Was Implementation Left Behind? A National Analysis of State and Federal School Accountability

Valentina Martinez-Pabon*

Most Recent Version Available Here

November 19, 2021

Abstract

No Child Left Behind (NCLB) and the state accountability systems implemented during the 1990s have included an increasingly intense series of interventions for low-performing schools. If the initial steps were insufficient, schools were supposed to be closed, taken over by other education organizations, or reconstituted. Using event study and difference-in-differences analyses, I find that state accountability and NCLB itself did not affect the frequency of closure and takeover of publicly funded schools in the country. Moreover, I find that NCLB generated an anticipatory but transitory increase in closures and led to a rise in closures among schools with small enrollments. Overall, my findings suggest that the infrequent and weak implementation of the most extreme sanctions on low-performing schools is partly behind the limited effects of accountability policies on student performance.

Keywords: education policy; school accountability; No Child Left Behind; school closures; school sanctions *(JEL: I21; I28; H520).*

*Tulane University. E-mail: vmartinezpabon@tulane.edu. I am very grateful to Douglas N. Harris for his time and continuous guidance throughout the different stages of this project. I also thank Wei Long, Nora Lustig, and the Tulane Economic Seminar participants for their valuable comments and suggestions. All errors remain my sole responsibility.

1 Introduction

Test-based school accountability policies spread across the states in the early 1990s to improve academic performance, especially of disadvantaged students. At the end of the decade, more than half of states had implemented accountability policies, and by 2001 the No Child Left Behind (NCLB) Act extended the policies across all states. NCLB, as well as some state policies, were designed to sanction low-performing schools with increasingly severe interventions, of which the most extreme were closures, takeovers, and reconstitutions.

Empirical evidence from state accountability before NCLB was promising, suggesting positive effects of accountability policies on student outcomes (Carnoy and Loeb, 2002; Hanushek and Raymond, 2005). However, evidence from the accountability policies implemented under NCLB suggests more mixed effects. The effects on test scores were small and often insignificant (Dee and Jacob, 2009; Dee and Dizon-Ross, 2019; Bonilla and Dee, 2020), while the effects on graduation rates were large (Harris et al., 2020). Evidence suggests that NCLB also changed decisions within schools. Per-pupil spending, teacher compensation, share of elementary school teachers with advanced degrees, time allocated by teachers to test preparation, and overall instructional time all increased (Dee, Jacob, and Schwartz, 2012; Rouse et al. 2013; Reback, Rockoff, and Schwartz, 2014).

No policy is likely to work, however, if it is not implemented as intended. Under NCLB, and some state accountability policies, closures and takeovers were part of the intended policy. Part of the reasoning was that the threat will pressure low-performance schools to make the improvements needed to avoid the sanctions and that closures will allow current and future cohorts of students to attend higher-quality schools (Bross, Harris and Liu, 2016; Sunderman, Coghlan, and Mintrop, 2017; Bifulco and Schwegman, 2020).

Whether the failure to implement this provision depends in part on the potential effects of closures and takeovers on students, this evidence is also mixed. Although NCLB-related sanctions in North Carolina seem to improve student performance, with the strongest positive effects associated with the most extreme sanctions (Ahn and Vig-

dor, 2014), a meta-analysis of 67 studies suggests null to positive effects of turnaround policies implemented after NCLB across the nation on student performance and no differential effects between the major types of interventions (including turnaround, labeling, chartering, and closure) (Schueler et al., 2020). Other research suggests that this may be because students do not end up in better schools after the closure (Engberg et al., 2012; Brummet, 2014; Kemple, 2015; Bross, Harris and Liu, 2016; CREDO, 2017; Han et al., 2017; Steinberg and MacDonald, 2018). The effects may also vary by the level of disruption for students (Bross, Harris and Liu, 2016; Gordon et al., 2018) and students' pre-intervention performance level (Bifulco and Schwegman, 2020).

Thus, even though school accountability was designed to identify, support, and, if needed, intervene in low-performing schools to generate improvements in the education system, the evidence suggests that the implementation of the interventions determines the policy's capacity to affect student achievement positively. That is, policies work under certain conditions, and they do not work if they are weakly implemented/enforced, or their implementation never occurs. In this case, the question is, were accountability policies implemented as intended?

To address this question, I explore the effect of state and federal (NCLB) test-based accountability policies on the most extreme intervention faced by low-performing public schools, closure. Specifically, I estimate the effect of state accountability policies implemented during the 1990s on the number of school closures using an event study approach that accounts for the time-varying implementation of state accountability policies. I also estimate the effect of NCLB itself on school closures using a difference-in-differences (DiD) approach in which, following Dee and Jacob (2010), I exploit the implementation of state accountability policies, similar to NCLB, before 2001. As a robustness check, I also explore the effect of test-based accountability policies on a measure that captures both school closures and takeovers. I define closure as the case in which the building is no longer operating as a school and where students and personnel have been forced to leave; while takeover occurs when school governance is transferred to a different organization

(e.g., chartering, state takeover).[1]

Overall, I find that state and federal performance-based accountability policies have an insignificant effect on public schools' closure and takeover rates. Moreover, in the case of NCLB, I find that, although the policy appears not to affect the closure of low-performing schools, it generated an anticipatory but transitory increase in closures and led to a rise in closures among small schools. Districts may have closed schools in advance because they expected the schools to fail on the NCLB metrics. Also, districts may have closed smaller schools because they had less political power or because the closure affected fewer students and required less reshuffling to other schools. Whatever the reason, these findings provide evidence that the heterogeneous (and weak) implementation of the most extreme sanctions (closures and takeovers among them) could be behind the limited effects of accountability policies to affect student performance positively.

I build on the prior literature in several ways. This study is among the first to address the broader question of the effect of two different accountability policies on nationwide school closures using causal inference. To my knowledge, only Bonilla and Dee (2017) have provided evidence of the effect of accountability policy on closures (and reconstitutions). However, their findings are limited to a specific state and explore the waivers program, which incorporated some flexibility to NCLB. Also, although Hess and Finn (2007) study the implementation of NCLB-related sanctions at the national level, they only do it descriptively for the 2006 school year.

