experiences and outcomes. In California, legislation adopted in 2003 (and in place in 2006) specified five new charter authorizer oversight duties "including, visiting each charter school annually, ensuring each charter school complies with

state reporting requirements, and monitoring each charter school's fiscal condition" (California State Statutes, 2003). In Oregon, the authorizer "at least annually shall visit the public charter school site and review the public charter school's compliance with the terms and provisions of the charter" (Oregon State Statutes, 2006). California and Oregon both receive a score of four on the annual performance reports criterion.

The Shober Accountability Index, while an imperfect measure for my study, provides a rigorous benchmark against which I can compare the impact of my own index. Student performance reporting is the only state charter law criterion that is included in the Shober index and my own. For each of the criterion included in the index, a score of one (1) indicates a mild accountability environment and a five (5) indicates a strict accountability environment. The criteria scores were averaged to generate an overall index score on a scale of 1-5. A summary of the Shober index scores for each state is in Appendix E.[55] Much like my own index, I use Shober's overall scores to rate states' charter law as mild, moderate, or strict in terms of accountability. I rate the states as mild, moderate, or strict depending on whether their overall score falls in the approximate lowest third, approximate middle third, or approximate highest third, respectively, of the overall score distribution. Table 10 outlines the categorical ratings of mild, moderate, and strict for each state. Figure 10 displays the distribution of schools across the Shober index score range. If the impact of the Shober et al, index differs from the impact of my own index, then further investigation into the drivers of these differences may be warranted.

[55] The mean rating was 3.71 with a standard deviation of 1.11.

Descriptive trends

In addition to differences in governance, charter schools and non-charter public schools were substantively different in terms of the demographics of their students, their AYP outcomes, and their likelihood of closure. Therefore, the Chapter II analysis of all public schools—including charters and non-charter publics—may obscure meaningful differences in the relationship between AYP outcomes and closure. The analysis in this chapter (Chapter III) is designed to illuminate how the relationship between AYP outcomes and school closure differs for charters and non-charter public schools. Table 11 displays the summary statistics of all variables for both charter schools and non-charter public schools. Descriptive variables, like student enrollment, student demographics, geographic location, and grades served are from the 2001-2002 school year, prior to the onset of NCLB.

Charter schools typically performed worse on state standardized exams, failed AYP at higher rates, and closed more often than non-charter public schools. Students at charter schools, on average, scored proficient or above on state standardized exams at lower rates than students at non-charter publics—the average proficiency rate for charter school students was 34% while the average proficiency rate for non-charter public school students was 51% ($p < 0.01$). Student proficiency rates are not directly related to AYP performance, yet charter schools and non-charter publics also fared differently on NCLB-related outcomes. Fifty-six percent of charter schools failed AYP at least once between 2003 and 2005 compared to 41 percent of non-charter public schools ($p < 0.01$). Yet, at the extremes, the differences between charter schools and non-charter schools is smaller than expected—16 percent of charter schools failed AYP three times versus 12 percent of non-charter public schools, still a statistically significant

difference ($p < 0.01$). Charter schools did not fare any better compared to non-charter publics in terms of their closure rates. Charter schools closed at three times the rate of non-charter publics despite charter schools only accounting for approximately 3 percent of all public schools nationally ($p < 0.01$).

Differences between charter schools and non-charter schools also extended to student demographics, geographical location, and student enrollment (point-in-time and change-over-time). In addition, charter schools, on average, served black students at nearly twice the rate of non-charter public schools ($p < 0.01$).[56] In 2002, 28 percent of charter school students were black compared to 16 percent in non-charter public schools. The difference for Hispanic students was smaller, 17 percent of charter school students were Hispanic while 15 percent of students in non-charter public schools were Hispanic. Surprisingly, the difference in the percentage of students eligible for free- and reduced-priced lunch was also small; Forty-seven percent of students in charter schools were eligible for free- and reduced-priced lunch compared to 41 percent in non-charter public schools ($p < 0.01$). Yet, this difference is statistically significant at the 0.01 level.

While most charter schools were located in urban areas (51 percent), most non-charter publics were located in suburban areas (35 percent) and were much less likely to be classified as "other" schools (meaning they serve non-traditional grade levels) and much more likely to be classified as elementary schools ($p < 0.01$, for both). Between 2002 and 2006, charter school enrollment, on average, grew by 20 percent while enrollment in non-charter public schools

[56] For descriptive characteristics of all public schools – including non-charter public schools and charter schools - during the same period, please see Table 4.

declined by one percent. The average total enrollment of charter schools in 2002, however, was significantly smaller than the average total enrollment in non-charter public schools. On average, charter schools enrolled 263 students in 2002 compared to 577 in non-charter public schools ($p < 0.01$). These disparities may be explained by the fairly typical charter school operational strategy of launching a school that serves relatively few grade levels and adding an additional grade level each year.

Table 12 displays differences in school characteristics between charter schools that closed between 2004-2011 and charter schools that remained open. Compared to charter schools that remain open, closed charter schools were more likely to enroll fewer students, have higher rates of black and Native American students, have lower rates of white and Hispanic students, and were more likely to serve high school grades ($p < 0.01$ for all). The difference in the percentage of black students between the two groups is large – on average, black students comprised 40.5 percent of the total school population in closed charter schools versus 25.3 percent of the total student population in charter schools that remained open ($p < 0.01$). This descriptive finding suggests that not only were black students disproportionately enrolled in charter schools but disproportionately impacted by charter school closure. Interestingly, 14 percent of students in closed charter schools were Hispanic while 17.6 percent of students were Hispanic in charter schools that remained open ($p < 0.01$). This finding reinforces the notion that black students are uniquely affected by charter school closures—black students seem to be affected to a greater degree than any other racial minorities.

While charter schools enrolled students eligible for free- and reduced-priced lunch at similar rates as all non-charter public schools, the differences between charters that closed and

charters that remained open are significant ($p < 0.01$). On average, students eligible for free- and reduced-priced lunch comprised 53 percent of total enrollment in closed charter schools compared to 43 percent in charters that remained open. As a review, Table 11 reports that enrollment in non-charter public schools declined by 1 percent, on average, between school years 2002 and 2006, and enrollment in charter schools grew 20 percent, on average, during the same period. Yet, even though enrollment in charter schools, on the whole, grew substantially between 2002 and 2006, much of that growth occurred in charter schools that did not close between 2006 and 2011 rather than in charter schools that did close during that same period. For charter schools that closed between 2006 and 2011, enrollment only grew by five percent between 2002 and 2006 while enrollment in charter schools that never closed grew by 25 percent during the same period ($p < 0.01$).

There are also significant differences in the characteristics of charter schools that failed AYP once, twice, or all three years between 2003 and 2005 and charter schools that never failed AYP during the same period. Table 13 shows that, compared to schools that never failed, charter schools that failed once, twice, or all three years were more likely to have significantly higher rates of students eligible for free- and reduced-priced lunch and higher shares of black and Hispanic students. Failing charter schools were also more likely to have significantly larger student-teacher ratios—23 students per teacher in charter schools that failed AYP all three years and 17 students per teacher in charter schools that never failed AYP. Failing charter schools also enrolled more students, on average, served different grade levels, and were much more likely to be located in an urban area than charter schools that never failed. Charter schools that failed AYP three times enrolled 130 more students, on average, than charter

schools that never failed AYP. This may be explained by differences between the groups in their grades served—elementary schools typically enroll fewer students than high schools, and elementary schools were the least likely to fail AYP. Sixty-two percent of charter schools that never failed AYP were elementary schools; only 10 percent of charter schools that never failed AYP were high schools. These rates converge the more times a school failed AYP—31 percent of schools that failed AYP all three years were elementary schools while 23 percent of schools that failed AYP all three years were high schools.

Methods

This paper is focused on the probability a charter school closed during a fixed period of time (in this paper, between 2006 and 2011), which I presume is influenced by a vector of regressors, *X*.

$$\Pr(Y = 1 \mid X) = \Phi(X'\beta)$$

Because the outcome variable, Y, has two possible outcomes—the charter school remained open throughout the fixed period or the charter school closed at some point during the fixed period—a probit estimation is appropriate.[57] My key predictor is an indicator of charter school performance during the period 2003 through 2005—specifically, whether or not a charter school failed to make Adequate Yearly Progress (AYP) as defined by state and federal regulations. I posit that charter schools that failed AYP at least once during the period 2003 through 2005 were more likely to close during the subsequent period (2006 through 2011) than

[57] Because I am using a probit estimation, I cannot determine the number of years during the fixed period a school "survives" before it closes. This paper, however, focuses on whether or not states and districts take action and not how swiftly they take action.

charter schools that did not fail AYP during the same period (2003 through 2005). In addition, the more often a charter school failed AYP (e.g., twice or three times between 2003 and 2005) the greater their likelihood of closure in the subsequent period (2006 through 2011).

I also control for other school-based factors as these factors likely influence the probability a charter school closed. For instance, charter schools that have experienced persistent and significant enrollment decline may be more likely to close than schools in which enrollment has increased or remained constant over time. To control for these factors, I include measures of schools' total enrollment and the percentage change in its enrollment between 2002 and 2006.[58] I also include school-level demographics and other school characteristics (e.g., grades served and a measure of the urbanicity of a school's setting).

Students' academic performance may also impact a school's likelihood of closure if states and districts use absolute measures of students' performance, like student proficiency rates,[59] to designate schools for closure. Failing AYP, however, is not a strict proxy for student and school standardized test performance. Due to the complexities of AYP calculations, schools with the lowest average student proficiency rates in their state are not necessarily the schools that fail AYP.[60] To isolate the effect of AYP failure on school closure, I include a measure of schools' student performance on state tests. Since states administer different tests, students' performance on those tests is incomparable across states. Therefore, I standardize

[58] To create a more stable measure of enrollment change over time, I average schools' total enrollment across the years 2002, 2003, and 2004. Then I compare this average enrollment with schools' 2006 enrollment to determine the percent change between the two periods.
[59] Proficiency rates are based on students' performance on state standardized exams.
[60] See Chapter I for additional discussion on the complexities of state AYP calculations.

performance across states by constructing a relative performance measure: schools' within-state percentile rank based on the average of its student proficiency rate in 2003 and 2004.

Education policy environments vary considerably across states and change frequently across years, especially with respect to their regulations regarding charter schools. Incorporating fixed effects into the probit estimation is the most appropriate remedy for variations in observable and non-observable state characteristics that may inadvertently confound the results. In this case, fixed effects account for cross-state differences in laws that govern charter school operations and evaluation, among other state characteristics. Presumably, state charter school laws influence the ways in which states, districts, and other authorities monitor charter school performance and intervene in cases of poor performance. Without a fixed effects approach, differences in state charter laws may inflate or deflate the impact of charter school AYP failure on charter school closure. Therefore, I estimate a probit regression where the unit of observation is the school, the dependent variable is an indicator of whether or not the school closed between 2006 and 2011, the key predictor is an indicator of the frequency of school AYP failure between school years 2003 and 2005 (at least once, at least twice, or at least three times), and state fixed effects control for average differences across states.

$$(1)\ \Pr(Close_{jq,2006\text{-}2011} = 1) = \alpha + FailAYP_{jq,2003\text{-}2005}\,\lambda + \beta_1(Enroll_{jq2006} - Enroll_{jq2002})/Enroll_{jq2002} + \beta_2 Rank_{jq,2003,2004} + X_{jq2002}\beta_3 + \delta_q + \zeta_{jqt}$$

Model 1 is executed in three distinct versions. In the first version, $FailAYP_{jq2003\text{-}2005}$ equals 1 if school *j* in state *q* failed at least once during school years 2003 through 2005. The null value represents schools that never failed during the same time period. In the second version of

Model 1, $FailAYP_{jq2003\text{-}2005}$ equals 1 if school *j* failed at least twice during school years 2003 through 2005. In this version, the null value represents schools that never failed AYP or failed only once. In the third version of Model 1, $FailAYP_{jq2003\text{-}2005}$ equals 1 if school *j* failed at least three times during school years 2003 through 2005; the null value represents schools that never failed AYP or failed once or failed twice. Like Chapter II, I also include $Enroll_{jq2006} - Enroll_{jq2002}/Enroll_{jq2002}$, a measure of the percentage change in school total enrollment between school years 2001-2002 and 2005-2006. $Rank_{jq,2003,2004}$ is the percentile rank of school *j* in state *q* based on the average of the school's overall student proficiency rate on state exams in 2003 and 2004. X_{jq2002} is a vector of school characteristics from 2002, the term, δ_a, captures state fixed effects, and ζ_{jqt} is a normally distributed error term.[61]

Under No Child Left Behind, public charter schools and traditional public schools are held to the same accountability standards and are subject to the same NCLB-related consequences. Federal and state requirements mandate charter school students participate in state testing programs and, within a given state, states are required to evaluate charter schools using the same AYP rules as traditional public schools. Even though charter schools may have autonomy in some operational areas, charter schools' AYP status is comparable to the AYP status of non-charter traditional public schools.[62] These similarities make it possible to compare the impact of school AYP failure on school closure for both charter schools and non-charter public schools. Therefore, I re-estimate Model 1 using only non-charter public schools in order

[61] This model may suffer from selection bias if schools that closed between 2002 and 2006 had relatively high or relatively low AYP failure rates during that period.

[62] In Chapter II, I estimate the impact of school AYP failure on school closure for all schools, including charter schools and non-charter traditional public schools. In this Chapter, I repeat the Chapter II analysis but exclude charter schools in order to construct a meaningful comparison group for charter schools.

to identify any meaningful differences in the relationship between AYP performance and school closure across these school types. In addition, non-charter public schools provide an interesting falsification test on the relationship between state charter laws and the likelihood of closure for failing charter schools. The extent to which non-charter public school outcomes are affected by state charter law is unclear; one might expect state charter law has no effect on non-charter public school closure, although it's possible that mild charter school laws increase chartering activity which may increase accountability and market pressures for non-charter public schools.

Interacting AYP Failure and Accountability Indices

While cross-state differences in charter laws are accounted for using state fixed effects, these differences in charter laws may impact charter schools that failed AYP and charter schools that did not fail AYP differently. Charter schools that fail AYP may be more likely to close if state charter law includes strict provisions related to performance monitoring and accountability. As a test, in Model 2, I include an interaction between $FailAYP_{jq2003\text{-}2005}$—an indicator of school failure (where the frequency of failure during the period depends on the model version)—and $Provisions_{q,2006}$–a vector of three indicators (mild, moderate, or strict) that describe the rigor of the accountability provisions of state's charter school law as of 2006. These three indicators are based on the ratings I assigned states using the accountability index (the EKD index) I constructed by reviewing state charter laws as of 2006.

$$(2)\ Pr(Close_{jq2006\text{-}08} = 1) = \alpha + (FailAYP_{jq2003\text{-}2005} \times Provisions_{q,2006})\lambda + \beta_1(Enroll_{jq2006} - Enroll_{jq2002})/Enroll_{jq2002} + \beta_2 Rank_{jq,2003,2004} + X_{jq2002}\beta_3 + \delta_q + \zeta_{jqt}$$

Like Model 1, Model 2 is executed in three distinct versions depending on the frequency of charter school failure during school years 2003 through 2005. In version 1, $FailAYP_{jq2003\text{-}2005}$ equals 1 if school *j* failed at least once; in version 2, $FailAYP_{jq2003\text{-}2005}$ equals 1 if school *j* failed at least twice; in version 3, $FailAYP_{jq2003\text{-}2005}$ equals 1 if school *j* failed at least three times. The comparison group is charter schools that never failed AYP, charter schools that never failed or failed only once, and charter schools that never failed or failed once or twice, respectively. The other control variables are identical to Model 1.

To further test the relationship between charter school AYP failure and charter school closure in different accountability contexts, I re-estimate Model 2 substituting the mild, moderate, and strict ratings based on the Shober Accountability Index. Admittedly, the Shober Accountability Index is an imperfect measure. The Shober Index is based on few criteria and draws those criteria from state law as of 2002, prior to the onset of NCLB and well in advance of the closure period I examine. The impact of the Shober Accountability Index, however, will provide useful context for my original analysis and may spur a deeper investigation of specific factors that impact the likelihood failing charter schools will close.

It is possible that certain accountability provisions may drive charter school closure decisions more than others. In the version of Model 2 that uses my own index, I rate states as mild, moderate, or strict according to the rigor of their charter school law along four dimensions – 1) criteria for charter renewal, 2) requirements for annual performance reports, 3) events that justify mid-contract termination, and 4) maximum charter length. Of these dimensions, the criteria for charter renewal and mid-charter termination are most directly associated with charter school closure, so in Model 3, I interact only these two dimensions with

charter school AYP failure. I am most interested in the impact of these two dimensions at the outer margin – where these two provisions are incorporated into charter law using the strictest terms. So, in Model 3, I group states into two categories—strict and non-strict[63]—based on their state's charter law provisions on the two dimensions.

$$(3)\ Pr(Close_{jq2006\text{-}08} = 1) = \alpha + (FailAYP_{jq2003\text{-}2005} \times RenewalTerm_{q,2006})\lambda + \beta_1(Enroll_{jq2006} - Enroll_{jq2002})/Enroll_{jq2002} + \beta_2 Rank_{jq,2003,2004} + X_{jq2002}\beta_3 + \delta_q + \zeta_{jqt}$$

Model 3 is also executed in three distinct versions depending on the frequency of charter school failure during school years 2003 through 2005. The other control variables are identical to Models 1 and 2.

Results

Model 1: Schools that failed at least once between 2003 and 2005

Charter Schools

The strong positive relationship between charter schools that failed AYP at least once and their likelihood of closure supports my hypothesis that states and districts use charter schools' AYP performance to inform their decisions about which schools to close. Table 14 displays the Model 1 results for both charter schools and non-charter public schools. Charter schools that failed AYP at least once were 5.8 percentage points more likely to close than charter schools that never failed AYP, holding all other values at their means ($p < 0.01$). This finding accounts for schools' within-state proficiency rate rank which suggests that states and districts were not just closing the worst-performing schools but rather, there is a distinct

[63] In this case, non-strict is comparable to mild and moderate ratings on the full indices.

association of AYP failure on closure. There is, however, a small but statistically significant relationship between charter schools' within-state relative performance on state exams and their likelihood of closure. All else equal, a charter school whose mean proficiency rate is in the 10th percentile in its state was one percentage point more likely to close than a charter school in the 20th percentile ($p < 0.01$). Even though they are statistically significant, these estimates are fairly small compared to a 5.8 percentage point increase in the likelihood of closure for charter schools that failed AYP at least once.[64]

As I predicted, charter schools with declining enrollment between 2002 and 2006 were more likely to close than schools that maintained or grew their enrollment during the same period. All else equal, a school with a 25 percent decline in enrollment between 2002 and 2006 was 0.64 percentage points more likely to close than a school with a 15 percent decline in enrollment during the same period ($p < 0.01$). Conversely, a school with a 25 percent increase in enrollment between 2002 and 2006 was 0.64 percentage points less likely to close than a school with a 15 percent increase in enrollment, all else equal. Charter school size, in terms of 2002 total student enrollment, is negatively associated with school closure. Charter schools with fewer students as of 2002 were more likely to close than charter schools with more students as of 2002, all else equal. For example, a charter school with 400 students was 1.9 percentage points more likely to close than a school with 500 students, all else equal ($p < 0.01$).

[64] Table 15 presents the results for the same model, but includes squared and cubic versions of a school's within-state proficiency rate ranking. While the inclusion of non-linear terms has a mixed and minor impact on the AYP failure estimates, it leads to a rather large increase in the negative coefficient on the within-state proficiency rate ranking term for both charter schools and non-charter public schools.

School type (i.e., grades served) and, potentially, schools' geographic location also seemed to influence a charter school's likelihood of closure. Charter elementary schools were the least likely to close. Compared to elementary schools, schools with alternative configurations (e.g., serving all grades K-12) were 10 percentage points more likely to close ($p < 0.01$). Similarly, middle schools and high schools were 8.2 and 7.0 percentage points, respectively, more likely to close than elementary schools ($p < 0.01$ and $p < 0.05$, respectively). Surprisingly, the results indicate that charter schools in cities were 4.3 percentage points *less likely* to close than charter schools in towns, yet, although substantive, this estimate is not significant at the 0.10 level ($p = 0.20$). Compared to being located in a town, charter schools in rural areas and suburban areas were not significantly more likely to close.

In most cases, schools' demographic composition was significantly related to their likelihood of closure, but the direction of the relationship varies across demographic groups. Charter schools that serve a larger share of black students were significantly more likely to close than schools with smaller shares of these students. All else equal, a 10-percentage point increase in the share of black students increased a school's likelihood of closure by 0.9 percentage points ($p < 0.01$). Conversely, charter schools that serve a larger share of Hispanic and Asian students were less likely to close. For Hispanic students, the estimate is fairly substantive and significant; for Asian students, the result is *very large* and nearly significant at the 0.10 level. All else equal, a 10-percentage point increase in the share of Hispanic students reduced the school's likelihood of closure by 1 percentage point ($p < 0.05$). All else equal, a 10-percentage point increase in the share of Asian students reduced the school's likelihood of closure by more than 4.2 percentage points ($p = 0.106$). The relationship between charter

school closure and a charter school's share of Native American students is positive and fairly substantive—a 10 percentage point increase in the share of Native American students is associated with a 0.7 percentage point increase in the likelihood of closure, but this result is statistically insignificant at the 0.10 level ($p = 0.198$).