The paper proceeds as follows. The next section describes the expansion of accountability policies in the country. Section 3 summarizes the data, and Section 4 discusses the methodology. Sections 5, presents the estimated results of the impact of test-based accountability on school closures. The final section summarizes my main conclusions and caveats.

[1]Reconstitution is another extreme sanction that occurs when the school is significantly restructured (e.g., changes personnel). I do not explore the effect of accountability policies on reconstitution due to data limitations.

2 The Expansion of Accountability Policies in the Education System

2.1 The Beginnings of the Test-Based Accountability Policies

Accountability policies based on threats and rewards to boost the quality of the education system spread across states in the early 1990s and then extended to all states under NCLB. Figure 1 presents the timeline of the implementation of state accountability policies across the nation based on work by Carnoy and Loeb (2002) and Dee and Jacob (2011).[2] As Figure 1 shows, from 1992 to 2001, 30 states implemented test-based accountability policies.[3] These included many of the components that subsequently became part of NCLB: standardized testing of students, school ratings, and performance-based sanctions and rewards.

[Figure 1]

Sanctions implemented under the statewide accountability, however, were heterogeneous. A survey conducted by Education Week in 2002 showed that only 23 of the 30 accountability states had penalties for failing schools: 19 states included reconstitution as a sanction, 11 included closures of failing schools, and 4 included conversions of traditional public schools into charters (Education Week, 2002).[4] Although details of the specific states are not available, we know from Carnoy and Loeb (2002) that only 9 of the 30 states that implemented these policies between 1992 and 2001 had accountability with strong repercussions in response to low performance (see states underlined in Figure 1). Whether these strong responses occurred remains a question that my subsequent analysis aims to address.

[2]The identification of states with test-based accountability before NCLB carried out by Dee and Jacob (2011) builds on Hanushek and Raymond (2005).

[3]See Figure A.1 in the Appendix for a geographical representation of the implementation of accountability policies across the nation.

[4]Unfortunately, the survey's underlying data are unavailable online, and Education Week could not locate them within their archive.

2.2 No Child Left Behind Act of 2001

Building on the policy expansion in the states, NCLB was signed into law in 2001 and was implemented during the 2003 school year.[5] NCLB focused on raising academic standards and ensuring that all students, especially those from low-income families, racial minorities, and students with disabilities, reached minimum academic proficiency. The federal policy required that each state implement annual standardized tests to students in grades 3 to 8 and that each school reach 100 percent proficiency level by 2014 and to make "Adequate Yearly Progress" (AYP) to that goal no later than the school year 2006.

The improvement process was rooted in a performance-based classification system with increasingly severe sanctions depending on the number of years schools failed to improve. Even though all public schools were rated for AYP, the sanctions under the improvement process only applied to schools receiving Title I funds.[6] The process included three stages of sanctions: school improvement (2-3 years of non-improvement), corrective action (4 years of non-improvement), and restructuring (5 or more years of non-improvement). Specifically, if a school did not make AYP for two consecutive years, it was identified for year one of school improvement and its Local Education Agency (LEA) was required to develop an improvement plan. If a school did not make AYP for three consecutive years, it was identified for year two of school improvement and its LEA was required to implement supplemental services. If a school did not improve for four consecutive years, it was identified for corrective action and its LEA was required to continue the supplemental services and implement at least one of the following actions: implement a new curriculum, extend the length of instruction, replace relevant staff, restructure the internal organization of the school, or appoint an outside advisor. If a school did not improve for five years, its LEA was required to continue the supplemental services and create a plan to restructure the school. Finally, if a school did not improve for six or more years, it was identified for restructuring and its LEA was required to

[5] For brevity, henceforth, I refer to school years 1990-1991 as "1991", 1991-1992 as "1992", and so on.

[6] Title I funds assist schools enrolling at least 40 percent of children from low-income families. In 2001, 90.8 percent of schools either reported receiving Title I funds or enrolled 40 percent or more children receiving Free or Reduced-Price Lunches.

implement one of the following interventions: closure, takeover, or reconstitution.[7,8]

Schools in any stage of the improvement process received technical assistance from the state and were required to give parents the option to transfer their child to another school. Furthermore, schools in the improvement process making AYP for two consecutive years guaranteed their exit from the process regardless of their stage in the improvement process.

As described above, NCLB aimed to establish a common accountability policy across the nation; nevertheless, it allowed states to define their own standards (e.g., how to classify schools making AYP) and allowed LEAs to determine the best sanction alternative to improve performance based on each school's characteristics (U.S. Department of Education, 2002; U.S. Department of Education, 2006). The approach resulted in considerable variation in how states defined schools not making progress. For example, states could minimize the impact of NCLB, if they wished, by lowering their standards (Loveless, 2003; Duncombe, Lukemeyer, and Yinger, 2008). Reback, Rockoff, and Schwartz (2014) show that the fact that each state defined their own standards generated numerous cases in which a school near the margin for satisfying its state's AYP requirements would have failed or made AYP if it were in another state. Also, Medler (2007) shows inconsistencies between state and federal definitions of struggling schools in Colorado in 2005: 134 of the 365 schools rated "unsatisfactory" or "low" in the School Accountability Reports (SAR) made AYP while 10 percent of schools that score "high" on the SAR failed to make AYP due to differences on how tests addressed diversity (that is, if the system disaggregates performance by students characteristics).

Moreover, the approach resulted in many states favoring the less radical sanctions for

[7]Remember that I define closure as the case in which the building is no longer operating as a school and where students and personnel have been forced to leave. Usually, school districts are the ones that oversee the decision to close traditional public schools. In the case of charter schools, the decision is on the authorizers. Takeover is when school governance is transferred to a different organization (e.g., chartering, state takeover) and reconstitution is when the school is significantly restructured or transformed (e.g., changes personnel).

[8]Until 2006, NCLB also allowed at the restructuring stage to implement any other form of a major restructuring that involve fundamental reforms (Kowal and Hassel, 2007). Manna (2007) shows that in California in 2005-06 this option was chosen by 76 percent of schools reaching the restructuring stage and involved less severe interventions like "hiring coaches" and "appointing a leadership team."

reasons related to lack of political will and state capacity (Manna, 2007).[9] For example, Casserly (2007) shows that districts with schools in corrective action or restructuring appeared to be most inclined to provide additional technical assistance and professional development than to implement somewhat tougher sanctions.

Despite the high variation and relative looseness of state standards, by the end of the 2010s decade, almost half of the schools in the country failed to meet AYP (Duncombe, Lukemeyer and Yinger, 2008). Thus, in 2011, when it was evident that states could not meet by 2014 the proficient level of academic achievement proposed in each state's plan, the federal government began offering waivers from NCLB requirements.[10] Under the NCLB waivers reform, LEAs still determined the sanctions to low-performing schools, but the categories of schools failing to improve changed from the ones described above to priority (schools in the bottom 5 percent of achievement) and focus (schools in the bottom 10 percent of achievement), and there was no a predefined set of sanctions anymore. As a result, evidence suggests that the waivers decreased the number of schools facing sanctions (Hyslop, 2013) and that being identified as a focus school reduced the probability of closure or reconstitution (Bonilla and Dee, 2017).

2.3 Every Student Succeeds Act of 2015

NCLB generated controversies regarding the loss of state independence and the implementation of measures that made the education system overly focused on standardized test scores. As a result, NCLB was eventually replaced by the Every Student Succeeds Act (ESSA) in 2015. ESSA further reduced the influence of the federal government in the states' education system and gave them more flexibility regarding the accountability system. It also opened the door to new assessments, considered additional academic and quality factors when evaluating schools, and changed the consequences of low performance.

[9]The latter, mainly associated with sanctions conflicting with state laws. For example, some states had no charter laws, some states do not authorize state takeover, in some states, collective bargaining makes it virtually impossible to fire staff, and in states like New Jersey, for a charter school to be established under law, the staff has to vote for it (McGuinn, 2007; Petrilli, 2007).

[10]By 2014, 34 states and the District of Columbia received waivers (U.S. Department of Education, 2013).

3 Data

3.1 Data Description

I use the panel of traditional public and charter schools of the National Longitudinal School Database (NLSD). The NLSD is a panel of all schools in the country hosted by the National Center for Research on Education Access and Choice (REACH). It combines a set of public and private data sources, including the Common Core of Data (CCD), the Private Schools Universe Survey (PSS), data on student achievement from the Stanford Education Data Archive (SEDA) (Reardon et al., 2021), data on parents/community ratings from the non-profit organization Great Schools, and data on NCLB-related school progress from the National Adequate Yearly Progress and Identification (NAYPI) database created by the American Institutes for Research (AIR). The NLSD captures a broad range of school characteristics, including enrollment size, closure/takeover status, and school quality, among others.[11]

I supplement the NLSD with information regarding each state's accountability policy prior to NCLB, from Dee and Jacob (2011) and Carnoy and Loeb (2002), and two additional datasets. The first dataset is the School-Level Achievement Database (SLAD) compiled and described in Harris (2007), from which I calculate the pre-NCLB average test score for schools in 32 states using each state's standardized tests to elementary and secondary school students from 1997 to 2000. The pre-NCLB school average test score is pooled across grades and subjects for the year closest to 2000 and is standardized and comparable across states. Thus, each school has a unique measure of school performance for the pre-NCLB period. See details of the tests type, grades, subjects, and years used to create the pre-NCLB school average test score in each state in Table A.1 in the Appendix.

The second dataset is a panel that I created with the schools identified in the NCLB's improvement process (that is, the schools identified as not making AYP) using each state's Consolidated School Performance Reports (CSPR) from 2006 to 2015 (U.S. Department of Education, 2017).[12] This information supplements the data from the NAYPI, which

[11]For a detailed presentation of the NLSD see Harris and Martinez Pabon (forthcoming).

[12]This excludes 2011 since the data was not available online for this year and I have not been able to

identifies schools in the NCLB's improvement process for 2004 and 2005.

3.2 Descriptive Statistics

Figure 2 presents the trends of the three most extreme sanctions schools face, from 1995 to 2019.[13] On average, during the last two decades, the closure rate of publicly funded schools ranged around 1.6 percent. Before the announcement of NCLB, the closure trend was relatively steady, with levels under 1.5 percent. After the announcement of NCLB, the closure rate showed a steady increase, reaching its peak in 2005 and staying on those levels until 2011, when it started its decline to the pre-NCLB levels. The trend of the number of closed schools follows a similar path as the closure rate, with an average of 1,550 schools closed each year (See Figure A.2 in the Appendix).[14] Takeovers and reconstitutions were less frequent than closures during this period. The takeover rate ranged steadily about 0.05 percent. The reconstitution rate was about 0.4 percent and showed a decreasing slope during the years with available information.

[Figure 2]

Given that, under NLCB's improvement process, extreme measures were to be implemented after five years in which schools did not improve (restructuring stage), one would have expected an increase in their rates starting five years after the law went into effect, in 2008. Figure 2 shows, however, a rise in the closure rate starting in 2001 that does not align with the stages of NCLB's improvement process but its announcement. The increase in the closure rate in 2001 suggests that districts carried out anticipatory closures of schools they would expect to close either way. Closing these schools early might have helped avoid even more costly processes later. Moreover, the observed steady decrease in the closure rate starting in 2011, also observed in the reconstitution rate, coincides with the announcement of the waivers from NCLB requirements, which introduced in-

access it.

[13] I do not present the reconstitution rate for the entire period because it is available in the data only starting in 2011.

[14] Figure A.2 also shows that the number of new schools per year was around 2,400 between 1995 and 2007. Starting in 2008, the number of new schools dropped and reached an average of 1,600 during the last decade.

creased flexibility that could have allowed districts to close even fewer schools. Takeovers, however, seem not to respond neither to NCLB's announcement or implementation.

Figure 3 illustrates the trend of the closure rate by implementation of state accountability, akin to Dee and Jacob (2011). The two plots show that before the implementation of NCLB in 2003, the school closure rate was lower in the group of states that implemented state accountability policies than in the group of states for which NCLB introduced test-based accountability for the first time. The closure rates between the two groups converged once NCLB was implemented, suggesting that either the states that implemented accountability during the 1990s did have but did not enforce closure sanctions until NCLB required them to increase accountability and/or that the states with a larger share of low-quality schools were less likely to adopt accountability policies.