Non-Charter Public Schools ("NCPS")

Unlike charter schools, there is a very small and statistically insignificant relationship between non-charter public schools that failed AYP at least once and their likelihood of closure. Compared to non-charter public schools that never failed AYP, non-charter public schools that failed AYP at least once were 0.2 percentage points more likely to close, holding all other values at their means, though this result is not statistically significant at the 0.10 level ($p = 0.178$). At the margin, the effect of failing AYP at least once for charter schools is 29 times larger than the effect for non-charter public schools. Additionally, the probability of closure for a charter school that failed AYP at least once is 5.6 percentage points higher than the probability of closure for a non-public charter school that failed AYP at least once. As stated in Table 14, the difference between the probit coefficients for charters ($\beta = 0.358$) and non-charter public schools ($\beta = 0.032$) is sizeable and statistically significant at the 0.01 level ($p=0.005$).

At the margin, the effect of non-charter public schools' within-state proficiency rate ranking on school closure seems trivial, if statistically significant at the 0.01 level. All else equal, a non-charter public school in the 20^{th} percentile of performance was 0.4 percentage points less likely to close than a non-charter public school in the 10^{th} percentile of performance in their state ($p < 0.01$). Yet, the difference in the likelihood of closure between the highest performing school and the lowest performing school in the state is 3.9 percentage points. In other words, a

non-charter public school at the very bottom of the distribution of performance (i.e., 1st percentile) was 3.9 percentage points more likely to close than a school at the very top of the distribution of performance (i.e., 99th percentile). For a comparison of the size of the marginal effects of schools' percentile rank on school closure for charters and non-charter public schools, see Figure 13. Even though the difference between the marginal effects seems rather large, the difference in probit coefficients for schools' within-state proficiency rank is statistically insignificant since there are so few charter schools compared to non-charter public schools in the U.S. ($p = 0.96$).

Like charters, the effect of the change in school enrollment between 2002 and 2006 for non-charter public schools is substantive and significant ($p < 0.01$). At the margin, a non-charter public school with a 15 percentage point decline in enrollment between 2002 and 2005 was 0.74 percentage points less likely to close than a school with a 25 percentage point decline in enrollment during the same period. Figure 14 displays the difference between the marginal effects of enrollment change on closure for charter schools and non-charter public schools. Not only is the marginal effect greater for charter schools than non-charter public schools, on average, the decline in the likelihood of closure is steeper. This finding indicates that, for charter schools, small *decreases* in enrollment over time had a much larger net impact on a school's likelihood of closure than similarly sized *increases* in enrollment over time.

Much like charter schools, the size of a non-charter public school impacted its likelihood of closure. Smaller non-charter public schools were more likely to close than larger non-charter public schools. In other words, a school with 400 students was 0.8 percentage points more likely to close than a school with 500 students all else equal ($p < 0.01$). Yet, this effect is less

than half the effect for charter schools (1.9 percentage points). Non-charter schools' geographic location—especially schools located in cities—is significantly related to the probability of closure, all else equal. Non-charter public schools located in cities and suburban areas were 1.4 and 0.4 percentage points more likely to close than non-charter public schools located in towns, respectively ($p < 0.01$ and $p < 0.10$). In addition, non-charter public schools located in rural areas were 0.9 percentage points less likely to close ($p < 0.01$).

While the relationships between school demographics variables and school closure are smaller in size for non-charter publics than for charters, schools' demographic composition is significantly related to their likelihood of closure. For non-charter publics, the share of students in a school is negatively associated with school closure for all demographic groups except for the share of a school's black students. Like charter schools, non-charter public schools that served a larger share of black students were significantly more likely to close than schools with smaller shares of these students. All else equal, a 10-percentage point increase in the 2002 share of black students is associated with a 0.3 percentage point increase in a school's likelihood of closure ($p < 0.01$). Conversely, non-charter public schools that served a larger share of Hispanic, Asian, and Native American students were *less likely* to close. All else equal, a 10-percentage point increase in the 2002 share of Hispanic students is associated with a reduction in the school's likelihood of closure by 0.17 percentage points ($p < 0.05$). All else equal, a 10-percentage point increase in the 2002 share of Asian students or Native American students is associated with a reduction in the school's likelihood of closure by approximately 0.4 percentage points in both cases ($p < 0.01$ for both).

Model 1: Schools that failed at least twice or all three Years between 2003 and 2005

Charter Schools and NCPS

Across all models, the effect of AYP failure on school closure is much larger for charter schools than non-charter public schools, which suggests that charter schools' performance was more highly scrutinized and the stakes for poor performance were higher. For charter schools, the largest and most precisely measured effect on school closure—5.8 percentage points—is for schools that failed at least once. Once schools that failed AYP *all three years*, however, were only 4.5 percentage points more likely to close than schools that failed once, twice, or not all ($p < 0.10$). On the other hand, for non-charter public schools, the effects of AYP failure on school closure grow increasingly larger as schools failed more frequently during 2003 to 2005 (i.e., failing at least twice and failing all three years). Non-charter public schools that failed AYP at least twice were 0.7 percentage points more likely to close than non-charter public schools that failed AYP only once or not at all—an increase of 0.5 percentage points over the result for schools that failed at least once ($p < 0.01$). Non-charter public schools that failed AYP all three years were 1.6 percentage points more likely to close than schools that failed once, twice, or not at all—an increase of 1.4 percentage points over the result for schools that failed AYP at least once and 0.9 percentage points over the result for schools that failed AYP at least twice ($p < 0.01$).

These results seem to suggest that charter schools were under more pressure to perform since the stakes for failing AYP at least once are as high or higher than failing AYP at least twice or three times. For non-charter publics, the stakes grew increasingly higher the more often a school fails AYP, but for schools that failed AYP at least once, the likelihood of closure was negligible. Yet, even though the estimates for charter schools are substantively

larger than for non-charter public schools, charter schools account for such a small proportion of all public schools that the difference in the probit estimates for AYP failure between charter schools and non-charter schools is significant only in Model 1—schools that failed AYP at least once.

Model 2: Charter Schools with Accountability Index

Charter schools that failed AYP at least once

Using my own index, there is no evidence that failing charter schools were more likely to close in states with strict accountability provisions than in states with mild accountability provisions. Table 16 displays the Model 2 results for charter schools under each of the indices. The result for charter schools that failed AYP at least once in states with strict accountability provisions is nearly equivalent to the result for charter schools that failed at least once in states with mild accountability provisions. Compared to all charter schools that never failed AYP in any state, regardless of its accountability provisions, charter schools that failed AYP at least once in states with mild accountability provisions were 8.8 percentage points more likely to close ($p < 0.01$). Compared to charter schools that never failed AYP, charter schools that failed AYP at least once in states with strict accountability provisions were 9.0 percentage points more likely to close ($p < 0.01$). There is a negligible and insignificant effect of moderate accountability provisions on failing charter schools' likelihood of closure. These findings suggest that multiple legal and accountability provisions may have worked together to influence charter school authorizers' decisions about closing failing charter schools.

There is compelling evidence, however, that failing charter schools were more likely to close in states with "strict" charter school laws as measured by Shober Accountability Index. Using the Shober Accountability Index, charter schools that failed AYP at least once in states with "strict" charter laws were 10 percentage points more likely to close than charter schools in any state that never failed AYP ($p < 0.01$). The relationship between failing AYP at least once and charter school closure in states with mild or moderate provisions is much smaller and imprecise; this finding suggests that accountability provisions identified as "strict" in the Shober Accountability Index may create the conditions under which charter school authorizers are more likely to rigorously monitor charter school performance and take action when deemed necessary.

Charter schools that failed AYP at least twice or three times

Using my own index, the findings for charter schools that failed AYP at least twice and all three years are similar to those of charter schools that failed AYP at least once. For charter schools that failed AYP at least twice, there was little difference in the impact of mild and strict accountability provisions on closure. Charter schools that failed AYP at least twice in states with mild accountability provisions were 9 percentage points more likely to close than charter schools in any state that never failed AYP or failed AYP only once ($p < 0.05$). Similarly, charter schools that failed AYP at least twice in states with strict accountability were 7.2 percentage points more likely to close ($p < 0.05$). For charter schools that failed AYP all three years, the effects of failure on closure in mild and strict accountability states is substantive but insignificant at the 0.10 level in both cases. Across all models, there is no evidence of a significant relationship between failing AYP and school closure for charter schools in states with

moderate provisions.[65] Since the effect sizes decline or become insignificant as a charter school fails AYP more frequently, it is unlikely that the accountability provisions in my index created an accountability environment that punished charter schools more harshly the more times it failed AYP.

The effects of the Shober Accountability Index are most consistent with my hypothesis that strict accountability provisions create the conditions for charter school authorizers to hold charter schools accountable, through closure, for their AYP performance. Furthermore, the provisions of the Shober Index seem to have had the greatest impact on charter schools that fail AYP the most. Compared to charter schools that failed AYP never, once, or twice in any state, charter schools that failed AYP three times in states with strict provisions were 16.3 percentage points more likely to close ($p < 0.01$). This result is between five and six percentage points greater than the result for schools that failed AYP at least twice ($\beta = 0.11$) or at least once ($\beta = 0.10$).

Model 3: Specific Accountability Provisions from the EKD Index

Unlike the Shober Accountability Index, my own index seems to have a mixed impact on failing charter schools' likelihood of school closure. For the most part, failing charter schools in states with mild provisions were equally as likely to close as failing charter schools in states with strict provisions. To further explore the relationship between accountability provisions and

[65] This result holds across all three models – 1) charter schools that failed at least twice in states with moderate provisions compared to schools in any state that never failed, 2) charter schools that failed at least twice in states with moderate provisions compared to schools that never failed or failed only once, 3) charter schools that failed all three years in states with moderate provisions compared to schools that never failed, failed only once, or failed only twice in any state.

charter schools' likelihood of closure, I isolate two provisions (out of the four) that, when defined "strictly" in state law, I think are most likely to drive closure. These two provisions are: the maximum length of a charter contract, in years; and whether authorizers could revoke the charter at any time due to charter schools' failure to satisfy academic progress set forth in the charter. In this model, however, I only use two ratings—"strict" and "non-strict"—to define states' restrictiveness along these two provisions. Table 17 displays the results from Model 3.

Much like the full index in Model 2, this narrower index also has a mixed impact on failing charter schools' likelihood of closure. Charter schools that failed AYP at least once in states with strict provisions were 8.4 percentage points more likely to close than charter schools in any state that never failed AYP ($p < 0.05$). Charter schools that failed AYP at least once in states with non-strict provisions were 5.7 percentage points more likely to close than charter schools in any state that never failed AYP ($p < 0.05$). The only other significant result is for charter schools that failed AYP all three years in states with non-strict provisions. Charter schools that failed AYP all three years are 6.8 percentage points more likely to close than charter schools in any state that never failed or failed AYP once or twice ($p < 0.10$). These results provide little additional evidence in support of my hypothesis and suggest that other provisions—perhaps those included in the Shober Accountability or others not included here—have a greater influence on closure decisions than the provisions in my full or narrow index.

Exploration of Additional Factors

So far, this paper has only examined the impact of accountability provisions on the likelihood that charter schools that fail AYP will close. These accountability provisions account for states' *ex-post* policy environment, or the accountability conditions that charter schools that

fail AYP face once they are already open and operating. These provisions, however, do not account for states' *ex-ante* policy environment, those conditions that influence the supply of charter schools in the market. Perhaps states with relatively flexible *ex-ante* charter law provisions encourage more chartering activity in ways that lead to increased churn. New charter schools open fairly frequently such that closing schools has a limited impact on the total number of charter schools in operation and perhaps political opposition or public outcry is relatively minimal. Mediocre operators may face fewer barriers to entry in states with relatively flexible *ex-ante* provisions, and if financial mismanagement or other operational deficiencies are associated with AYP failure, then *ex-ante* provisions, like fewer barriers to entry, may inflate my *ex-post* estimates.

To test the impact of ex-ante charter law provisions on the likelihood that charter school that fails AYP will close, I add a measure of states' *ex-ante* charter law provisions to the model that includes the *ex-post* EKD index. This measure, developed in 2006 by the Center for Education Reform (CER), categorizes state charter school laws based on 10 provisions related to autonomy and flexibility. The provisions were scored on a scale of 0-5 and then summed to determine an overall "score" which served as the basis for a state's ranking. See Appendix E for a summary of the CER overall scores for each state and a discussion of the challenges presented by using the CER index.[66] States were then assigned a letter grade from A-F depending on the fifth of the overall score distribution in which they fall. States with an A or B rating were labeled as "Strong" while states with a C, D, or F rating were labeled as "Weak." Instead of the CER's binary labels or the too-disaggregated letter grades, I divide the states into three groups

[66] Raw scores range from 5.5 to 46.5 with a mean of 28.9 and a standard deviation of 11.0.

depending on where they fall in the score distribution. States with raw scores in the approximate highest third of the distribution, I rate as "highly flexible"; states with raw scores in the approximate middle third of the distribution are "moderately flexible"; states with raw scores in the approximate lowest third of the distribution are "mildly flexible." Table 10 outlines the categorical ratings of mild, moderate, and high flexibility for each state. Figure 11 displays the distribution of schools across the index score range.

To similarly test the impact of *ex-ante* provisions in the model that includes the Shober Index rather than my own, I include a flexibility measure that Shober et al., constructed as part of their review of states' 2002 charter school laws. This measure rates charter school laws on state charter law provisions related to the ease of opening a charter school, the influence of local control, the types of organizations that can serve as authorizers, funding, and restrictions on employees. See Appendix D for a full description of the provisions included in the flexibility scale. As I did with the CER Index, I divide states amongst three groups—highly flexible, moderately flexible, mildly flexible—based on whether they fall in the approximate highest third, middle third, or bottom third of the rating's distribution, respectively. Table 10 outlines the categorical ratings of mild, moderate, and high flexibility for each state. Figure 12 displays the distribution of schools across the index score range. Appendix E lists each state's rating on the Shober flexibility scale.

As displayed in Table 18, adding the *ex-ante* CER measures to the model has little impact on the size or direction of the coefficients on the *ex-post* EKD index. The estimates on the CER measures are imprecisely measured and do not consistently indicate the direction of the relationship between ex-ante provisions and the likelihood a failing charter school will close.

The inclusion of the Shober Flexibility Index, on the other hand, does reveal a sizeable increase in the relationship between mild accountability provisions and the likelihood of closure for a charter school that failed AYP at least once (Table 19). The inclusion of the Shober Flexibility Index also leads to a small increase in the size of the coefficient for charter schools that failed AYP at least once and all three years in states with strict accountability provisions. In both of the models with ex-ante and ex-post measures, however, the highly flexible indicator is dropped due to collinearity with the other index measures.

Discussion

Opposition to charter schools has existed since their inception. Recent reports, however, indicate that opposition to charter schools may be increasing, even among groups like the NAACP that have traditionally supported school choice policies (Nix, 2016). And there is some evidence that supports the NAACP's case. My findings suggest that charter school closures disproportionately impact Black students but not students from other minority groups. Even though Hispanic students are overrepresented in schools that failed AYP, they are *underrepresented* in closed schools. Black students, on the other hand, are enrolled in closed charter schools at 1.5 times the rate at which they are enrolled in all charter schools. Yet the demand from families for high-quality charter schools remains high (NYC Charter Center, 2016; Dynarksi, et al., 2010). Efforts to improve the quality of all charters, remedy student segregation, and refine practices would benefit students and communities alike, especially Black students.

To the extent that charter school authorizers are directly or indirectly involved in charter school closures,[67] these findings suggest that the response of charter school authorizers to AYP failure was stronger than the response of states and districts to AYP failure in non-charter public schools. Charter school performance may be subject to stricter scrutiny than non-charter public schools or charter school authorizers may have clearer processes for closing charter schools than states and districts have for closing non-charter public schools. Further, unlike traditional districts, charter school authorizers generally serve in an oversight capacity only, so they infrequently provide intensive technical support or intervene in charter school operations; their primary mechanisms for intervention is non-renewal or termination. Charter schools, on average, may operate in districts with greater levels of competition and school choice than non-charter public schools.[68] In these environments, closing schools may be a more feasible reform strategy as students have sufficient options once their school closes.

My findings support my hypothesis that strict accountability provisions, as measured by the EKD index, are associated with higher rates of school closure for schools that failed AYP. Yet, mild accountability provisions seem to have an equivalent effect. If states with mild accountability provisions, as measured by the EKD index, also have authorizers whose monitoring and evaluation of school performance extends further than the law requires, then low-performing schools may be more likely to close in these states. It is also possible that mild

[67] Alternatively, authorizers terminate or non-renew charter contracts in response to charter school closure rather than to force charter schools to close.

[68] In each of my models, I include an indicator of whether the charter school is located in a city, suburb, rural area, or town. There may be variation within those designations that contributes to higher rates of closure amongst charter schools.

provisions attract more low-quality charter operators, so these results may reflect a culling of poor performing charter operators.

The inclusion of the Shober Accountability Index provides an interesting counterpoint to the somewhat mixed impact of the EKD index on closure for failing charter schools. Even though it is less relevant than the EKD Index to this paper, the results of the Shober Accountability Index are more in-line with my original hypothesis and appear to only be strengthened by the inclusion of measures that account for states' *ex-ante* policy environment. These findings also suggest that increased flexibility may not come at the expense of accountability—I find no evidence that controlling for the flexibility of states' provisions has any impact on the relationship between accountability provisions and the likelihood of closure for schools that failed AYP.

Some policymakers believe that increased authorizer accountability will improve charter sector outcomes and improve the quality of charter schools on the whole. Efforts to increase authorizer accountability, however, may have unintended consequences. Increased accountability may have a negative effect on the supply of high-quality authorizers and reduce flexibility and autonomy in ways that inhibit innovation and replication. Serious questions remain about the best ways to increase the rigor of the charter school application process, predict which operators will maximize student outcomes, and change the incentives authorizers face to open and rigorously monitor charter schools. While the results herein provide some evidence that accountability provisions in state charter school laws matter, the exact mechanisms and policy factors remain unclear. Further exploration is needed to identify those

ex-ante and ex-post policy factors that encourage innovation and activity but also promote a rigorous accountability process that promotes excellence.

CONCLUSION

No Child Left Behind drastically changed the education landscape and expanded the federal role in education. While some significant progress was made, states and districts have more work to do to close achievement and opportunity gaps. The studies herein provide rich context about states' implementation of federal education policy that policy-makers can use to inform future reform efforts. My findings suggest that states' esoteric NCLB implementation decisions led to extreme differences in the number of states' schools that failed AYP and therefore, were subject to NCLB consequences. I find some evidence that states use school closure to avoid implementing NCLB consequences in failing schools, and that, for the most part, decisions to close failing schools are unaffected by political factors, as measured by the strong presence of teachers' unions. Failing charter schools are particularly susceptible to school closure, and black charter school students experience closure at much higher rates than students of any other racial or ethnic group.

In response to these findings, policy-makers should carefully consider using closure as the primary strategy for reforming schools, especially schools with large shares of black students. Charter school closure, while more prevalent, is, in some ways, more preventable. Lack of authorizer oversight and perverse incentives for authorizers are serious concerns. I find that, in some instances, stronger charter laws lead to increased rates of closure for failing charter schools, yet instead of focusing on tweaking their laws, perhaps policy-makers should focus on replicating effective practices, screening and overseeing authorizers, encouraging authorizers to more rigorously screen charter applicants, and changing authorizers' incentives. Future research should further explore the policy factors that attract high-performing charter

school operators and result in a relatively high-performing charter school sector. Due to data constraints, I am unable to identify the specific reasons that charter schools closed and the mechanisms that led to that closure. Future research should continue to explore these reasons and mechanisms to determine how they shape the charter school landscape and stem from specific policy provisions.

As with any education policy, the return of control to the states under ESSA has benefits and drawbacks. On the one hand, states can develop accountability systems that provide rich, nuanced information on school and student performance across multiple dimensions. This information helps states and districts tailor support, prioritize interventions, identify schools that are "beating the odds," and replicate successful strategies to lift the performance of all schools. On the other hand, as discussed in Chapter I, without careful attention to the precise definitions and formulae states use, school outcomes and subsequent consequences may a function of state idiosyncrasies rather than substantive differences in school performance. Therefore, policy-makers will have less-comparable information on which to assess state and district performance. Just as there are schools that lag or beat the odds, so are there districts and states that should be held accountable or rewarded for the performance of their students.

TABLES

Table 1: Chapter I: Characteristics of Schools by Whether They Failed to Make AYP

	2003-2005		
	Failed all three years	Failed at least once	Never failed
Number of Schools	9,382	37,909	42,883
Average Enrollment	891	681	469
Student/Teacher Ratio	17.6	16.5	15.7
Percent of Students…			
Eligible for Free/Reduced Lunch	55.0%	49.5%	34.1%
White	39.3%	52.1%	73.9%
Black	29.9%	23.3%	9.9%
Hispanic	23.8%	18.3%	11.4%
Asian	4.0%	3.4%	3.4%
Percent of Schools…			
Eligible for Title I	67.9%	61.0%	44.9%
Serving Primary Grades	32.8%	46.7%	71.5%
Serving Middle Grades	35.2%	25.7%	14.2%
Serving High Grades	31.9%	27.6%	14.3%
Located in City	41.2%	31.1%	18.3%
Located in Suburb	32.8%	30.5%	33.9%
Located in Town or Rural Area	24.4%	33.6%	46.7%

Notes to Table 1: The data on school characteristics are from the Common Core of Data, 2001-2002. For schools in Tennessee, data on student ethnicity comes from 1998-1999 instead of 2001-2002 and data on free/reduced price lunch eligibility is unavailable. Aside from the Percent of Students who are Asian, all differences in means between the second and third columns are statistically significant at the 0.01 level.