[Figure 3]

Regarding NCLB, Figure 4 presents the trend of the NCLB-related sanctions from 2004 to 2015.[15] In Panel (a), sanctioned schools incorporate all schools under NCLB's improvement process. Panel (a) shows that the sanction rate was stable and around 11 percent between 2004 and 2008 and that in 2010 the sanction rate hit a peak, after which it started declining. Panel (b), on the other hand, presents the sanction rate by stage of the improvement process. As mentioned earlier, the first stage of the process was called school improvement. The school improvement was meant for schools undergoing 2 to 3 years of non-improvement and involved the less severe interventions. The corrective action stage was meant for schools undergoing 4 years of non-improvement. And the restructuring stage was meant for schools undergoing 5 or more years of non-improvement and involved the most severe interventions. As expected, Panel (b) shows that most schools were identified for school improvement in the earlier years. In 2008, once corrective action and restructuring stages took effect, the share of schools facing restructuring increased significantly. Again, as in Figure 2, it seems as if the introduction of the waivers program in 2011 led to a steady decrease in the share of schools facing NCLB-related sanctions.

[Figure 4]

[15] Information regarding NCLB-related sanctions is only available starting from 2004.

In the analysis that follows, I focus on the years from 1995 to 2010 since after 2010, as shown in the previous figures, either NCLB was not entirely in force or ESSA had replaced it. Note that omitting the years after 2010 will not hurt the results since even delayed implementation of NCLB would have produced closures, takeovers, and reconstitutions before 2010. Table 1 presents the descriptive statistics from 1995 to 2010 for the universe of publicly funded schools, for the subsample of those schools in states that implemented state accountability, and for the subsample of those schools in states where NCLB accountability introduced test-based accountability for the first time. Table 1 shows that schools in states with test-based accountability before NCLB, representing 62 percent of total schools, are more similar in their student and school characteristics to the universe of publicly funded schools in the nation than schools in other states.

[Table 1]

4 Methodology

4.1 The Effect of Pre-NCLB State Accountability Policies on School Closures

To estimate the effect of state (pre-NCLB) accountability policies on school closures, I use an event study approach that accounts for the time-varying implementation of accountability policies during the 1990s (see Figure 1). The estimation focuses on the years from 1995 to 2000 to avoid capturing any NCLB-related effect. The regression model for the event study is:

$$Closed_{ist} = \beta_0 + \sum_{t \neq 0} \theta_t(Accountability_{st}) + X_{ist} + \mu_s + \gamma_t + \epsilon_{ist} \quad (1)$$

where $Closed_{ist}$ is an indicator variable of closure for school i in state s and school year t. I also explore the effect of state accountability policies on a broader definition of the dependent variable that captures both closures and takeovers as a robustness check. $Accountability_{st}$ represents the treatment status defined based on the year in which

each state implemented its accountability policy. That is, $Accountability_{st}$ indicates the time relative to the implementation of the state accountability policy, of which the last year before the implementation of the policy is omitted (t=0). X_{ist} are the school and student characteristics (grades offered, enrollment size, share of minority students, share of students who received free or reduced-price lunches, and urban location) that control for observable differences between schools in control and treatment groups that are correlated with the closure rate. μ_s and γ_t are state and time fixed effects that control for state-invariant and time-invariant characteristics. In this event study, θ_t captures the evolution of the treatment effect over time. Standard errors are clustered at the state level given that the policy was implemented at this level.

I use a Two-Way Fixed Effects (TWFE) regression model to estimate the effect of state accountability on school closures. To address the concerns about the possibility that the TWFE estimator could be biased, given that there is variation in the treatment timing (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021), I use a balanced panel of states that implemented accountability policies between 1997 and 2000. Moreover, I also estimate the event study using the estimator proposed by Callaway and Sant'Anna (2021) and Sun and Abraham (2021) to test the robustness of my results.

The Callaway and Sant'Anna and the Sun and Abraham estimators address the issues with the TWFE estimator when the treatment occurs at different points in time. The Callaway and Sant'Anna estimator is based on a group-time analysis in which each group-time Average Treatment on the Treated (ATT) is unbiased and consistent and is estimated for units treated at the same time. For example, schools in Alabama, Rhode Island, and West Virginia are in the group "1997" because those states implemented test-based accountability in 1997 (see Figure 1). The Sun and Abraham estimator is a special case of the Callaway and Sant'Anna estimator, and is based on a weighted average of the group-time ATT.

In all estimation methods, I exclude from the analysis the always treated units and use the not yet treated (and never treated for Sun and Abraham) ones as the control

group.

4.2 The Effect of NCLB on School Closures

To estimate the effect of NCLB on school closures, I follow the identification strategy proposed by Dee and Jacob (2011) and rely on the assumption that NCLB requirements were irrelevant to states with well-established test-based accountability policies before its implementation. Specifically, Dee and Jacob (2011) proposed identifying NCLB's effects by comparing schools in states that never implemented accountability before NCLB to schools in states that implemented test-based accountability policies with elements similar to those promoted under NCLB before 2001. Thus, in this case, the former is the treated group, and the latter is the control group, turning the treated group in the estimation of the state accountability era into the control group in the estimation of the NCLB era.

To estimate the effect of NCLB on school closures, I restrict the sample to the years from 1998 to 2010 to avoid capturing any effect from the years where NCLB was not entirely in force due to the waivers reform or ESSA. In addition, the sample only includes schools that had ever received Title I funds because, as mentioned above, the sanctions implemented under the improvement process only applied to those schools.[16]

I estimate the overall effect of NCLB on school closures using a Difference-in-Differences (DiD) approach. The regression model for the DiD is:

$$Closed_{ist} = \beta_0 + \beta_1(NCLB_t \times NoAccountability_s) + X_{ist} + \mu_s + \gamma_t + \epsilon_{ist} \qquad (2)$$

where $NoAccountability_s$ is a dummy variable that identifies whether a given state did not implemented accountability policies before NCLB. $NCLB_t$ is a dummy variable that equals 1 for observations after the implementation of NCLB (starting on 2003). X_{ist} is a vector of school and student characteristics (grades offered, enrollment size,

[16]Since the Title I status was not available for the entire period in the NLSD, I identify Title I schools as those ever reporting receiving Title I funds (variable included in the survey starting in 1999 but with a low response rate until 2002) and/or those ever having at least 40 percent of students receiving Free or Reduced-Price Lunches.

enrollment growth rate between 1999 and 2001 in neighboring schools,[17] share of minority students, share of students who received free or reduced-price lunches, and urban location) that control for observable differences between the schools. In the DiD approach, β_1 is interpreted as the estimated average policy effect. Standard errors are clustered at the state level.