Table 2: Chapter I: States' Early Policies for Determining AYP, with States Sorted by the Fraction of Schools Failing to Make AYP in 2004

STATE	% of Schools Failing to Make AYP in…			Confidence Intervals Applied to Proficiency Rates During Early Years of NCLB	Grades Tested in 2004		Avg. # of Student Subgroups Contributing Proficiency Rates Toward AYP Ratings, 2004[1]
	2003	2004	2005		Math	Reading	
Iowa	0.8%	**4.7%**	7.3%	98%	4,8,11	4,8,11	N/A
Wisconsin	4.5%	**4.8%**	2.4%	99%	4,8,10	4,8,10	1.9
Louisiana	6.4%	**5.0%**	17.2%	99%	4,8,10	4,8,10	4.2
Texas	8.2%	**5.7%**	11.6%	95%	3-8,10	3-8,10	5.2
Wyoming	15.1%	**7.1%**	18.9%	95%	4, 8, 11	4, 8, 11	1.7
North Dakota	31.7%	**7.8%**	11.5%	99%	4,8,12	4,8,12	N/A
North Carolina	9.1%	**9.1%**	9.2%	99%	3-8,10	3-8,10	3.9
Kansas	29.3%	**9.2%**	8.8%	99%	4,7,10	5,8,11	2.0
Vermont	12.7%	**12.7%**	3.3%	99%	4,8,10	2,4,8,10	1.8
Washington	22.0%	**13.8%**	19.4%	99%	4,7,10	4,7,10	2.6
Tennessee	46.7%	**14.6%**	7.6%	95%	3,5,8,HS[2]	3,5,8,HS[2]	2.8
Montana	20.1%	**14.6%**	6.1%	95%	4,8,10	4,8,10	4.8
Pennsylvania	35.5%	**14.8%**	19.3%	95%	5,8,11	5,8,11	2.4
Michigan	24.0%	**15.5%**	7.7%	None[3]	4,8,11	4,7,11	2.2
Maine	26.5%	**15.7%**	26.5%	95%	4,8,11	4,8,11	3.5
Ohio	24.2%	**15.8%**	24.2%	None	4,6,9	4,6,9	2.5
Arizona	23.3%	**16.7%**	13.2%	99%	3,5,8,10	3,5,8,10	3.3
Rhode Island	31.1%	**17.1%**	11.4%	None[4]	4,8,11	4,8,11	3.4
Connecticut	14.7%	**17.9%**	20.4%	99%	4,6,8,10	4,6,8,10	2.7
Utah	35.8%	**18.2%**	13.1%	99%	3-8,11	3-8,10	5.5
Idaho	35.3%	**18.2%**	42.8%	None[4]	3,4,7,8,10	3,4,7,8,10	3.1
Georgia	36.2%	**20.2%**	18.1%	95%	3-8,11	3-8,11	4.1
Maryland	35.2%	**20.6%**	23.1%	99%	3-8,10	3-8,10	5.8
New Mexico	20.7%	**20.7%**	52.6%	99%	4,8,11	4,8,11	N/A
Minnesota	7.8%	**22.3%**	13.1%	95-99%	3,5,7,11	3,5,7,10	3.2
Arkansas	22.2%	**22.7%**	42.5%	95%	4,6,8,HS[5]	4,6,8,11	2.9
Colorado	37.6%	**23.1%**	27.5%	95%	5-10	3-10	3.1
Mississippi	23.1%	**23.6%**	11.8%	99%	3-8,10[6]	3-8,10	3.7
Delaware	54.0%	**23.8%**	26.4%	98%	3,5,8,10	3,5,8,10	4.1
Kentucky	40.7%	**24.3%**	25.7%	99%	5,8,11	4,7,10	3.4
Indiana	23.2%	**24.4%**	40.8%	99%	3,6,8,10	3,6,8,10	3.9
Missouri	48.3%	**25.2%**	34.8%	99%	4,8,10	3,7,11	2.5
New York	25.9%	**25.9%**	18.7%	90%	4,8,HS[7]	4,8,HS[7]	3.2

Massachusetts	44.3%	**26.5%**	29.0%	95%	4,6,8,10	3,4,7,10	2.6
New Jersey	42.4%	**28.4%**	37.8%	95%	4,8,11	4,8,11	3.3
West Virginia	40.5%	**28.5%**	16.9%	99%	3-8,10	3-8,10	3.1
Illinois	32.4%	**28.6%**	26.3%	95%	3,5,8,11	3,5,8,11	2.7
Nebraska	52.6%	**29.2%**	42.6%	95%	4,8,11	4,8,11	N/A
Oregon	29.7%	**29.2%**	32.6%	99%	3,5,8,10	3,5,8,10	3.9
New Hampshire	31.4%	**29.6%**	46.8%	99%	3,6,10	3,6,10	3.3
South Dakota	33.6%	**33.6%**	13.8%	99%	3-8,11	3-8,11	3.5
California	45.9%	**34.4%**	38.8%	99%	2-8,10[8]	2-8,10[8]	3.7
Nevada	42.6%	**34.5%**	60.0%	95%	3,5,8,11	3,5,8,11	4.7
Oklahoma	22.8%	**35.9%**	1.5%	95% for campus-wide group only[4]	3,5,8,HS[9]	3,5,8,HS[9]	1.5
Alaska	57.7%	**39.0%**	40.9%	99%	3-10	3-10	3.4
Virginia	40.5%	**40.7%**	24.9%	None	3,5,8, HS[10]	3,5,8, HS[10]	2.9
South Carolina	79.7%	**42.5%**	51.7%	68% starting in 2005	3-8,10	3-8,10	4.0
Hawaii	60.6%	**47.5%**	65.9%	68%	3,5,8,10	3,5,8,10	3.2
Alabama	4.2%	**68.3%**	46.7%	99%	4,6,11	4,6,8,11	2.8
Florida	82.2%	**76.3%**	64.0%	None	3-10	3-10	5.2

1. The number of subgroups reported here are averaged across math and reading.
2. Proficiency and participation rates are based on the cohort of students enrolled in Algebra I and English II courses which may be taken at varying grade levels in high school.
3. Only very small schools were allowed to use confidence interval adjustments.
4. Although these states did not use confidence interval adjustments for subgroups, they used relatively large minimum required subgroup sizes.
5. To calculate proficiency and participation rates, Arkansas officials combine students' assessment results on End-of-Course (EOC) exams in Algebra I and Geometry.
6. To calculate proficiency and participation rates, Mississippi matches 10th grade students with their Algebra I test scores whether or not they take the exam in 10th grade.
7. Proficiency and participation rates are based on the cohort of students who are in courses culminating in State math and ELA High School Regents Exams.
8. California requires all 10th grade students to take the California High School Exit Exam (CAHSEE).
9. Proficiency and participation rates are based on the cohort of students who are in courses culminating in State math and ELA End-of-Instruction (EOI) Exams.
10. Proficiency and participation rates are based on the cohort of students who are in courses culminating in math and ELA End-of-Course (EOC) exams.

Table 3: Chapter II: Summary Statistics for All Variables

Variable	All Publics Mean	SD
School Failed at least Once	42%	49%
School Failed at least Twice	22%	41%
School Failed Three Times	12%	33%
School Closed between 2004 and 2011	9%	28%
School Closed between 2006 and 2011	6%	24%
Proficiency Rate	50%	29%
Percent Change in Enrollment Between 2002 and 2006	-0.4%	21%
Total School Enrollment	569	446
School Eligible for Title I	56%	50%
Pupil/Teacher Ratio	16.3	11.2
Percent of States' Districts Covered by Collective Bargaining	70%	36%
Percent of States' with more than 75% of Districts Covered by Collective Bargaining	67%	47%
High Schools	17%	38%
Middle Schools	18%	39%
Other Schools	4%	19%
Elementary Schools	61%	49%
Located in a City	25%	44%
Located in a Rural Area	28%	45%
Located in a Suburb	35%	48%
Located in a Town	12%	32%
Percent Black Students	17%	26%
Percent Hispanic Students	15%	23%
Percent Native American Students	2%	9%
Percent Asian-American Students	3%	8%
Percent Economically Disadvantaged	41%	28%
Percent of Residents with High School Diplomas	80%	10%
Percent of Residents with Bachelor's Degrees	22%	12%
Percent of Residents who are Black	10%	15%
Percent of Residents who are White	73%	23%
Percent of Residents who are Hispanic	9%	11%

Note: Other schools are defined by the CCD as schools whose configuration doesn't fall within the elementary, middle, or high school definition. Also includes ungraded schools.

Table 4: Chapter II: Characteristics of Schools (2002) Nationally, by Closure 2006-2011

	Closed	2006-2011 Remained Open	*p-values*
Number of Schools (2006-2011)	4,354	66,952	
Number of Schools (2004-2011)	*6,458*	*67,106*	
School Failed at least Once	48.6%	41.0%	< 0.001
School Failed at least Twice	27.3%	22.1%	< 0.001
School Failed all Three Years	17.6%	11.8%	< 0.001
Percent Change in Enrollment (2002/2004 to 2006)	-7.3%	0.2%	< 0.001
Average Enrollment (2002)	369	598	< 0.001
Student/Teacher Ratio	16.5	16.3	0.298
Percent of State's Districts with Union Contracts > 75%	73.0%	65.9%	< 0.001
Percent of Students…			
Eligible for Free/Reduced Lunch	52.0%	40.0%	< 0.001
White	55.5%	64.5%	< 0.001
Black	28.4%	15.1%	< 0.001
Hispanic	12.2%	14.9%	< 0.001
Native American	1.7%	1.8%	0.387
Asian	2.1%	3.5%	< 0.001
Percent of Schools…			
Eligible for Title I	67.1%	55.2%	< 0.001
Located in City	36.3%	23.8%	< 0.001
Located in Rural Area	25.7%	28.8%	< 0.001
Located in a Suburb	26.5%	35.9%	< 0.001
Located in a Town	11.6%	11.5%	0.780
Serving Primary Grades	65.8%	62.6%	< 0.001
Serving Middle Grades	23.7%	19.9%	< 0.001
Serving High Grades	11.8%	19.2%	< 0.001
Percent of District…			
with HS diplomas	78.6%	80.1%	< 0.001
with Bachelor's Degrees	19.4%	22.6%	< 0.001
White	70.4%	73.6%	< 0.001
Black	15.5%	9.5%	< 0.001
Hispanic	7.6%	9.1%	< 0.001
District Median Income	$38,153	$43,878	< 0.001

Note: Data on school characteristics taken from the Common Core of Data, 2001-2002. For schools in Tennessee, data on student ethnicity taken from 1998-99 and data on free/reduced price lunch eligibility is unavailable. Sample is restricted to those schools operational and identified as charters in 2002.

Table 5: Chapter II: Characteristics of Schools (2002) Nationally, by AYP Status

		2003-2005		
	Never failed	**Failed at least Once**	**Failed at least Twice**	**Failed Three Years**
Number of Schools	37,841	30,239	16,583	8,132
Closed between 2006 and 2011	4.6%	6.9%	7.4%	7.8%
Change in Enrollment (2002/2004 to 2011)	0.2%	-1.6%	2.6%	-3.3%
Average Enrollment	490	734	836	949
Student/Teacher Ratio	16.0	16.8	17.0	17.5
Percent of Students…				
Eligible for Free/Reduced Lunch	34.0%	50.8%	54.2%	56.7%
White	72.5%	50.1%	42.9%	37.8%
Black	10.5%	24.6%	29.3%	31.8%
Hispanic	12.2%	19.2%	21.6%	24.0%
Asian	3.5%	3.6%	3.9%	4.1%
Native American	1.2%	2.4%	2.2%	2.2%
Percent of Schools…				
Eligible for Title I	54.0%	59.2%	59.2%	58.5%
Serving Primary Grades	71.5%	49.1%	42.9%	35.9%
Serving Middle Grades	15.4%	27.3%	31.8%	37.1%
Serving High Grades	14.9%	25.0%	27.0%	29.1%
Located in City	19.3%	34.2%	39.3%	43.0%
Located in Rural Area	11.0%	11.2%	9.5%	8.4%
Located in Suburb	33.6%	21.2%	16.9%	14.6%
Located in Town	36.2%	33.4%	34.3%	34.0%
Percent of District…				
with HS diplomas	81.6%	77.4%	76.4%	75.4%
with Bachelor's Degrees	23.5%	21.1%	21.0%	20.8%
White	78.2%	65.0%	60.5%	57.1%
Black	7.2%	14.3%	16.6%	17.3%
Hispanic	8.0%	11.0%	12.2%	13.8%
District Median Income	$45,807	$40,324	$39,811	$39,126

Note: Unweighted by enrollment. Data on school characteristics taken from the Common Core of Data, 2001-2002. For schools in Tennessee, data on student ethnicity taken from 1998-99 and data on free/reduced price lunch eligibility is unavailable.

Table 6: Chapter II: Main Results

	Baseline	Baseline + Interaction	Baseline + Prof Rank	Baseline + Interaction + Prof Rank
School Failed Once	**0.005***** 0.068 (0.006)	**0.013***** 0.175 (0.000)	**0.0002** 0.004 (0.886)	**0.007**** 0.098 (0.025)
School Failed Twice	**0.011***** 0.145 (0.000)	**0.016***** 0.207 (0.000)	**0.003** 0.045 (0.174)	**0.007*** 0.104 (0.067)
School Failed Three Times	**0.029***** 0.337 (0.000)	**0.031***** 0.351 (0.000)	**0.014***** 0.189 (0.000)	**0.016***** 0.209 (0.002)
School Failed Once*Union Coverage > 75%		**-0.01***** -0.157 (0.002)		**-0.008***** -0.139 (0.008)
School Failed Twice*Union Coverage > 75%		**-0.006** -0.088 (0.170)		**-0.005** -0.084 (0.196)
School Failed Thrice *Union Coverage > 75%		**-0.002** -0.024 (0.741)		**-0.002** -0.031 (0.668)
Percent Change in Enrollment (2002-2006)	**-0.065***** -0.965 (0.000)	**-0.065***** -0.966 (0.000)	**-0.062***** -0.954 (0.000)	**-0.062***** -0.955 (0.000)
Total Enrollment in 2002 (in hundreds)	**-0.009***** -0.136 (0.000)	**-0.01***** -0.136 (0.000)	**-0.01***** -0.125 (0.000)	**-0.01***** -0.125 (0.000)
Proficiency Rank (in tens)			**-0.004***** -0.056 (0.000)	**-0.004***** -0.056 (0.000)

Table 6: Chapter II: Main Results (Continued from Previous Page)

	Baseline	Baseline + Interaction	Baseline + Prof Rank	Baseline + Interaction + Prof Rank
	0.0004	**0.0003**	**-0.003**	**-0.003**
High Schools	0.006	0.005	-0.051	-0.051
	(0.856)	(0.878)	(0.140)	(0.138)
	0.019* **	**0.019*** **	**0.017*** **	**0.017*** **
Middle Schools	0.244	0.242	0.229	0.228
	(0.000)	(0.000)	(0.000)	(0.000)
	0.025* **	**0.025*** **	**0.018*** **	**0.018*** **
Other Schools	0.284	0.284	0.221	0.222
	(0.000)	(0.000)	(0.000)	(0.000)
	0.018* **	**0.018*** **	**0.018*** **	**0.018*** **
Located in a City	0.234	0.237	0.240	0.243
	(0.000)	(0.000)	(0.000)	(0.000)
	-0.011* **	**-0.0105*** **	**-0.01*** **	**-0.009*** **
Located in a Rural Area	-0.165	-0.165	-0.156	-0.155
	(0.000)	(0.000)	(0.000)	(0.000)
	0.005 **	**0.005** **	**0.006** **	**0.006** **
Located in a Suburb	0.075	0.076	0.083	0.085
	(0.020)	(0.018)	(0.014)	(0.012)
	0.005* **	**0.005*** **	**0.004** **	**0.004** **
Title I School	0.069	0.069	0.059	0.060
	(0.002)	(0.002)	(0.011)	(0.010)
	0.00003	**0.00003**	**0.0001**	**0.0001**
Pupil: Teacher Ratio	0.000	0.000	0.059	0.001
	(0.487)	(0.478)	(0.212)	(0.214)
Percent of students who are…				
	0.008 **	**0.008** **	**0.001**	**0.002**
Eligible for Free or Reduced-Priced Lunch	0.117	0.123	0.019	0.026
	(0.037)	(0.029)	(0.747)	(0.668)
	0.045* **	**0.045*** **	**0.031*** **	**0.031*** **
Black	0.665	0.658	0.477	0.473
	(0.000)	(0.000)	(0.000)	(0.000)
	0.012*	**0.012***	**-0.001**	**-0.001**
Hispanic	0.177	0.179	-0.018	-0.017
	(0.072)	(0.071)	(0.856)	(0.870)
	-0.032* **	**-0.032*** **	**-0.04*** **	**-0.041*** **
Native American	-0.469	-0.478	-0.619	-0.629
	(0.001)	(0.001)	(0.000)	(0.000)
	-0.03 **	**-0.03** **	**-0.034*** **	**-0.034*** **
Asian	-0.443	-0.446	-0.518	-0.520
	(0.015)	(0.014)	(0.006)	(0.005)

Table 6: Chapter II: Main Results (Continued from Previous Page)

	Baseline	Baseline + Interaction	Baseline + Prof Rank	Baseline + Interaction + Prof Rank
Districts' percent of residents who are...				
	-0.059* **	**-0.059*** **	**-0.043*** **	**-0.043*** **
Bachelor's Degree Recipients	-0.873	-0.878	-0.662	-0.667
	(0.000)	(0.000)	(0.000)	(0.000)
	0.007	**0.008**	**0.014** **	**0.015** **
Black	0.105	0.117	0.219	0.229
	(0.334)	(0.282)	(0.048)	(0.039)
	-0.067* **	**-0.067*** **	**-0.055*** **	**-0.055*** **
Hispanic	-0.988	-0.995	-0.841	-0.846
	(0.000)	(0.000)	(0.000)	(0.000)

Table 7: Chapter II: Main Results with Non-linear Proficiency Rank Measures

	Baseline	Baseline + Interaction	Baseline + Prof Rank	Baseline + Interaction + Prof Rank
School Failed Once	**0.005***** 0.068 (0.006)	**0.013***** 0.175 (0.000)	**0.002** 0.035 (0.186)	**0.008**** 0.110 (0.012)
School Failed Twice	**0.011***** 0.145 (0.000)	**0.016***** 0.207 (0.000)	**0.004** 0.053 (0.114)	**0.006** 0.087 (0.132)
School Failed Three Times	**0.029***** 0.337 (0.000)	**0.031***** 0.351 (0.000)	**0.01***** 0.135 (0.001)	**0.011**** 0.148 (0.029)
School Failed Once*Union Coverage > 75%		**-0.01***** -0.157 (0.002)		**-0.007**** -0.112 (0.032)
School Failed Twice*Union Coverage > 75%		**-0.006** -0.088 (0.170)		**-0.003** -0.048 (0.466)
School Failed Three Times*Union Coverage > 75%		**-0.002** -0.024 (0.741)		**-0.001** -0.021 (0.775)
Percent Change in Enrollment (2002-2006)	**-0.065***** -0.965 (0.000)	**-0.065***** -0.966 (0.000)	**-0.062***** -0.949 (0.000)	**-0.062***** -0.949 (0.000)
Total Enrollment in 2002 (in hundreds)	**-0.009***** -0.136 (0.000)	**-0.01***** -0.136 (0.000)	**-0.01***** -0.119 (0.000)	**-0.01***** -0.119 (0.000)
Prof Rank (in tens)			**-0.025***** -0.382 (0.000)	**-0.025***** -0.379 (0.000)

	0.0004* **	**0.0004*** **
Prof Rank Squared (in tens)	0.007	0.007
	(0.000)	(0.000)
	-0.000002* **	**-0.000002*** **
Prof Rank Cubed (in tens)	0.000	0.000
	(0.000)	(0.000)

Table 8: Chapter II: Individual Year Results

	Closed after 2004		Closed after 2005		Closed after 2006	
	Baseline	**Interaction**	**Baseline**	**Interaction**	**Baseline**	**Interaction**
School Failed in…	**2003**		**2004**		**2005**	
	0.015***	0.018***	0.022***	0.022***	0.016***	0.019***
	0.152	0.182	0.245	0.248	0.218	0.254
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Percent Change in Enrollment between…	**2002-2004**		**2002-2005**		**2002-2006**	
	-0.006***	-0.006***	-0.004***	-0.004***	-0.064***	-0.064***
	-0.064	-0.064	-0.051	-0.051	-0.956	-0.956
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***	-0.01***
Total Enrollment in 2002 (in hundreds)	-0.13	-0.13	-0.13	-0.13	-0.13	-0.13
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
		-0.004		-0.0004		-0.003
School Failed (in year)*Union Coverage > 75%		-0.042		-0.005		-0.051
		(0.335)		(0.919)		(0.274)

Table 9: Chapter III: EKD Accountability Index Criteria and Score Definitions

Rating	Renewal Criteria	Annual Performance Reports	Termination Clause	Maximum Charter Length
1	Law lists no renewal criteria	Annual performance reports are not required in addition to any state-required reports for all schools	No termination clause	More than 10 years; No specified maximum length
2	General criteria outlined in the charter; Renewal contingent on a vote by parties to original charter, staff and parents, etc.	State board, local agency, or school submits high-level performance report	Termination clause exists but requires a public hearing, probationary period, or vote prior to termination	Between 5 years (exclusive) and 10 years (inclusive)
3	Authorizer **may** deny renewal if obligations of charter, including performance outcomes, are not met	State board, local agency, or school submits report; an in-depth evaluation is completed prior to renewal or at a scheduled time during the charter;	Termination clause exists for failing to meet educational goals (may include an informal hearing)	5 years
4	Authorizer **must** deny renewal if obligations of charter, including performance outcomes, are not met; Criteria for renewal include minimum performance requirements outlined by the state (may specifically include AYP performance); Renewal may also require an in-depth evaluation by state.	An in-depth evaluation is conducted annually (includes school visits)	*N/A*	Fewer than 5 years

Table 10: Chapter III: Summary of State Ratings on Four Indices

State	EKD Index	Shober Index - Accountability	Shober Index - Flexibility	CER Index
AK	mild	mild	moderate	mild
AR	strict	mild	mild	moderate
AZ	mild	strict	high	strict
CA	strict	strict	moderate	mild
CO	strict	strict	high	moderate
CT	strict	mild	mild	moderate
DC	strict	strict	high	strict
DE	mild	strict	mild	strict
FL	moderate	strict	high	mild
GA	mild	moderate	high	moderate
HI	mild	mild	mild	moderate
IA	strict	mild	high	mild
ID	mild	moderate	mild	mild
IL	moderate	mild	high	mild
IN	mild	strict	mild	moderate
KS	moderate	mild	moderate	mild
LA	moderate	mild	high	mild
MA	moderate	strict	mild	strict
MD	mild	*missing*	*missing*	mild
MI	mild	strict	moderate	strict
MN	moderate	strict	moderate	mild
MO	mild	moderate	moderate	strict
MS	moderate	mild	moderate	moderate
NC	moderate	moderate	high	strict
NH	moderate	mild	moderate	moderate
NJ	strict	moderate	mild	strict
NM	moderate	moderate	moderate	moderate
NV	moderate	mild	mild	strict
NY	strict	moderate	mild	strict
OH	moderate	strict	moderate	mild
OK	strict	moderate	high	strict
OR	moderate	moderate	moderate	moderate
PA	strict	strict	high	strict
RI	mild	mild	mild	strict
SC	moderate	moderate	moderate	strict
TN	moderate	mild	mild	strict
TX	strict	moderate	high	strict
UT	mild	moderate	moderate	mild
VA	moderate	mild	moderate	moderate
WI	mild	moderate	mild	mild
WY	moderate	mild	moderate	moderate

Table 11: Chapter III: Summary Statistics for All Variables

Variable	Charters		Non-Charter Publics		*p-values*
	Mean	**SD**	**Mean**	**SD**	
School Failed at least Once	56%	50%	41%	49%	< 0.001
School Failed at least Twice	27%	45%	22%	41%	< 0.001
School Failed Three Times	16%	37%	12%	33%	< 0.001
School Closed between 2004 and 2011	25%	43%	8%	28%	< 0.001
School Closed between 2006 and 2011	17%	38%	6%	23%	< 0.001
Proficiency Rate	34%	32%	51%	29%	< 0.001
Percent Change in Enrollment Between 2002 and 2006	20%	50%	-1%	20%	< 0.001
Total School Enrollment	263	319	577	446	< 0.001
School Eligible for Title I	42%	49%	56%	50%	< 0.001
Pupil/Teacher Ratio	18.2	15.0	16.3	11.1	< 0.001
High Schools	18%	38%	17%	38%	0.177
Middle Schools	9%	28%	19%	39%	< 0.001
Other Schools	25%	43%	3%	18%	< 0.001
Elementary Schools	48%	50%	61%	49%	< 0.001
Located in a City	51%	50%	25%	43%	< 0.001
Located in a Rural Area	15%	36%	29%	45%	< 0.001
Located in a Suburb	26%	44%	35%	48%	< 0.001
Located in a Town	7%	26%	12%	32%	< 0.001
Percent Black Students	28%	35%	16%	26%	< 0.001
Percent Hispanic Students	17%	24%	15%	23%	< 0.001
Percent Native American Students	3%	13%	2%	8%	< 0.001
Percent Asian-American Students	3%	10%	3%	8%	0.145
Percent Economically Disadvantaged	47%	38%	41%	28%	< 0.001
Percent of Schools - Mild Accountability (EKD)	35%	48%	22%	41%	< 0.001
Percent of Schools - Moderate Accountability (EKD)	29%	45%	37%	48%	< 0.001
Percent of Schools - Strict Accountability (EKD)	36%	48%	41%	49%	< 0.001
Percent of Schools - Mild Accountability (Shober)	36%	48%	39%	49%	0.028
Percent of Schools - Moderate Accountability (Shober)	10%	30%	14%	35%	< 0.001
Percent of Schools with Strict Accountability (Shober)	53%	50%	44%	50%	< 0.001
Percent of Schools in "Weak" States (CER)	6%	23%	25%	43%	< 0.001
Percent of Schools in "Moderate" State (CER)	22%	41%	37%	48%	< 0.001
Percent of Schools in "Strong" States (CER)	72%	45%	39%	49%	< 0.001

Table 12: Chapter III: Characteristics of Charter Schools (2002) Nationally, by School Closure from 2006-2011

	Closed	2006-2010 Remained Open	*p-values*
Number of Schools	293	1,402	
School Failed at least Once	73.1%	51.5%	< 0.001
School Failed at least Twice	41.0%	26.7%	< 0.001
School Failed all Three Years	28.2%	13.0%	< 0.001
Percent Change in Enrollment (2002/2004 to 2006)	11.9%	24.4%	< 0.001
Average Enrollment (2002)	191	298	< 0.001
Student/Teacher Ratio	16.5	18.3	0.030
Percent of Students…			
Eligible for Free/Reduced Lunch	52.8%	42.6%	< 0.001
Asian	1.2%	3.8%	< 0.001
Black	41.6%	24.8%	< 0.001
Hispanic	12.5%	17.6%	0.001
Native American	5.0%	2.3%	0.001
White	39.1%	50.9%	< 0.001
Percent of Schools…			
Eligible for Title I	50.4%	41.3%	0.006
Located in City	54.3%	50.6%	0.250
Located in Rural Area	16.0%	14.6%	0.535
Located in a Suburb	22.5%	27.6%	0.074
Located in a Town	7.2%	7.2%	0.982
Elementary Schools	40.6%	53.3%	0.002
Middle Schools	9.2%	8.6%	0.187
High Schools	21.5%	16.0%	0.004
Other Schools	28.7%	22.0%	0.015

Note: Data on school characteristics taken from the Common Core of Data, 2001-2002. For schools in Tennessee, data on student ethnicity taken from 1998-99 and data on free/reduced price lunch eligibility is unavailable. Sample is restricted to those schools operational and identified as charters in 2002.

Table 13: Chapter III: Characteristics of Charter Schools (2002) Nationally, by AYP Status

	2003-2005			
	Never failed	**Failed at least Once**	**Failed at least Twice**	**Failed Three Years**
Number of Schools	625	1,036	578	237
Closed between 2006 and 2011	8%	22%	24%	27%
Change in Enrollment (2002/2004 to 2006)	21%	20%	18%	18%
Average Enrollment	276	298	321	406
Student/Teacher Ratio	17.4	19.2	19.7	22.8
Percent of Students…				
Eligible for Free/Reduced Lunch	9.4%	9.4%	11.9%	13.1%
Asian	3.5%	3.4%	3.6%	2.9%
Black	17.2%	35.6%	38.6%	38.2%
Hispanic	16.5%	19.0%	20.8%	24.1%
Native American	1.6%	3.4%	3.9%	2.8%
White	60.7%	38.0%	32.6%	31.3%
Percent of Schools…				
Eligible for Title I	40.0%	44.6%	45.0%	38.1%
Elementary Schools	61.8%	42.2%	35.8%	31.2%
Middle Schools	9.4%	9.4%	11.9%	13.1%
High Schools	9.9%	21.7%	23.7%	22.8%
Other Schools	18.9%	26.7%	28.5%	32.9%
Located in City	43.7%	56.0%	56.4%	57.8%
Located in Rural Area	17.3%	13.1%	12.6%	13.1%
Located in Suburb	30.7%	25.0%	24.6%	24.9%
Located in Town	8.3%	5.9%	6.4%	4.2%

Note: Unweighted by enrollment. Data on school characteristics taken from the Common Core of Data, 2001-2002. For schools in Tennessee, data on student ethnicity taken from 1998-99 and data on free/reduced price lunch eligibility is unavailable.

Table 14: Chapter III: Main Results (Charter Schools and Non-Charter Publics Schools)

	Failed at least once		Failed at least twice		Failed all three years	
	Non-Charters	**Charters**	**Non-Charters**	**Charters**	**Non-Charters**	**Charters**
School failed….	**0.002** 0.032 (0.178)	**0.058***** 0.358 (0.002)	**0.007***** 0.097 (0.000)	**0.043**** 0.238 (0.043)	**0.016***** 0.229 (0.000)	**0.045*** 0.258 (0.093)
TTest of Difference in Estimates	p = 0.0054		p = 0.2426		p = 0.8087	
Percentage change in Enrollment (2002-2006)	**-0.074***** -1.151 (0.000)	**-0.064***** -0.389 (0.001)	**-0.074***** -1.148 (0.000)	**-0.066***** -0.391 (0.001)	**-0.073***** -1.207 (0.000)	**-0.065***** -0.425 (0.002)
Prof Rank (in tens)	**-0.004***** -0.006 (0.000)	**-0.01***** -0.006 (0.005)	**-0.004***** -0.006 (0.000)	**-0.011***** -0.007 (0.003)	**-0.003***** -0.005 (0.000)	**-0.011***** -0.007 (0.002)
Total Enrollment in 2002 (in hundreds)	**-0.008***** -0.001 (0.000)	**-0.019***** -0.001 (0.000)	**-0.008***** -0.001 (0.000)	**-0.019***** -0.001 (0.000)	**-0.008***** -0.001 (0.000)	**-0.017***** -0.001 (0.000)
High School	**-0.001** -0.022 (0.522)	**0.07**** 0.362 (0.012)	**-0.002** -0.027 (0.432)	**0.078***** 0.392 (0.006)	**0.001** 0.012 (0.750)	**0.021** 0.127 (0.444)
Middle School	**0.019***** 0.255 (0.000)	**0.082**** 0.403 (0.014)	**0.018***** 0.245 (0.000)	**0.077**** 0.378 (0.021)	**0.016***** 0.233 (0.000)	**0.063*** 0.340 (0.051)
Other School	**0.01**** 0.137 (0.022)	**0.102***** 0.508 (0.000)	**0.009**** 0.128 (0.032)	**0.106***** 0.519 (0.000)	**0.007*** 0.111 (0.094)	**0.089***** 0.476 (0.001)
City	**0.014***** 0.190 (0.000)	**-0.043** -0.257 (0.204)	**0.013***** 0.187 (0.000)	**-0.04** -0.237 (0.238)	**0.012***** 0.184 (0.000)	**-0.027** -0.176 (0.431)
Rural Area	**-0.009***** -0.140 (0.000)	**-0.015** -0.095 (0.654)	**-0.009***** -0.142 (0.000)	**-0.018** -0.112 (0.594)	**-0.008***** -0.136 (0.000)	**-0.016** -0.113 (0.635)

	0.004*	**0.006**	**0.004***	**0.007**	**0.004***	**0.00005**
Suburb	0.059	0.037	0.058	0.039	0.060	0.000
	(0.077)	(0.858)	(0.082)	(0.848)	(0.086)	(0.999)
	0.005***	**0.023**	**0.005*****	**0.024**	**0.005*****	**0.019**
Title I School	0.074	0.136	0.076	0.140	0.080	0.125
	(0.002)	(0.241)	(0.001)	(0.227)	(0.001)	(0.336)
	0.0001	**-0.002**	**0.0001**	**-0.002**	**0.0001**	**-0.001**
Pupil:Teacher Ratio	0.002	-0.013	0.002	-0.012	0.002	-0.009
	(0.118)	(0.122)	(0.126)	(0.142)	(0.101)	(0.302)
Percent of students who are…						
Eligible for Free or Reduced-Priced Lunch	**0.012*****	**0.047**	**0.012*****	**0.049**	**0.013*****	**0.035**
	0.193	0.283	0.185	0.290	0.216	0.225
	(0.001)	(0.124)	(0.001)	(0.114)	(0.000)	(0.289)
	0.029***	**0.094*****	**0.028*****	**0.098*****	**0.024*****	**0.091****
Black	0.452	0.570	0.438	0.583	0.393	0.591
	(0.000)	(0.005)	(0.000)	(0.004)	(0.000)	(0.012)
	-0.017***	**-0.098****	**-0.018*****	**-0.104****	**-0.018*****	**-0.082***
Hispanic	-0.267	-0.594	-0.278	-0.621	-0.298	-0.531
	(0.000)	(0.028)	(0.000)	(0.023)	(0.000)	(0.063)
	-0.04***	**0.077**	**-0.041*****	**0.083**	**-0.045*****	**0.13****
Native American	-0.627	0.464	-0.643	0.492	-0.745	0.845
	(0.000)	(0.198)	(0.000)	(0.170)	(0.000)	(0.034)
	-0.046***	**-0.419**	**-0.046*****	**-0.439***	**-0.041*****	**-0.287**
Asian	-0.717	-2.531	-0.718	-2.616	-0.675	-1.868
	(0.000)	(0.106)	(0.000)	(0.099)	(0.000)	(0.220)

Table 15: Chapter III: Main Results (Charter Schools and Non-Charter Publics Schools) with non-linear Proficiency Rank measures

	Failed at least once		**Failed at least twice**		**Failed all three years**	
	Non-Charters	**Charters**	**Non-Charters**	**Charters**	**Non-Charters**	**Charters**
School failed….	**0.003****	**0.062*****	**0.005****	**0.044****	**0.012*****	**0.038**
	0.052	0.379	0.070	0.244	0.172	0.222
	(0.034)	(0.001)	(0.012)	(0.040)	(0.000)	(0.154)
TTest of Difference in Estimates	p = 0.0057		p = 0.1526		p = 0.6766	
Percentage change in Enrollment (2002-2006)	**-0.074*****	**-0.062*****	**-0.074*****	**-0.064*****	**-0.073*****	**-0.063*****
	-1.143	-0.376	-1.140	-0.379	-1.197	-0.412
	(0.000)	(0.001)	(0.000)	(0.001)	(0.000)	(0.002)
Prof Rank (in tens)	**-0.025*****	**-0.058*****	**-0.024*****	**-0.055****	**-0.019*****	**-0.066*****
	-0.038	-0.035	-0.038	-0.033	-0.031	-0.043
	(0.000)	(0.007)	(0.000)	(0.011)	(0.000)	(0.004)
Prof Rank Squared (in tens)	**0.0004*****	**0.001****	**0.0004*****	**0.001***	**0.0003*****	**0.001****
	0.001	0.001	0.001	0.001	0.001	0.001
	(0.000)	(0.049)	(0.000)	(0.074)	(0.000)	(0.026)
Prof Rank Cubed (in tens)	**-0.000002*****	**-0.00001***	**-0.000002*****	**-0.00001**	**-0.000002*****	**-0.00001****
	0.000	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.091)	(0.000)	(0.126)	(0.000)	(0.046)
Total Enrollment in 2002 (in hundreds)	**-0.008*****	**-0.018*****	**-0.008*****	**-0.018*****	**-0.007*****	**-0.016*****
	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)

Table 16: Chapter III: Accountability Provisions Results (Charters only)

School Failed…	**Failed at least once**		**Failed at least twice**		**Failed all three years**	
	EKD Index	Shober Index	EKD Index	Shober Index	EKD Index	Shober Index
Mild Accountability	**0.095****	**0.035**	**0.093****	**-0.013**	**0.113**	**0.01**
	0.467	0.198	0.434	-0.079	0.531	0.065
	(0.012)	(0.263)	(0.040)	(0.632)	(0.280)	(0.730)
Moderate Accountability	**0.008**	**0.018**	**-0.023**	**0.041**	**0.023**	**-0.025**
	0.050	0.105	-0.145	0.214	0.140	-0.191
	(0.817)	(0.749)	(0.455)	(0.615)	(0.551)	(0.733)
Strict Accountability	**0.086****	**0.104*****	**0.068***	**0.107*****	**0.042**	**0.156****
	0.442	0.534	0.342	0.500	0.241	0.690
	(0.013)	(0.001)	(0.054)	(0.002)	(0.245)	(0.012)

Table 17: Chapter III: Specific Accountability Policies Results

School Failed…	**Failed at least once**	**Failed at least twice**	**Failed all three years**
Renewal Term and Termination Clause – Non-Strict	**0.057****	**0.04**	**0.068***
	0.323	0.214	0.364
	(0.020)	(0.130)	(0.066)
Renewal Term and Termination Clause - Strict	**0.084****	**0.042**	**0.02**
	0.422	0.222	0.121
	(0.023)	(0.227)	(0.589)

Table 18: Chapter III: CER Flexibility and EKD Accountability Provisions Results (Charters only)

School Failed…	**Failed at least once**		**Failed at least twice**		**Failed all three years**	
	CER Flex Index	EKD Acct Index	CER Flex Index	EKD Acct Index	CER Flex Index	EKD Acct Index
Mild	**0.043** 0.230 (0.733)	**0.089**** 0.443 (0.017)	**0.002** 0.012 (0.983)	**0.096**** 0.451 (0.034)	**-0.052** -0.479 (0.473)	**0.114** 0.534 (0.278)
Moderate	**-0.015** -0.099 (0.722)	**0.01** 0.059 (0.810)	**0.028** 0.155 (0.603)	**-0.024** -0.153 (0.483)	**0.014** 0.089 (0.833)	**0.034** 0.196 (0.444)
High/Strict[69]		**0.094**** 0.478 (0.012)		**0.064*** 0.331 (0.076)		**0.051** 0.284 (0.178)

[69] This category is omitted due to collinearity.

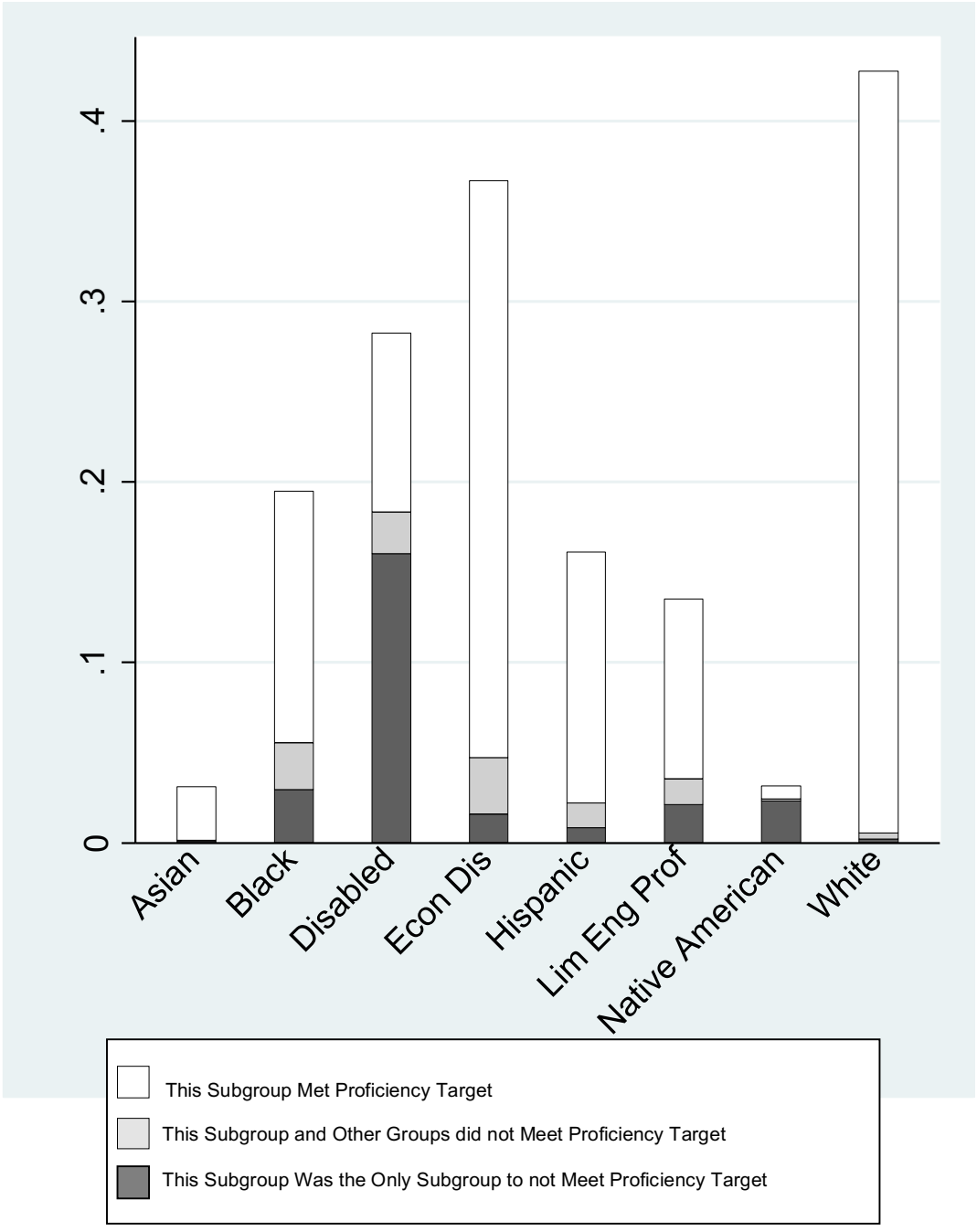
Table 19: Chapter III: Shober Flexibility and Accountability Provisions Results (Charters only)

School Failed…	**Failed at least once**		**Failed at least twice**		**Failed all three years**	
	Shober Flex Index	Shober Acct Index	Shober Flex Index	Shober Acct Index	Shober Flex Index	Shober Acct Index
	0.119	**0.104***	**0.03**	**-0.021**	**-0.014**	**0.029**
Mild	0.545	0.536	0.164	-0.137	-0.096	0.175
	(0.319)	(0.077)	(0.726)	(0.599)	(0.877)	(0.567)
	-0.053	**0.03**	**0.019**	**0.032**	**-0.015**	**-0.015**
Moderate	-0.379	0.168	0.108	0.174	-0.109	-0.104
	(0.160)	(0.625)	(0.673)	(0.691)	(0.760)	(0.861)
		0.118* **		**0.099** **		**0.168** **
High/Strict[70]		0.601		0.479		0.733
		(0.002)		(0.010)		(0.014)

[70] This category is omitted due to collinearity.