In addition, since effects might be dynamic, I use an event study model to estimate the dynamic effects of the policy and to verify the parallel-trends assumption. The regression model for the event study is:

$$Closed_{ist} = \beta_0 + \sum_{t \neq 2002} \theta_t (NCLB_t \times NoAccountability_s) + X_{ist} + \mu_s + \gamma_t + \epsilon_{ist} \qquad (3)$$

Again, in the event study approach, θ_t captures the evolution of the treatment effect over time and standard errors are clustered at the state level.

I identify the states in each comparison group based on the classification of states with test-based accountability before NCLB developed by Dee and Jacob (2011) and Carnoy and Loeb (2002). As shown in Figure 1, Dee and Jacob (2011) identify the thirty states that before NCLB implemented test-based accountability, while Carnoy and Loeb (2002) provide more fine-grained distinctions and identify nine states where the repercussions for low performance are strong. This yields to two possible definitions of the control group (30 states with any accountability from Dee and Jacob (2011) and nine strong accountability states from Carnoy and Loeb (2002)) and one definition of the treatment group (schools in the 21 states that never implemented test-based accountability before NCLB).

Furthermore, under the assumption that states implementing accountability policies close to the announcement of NCLB did not reach to put into operation the threat and reward measures that come along with accountability policies, I consider that the four states implementing accountability policies between 2000 and 2001 are part of the

[17] This variable controls for demographic changes (under-enrollment) in the school district that could also influence the closure decision.

treatment group. That is, Alaska, Georgia, Oregon, and Tennessee.

5 Results

5.1 The Effect of Pre-NCLB State Accountability Policies on School Closures

Figure 5 presents the event study estimates of the effect of state accountability policies on school closures based on the TWFE, Callaway and Sant'Anna, and Sun and Abraham estimators. For the TWFE estimates, the pre-treatment effects are small and statistically insignificant, showing that the parallel trends test is satisfied.[18] In the case of the Callaway and Sant'Anna and Sun and Abraham estimates, the parallel trends test is generally satisfied. An exception is the Callaway and Sant'Anna case in the "Dee and Jacob" classification, where there is evidence of an effect in the pre-period. The latter threatens the validity of the results and suggests that the estimates for this classification should be taken with more precaution.

Overall, for the two sets of results and the different estimates analyzed, the post-treatment effects suggest that states did not change the frequency of closure sanctions on public schools after state accountability policies were in effect. In addition, the results of the effect of state accountability on school closures/takeovers are similar in magnitude and significance to the main findings, as expected given that school takeovers are very unlikely, confirming the estimated null effect of state accountability on the implementation of other extreme interventions (see Figure A.3 in the Appendix).

[Figure 5]

I also explore the possible heterogeneity in the effect of state accountability on school closures by the characteristics of schools and students, such as the school level (elemen-

[18] In addition, the decomposition of the estimated average effect, following Goodman-Bacon (2021), supports the validity of the TWFE estimate. In particular, Figure A.4 in the Appendix shows that its different components are close to each other in magnitude (almost all negative and close to zero). Likewise, the decomposition suggests that none of the components is flipping the sign/magnitude of the estimate, and the weights of each component of the estimate show that later treatment/earlier control variation contributes to the average effect more than the earlier treatment/later control group one.

tary, middle, and high), the share of minority students, and the school size. As in the main estimates, the results in Figure 6 show the effect do not vary by school or student characteristics.

[Figure 6]

While inconclusive in spots due to the parallel trend tests, this analysis generally suggests that state accountability did not alter the frequency of school closures and takeovers. In addition, the negative, but null, estimates rule out the possibility that state accountability policies lead to more school closures or takeovers. Nevertheless, an important limitation of this analysis is that I cannot analyze the effect of state accountability on low-performing school closures because data from before the start of these accountability systems is not available. Also, information relative to school quality at the national level is scarce, and when it exists, it is available for more recent years. In the next section, I will use available information from school performance in 32 states circa 2000 to estimate the effect of NCLB on the closure of low-performing schools.

5.2 The Effect of NCLB on School Closures

The analysis above focuses on the state accountability systems implemented during the 1990s. Below, I explore the average and heterogeneous effects of NCLB on school closures. As a robustness check, I also analyze the impact of NCLB on school closures and takeovers. Moreover, I use additional information from NCLB-related sanctions to describe how the probability of closure changes with each stage of NCLB's improvement process.

5.2.1 Average Treatment Effects

The identification strategy to estimate the effect of NCLB on school closures, following Dee and Jacob (2011), lies under the assumption that NCLB requirements were irrelevant in states that had test-based accountability policies before NCLB. Also, that observable and unobservable characteristics did not both affect the pattern in which states adopted (or not) accountability policies during the 1990s and the subsequent pattern of school

closures. Prior work by Hanushek and Raymond (2005) suggests that states adopted accountability policies during the 1990s in a non-systematic way. The authors find no significant differences of early (1996 or before) and late (1997 to 2002) implementers of school accountability policies in their economic, political, and educational characteristics. However, Table 1 shows that, on average, schools in states that implemented accountability during the 1990s are significantly different in all school and student characteristics to schools in other states.

Furthermore, the results presented in the previous sections suggest some limitations of the identification strategy. First, the estimated effects of state accountability on school closures and takeovers, shown in Figures 5 and A.3, suggest that states did not enforce extreme sanctions after implementing state accountability. On the contrary, as shown in Figures 2 and 3, the closure rate was relatively steady during the 1990s and only started increasing once NCLB was announced. Thus, the evidence suggests that NCLB might have been relevant not only for states for which NCLB introduced test-based accountability for the first time, but also for at least some of those that already had implemented state accountability policies. This limitation will attenuate the estimated effects of NCLB using the Dee and Jacob (2011) identification strategy and will make the following analysis more descriptive than causal, but still informative about NCLB.

Figure 7 presents the event study results of the NCLB effects on school closures. The pre-treatment estimates show that the parallel trends test is satisfied, and the estimates for the post-treatment period suggest that NCLB does not affect school closures.[19] It is also important to highlight that the positive but insignificant effect estimated for 2008 coincides with the first year the closure sanction (and the other extreme sanctions of the restructuring stage) went into effect under NCLB. Recall that closure was not recommended until schools were failing for five consecutive years and the school was under the restructuring stage. Again, the estimates of the effect of the policy when the dependent variable captures both school closures and takeovers are similar in magnitude

[19]Figure A.5 presents the event study with a longer pre-treatment period as robustness check. I also estimate the effect of the policy on the state's cumulative number of school closures as a robustness check. I find that NCLB also had a null effect on the aggregate number of school closures at the state level (see Table A.2 in the Appendix).

and significance to the main findings (Figure A.6 and Table A.2 in the Appendix).

[Figure 7]

Table 2 presents the DiD estimates of the effect of NCLB on school closures.[20] Odd columns show the estimation with state and year fixed effects only, and even columns add school characteristics to the model in the previous columns. Moreover, columns 1-4 show the main results assuming, as presented in Equation 2, that the post-treatment period started with the implementation of NCLB (2003). The coefficients on school closures are small and insignificant, as expected, given the earlier event study results.

To test the hypothesis that closures could have responded to NCLB's announcement rather than its implementation, columns 5 to 8 present the results considering that the post-treatment period started in 2001. I define the policy announcement in 2001 because NCLB was debated mainly this year and because starting in this year, the closure rate presents a steady increase (see Figure 2). Here, the estimates are marginally significant. These results suggest an anticipatory effect in which districts responded to the announcement of the policy by carrying out closures of schools they knew were not going to succeed under NCLB.[21] Thus, closing these schools earlier might have helped them avoid a more costly process later given that NCLB implied a high burden on states which were required to report in detailed their progress in meeting the NCLB standards and requirements. That is, states were required to submit each year to the federal government a performance report with the proficiency by subgroup of students (including race, ethnicity, disability status, English proficiency, among others) participating in the assessment, graduation and dropout rates, the list of schools and districts identified for the improvement process, and more (U.S. Department of Education, 2005; U.S. Department of Education, 2017).

[Table 2]

[20] To rule out the possibility that NLCB conduced to compositional shifts in the remaining schools, I estimate the effect of the policy on enrollment size. The results in Table A.3 suggest no effect of the policy on this variable.

[21] Figure A.7 in the Appendix shows that the distribution of the school performance for schools closed between 2001 and 2003 and that for schools closed between 2001 and 2010 overlap almost perfectly, suggesting that schools closed in response to the announcement of NCLB seem not to be significantly different from the rest of the schools closed during this period. Nevertheless, as expected, the distribution of the school performance for schools that eventually closed is significantly lower than the distribution of the school performance for schools that were always open.

5.2.2 Effect Heterogeneity

Until this point, estimates account for the effect on all types of closures because identifying the reason behind each closure in the NLSD is not possible. To estimate the effect of NCLB on the closure of low-performing schools, as mentioned before, I supplement the NLSD with pre-NCLB average performance levels (circa 2000) of schools in 32 states from SLAD (see Table A.1 for more details). I define low-performing schools as those in the bottom 5, 10, 15, and 15 percent of the pre-NCLB performance distribution. In Table 3, I present the estimates of the effect of NCLB on low-performing school closures. Once again, results suggest that the policy did not change the frequency of school closures, not even for those with low performance.

[Table 3]

I also explore heterogeneous responses to school accountability. Table 4 presents the results in which I estimate the effect of NCLB on school closures by characteristics of schools and students. Even when one of the purposes of the policy was to focus on student subgroups (Loeb and Figlio, 2011), I find that there is no effect of NCLB on school closures by school level or students' demographic characteristics. However, Table 4 suggests that the policy increased small-size school closures. This is an interesting result since there is no direct connection between school size and NCLB sanctions. One possible explanation is that it was easier for districts to close smaller schools because their constituents had less political power and/or this required reassigning fewer students, compared with high enrollment schools.

[Table 4]

In addition, I test the sensitivity of the results to the time-length during which the comparison states were exposed to test-based accountability policies before NCLB. Here, I explore in more detail my assumption that states implementing accountability policies close to the announcement of NCLB (2000 and 2001 assumed for my main results) were unable to put into operation their state accountability before NCLB. Thus, for these

states and those that never implemented state accountability NCLB introduced test-based accountability for the first time. The results in Table 5 show that as the length of time during which the control group was exposed to state accountability increases, it also increases the significance of the estimated effect of NCLB on school closures. This result suggests that NCLB might have treated not only states that never implemented accountability but also those implementing accountability policies during the late 1990s.

[Table 5]

In sum, the main analysis of the effect of NCLB on school closures aligns with the findings of the effect of state (pre-NCLB) accountability on school closures and suggest that test-based accountability did not alter the frequency of school closures and takeovers. In addition, the NCLB analysis contributes to the literature by identifying where the effects of the policy were observed. First, I find that the announcement of NCLB led to an increase on school closures, however, this was a transitory effect that was not sustained after implementing the policy. Second, I find that the policy increased the closure of small schools. Overall, these results are in line with Hess and Finn (2007), who show, descriptively, that by 2005-2006 states were not much applying NCLB-related sanctions or, if used, were deploying them in their mildest forms. These results are also helpful to understand Bonilla and Dee's (2017) finding that schools identified as low performing under the waivers reform in Kentucky had a lower probability of being closed or reconstituted than schools not identified as low performing.

5.2.3 Additional Evidence

As a final step, I use data on schools undergoing NLCB's improvement process from 2004 to 2010 to complement my findings and explore if the probability of closure changes with each stage in the improvement process. Note, however, that the following analysis is purely descriptive since I do not have information regarding the improvement process before 2004, and the data does not include the specific interventions implemented on schools within each stage (except for closure/takeover).