FIGURES

Figure 1: Chapter I: Subgroup Accountability and Likelihood of Failure in Math, 2004



Notes to Figure 1: The total height of each bar illustrates the fraction of schools where that subgroup's proficiency rate counts towards the AYP determination, while the shaded areas of the bars represent the fraction of schools where that subgroup failed to make AYP. The figure is based on 46 states with available data. Iowa, North Dakota, Nebraska, and New Mexico are missing subgroup-level proficiency data in 2004.

Figure 2: Chapter I: Distribution of State Failure Rates, 2003 – 2005

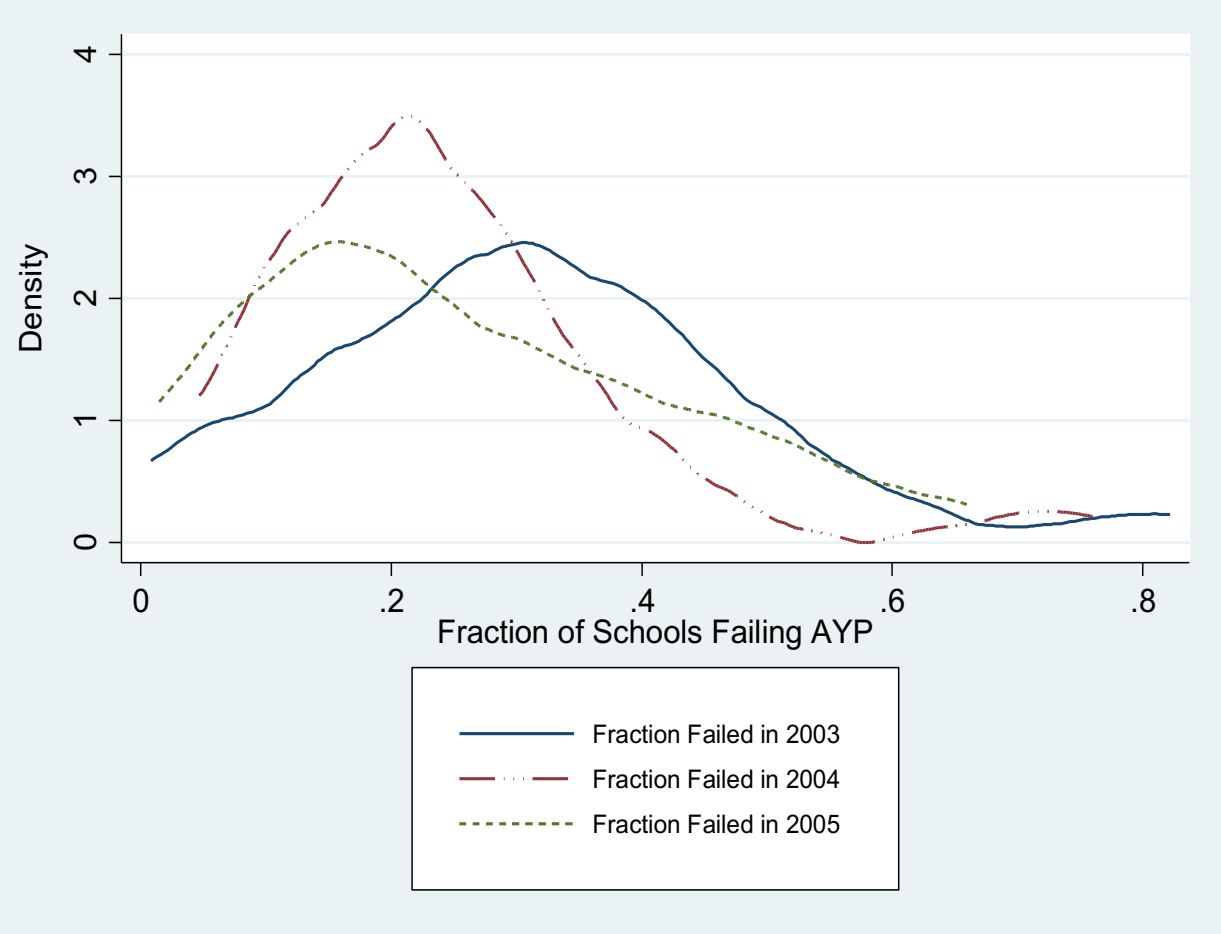


Figure 3: Chapter I: School Failure Rates vs. State Proficiency Rates in Math, 2004

Notes to Figures 3: N = 46 states. Alabama, Nebraska, and New Hampshire are missing proficiency rate data. Vermont reports a performance index in lieu of proficiency rates. When we aggregate proficiency rates to the state level for the x-axis, we weight schools by their number of tested students. For 12 states that failed to report the number of tested students by school, we use schools' student enrollment in tested grades as reported in the Common Core of Data as a proxy for the number of students tested.

Figure 4: Chapter I: School Failure Rates vs. Average Number of Accountable Groups in Schools, 2004

Notes to Figure 4: Based on 46 states with available data. Iowa, North Dakota, Nebraska, and New Mexico are missing subgroup-level proficiency data in 2004. Accountable groups include both student subgroups and the overall student population. For each state, we take the average of the number of accountable groups for math achievement and the number of accountable groups for ELA achievement. For states that hold schools accountable separately for the grade-level performance of student subgroups, we accordingly treat each subgroup-by-grade-level as a separate group.

Figure 5: Chapter II: Fraction of States' Schools that Closed between 2004-2011

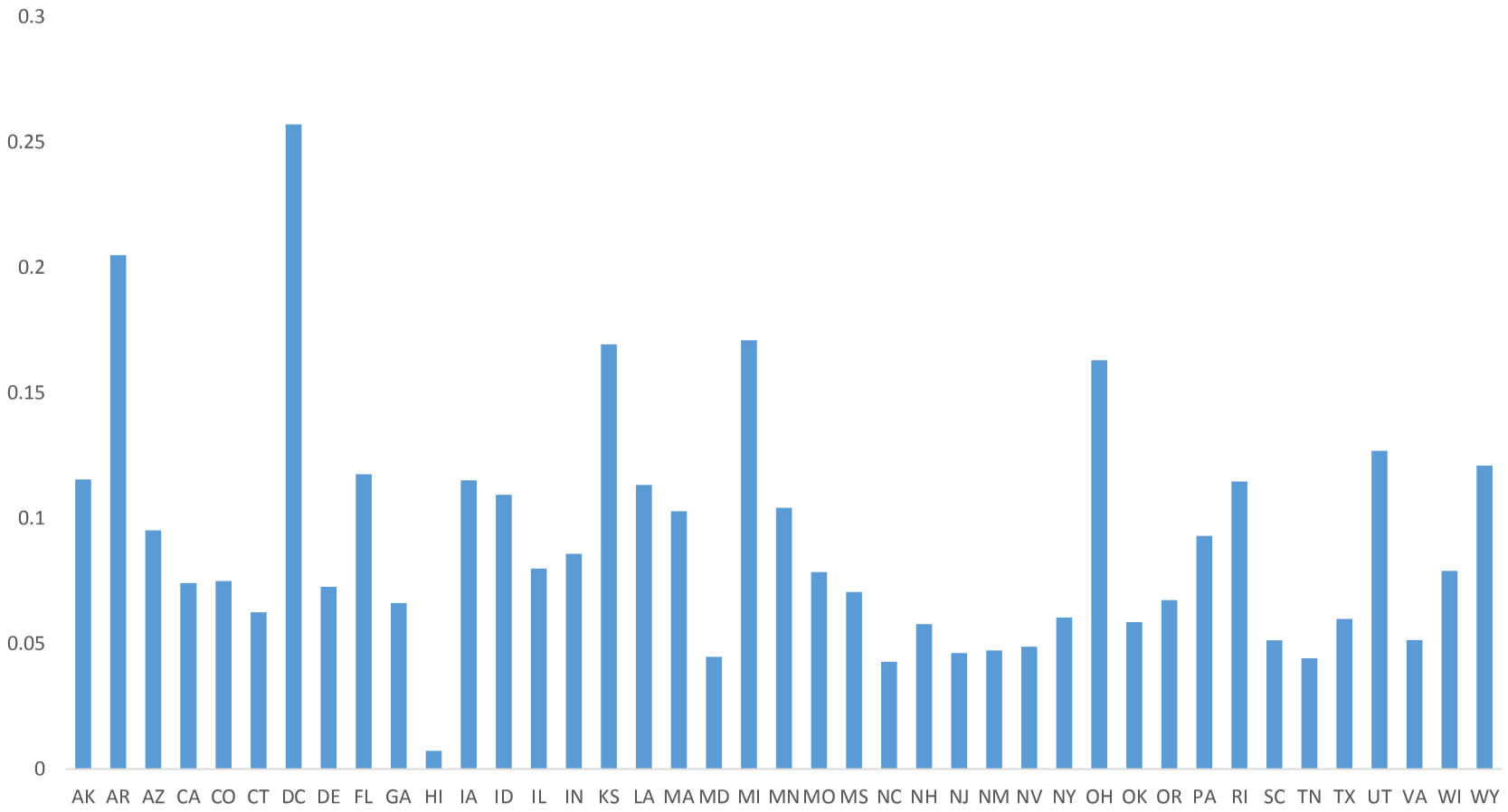
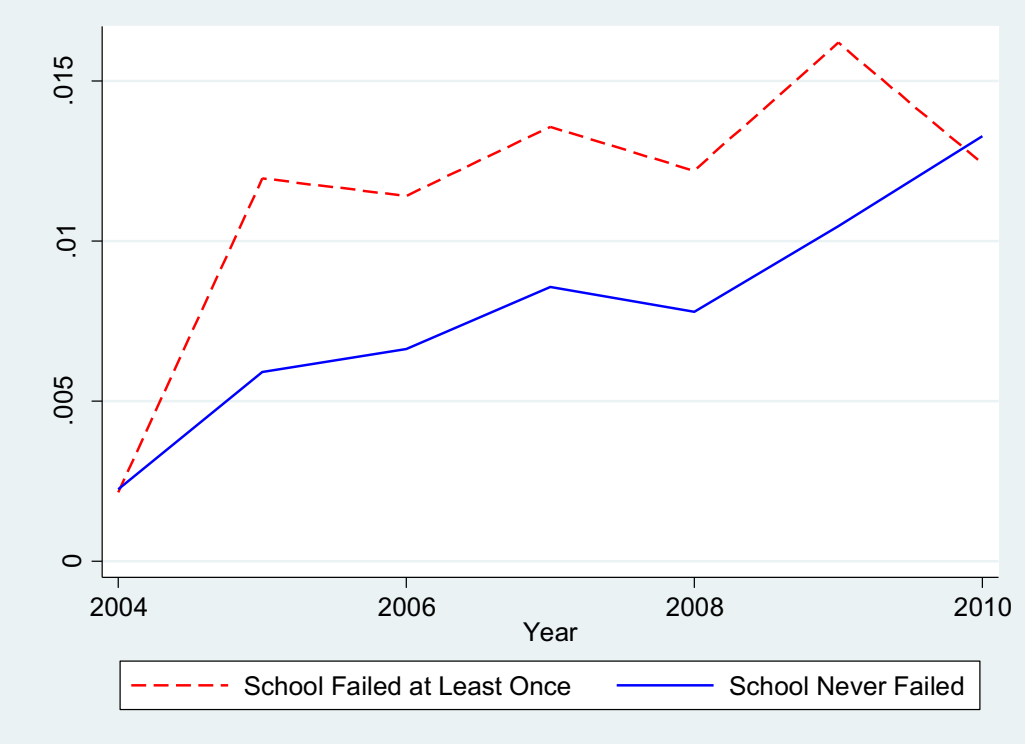


Figure 6: Chapter II: Annual School Closure Rates across Years, Nationally



*Note: Since schools may have dropped from the data, this analysis includes schools with missing AYP indicators for 2004 or 2005.

Figure 7: Chapter II: Cumulative School Closure Rates across Years, Nationally

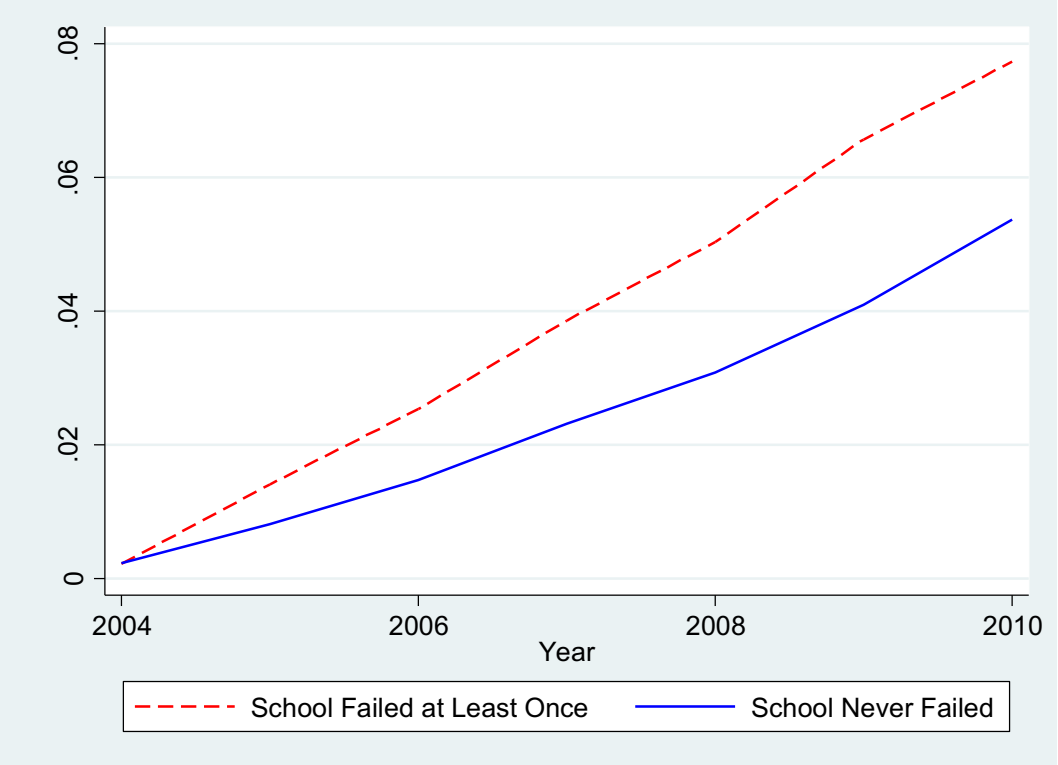


Figure 8: Chapter II: Fraction of States' Schools that Failed AYP all Three Years vs. Fraction of States' Schools that Closed between 2006-2011

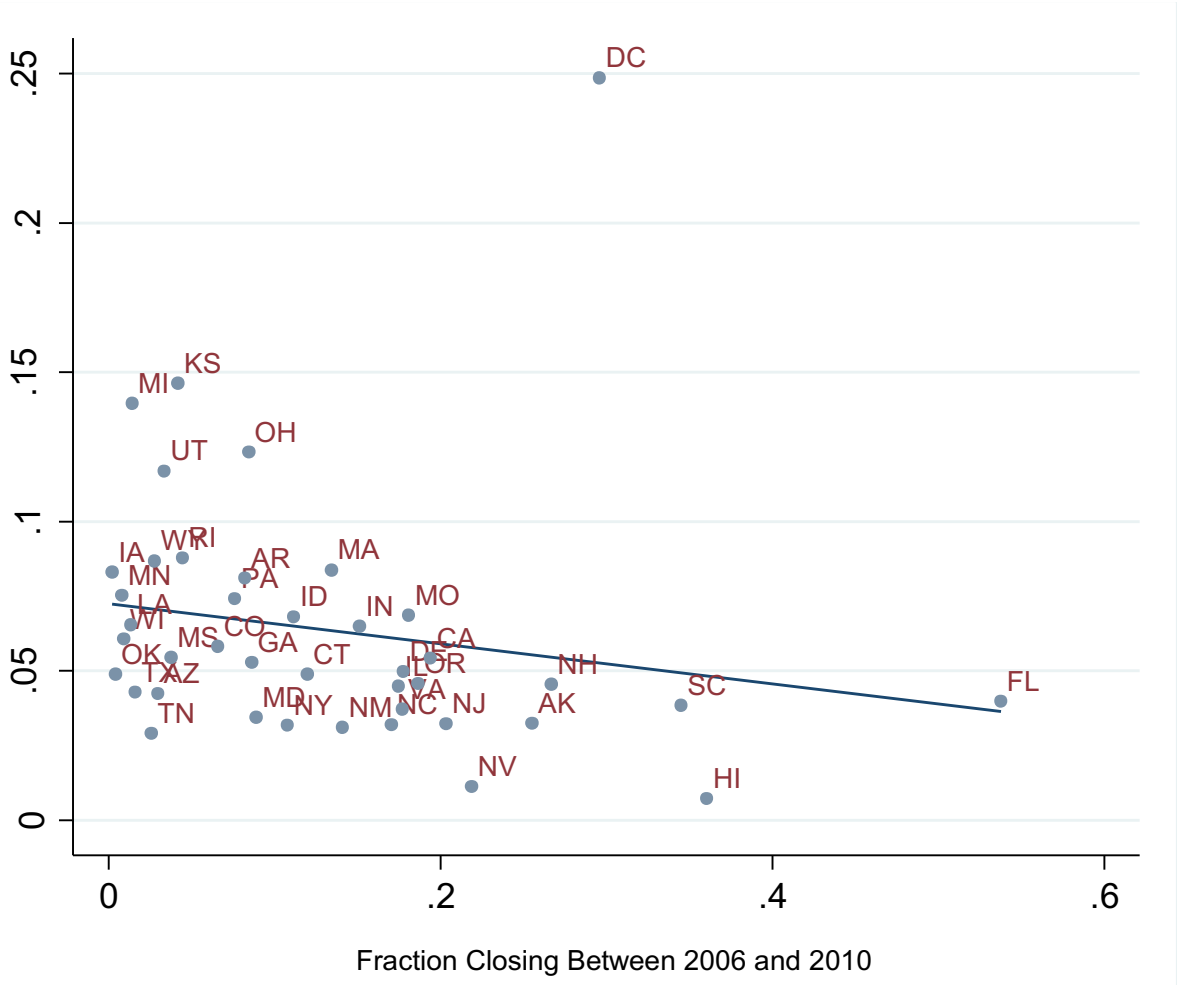


Figure 9: Chapter III: Number of Charter Schools in each EKD Index Score

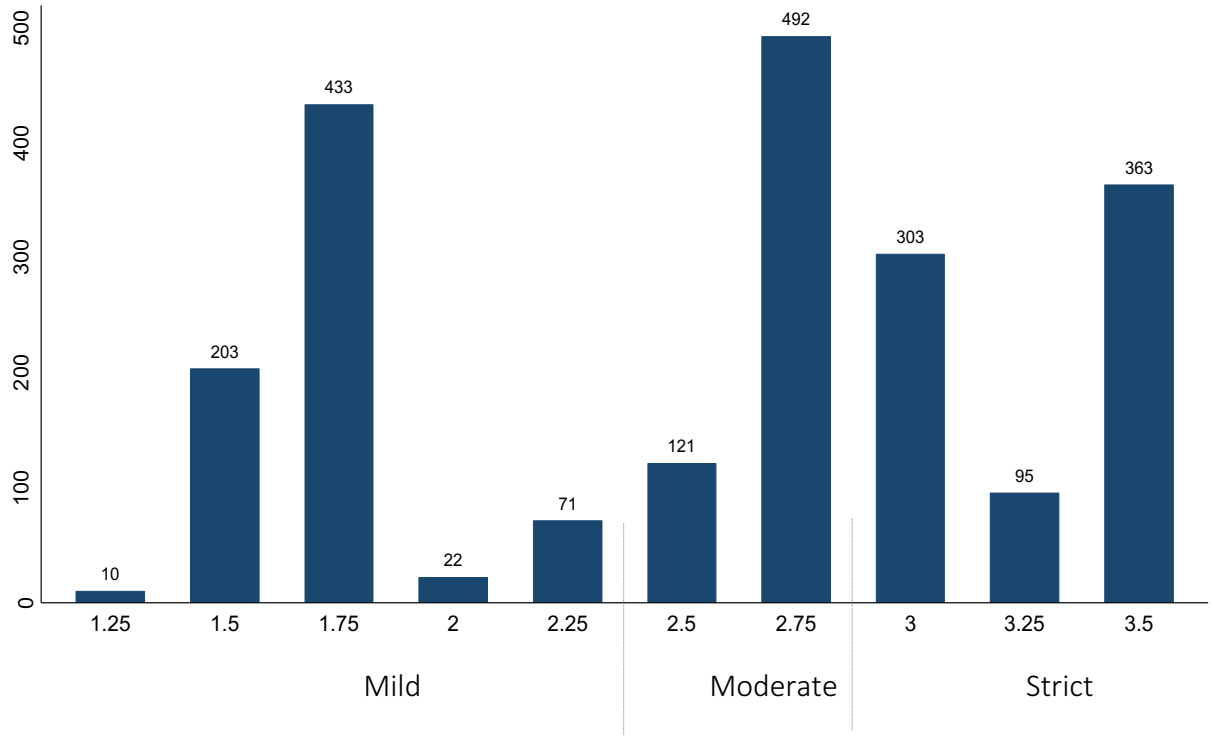


Figure 10: Chapter III: Number of Charter Schools in each Shober Accountability Index Score

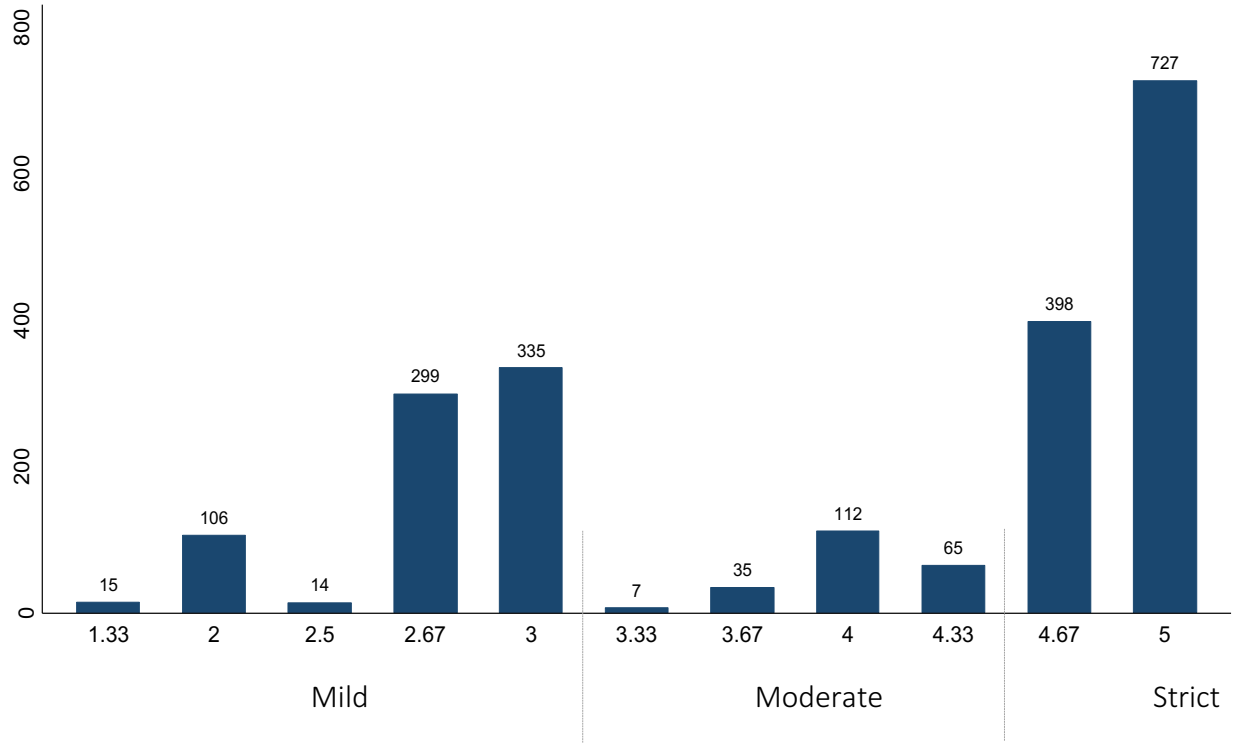


Figure 11: Chapter III: Distribution of Charter Schools across CER Index Scores

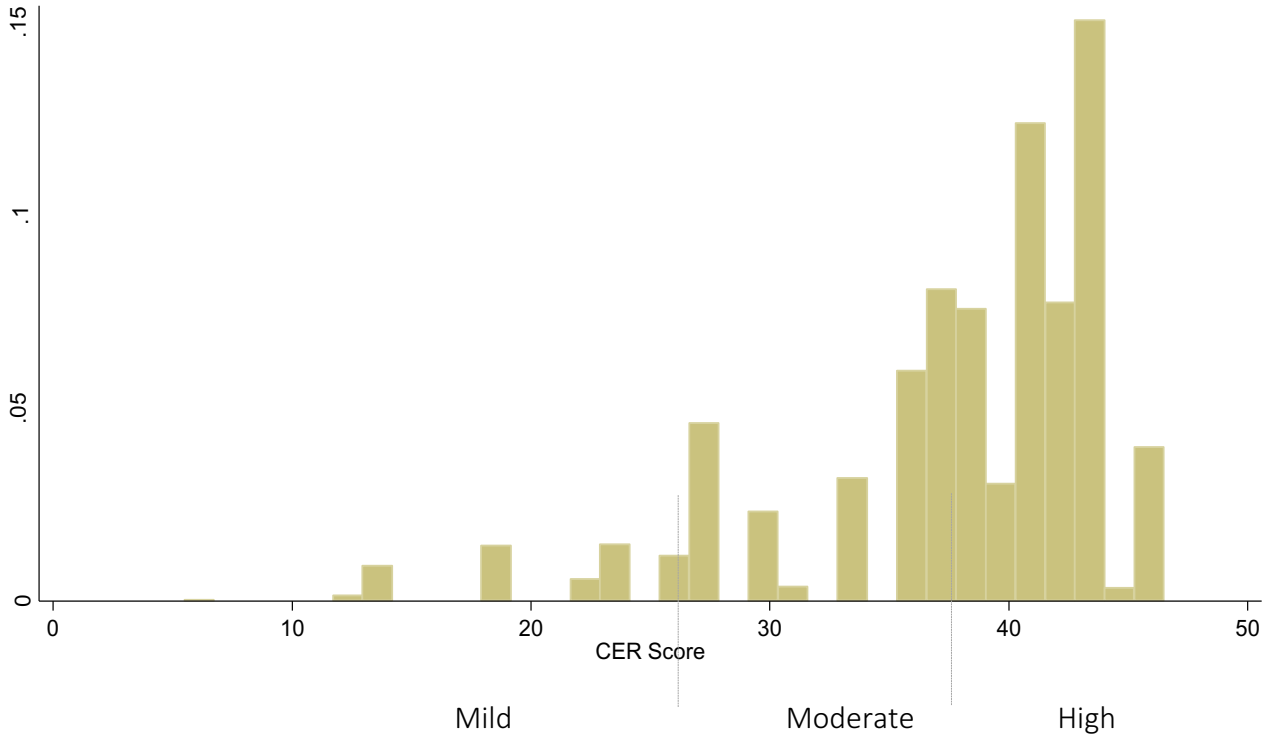


Figure 12: Chapter III: Distribution of Charter Schools across Shober Flexibility Index Scores

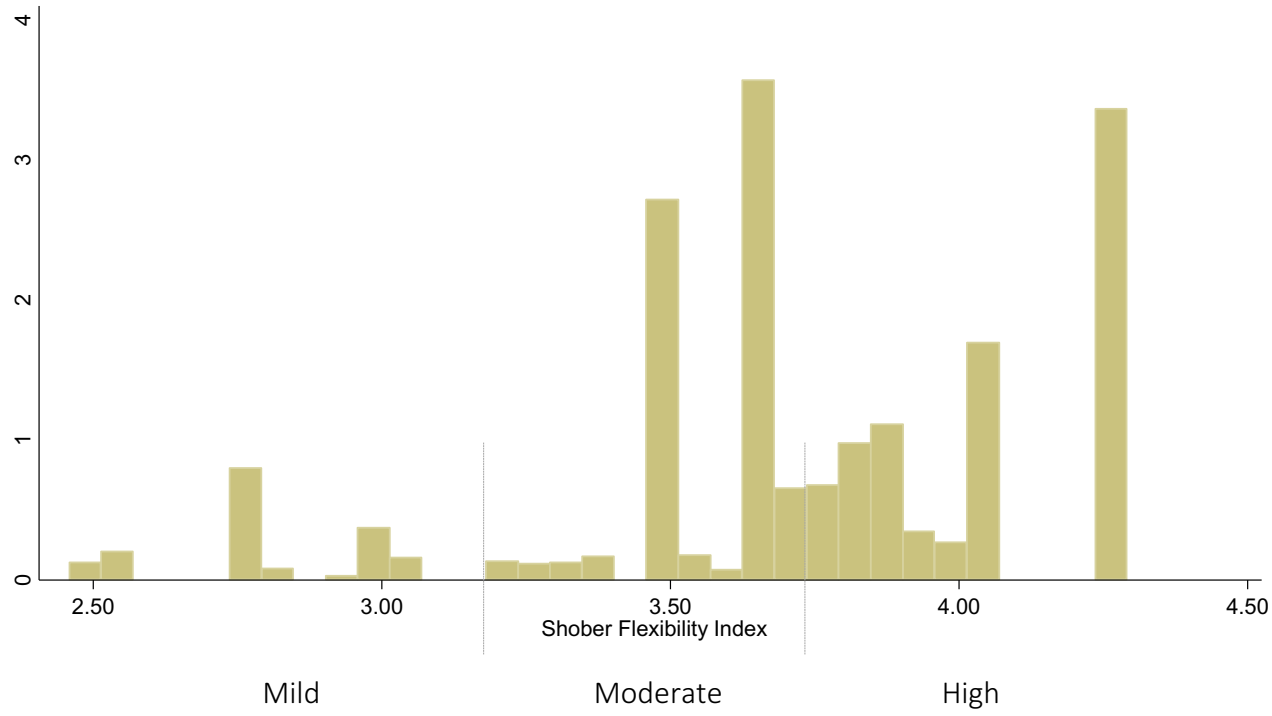


Figure 13: Chapter III: Marginal Effects of Schools' Relative Test Score Performance on the Likelihood of Closure

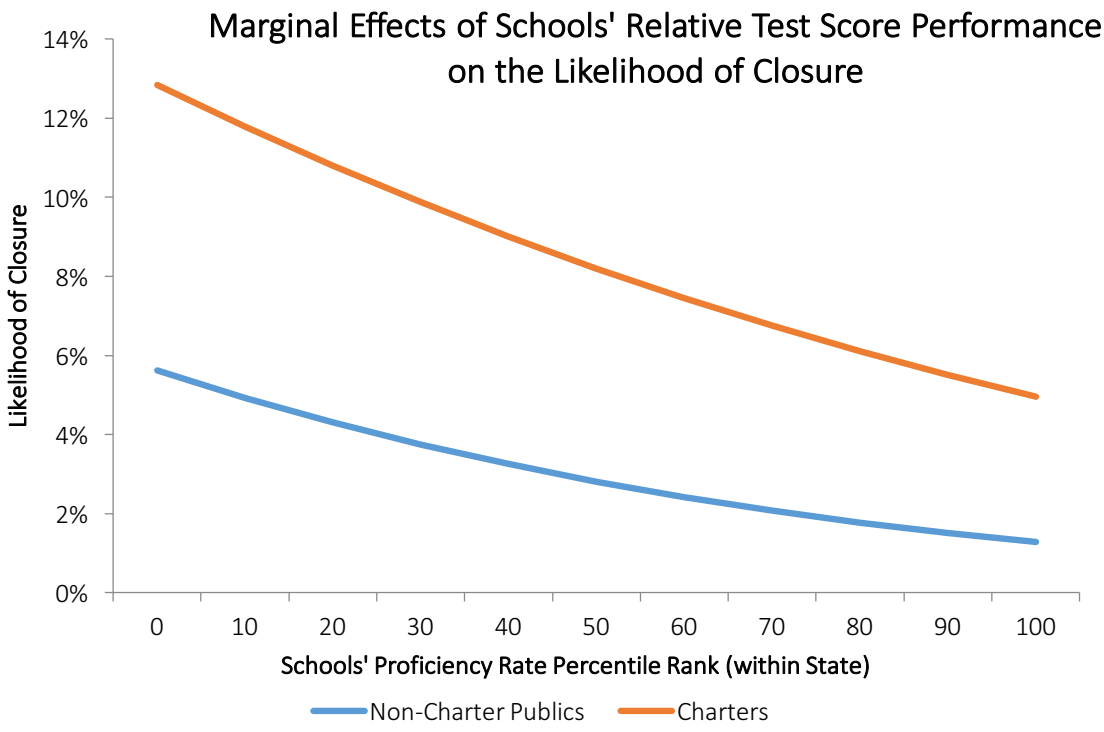
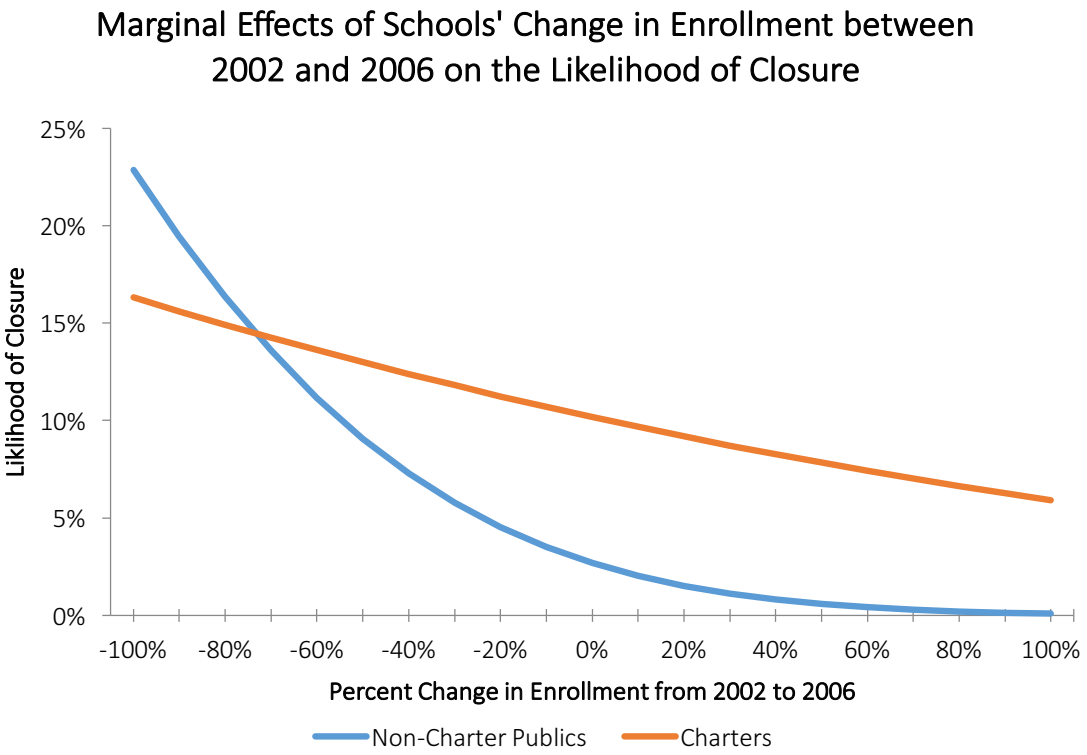


Figure 14: Chapter III: Marginal Effects of Schools' Change in Enrollment between 2002 and 2006 on the Likelihood of Closure



REFERENCES

Abdulkadiroglu, A., Angrist, J., Dynarski, S., Kane, T., & Pathak, P. (2011). "Accountability and Flexibility in Public Schools: Evidence from Boston's Charters and Pilots" Quarterly Journal of Economics, 126(2), 699-748.

Abdulkadiroglu, A., Hu, W., & Pathak, P. (2013). Small High Schools and Student Achievement: Lottery-Based Evidence from New York City. NBER Working Papers 19576. National Bureau of Economic Research.

Ahn, T., & Vigdor, J. (2014). The Impact of No Child Left Behind's Accountability Sanctions on School Performance: Regression Discontinuity Evidence from North Carolina. NBER Working Paper No. w20511.

American Federation of Teachers. (1996). Charter School Laws: Do They Measure Up?

American Institutes for Research. (2005). National AYP and Identification Database (NAYPI). Washington, D.C. Data retrieved via webcrawl on November 12, 2008 from: http://www.air.org/publications/naypi.data.download.aspx.

Angrist, J., Dynarski, S., Kane, J., Pathak, P., & Walters, C. (2011). "Who Benefits from KIPP?" Discussion Paper prepared for the Institute for the Student of Labor (IZA). DP 5690. Available at: http://economics.mit.edu/files/6965, accessed December 17, 2015.

Authors. (2011). Barnard/Columbia No Child Left Behind Database, 2002-2003 and 2003-2004. http://www.gsb.columbia.edu/nclb.

Aviv, R. (2014) Wrong Answer. In an era of high-stakes testing, a struggling school made a shocking choice. The New Yorker. New York, NY. Retrieved on August 2, 2016 from: http://www.newyorker.com/magazine/2014/07/21/wrong-answer.

Balfanz, R., Herzog, L., & MacIver, D. (2007). Preventing student disengagement and keeping students on the graduation path in urban m iddle grade schools: Early identification and effective interventions. *Educational Psychologist, 42*(4), 223-235.

Ballou, D. & Springer, M.G. (2008). Achievement Trade-offs and No Child Left Behind. Working Paper. Urban Institute.

Barrow, L., Claessens, A., & Schanzenbach, D. (2015). The Impact of Chicago's Small High School Initiative. *Journal of Urban Economics, 87,* 100-113.

Belfield, C. & Levin, H.M. (2005). Privatizing Educational Choice: Consequences for Parents, Schools, and Public Policy. Paradigm Publishers. Boulder, CO.

Betts, J. & Tang, Y.E. (2011). The effect of charter schools on student achievement: A meta-analysis of the literature. Bothell, WA: Center on Reinventing Public Education.

Bierlein Palmer, L. and Gau, R. (2003). Charter School Authorizing: Are States Making the Grade? Report. Washington, DC: Thomas Fordham Institute.

Bifulco, R., and Ladd, H. (2006). The Impact of Charter Schools on Student Achievement: Evidence from North Carolina. American Education Finance Association, 1(1), 50-90, Winter.

Bloom, H. & Unterman, R. (2012). Sustained Positive Effects on Graduation Rates Produced by New York City's Small Public High Schools of Choice. MDRC Policy Brief.

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

Boyd, W. L. (1979) Educational policy making in declining suburban school districts: some preliminary findings. Education and Urban Society 11: 333-366.

Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Education Policy Analysis Archives, 12*(1), 1-43.

Brummet, Q. (2014). The effect of school closings on student achievement. Journal of Public Economics, 119, 108–24.

Budde, R. (1988). Education by Charter: Restructuring School Districts: Key to long-term improvement in American Education. Regional Laboratory for Educational Improvement of the Northeast and Islands. Retrieved on October 8, 2016 from http://eric.ed.gov/?id=ED295298.

Buddin, R. & Zimmer, R. (2005). Student achievement in charter schools: A complex picture. *Journal of Policy Analysis and Management* (24), 351–371.

Bulkley, K. (2001). Educational Performance and Charter School Authorizers: The Accountability Bind. *Education Policy Analysis Archives, 9(37)*. Retrieved on May 1, 2014 from http://epaa.asu.edu/epaa/v9n37.html.

Bulkley, K. (2005). Understanding the Charter School Concept in Legislation: The Cases of Arizona, Michigan, and Georgia. International Journal of Qualitative Studies in Education, 18: 527-54.

Bulman, R., & Kirp, D. (1999). From vouchers to charters: The shifting politics of school choice. In S. D. Sugarman & F. R. Kemerer (Eds.), School choice and social controversy: Politics, policy, and law (pp. 36–67). Washington, DC: Brookings Institution.

Burlingame, M. (1979). Declining Enrollments and Small Rural Cities and Districts: An Exploratory Analysis. *Education and Urban Society* (11)*,* 313-332.

California State Statutes. (2006). Retrieved on October 1, 2014 from: http://www.procopio.com/userfiles/file/user/California%20Charter%20School%20Laws%20and%20Regulations%20-%20An%20Annual%20Compilation%20of%20Selected%20Provisions.pdf and http://www.lao.ca.gov/handouts/education/2006/Charter_School_Policy_080106.pdf.

Card, D. & Krueger, Alan B. (1996). Labor Market Effects of School Quality: Theory and Evidence. NBER Working Papers 5450. National Bureau of Economic Research.

Carlson, D., & Lavertu, S. (2015). School Closures and Student Achievement: An Analysis of Ohio's Urban District and Charter Schools. Working Paper. Thomas B. Fordham Institute.

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

CER (Center for Education Reform). (2006). Charter Schools Today: Changing the Face of American Education. Washington, DC: CER.

CER (Center for Education Reform). (2011). The State of Charter Schools: What We Know-and What We Do Not-About Performance and Accountability. Washington, DC: CER. Retrieved on May 1, 2014 from: https://www.edreform.com/wp-content/uploads/2011/12/StateOfCharterSchools_CER_Dec2011-Web-1.pdf.

Chetty, R., Friedman, J., & Rockoff, J. (2011) The Long-Term Impacts of Teachers: Teacher Value-Added and Student Outcomes in Adulthood. NBER Working Papers 17699. National Bureau of Economic Research.

Chiang, H. (2009). How accountability pressure on failing schools affects student achievement. *Journal of Public Economics, 93*(9-10), 1045-1057.

Chubb, J., & Moe, T. (1990). Politics, Markets and America's Schools. Washington, D.C.: Brookings Institution Press.

Cohen, D.K., & Moffitt, Susan L. (2009). The Ordeal of Equality: Did Federal Regulation Fix the Schools? Cambridge, Massachusetts: Harvard University Press.

Coleman, J. (1966). Equality of educational opportunity. Washington, D.C. U.S. Government Printing Office.

Colton, D., & Frelich, A. (1979). Enrollment decline and school closings in a large city. *Education and Urban Society,* 11: 396-417.

Congressional Budget Office. (1986). Trends in Educational Achievement. Washington DC.