As described in section II, based on the consecutive number of years in which schools did not make progress, they were classified into the following stages of the NCLB improvement process: school improvement (2-3 years of non-improvement), corrective action (4 years of non-improvement), and restructuring (5 or more years of non-improvement).[22] Table 6 shows that the vast majority of schools closing between 2004 and 2010 was never identified in the improvement process. That is not surprising, given the array of conditions that lead to closure even without accountability policies (Harris and Martinez Pabon, forthcoming). Also, the proportion of schools closed in the least severe sanction stage (school improvement) is higher than the proportion of schools closed at the most severe stage (restructuring), which suggests that the NCLB sanctions may not have been implemented as intended. The share of schools closed in states with and without test-based accountability is similar for all stages of the improvement process, suggesting, once again, no differences in the probability of closure for schools in states with state accountability than for schools in states where NCLB introduced test-based accountability for the first time.

[Table 6]

6 Conclusions

The introduction of test-based accountability in the education system sought to improve the education quality through sanctions and rewards based on school performance. Under accountability policies, sanctions were designed with increasing interventions, of which the most extreme were closure, takeover, and reconstitution. Part of the reasoning was that the threat would pressure low-performance schools to make the improvements needed to avoid the sanctions and closures would allow current and future cohorts of students to attend higher-quality schools, which would boost their performance improvement (Bross, Harris and Liu, 2016; Sunderman, Coghlan, and Mintrop, 2017; Bifulco and Schwegman,

[22]Schools identified under the improvement process during 2004 and 2010 had an average of 3.3 years of non-improvement, with more than 30 percent of schools being part of the improvement process for five or more years.

2020). However, empirical evidence suggests that state and federal accountability policies have a null to positive effect on student performance (Carnoy and Loeb, 2002; Hanushek and Raymond, 2005; Dee and Jacob, 2009; Dee and Dizon-Ross, 2019; Bonilla and Dee, 2020).

Overall, my findings show that both state and federal accountability policies do not affect the frequency of school closures and takeovers. My results on the effect of NCLB on school closures suggest no impact on low-performing school closures. However, I find evidence that the announcement of the policy generated an anticipatory (but transitory) increase in school closures which might suggest that districts responded to the announcement of the policy by carrying out closures of schools they knew were not going to succeed under NCLB. In addition, my results also suggest that the policy increased the closure of small schools that potentially have less political power and whose closure affects fewer students.

While my analysis comes with some caveats regarding data limitations and the causality of some of the estimated effects, still, my findings provide suggestive evidence that accountability policies may not have been implemented as intended. Therefore, the limited capacity of test-based accountability to affect performance may be related to the weak (almost null) implementation of the most extreme sanctions. Here, the anticipated concern of West and Peterson (2003) that the impact of NCLB (and any accountability policy at this point) would depend upon the way states administer the policy and the strategies they would devise to promote improvement takes hold once more.

My results are meaningful because they rule out the idea that test-based accountability policies had led to more closures and contribute to the literature by identifying a significant element behind the seeming failure of the test-based accountability to improve student outcomes. Although it is unclear that closures implemented under test-based accountability policies would produce significant gains in learning, undoubtedly, interventions on low-performing schools are required. For this reason, districts must work to produce better results for students and those who benefit from an open school in their community.

References

Ahn, T., and J. Vigdor. (2014). The impact of No Child Left Behind's accountability sanctions on school performance: Regression discontinuity evidence from North Carolina (No. w20511). National Bureau of Economic Research.

Bifulco, R., and Schwegman, D. J. (2020). Who Benefits from Accountability-Driven School Closure? Evidence from New York City. Journal of Policy Analysis and Management, 39(1), 96-130.

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

Bross, W., Harrris, D., and L. Liu. (2016). Extreme measures: When and how school closures and charter takeovers benefit students. New Orleans: Education Research Alliance for New Orleans, Tulane University.

Brummet, Q. (2014). The effect of school closings on student achievement. Journal of Public Economics 119 (2014) 108–124.

Callaway, B. and P. HC. Sant'Anna. (2021). Difference-in-differences with multiple time periods. Journal of Econometrics 225 (2): 200-230.

Carnoy, M. and S. Loeb. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. Educational Evaluation and Policy Analysis Winter 2002, Vol. 24, No. 4, pp. 305–331.

Casserly, M. (2007). America's Great City Schools: Moving in the Right Direction in No remedy left behind: Lessons from a half-decade of NCLB, Hess, Frederick M., and Chester E. Finn (editors). AEI Press.

CREDO. (2017). Lights Off: Practice and Impact of Closing Low-Performing Schools. Center for Research on Education at Stanford University. Volume I and II.

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

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

Dee, T. S., Jacob, B.A., and Schwartz, N. L. (2012). The effects of NCLB on school resources and practices. Educational Evaluation and Policy Analysis, Vol. 35, No. 2, 252-279.

Duncombe, W. D., A. Lukemeyer and J. Yinger (2008). Dollars Without Sense: The Mismatch Between the No Child Left Behind Act Accountability System and Title I Funding, in Improving on No Child Left Behind: Getting Education Reform Back on Tract, R. D. Kahlenber (editor). The Century Fundation Press.

Education Week. (2002). Where States Stand. Volume 22, Number 15, Page 19.

Engberg, J., Gill, B., Zamarro, G. and R. Zimmer. (2012). Closing schools in a shrinking district: Do student outcomes depend on which schools are closed? Journal of Urban Economics 71(2):189–203.

Goodman-Bacon, Andrew. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics 225 (2): 254-277.

Gordon, M., de la Toree, M., Cowhy, J., Moore, P. T., Sartain, L. and D. Knight. (2018). School closings in Chicago: Staff and student experiences and academic outcomes. Chicago, IL: University of Chicago Consortium on School Research.

Hanushek, E. A., and Raymond, M. E. (2005). Does school accountability lead to improved student performance? Journal of Policy Analysis and Management, 24, 297–327.

Harris, D. (2007). High-Flying Schools, Student Disadvantage, and the Logic of NCLB. American Journal of Education 113.

Harris, D., Liu, L., Barrett, N. and R. Li. (2020). Is the Rise in High School Graduation Rates Real? High-Stakes School Accountability and Strategic Behavior. Brown Center on Education Policy, Brookings.

Harris, D., Liu, L., Gerry, A. and P. Arce-Trigatti. (2019). From revolution to evolution: Market dynamics in school value-added and marketed program offerings under the post-Katrina school reforms in New Orleans. New Orleans: Education Research Alliance for New Orleans, Tulane University.