Council of Chief State School Officers. (2005). School Data Direct. Washington, D.C. Data retrieved on December 2, 2007 from: http://www.schooldatadirect.org.

CREDO. (2009). Multiple Choice: Charter School Performance in 16 States. Center for Research on Education Outcomes (CREDO). Stanford, CA. Retrieved on May 31, 2016 from: http://credo.stanford.edu/reports/MULTIPLE_CHOICE_CREDO.pdf.

CREDO. (2012). Charter School Performance in New Jersey. Center for Research on Education Outcomes (CREDO). Stanford, CA. Retrieved on May 31, 2016 from: http://credo.stanford.edu/pdfs/nj_state_report_2012_FINAL11272012.pdf.

CREDO. (2014). Charter School Performance in Ohio. Center for Research on Education Outcomes (CREDO). Stanford, CA. Retrieved on May 31, 2016 from: https://credo.stanford.edu/pdfs/OHReport12182014_FINAL.pdf.

CREDO. (2015). National Charter School Study. Center for Research on Education Outcomes (CREDO). Stanford, CA. Retrieved on May 31, 2016 from: https://credo.stanford.edu/documents/NCSS%202013%20Final%20Draft.pdf.

Dardick, H., & Ahmed-Ullah, N. (2012). Alderman Want Hearings on Chicago School Closings. articles.chicagotribune.com/2012-10-05/news/ct-met-chicago-aldermen-schools-1005-20121005_1_school-closings-charter-schools-alderman.

Darling-Hammond, L., (2006). No Child Left Behind and High School Reform. Harvard Educational Review, 76(4): 642-667.

Davidson, E., Reback, R., Rockoff, J., & Schwartz, H.L. (2015). Fifty Ways to Leave a Child Behind: Idiosyncrasies and Discrepancies in States' Implementation of NCLB. Educational Researcher, 44(6), 347-358.

De La Torre, M., & Gwynne, J. (2009). When schools close: Effects on displaced students in Chicago Public Schools." Chicago: Consortium on Chicago School Research.

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. (2011). The impact of No Child Left Behind on Student Achievement. Journal of Policy Analysis and Management, John Wiley & Sons, Ltd., vol. 30(3), pages 418-446, Summer.

Dee, T., Jacob, B., & Schwartz, N.L. (2013). The effects of NCLB on school resources and practices. Education Evaluation and Policy Analysis, 35(20), 252-279.

Deeter, Tom. (2013). Iowa Department of Education. Personal communication via phone on 3/5/13.

Deruy, E. (2016). Unequal Discipline at Charter Schools. The Atlantic Magazine. New York, NY. Retrieved on May 31, 2016 from: http://www.theatlantic.com/education/archive/2016/03/unequal-discipline-at-charter-schools/474459/.

Dynarski, S., Hoxby, C., Loveless, T., Schneider, M., Whitehurst, G., & Witte, J. (2010). Charter schools: A report on rethinking the federal role in education. Washington, DC: The Brown Center on Education Policy at the Brooking Institute.

Education Trust-Midwest. (2016). Accountability for all: 2016, The Broken Promise of Michigan's Charter Sector. The Education Trust-Midwest. Royal Oak, MI. Retrieved on August 10, 2016 from: https://midwest.edtrust.org/accountability-for-all/.

Engberg, J., Gill, B., Zamarro, G., & Zimmer, R. (2011). Closing Schools in Shrinking Districts: Do Student Outcomes Depend on Which Schools Are Closed? Paper presented at the annual meeting of the Association for Public Policy Analysis and Management, Washington, D.C.

Feng, L. (2010). Hire Today, Gone Tomorrow: New Teacher Classroom Assignments and Teacher Mobility. Education Finance and Policy, Vol. 5, No. 3: 278–316.

Feng, L., & Sass, T. (2011). Teacher Quality and Teacher Mobility. CALDER Working Paper #57.Urban Institute, Washington, D.C.

Figlio, D. (1999). Functional Form and the Estimated Effects of School Resources. Economics of Education Review, 18 (2): 242 – 252.

Figlio, D., & Loeb, S. (2011). School Accountability. In Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, editor: Handbooks in Economics, The Netherlands: North-Holland, 3: 383-421.

Figlio, D., & Rouse, C. (2006). Do accountability and voucher threats improve low-performing schools? *Journal of Public Economics, 90*(1-2), 239-255.

Finn, C., Manno, B., & Vanourek, G. (2000). Charter schools in action: Renewing public education. Princeton, NJ: Princeton University Press. ED 451 273.

Firestone, W.A., Fuhrman, S.H., & Kirst, M.W. (1991). State educational reform since 1983: Appraisal and the Future. Educational Policy, 5(3): 233-250.

Fredricks, J., Blumenfeld, P., Friedel, J., & Paris, A. (2005). School engagement. In K. Moore, & L. Lippman, *Conceptualizing and measuring indicators of positive development: what do children need to flourish?* New York: Kluwer Academic/Plenum Press.

Gaddis, S., & Lauen, D. (2014). School accountability and the black–white test score gap. Social Science Research, 44(0), 15-31.

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

Gadlin, S. (2013). Thousands prepare for a three-day march against school closings. Chicago Teachers Union. Retrieved from: http://www.ctunet.com/blog/thousands-prepare-for-a-three-day-march-against-school-closings.

Galletta, A., & Ayala, J. (2008). Erasure and survival: Creating a future and managing a past in a restructuring high school. The Teachers College Record, 110(9), 1959- 1985.

Giordano, M.A. (2012). A Final Decision for 24 City Schools. WNYC's Schoolbook. Retrieved from: http://www.schoolbook.org/2012/04/27/a-final-decision-for-24-city-schools/.

Gleason, P., Clark, M., Tuttle, C., & Dwoyer, E. (2010). The evaluation of charter school impacts. Washington DC: US Department of Education. Available at: http://eric.ed.gov/?id=ED560080.

Goldstein, J., Kelemen, M. and Koski, W. (1998). Reconstitution in Theory and Practice: The Experience of San Francisco. Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).

Gordon, N., & Reber, S. (2015). The quest for a targeted and effective Title I ESEA: Challenges in designing and implementing fiscal compliance rules, mimeo, Georgetown University.

Graham, K. (2012). Philadelphia Superintendent Identifies Schools He Intends to Close. The Philadelphia Inquirer. Available at: http://www.philly.com/philly/news/20121214_Philadelphia_superintendent_identifies_sc.

Hamilton, L.S., Stecher, B.M., Marsh, J.A., McCombs, J.S., Robyn, A., Russell, J., Naftel, S., & Barney, H. Standards-Based Accountability Under No Child Left Behind: Experiences of Teachers and Administrators in Three States. Santa Monica, California: RAND Corporation.

Hanushek, E. A. (1997). Assessing the Effects of School Resources on Student Performance: An Update. Educational Evaluation and Policy Analysis. 19(2): 141-164.

Hanushek, E., & Raymond, M. (2005). Does school accountability lead to imprvoed student performance? *Journal of Policy Analysis and Management*(24), 297-327.

Hanushek, E., Kain, J., Rivkin, S. & Branch, G. (2007). Charter School Quality and Parental Decision Making with School Choice. Journal of Public Economics, Elsevier, vol. 91(5-6), pages 823-848, June.

Hanushek, E.A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. Journal of Economic Literature, American Economic Association, vol. 24(3): 1141-77.

Hanushek, E.A., & Raymond, M.E. (2002). Sorting Out Accountability Systems. In Williamson M. Evers, Herbert J. Walberg (Eds.), School Accountability, 2002, pp. 75-104.

Hanushek, E.A., Kain, J.F., & Rivkin, S.G. (2004). Disruption versus Tiebout improvement: the Costs and Benefits of Switching Schools." Journal of Public Economics (88): 1721-1746.

Harris, D., & Herrington, C. (2006). Accountability, Standards, and the Growing Achievement Gap: Lessons from the Past Half-Century. *American Journal of Education, 112*(2), 209-238.

Hassel, B. C. (1999). The charter school challenge: Avoiding the pitfalls, fulfilling the promise. Washington, DC: The Brookings Institution.

Hassel, B., Terrell, M., & Kowal, J. (2006). Florida Charter Schools: Hot and Humid with Passing Storms. Washington, DC: Education Sector.

Hedges, L.V., Laine, R.D. & Greenwald, R. (1994). Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes. Educational Researcher, 23(3): 5-14.

Henig, J. (1994). Rethinking School Choice: Limits of the Market Metaphor. Princeton University Press.

Henig, J. (2013). The End of Exceptionalism in American Education: The Changing Politics of School Reform. Cambridge, Massachusetts: Harvard Education Press.

Hess, F., & Finn, C. (2007). Crash Course: NCLB is driven by education politics. Education Next, 7(4): 40-45. Retrieved on August 1, 2016 from: http://educationnext.org/crash-course/.

Hess, G. A. (2003). Reconstitution—three years later: Monitoring the effect of sanctions on Chicago high schools. Education and Urban Society, 35(3), 300-327.

Holyoke, T.T., Henig, J.R., Brown, H., & Lacireno-Paquet, N. (2009). Policy dynamics and the evolution of state charter school laws. Policy Sciences, 42 (1), 33-55.

Hoxby, C., & Rockoff, J. (2004). "The Impact of Charter Schools on Student Achievement," Department of Economics, Harvard University, mimeo.

Hoxby, C.M., Murarka, S., & Kang, J. (2009). The New York City charter schools evaluation project. NBER Working Papers 14852. National Bureau of Economic Research.

Jennings, J. & Crosta, P. (2011). The Unaccountables. Paper presented at the 2011 conference of the Association for Education Finance and Policy.

Kane, T., Abdulkadiroglu, A., Angrist, J., Cohodes, S., Dynarski, S., Fullerton, J., & Pathak, P. (2009). Informing the debate: Comparing Boston's charter, pilot, and traditional schools. Boston, MA: Boston Foundation.

Kane, T.J. & Staiger, D. (2003). Unintended consequences of racial subgroup rules. in P.E.

Kane, T.J., & Staiger, D. (2002). The Promise and pitfalls of using imprecise school accountability measures. Journal of Economic Perspectives, 16, 91-114.

Kemple, J.J. (2016). High School Closures in New York City: Impacts on Students' Academic Outcomes, Attendance, and Mobility. The Research Alliance for New York City Schools. Retrieved in Aug. 2016 from: http://steinhardt.nyu.edu/scmsAdmin/media/users/sg158/PDFs/hs_closures/HighSchoolClosuresinNewYorkCity_ResearchAllianceforNYCSChools_pdf.pdf.

Kerbow, D. (1996). Patterns of urban student mobility and local school reform. Journal of Education for Students Placed at Risk, 1, 147–169.

Kirshner, B., Gaertner, M., & Pozzoboni, K., Tracing transitions: A multiple methods analysis of the impact of school closures on displaced students. Paper presented at the American Educational Research Association, 2009.

Kolderie, T. (1990). Beyond choice to new public schools: Withdrawing the exclusive franchise in public education (8). Washington, DC: Progressive Policy Institute.

Krueger, A. B., & Whitmore, D. (2001, March). Would smaller classes help close the Black–White achievement gap? (Princeton University Industrial Relations Section Working Paper No. 451). Retrieved August 1, 2016 from: http://dataspace.princeton.edu/jspui/handle/88435/dsp01w66343627.

Kuziemko, I. (2006). Using Shocks to School Enrollment to Estimate the Effect of School Size on Student Achievement. Economics of Education Review, 25 (1): 63-75.

Ladd, H. (1999). The Dallas school accountability and incentive program: an evaluation of its impacts on student outcomes. Economics of Education Review, vol. 18(1): 1-16.

Ladd, H., & Walsh, R. (2002). Implementing Value-Added Measures of School Effectiveness: Getting the Incentives Right. Economics of Education Review, 25(1): 1-17.

Ladd, H.F. & Lauen, D.L. (2009). Status versus growth: The distributional effects of school accountability policies. Working paper. Urban Institute.

Larsen, M. (2014). Does Closing Schools Close Doors? The Effect of High School Closings on Achievement and Attainment. Job Market Paper. Tulane University. New Orleans, LA.

Lauen, D., & Gaddis, S. (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.

Linn, R. L. (2003) “Accountability: Responsibility and Reasonable Expectations,” Educational Researcher, 32, pages 3-13.

Linn, R. L., Graue, M. E., & Sanders, N. M. (1990). Comparing state and district results to national norms: The validity of claims that "everyone is above average. Educational Measurement: Issues and Practice (Fall), 5-14.

Lipman, P., & Haines, N. (2007). From education accountability to privatization and African American exclusion—Chicago Public Schools: Renaissance 2010. Educational Policy, 21, 471-502.

Losen, D., Keith, M., Hodson, C., & Martinez, T. (2016). Charter Schools and School Discipline: A Comprehensive Review. Center for Civil Rights Remedies at the University of California, Los Angeles. Available at: https://civilrightsproject.ucla.edu/resources/projects/center-for-civil-rights-remedies/school-to-prison-folder/federal-reports/charter-schools-civil-rights-and-school-discipline-a-comprehensive-review.

Louisiana State Statutes. (2006). Retrieved on October 1, 2014 from: http://law.justia.com/codes/louisiana/2006/143/143.html.

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

Manna, P. (2010). Collision Course: Federal Education Policy Meets State and Local Realities. CQ Press.

McDonnell, L. (2005). No Child Left Behind and the Federal Role in Education: Education or Revolution? Peabody Journal of Education 80. (2): 19-38.

McGuinn, P. (2012). Stimulating Reform: Race to the Top, Competitive Grants and the Obama Education Agenda. Education Policy, 26: 136 – 159.

Mintrom, M. & Vergari, S. (2009). Policy Networks and Innovation Diffusion: The Case of State Education Reforms. The Journal of Politics, 60 (01); 126-148.

Miron, Gary. 2005. “Strong Charter School Laws Are Those That Result in Positive Outcomes.” Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, Canada.

Nathan, J. (1996). Charter Schools: Creating Hope and Opportunity for American Education. San Francisco, CA: Jossey-Bass Publishers.

National Alliance for Public Charter Schools. (2016). Charter School Data Dashboard. Washington, D.C. Data retrieved on May 1, 2016 from: http://dashboard.publiccharters.org/Home/?p=Home.

National Center for Education Statistics. (2000). School District Demographic System. U.S. Department of Education. Washington, D.C.

National Center for Education Statistics. (2003). Schools and Staffing Survey. U.S. Department of Education. Washington, D.C.

National Center for Education Statistics. (2013). Common Core of Data. U.S. Department of Education. Washington, D.C. Retrieved on August, 1 2013 from: https://nces.ed.gov/ccd/.

National Commission on Excellence in Education. (1983). A Nation at Risk: The Imperative for Educational Reform. U.S. Department of Education. Washington, DC.

National Education Association (2008): Class Size Reduction: A Proven Reform Strategy. Policy Brief. Retrieved on August 2, 2016 from: http://www.nea.org/assets/docs/PB08_ClassSize08.pdf.

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

New York City Charter School Center. (2016). 2015-2016 Charter School Facts: Demand and Growth. Retrieved on August 2, 2016 from: http://www.nyccharterschools.org/facts.

Nix, N. (2016). NAACP May Double Down on Charter School Opposition as Civil Rights Allies Strongly Disagree. The Seventy-Four Million. New York, NY. Retrieved on May 31, 2016 from: https://www.the74million.org/article/naacp-may-double-down-on-charter-school-opposition-some-civil-rights-allies-strongly-disagree.

North Carolina State Statutes. (2005). Retrieved on October 1, 2014 from: http://law.justia.com/codes/north-carolina/2005/chapter_115c/chapter_115c.html.

Oregon State Statutes. (2005). Retrieved on October 1, 2014 from: http://law.justia.com/codes/oregon/2005/vol9/338.html.

Ozek, U., Hansen, M., & Gonzalez, T. (2012). A Leg Up or a Boot Out? Student Achievement Data in Education Research. CALDER Center. Working Paper 78.

Peterson & M.R. West (Eds.), No Child Left Behind? The Politics and Practice of accountability (pp. 152-176). Washington, DC: Brookings Institution Press.

PEW Charitable Trusts. (2011). Philadelphia Research Initiative: Closing Public Schools in Philadelphia: Lessons from Six Urban Districts.

Phillips, A. (2011). 12 with Low State Test Scores Are Put on School Closing List. The New York Times. Available at: http://www.nytimes.com/2011/12/09/nyregion/12-new-york-.

Plank, S., Farley-Ripple, E., Durham, R., & Norman, O. (2009). *First grade and forward: A seven-year examination within the Baltimore City Public School System.* Baltimore MD: Baltimore Education Research Consortium.

Polikoff, M., & Wrabel, S. (2013). When is 100% not 100%? The Use of safe harbor to make Adequate Yearly Progress. Education Finance and Policy, 8(2), 251-270.

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

Pritchett, L. & Filmer, D. (1997). What educational production functions really show: a positive theory of education spending. Policy Research Working Paper Series 1795, The World Bank.

Reardon, S. (2011). The Widening Academic-Achievement Gap between the Rich and the Poor: Disadvantage. Murnane, R.M. & Duncan, G. (Eds.). Washington, DC; Brookings Institution.

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

Reback, R., Rockoff, J., & Schwartz, H. (2014). Under pressure: Job security, resource allocation, and productivity in schools under NCLB. *American Economic Journal: Economic Policy, 6*(3).

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

Renzulli, L., & Roscigno, V. (2007). Charter schools and the public good. Contexts, 6(1):31-36.

Rice, J. K. & Malen, B. (2003). The human costs of education reform: The case of school reconstitution. Educational Administration Quarterly, 39(5), 635.

Rich, M. (2012). Enrollment Off in Big Districts, Forcing Layoffs. The New York Times. New York, NY. Available at: www.nytimes.com/2012/07/24/education/largest-school-districts-see-steady-drop-in-enrollment.html?pagewanted=all.

Riddle, W., & Kober, N. (2012). What impact will NCLB waivers have on the consistency, complexity and transparency of state accountability systems? Center on Education Policy: Washington, D.C.

Rockoff, J., & Turner, L. (2010). Short-run Impacts of Accountability on School Quality, American Economic Journal, Economic Policy 2(4): 119-147.

Rumberger, R. (2003). The causes and consequences of student mobility. Journal of Negro Education, 72, 6-2.

Santos, F. (2011). Teachers' Union Sues to Stop School Closings. New York Times. Retrieved on August 2, 2016 from: http://cityroom.blogs.nytimes.com/2011/05/18/teachers-union-sues-to-stop-school-closings/?_r=0.

Santos, F. (2012). City Pushes Ahead to Plan to Close, Then Reopen 33 Schools. WNYC's School Book. New York, New York. Available at: www.schoolbook.org/2012/02/28/city- pushes-ahead-on-plan-to-close-then-reopen-33-schools.

Schanzenbach, D. W. (2014). "Does Class Size Matter?" In: Policy Briefs, National Education Policy Center, School of Education, University of Colorado, Boulder.

Scott, C. (2008). A call to restructure restructuring. Washington DC: Center on Education Policy.

Shanker, A. (1988). Speech at the National Press Club on March 31, 1988. Retrieved on October 10, 2016 from the Walter P. Reuther Library at http://reuther.wayne.edu/files/64.43.pdf.

Shober, A., Manna, P., & Witte, J. (2006). Flexibility Meets Accountability: State Charter School Laws and Their Influence on the Formation of Charter Schools in the United States. Policy Studies Journal, 34(4): 563-587.

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

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.

SRI International. (2000). Evaluation of the public charter schools program: Year one evaluation report. Washington, DC: SRI International.

SRI International. (2004). *Evaluation of the Public Charter Schools Program: Final Report*. Washington, DC: U.S. Department of Education.

Srikantaiah, D. (2009). How State and Federal Accountability Policies Have Influenced Curriculum and Instruction in Three States: Common Findings from Rhode Island, Illinois, and Washington. Washington, D.C.: Center for Education Policy.

Stedman, L. C. & Kaestle, C. F. (1991). The great test score decline: A closer look. In C. F. Kaestle, H. Damon-Moore, L. C. Stedman, K. Tinsley, & W. V. Trollinger (Eds.), Literacy in the United States (Chapter 4). New Haven: Yale University Press.

Stedman, L. C. (1993). The condition of education: Why school reformers are on the right track. Phi Delta Kappan, 75 (3), 215-225.

Steiner, L. (2009). Tough decisions: Closing persistently low-performing schools. Lincoln, NE: Academic Development Institute.

Temple, J. A., & Reynolds, A. J. (1999). School mobility and achievement: Longitudinal findings from an urban cohort. Journal of School Psychology, 37 (4), 355-377.

Thomas B. Fordham Institute (2008). Accelerating Student Learning in Ohio" Dayton: Thomas B. Fordham Institute.

Thomas B. Fordham Institute with Palmer, L., Terrel, M., Hassel, B., & Svahn, C. (2006). Turning the Corner to Quality: Policy Guidelines for Strengthening Ohio's Charter Schools. Retrieved on August 15, 2016 from: http://www.edexcellence.net/doc/TurningtheCornertoQualityOH2006.pdf.

Toch, T., & Harris, D. (2008). Salvaging accountability. Education Week.

Tuttle, C., Gleason, P., Knechtel, V., Nichols-Barrer, I., Booker, K., Chojnacki, G., Coen, T., & Goble, L. (2015). Understanding the Effect of KIPP as It Scales: Volume I, Impacts on Achievement and Other Outcomes. Final Report of KIPP's Investing in Innovation Grant Evaluation. Executive Summary.

Tyack, D., & Cuban, L. (1995). Tinkering toward Utopia: A Century of Public School Reform. Cambridge, MA: Harvard University Press.

U.S. Department of Education (1994). Biennial Evaluation Report FY 93-94, Chapter 101, Education of Disadvantaged Children (Chapter 1, ESEA) Formula Grants to Local Education Agencies. Available at https://www2.ed.gov/pubs/Biennial/101.html.

U.S. Department of Education (2001). Digest of Education Statistics, Table 369, Available at https://nces.ed.gov/programs/digest/d01/dt369.asp.

U.S. Department of Education (2006). Digest of Education Statistics, Table 368, Available at https://nces.ed.gov/programs/digest/d06/dt06_368.asp.

U.S. Department of Education (2012). Digest of Education Statistics, Table 205, Summary of expenditures for public elementary and secondary education, by purpose, 1919-20 through 2009-10, Available at https:/nces.ed.gov/programs/digest/d12/tables/dt12_205.asp.

U.S. Department of Education. (2002). Evaluation of the Public Charter Schools Program: Final Report. Retrieved on April 24, 2016, from https://www2.ed.gov/rschstat/eval/choice/pcsp-final/finalreport.pdf.

U.S. Department of Education. (2008). Differentiated accountability: Targeting resources to schools with the greatest need. Retrieved on April 24, 2009, from www.ed.gov/nclb/accountability/differentiated/factsheet02.pdf.

United Federation of Teachers (2016). Legislative Priorities. Retrieved on August 2, 2016 from: http://www.uft.org/political-action/legislative-priorities.

United States Department of Education (1994a). The Reauthorization of the Elementary and Secondary Education Act. Available at: http://www2.ed.gov/legislation/ESEA/toc.html.

United States Department of Education (1994b). Goals 2000: Educate America Act. Available at: http://www2.ed.gov/PDFDocs/g2kfinal.pdf.

United States Department of Education (2001). No Child Left Behind Act of 2001. Available at: http://www2.ed.gov/policy/elsec/leg/esea02/index.html.

United States Department of Education (2009). Race to the Top Fund. Available at: http://www2.ed.gov/programs/racetothetop/index.html.

Valencia, R. (1984). School Closures and Policy Issues. Policy Paper No. 84-C3, ERIC No. ED323040.

Vergari, S. (2001). Charter School Authorizers: Public Agents for Holding Charter Schools Accountable. Education and Urban Society 33, 2: 129-140.

Whitehurst, G., & Chingos, M. (2011). Class size: What research says and what it means for state policy. Washington, DC: Brown Center for Education Policy at Brookings.

Winters, M., Trivitt, J., & Greene, J. (2010). The impact of high-stakes testing on student proficiency in low-stakes subjects: Evidence from Florida's elementary science exam. *Economics of Education Review, 29*(1), 138-146.

Witte, J. F., Shober, A. F., & Manna, P. (2003). Analyzing state charter school laws and their influence on the formation of charter schools. Presented at the Annual Meeting of the American Political Science Association, Philadelphia, PA.

Wohlstetter, P., Smith, J., & Farrel, C. (2013). Choices and Challenges: Charter School Performance in Perspective. Cambridge, MA: Harvard Education Press.

Wohlstetter, P., Wenning, R., & Briggs, K. (1995). Charter Schools in the United States: The Question of Autonomy. Education Policy (4), 331-358.

Xu, Z., Hannaway, J., & D'Souza, S. (2009). Student Transience in North Carolina: The Effect of Mobility on Student Outcomes Using Longitudinal Data. Washington, DC: National Center for Analysis of Longitudinal Data in Education Research, Working Paper 22.

Young, V., Humphrey, D., Wang, H., Bosetti, K., Cassidy, L., & Wechsler, M. (2009). Renaissance schools fund-supported schools: Early outcomes, challenges, and opportunities. Menlo Park, CA: SRI International.

Zimmer, R., Gill, B., Booker, K., Lavertu, S., Sass, T., & Witte, J. (2009). Charter Schools in Eight States: Effects on Achievements, Attainment, Integration, and Competition. RAND Corporation. Santa Monica, CA.

APPENDICES

Appendix A: Post-NCLB Accountability Changes

Waiver Program

To receive a waiver, states agreed to further differentiate amongst low-performing schools by both expanding their ratings categories and including additional measures of performance. The submission and approval of waivers was widespread. By October 2013, the USDOE approved waiver applications for 43 states. Polikoff, et al., (2014) examined these waiver applications and determined that some states "strengthened" their accountability systems by improving upon some of NCLB's flaws while other states maintained or "weakened" their accountability systems by either retaining flawed NCLB components or introducing new complexities that make interpretations of school performance much more difficult. On the whole, the introduction of the waiver program reset the threshold for NCLB intervention which alleviated the burden faced by states and districts of increasing numbers of schools failing AYP.

Unlike NCLB's designations of "passing" and "failing," state waiver applications needed to include states' plans to categorize schools as "priority," "focus," or "reward," in addition to the binary designation of passing or failing AYP. Priority schools are those in the bottom 5 percent in terms of an index measure that includes proficiency rates (and in some cases, measures of annual growth), and graduation rates or attendance rates. Focus schools are those with large within-school gaps between racial/ethnic groups and low- and high-income students on any of the aforementioned measures. Reward schools are the highest performers on the defined index as long as the school does not have large outcome gaps between student groups. Waivers also allowed states to expand the measures they use to determine school performance. Some states

added measures like educator effectiveness and students' college and career readiness to their school performance calculations.

School Improvement Grant School Turnaround Options

In addition to the waiver program, RTT appropriated $4.35 billion to education through the provision of competitive School Improvement Grants (SIG) (USDOE, 2009). To qualify for SIGs, states committed to employing a specified school improvement model to improve its lowest performing schools (USED, 2009).[71] Unlike NCLB, RTT specifically lists school closure as one of the five reform models that states can use to comply with the terms of the grant. RTT, much like NCLB, incentivized states and districts to abandon their reform efforts that tinker with school operations, staffing, or resources at the margin, and instead embrace efforts that attempt to dismantle and rebuild schools that persistently failed AYP.

1. *Turnaround Model:* This option requires that states replace an existing principal with a new principal who is granted additional autonomy to substantially change the operations of the school.
2. *Transformation Model*: This option is similar to the turnaround model, but also includes the implementation of teacher and principal evaluation and support systems.
3. *Restart Model*: This option requires that states to close or "restart" a school under the management of a Charter Management Organization (CMO)
4. *Closure Model:* This option requires that states close a school and enroll the students of the closed school in other, higher-achieving district schools.
5. *Evidence-based Whole-School Reform Model:* This option, added in the 2016 school year, requires states to implement one of four proprietary and non-proprietary intervention models, including Success for All, Institute for Student Achievement, Positive Action, and Small Schools of Choice.

[71] For a complete description of the school improvement models, see Appendix B.

6. *Early Learning Model:* This option, added in the 2016 school year, requires states to establish or expand schools' offerings to include full-day kindergarten and high-quality preschool.

7. *State-determined Intervention Model:* This option, added in the 2016 school year, requires states to submit a school turnaround model for approval by the USDOE.

Appendix B: Summary of Data Corrections in the CCD Data

Tennessee

Tennessee was missing CCD for school year 2001-2002 for the following variables:

- Total school enrollment – replaced with data from 1998-1999 school year
- Black student enrollment – replaced with data from 1998-1999 school year
- Hispanic student enrollment – replaced with data from 1998-1999 school year
- White student enrollment – replaced with data from 1998-1999 school year
- Native American student enrollment – replaced with data from 1998-1999 school year
- Asian American student enrollment – replaced with data from 1998-1999 school year
- Pupil:teacher ratio – replaced with data from the 2004-2005 school year
- FRPL student enrollment – replaced with data from the 2004-2005 school year
- School eligibility for Title I funds – replaced with data from 2004-2005 school year

Arkansas

Arkansas was missing CCD indicators on school locale (i.e., city, suburb, town, rural) for 2001-2002, so I replaced those missing values with locale data from 2002-2003 school year.

Arizona

Arizona was missing CCD information on the enrollment of students eligible for free- or reduced-priced lunch for 2002, and indicators for whether the school was eligible for Title 1 funds in 2002. I replaced these missing values with values from 2006 and 2003, respectively.

Connecticut

Connecticut was missing CCD information on the enrollment of students eligible for free- or reduced-priced lunch for 2002, so I replaced these missing values with values from 2003.

Hawaii

Hawaii was missing CCD indicators on school locale (i.e., city, suburb, town, rural) for 2001-2002, so I replaced those missing values with locale data from 2002-2003 school year.

Massachusetts

Massachusetts was missing CCD information on pupil:teacher ratio for 2001-2002, so I replaced those missing values with locale data from 2003-2004 school year.

Michigan

Michigan was missing CCD information on whether the school was eligible for Title 1 funds in 2002. I replaced these missing values with values from 2005.

New Jersey

New Jersey was missing CCD indicators on school locale (i.e., city, suburb, town, rural) for 2001-2002, so I replaced those missing values with locale data from 2002-2003 school year.

Wyoming

Wyoming was missing CCD information on the enrollment of students eligible for free- or reduced-priced lunch for 2002, so I replaced these missing values with values from 2003.

Appendix C: Chapter II: Main Results for Non-charter Public Schools

	Baseline	Baseline + Interaction	Baseline + Prof Rank	Baseline + Interaction + Prof Rank
School Failed Once	**0.004**** 0.058 (0.021)	**0.012***** 0.167 (0.000)	**-0.0001** -0.001 (0.964)	**0.007**** 0.105 (0.018)
School Failed Twice	**0.009***** 0.129 (0.000)	**0.014***** 0.188 (0.001)	**0.002** 0.027 (0.422)	**0.005** 0.080 (0.170)
School Failed Three Times	**0.028***** 0.331 (0.000)	**0.032***** 0.368 (0.000)	**0.013***** 0.181 (0.000)	**0.016***** 0.214 (0.002)
School Failed Once*Union Coverage > 75%		**-0.009***** -0.160 (0.002)		**-0.009***** -0.157 (0.003)
School Failed Twice*Union Coverage > 75%		**-0.005** -0.083 (0.202)		**-0.004** -0.075 (0.258)
School Failed Three Times*Union Coverage > 75%		**-0.003** -0.055 (0.444)		**-0.003** -0.048 (0.511)
Percent Change in Enrollment (2002-2006)	**-0.077***** -1.176 (0.000)	**-0.077***** -1.176 (0.000)	**-0.072***** -1.142 (0.000)	**-0.072***** -1.143 (0.000)
Total Enrollment in 2002 (in hundreds)	**-0.009***** -0.131 (0.000)	**-0.01***** -0.131 (0.000)	**-0.01***** -0.120 (0.000)	**-0.01***** -0.120 (0.000)
Proficiency Rank (in tens)			**-0.003***** -0.054 (0.000)	**-0.003***** -0.053 (0.000)

	-0.00106	**-0.0011**	**-0.004***	**-0.004***
High Schools	-0.016	-0.017	-0.063	-0.063
	(0.622)	(0.602)	(0.073)	(0.071)
	0.018***	**0.018*****	**0.016*****	**0.016*****
Middle Schools	0.236	0.235	0.226	0.224
	(0.000)	(0.000)	(0.000)	(0.000)
	0.014***	**0.014*****	**0.008***	**0.008****
Other Schools	0.177	0.178	0.118	0.119
	(0.002)	(0.002)	(0.050)	(0.047)
	0.016***	**0.016*****	**0.016*****	**0.016*****
Located in a City	0.218	0.221	0.227	0.230
	(0.000)	(0.000)	(0.000)	(0.000)
	-0.01***	**-0.0098*****	**-0.009*****	**-0.009*****
Located in a Rural Area	-0.159	-0.158	-0.150	-0.149
	(0.000)	(0.000)	(0.000)	(0.000)
	0.005**	**0.005****	**0.005****	**0.006****
Located in a Suburb	0.076	0.077	0.084	0.085
	(0.019)	(0.018)	(0.014)	(0.013)
	0.005***	**0.005*****	**0.004*****	**0.004*****
Title I School	0.070	0.070	0.065	0.065
	(0.002)	(0.002)	(0.006)	(0.006)
	0.00003	**0.00003**	**0.0001**	**0.0001**
Pupil:Teacher Ratio	0.000	0.000	0.065	0.001
	(0.468)	(0.457)	(0.215)	(0.218)
Percent of students who are…				
	0.011***	**0.011*****	**0.004**	**0.004**
Eligible for Free or Reduced-Priced Lunch	0.165	0.173	0.063	0.071
	(0.004)	(0.003)	(0.303)	(0.248)
	0.04***	**0.04*****	**0.028*****	**0.028*****
Black	0.608	0.601	0.446	0.441
	(0.000)	(0.000)	(0.000)	(0.000)
	0.015**	**0.015****	**0.003**	**0.003**
Hispanic	0.222	0.223	0.040	0.042
	(0.029)	(0.028)	(0.706)	(0.689)

	-0.032* **	**-0.033*** **	**-0.04*** **	**-0.041*** **
Native American	-0.490	-0.502	-0.628	-0.640
	(0.001)	(0.001)	(0.000)	(0.000)
	-0.024*	**-0.024** **	**-0.027** **	**-0.027** **
Asian	-0.357	-0.360	-0.429	-0.430
	(0.051)	(0.049)	(0.023)	(0.023)
Districts' percent of residents who are...				
	-0.056* **	**-0.057*** **	**-0.041*** **	**-0.041*** **
Bachelor's Degree Recipients	-0.856	-0.860	-0.645	-0.651
	(0.000)	(0.000)	(0.000)	(0.000)
	0.008	**0.009**	**0.015** **	**0.016** **
Black	0.126	0.137	0.236	0.246
	(0.252)	(0.215)	(0.036)	(0.029)
	-0.073* **	**-0.073*** **	**-0.06*** **	**-0.06*** **
Hispanic	-1.107	-1.111	-0.944	-0.950
	(0.000)	(0.000)	(0.000)	(0.000)

Appendix D: Shober, Manna, and Witte (2006) Flexibility and Accountability Index Rules, CER Criteria, and Measuring Up Criteria

Shober Flexibility Index Criteria

- School Cap (note the actual caps)
    1. Charter schools limited by geography.
    2. Limit on total number of charter schools with no distinction between at-risk, urban, or other.
    3. Limits designated per charter school class (e.g. limit on non-at-risk schools and a different limit on at-risk schools).
    4. Limit on one class of charters (e.g. non-at-risk/urban), but no limit on others (e.g. at-risk).
    5. No limit on schools in statute.
- Who may hold a charter for a start-up school? Add one point for each group that may hold a charter: Teacher groups; groups of parents or other community members; non-profits; for-profits; and individuals.
- Who may hold a charter for a conversion school? Add one point for each group that may hold a charter: Teacher groups; groups of parents or other community members; non-profits; for-profits; and individuals.
- Number of first-application authorizers：
    1. Local school districts only, or approval by both local district and a state board required.
    2. Local school districts, but state board may sponsor if local board denies OR one special state board for chartering purposes.
    3. State board of education and one other entity (e.g. local districts).
    4. State board of education and two other entities (e.g. public universities and districts).
    5. State board of education and three or more other entities (e.g. public universities, local districts, and political officials).
- Local support for conversion school applications?
    1. Referendum at the next scheduled school election.
    2. Super-majority of parents and teachers in school.

3. Majority of parents and majority of teachers in school or other local residents.
4. Evidence of significant support from teachers, parents, or other local residents.
5. Approval of responsible school board, school leadership, or city body only.

- Local support for new charter school applications?
    1. Referendum at the next scheduled school election.
    2. Percentage of support from local teachers, parents, or other local residents.
    3. "Adequate" evidence of support from local teachers and parents or other residents.
    4. Interest from local teachers, parents, or other local residents (usually x signatures).
    5. No evidence of local support required.

Shober Accountability Index Criteria

- How much control does the school have over students?
    1. Schools may limit admissions based on admissions criteria outlined in its charter (and the statute does not specify that academic discrimination is illegal).
    2. Schools may limit admissions to a single sex of at-risk students or to those who would "benefit" from a particular academic program — broader than just "at risk."
    3. Schools may limit admissions to a particular geographic area or to at-risk students.
    4. School must admit any person that resides in the district, but may give preference to certain classes of students (e.g. teachers' children, board members' children, siblings).
    5. School must admit any person that resides in the district who would be eligible for the ages or grades served by the school.
- Do state reporting requirements exist for school performance?
    1. No performance reporting required, or unspecified.
    2. Yes, on renewal only.
    3. Yes, at the request of district or state officials.
    4. Yes, reports are required but the form is left to the charter school.
    5. Yes, annual, public report cards (or similar) are required to be submitted to the state.
- Must charter school use state standards?

1. Not required by state statute.
2. Standards may be waived.
3. The school is responsible for course-taking requirements only.
4.
5. Yes

CER Criteria (Summary from Holyoke, et al., 2009, p. 50-51)

1. Number of schools: States that permit a number of autonomous charter schools encourage more activity than states that limit the number of autonomous schools.
2. Multiple chartering authorities/binding appeals process: States that permit a number of entities in addition to or instead of local school boards to authorize charter schools, or that provide applicants with a binding appeals process, encourage more activity.
3. Variety of applicants: States that permit a variety of individuals and groups both inside and outside the existing public school system to start charter schools encourage more activity than states that limit eligible applicants to public schools or public school personnel.
4. New starts: States that permit new schools to start up encourage more activity than those that permit only public school conversions.
5. Schools may start without third-party consent: States that permit charter schools to form without needing consent from competing districts or the general public encourage more activity than those that do not.
6. Automatic waiver from laws and regulations: States that provide automatic blanket waivers from most or all state and district education laws, regulations, and policies encourage more activity than states that provide no waivers or require charter schools to negotiate waivers on an issue-by issue basis.
7. Legal/operational autonomy: States that allow charter schools to be independent legal entities that can own property, sue and be sued, incur debt, control budget and personnel, and contract for services, encourage more activity than states in which charter schools remain under district jurisdiction. In addition, legal autonomy refers to the ability of charter schools to control their own enrollment numbers.
8. Guaranteed full funding: States where 100% of per-pupil funding automatically follows students enrolled in charter schools encourage more activity than states where the amount is automatically lower or negotiated with the district.
9. Fiscal autonomy: States that give charter schools full control over their own budgets, without the district holding the funds, encourage more activity than states that do not.
10. Exemption from collective bargaining agreements/district work rules: States that give charter schools complete control over personnel decisions encourage more activity than states where charter school teachers must remain subject to the terms of district collective bargaining agreements or work rules.

National Alliance for Public Charter Schools: Measuring Up Ranking Criteria (2011)

1. No Caps
2. A Variety of Public Charter Schools Allowed
3. Multiple Authorizers Available
4. Authorizer and Overall Program Accountability System Required
5. Adequate Authorizer Funding
6. Transparent Charter Application, Review and Decision-Making Processes
7. Performance-Based Charter Contracts Required
8. Comprehensive Charter School Monitoring and Data Collection Processes
9. Clear Processes for Renewal, Nonrenewal and Revocation Decisions
10. Educational Service Providers Allowed
11. Fiscally and Legally Autonomous Schools with Independent Public Charter School Boards
12. Clear Student Recruitment, Enrollment and Lottery Procedures
13. Automatic Exemptions from Many State and District Laws and Regulations
14. Automatic Collective Bargaining Exemption
15. Multi-School Charter Contracts and/or Multi-Charter Contract Boards Allowed
16. Extra-Curricular and Interscholastic Activities Eligibility and Access
17. Clear Identification of Special Education Responsibilities
18. Equitable Operational Funding and Equal Access to All State and Federal Categorical Funding
19. Equitable Access to Capital Funding and Facilities
20. Access to Relevant Employee Retirement Systems

Appendix E: Index Scores for each Index by State

State	EKD Index	Shober Index - Accountability	Shober Index - Flexibility	CER Index
AL	-9	-9	-9	-9
KY	-9	-9	-9	-9
ME	-9	-9	-9	-9
MT	-9	-9	-9	-9
ND	-9	-9	-9	-9
NE	-9	-9	-9	-9
SD	-9	-9	-9	-9
VT	-9	-9	-9	-9
WA	-9	-9	-9	-9
WV	-9	-9	-9	-9
MD	1	-9	-9	13.5
ID	1.25	3	3.04	26.5
MI	1.5	5	3.47	42
AK	1.75	2.5	3.25	19
HI	1.75	4.33	2.54	18
AZ	1.75	5	4.29	44
WI	2	2	3.19	31.5
UT	2	2.67	3.63	25.5
RI	2	4.67	2.94	12
IN	2.25	4.33	3.06	41.5
GA	2.25	4.33	3.92	34
DE	2.25	4.67	3.04	44.5
MO	2.25	4.67	3.56	36.5
LA	2.5	1	4.17	25.25
NM	2.5	3.67	3.35	34
WY	2.5	4	3.42	19.75
VA	2.5	4	3.6	13
NC	2.5	4.67	3.81	35.5
KS	2.75	1.33	3.33	13
MN	2.75	2	3.67	45.5
IL	2.75	2	3.83	23.75
OH	2.75	2.67	3.46	37
FL	2.75	2.67	4.02	38.75
NH	2.75	3.33	3.4	23
MS	2.75	3.33	3.46	5.5
OR	2.75	4	3.48	34
SC	2.75	4.67	3.67	25.75
MA	2.75	5	2.75	37.5

NV	2.75	5	2.83	24
TN	2.75	5	2.9	20.75
IA	3	3	3.98	6
NJ	3	4.67	2.75	30
PA	3	4.67	3.71	37.25
TX	3	4.67	3.88	27.5
OK	3	4.67	3.9	30
NY	3	5	3	36.5
CT	3.25	3.67	2.46	22
CO	3.25	4	3.77	39.5
DC	3.5	-9	-9	46.5
CA	3.5	3	3.67	40.5
AR	3.5	3.33	3.21	23