Harris, D. and V. Martinez Pabon. (Forthcoming). Extreme Measures: A National Analysis of Closure and Takeover on Traditional Public, Charter, and Private Schools.

Hess, F. and C. E. Finn. (2007). Conclusion: Can This Law Be Fixed? A Hard Look at the NCLB Remedies in No remedy left behind: Lessons from a half-decade of NCLB, Hess, Frederick M., and Chester E. Finn (editors). AEI Press.

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

Kemple, J. (2015). High School Closures in New York City: Impacts on Students' Academic Outcomes, Attendance, and Mobility. New York: The Research Alliance for New York City Schools.

Kowal, J. and B. C. Hassel. (2007). Remedies in Action: Four "Restructured" Schools in No remedy left behind: Lessons from a half-decade of NCLB, Hess, Frederick M., and Chester E. Finn (editors). AEI Press.

Loeb, S., and Figlio, D. (2011). School accountability. In E. A. Hanushek, S. Machin, and L. Woessmann (Eds.), Handbook of the Economics of Education, Vol. 3 (pp. 383-423). San Diego, CA: North Holland.

Loveless, T. (2003). Charter School Achievement and Accountability, in No Child Left Behind? The Politics and Practice of School Accountability, P. E. Peterson and M. R. West (editors). Brookings Institution Press.

Manna, P. (2007). NCLB in the States: Fragmented Governance, Uneven Implementation in No remedy left behind: Lessons from a half-decade of NCLB, Hess, Frederick M., and Chester E. Finn (editors). AEI Press.

McGuinn, P. (2007). New Jersey: Equity Meets Accountability in No remedy left behind: Lessons from a half-decade of NCLB, Hess, Frederick M., and Chester E. Finn (editors). AEI Press.

Medler, A. (2007). Colorado: The Misapplication of Federal Power in No remedy left behind: Lessons from a half-decade of NCLB, Hess, Frederick M., and Chester E. Finn (editors). AEI Press.

Petrilli, M. J. (2007). The Problem with "Implementation" is the Problem in No remedy left behind: Lessons from a half-decade of NCLB, Hess, Frederick M., and Chester E. Finn (editors). AEI Press.

Reardon, S. F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., Jang, H., and Chavez, B. (2021). Stanford Education Data Archive (Version 4.0).

Reback, R., J. Rockoff, and H. L. Schwartz. (2014). Under Pressure: Job Security, Resource Allocation, and Productivity in Schools under No Child Left Behind. American Economic Journal: Economic Policy, 6 (3): 207-41.

Rouse, C. E., J. Hannaway, D. Goldhaber, and D. Figlio. (2013). Feeling the Florida Heat? How Low-Performing Schools Respond to Voucher and Accountability Pressure. American Economic Journal: Economic Policy, 5 (2): 251-81.

Schueler, B. E., C. A. Asher, K. E. Larned, S. Mehrotra, and C. Pollard. (2020). Improving Low-Performing Schools: A Meta-Analysis of Impact Evaluation Studies. EdWorkingPaper: 20-274.

Steinberg, M. and J. MacDonald. (2018). The Effects of Closing Urban Schools on Students' Academic and Behavioral Outcomes: Evidence from Philadelphia, Economics of Education Review (2018).

Sun, L., and S. Abraham. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics 225 (2): 175-199.

Sunderman, G. L., E. Coghlan, and R. Mintrop. "School Closure as a Strategy to Remedy Low Performance." National Education Policy Center (2017).

U.S. Department of Education. (2002). No Child Left Behind: A Desktop Reference. US Department of Education, Office of the Under Secretary, Office of Elementary and Secondary Education.

U.S. Department of Education. (2005). No Child Left Behind: A road map for state implementation. US Department of Education, Office of the Under Secretary, Office of Elementary and Secondary Education.

U.S. Department of Education. (2006). LEA and School Improvement Non-Regulatory Guidance. US Department of Education, Office of the Under Secretary, Office of Elementary and Secondary Education.

U.S. Department of Education. (2013). No Child Left Behind: Early Lessons from State Flexibility Waivers. US Department of Education, Office of the Under Secretary, Office of Elementary and Secondary Education.

U.S. Department of Education. (2017). Consolidated State Performance Reports. US Department of Education, Office of the Under Secretary, Office of Elementary and Secondary Education.

West, M. R. and P. E. Peterson. (2003). The Politics and Practice of Accountability, in No Child Left Behind? The Politics and Practice of School Accountability, P. E. Peterson and M. R. West (editors). Brookings Institution Press.

Tables and Figures

Table 1: Summary Statistics of Schools Characteristics

Variables	All Schools		Schools in states with accountability before NCLB		Schools in states where NCLB introduced accountability		Diff.
	Mean (1)	S.D. (2)	Mean (3)	S.D. (4)	Mean (5)	S.D (6)	(3) - (5)
Panel A. Outcome Variables							
Closure	0.0162	0.1263	0.0154	0.1230	0.0176	0.1316	-0.0023***
Closure or takeover	0.0166	0.1278	0.0158	0.1248	0.0179	0.1327	-0.0021***
Panel B. Control Variables							
Charter	0.0271	0.1623	0.0263	0.1599	0.0285	0.1663	-0.0022***
Number of students	513	439	547	462	455	391	91.8466***
% Black students	0.1591	0.2530	0.1771	0.2560	0.1279	0.2445	0.0492***
% Hispanic students	0.1465	0.2341	0.1847	0.2611	0.0801	0.1570	0.1046***
% FRPL	0.4138	0.2776	0.4379	0.2817	0.3694	0.2642	0.0684***
Title I elegible	0.9065	0.2912	0.9154	0.2783	0.8916	0.3109	0.0238***
Urban	0.7031	0.4569	0.7327	0.4426	0.6538	0.4758	0.0789***
Magnet	0.0270	0.1620	0.0372	0.1892	0.0097	0.0980	0.0275***
Offers grade 1	0.5385	0.4985	0.5444	0.4980	0.5286	0.4992	0.0158***
Oferrs grade 2	0.5394	0.4984	0.5453	0.4979	0.5295	0.4991	0.0158***
Oferrs grade 3	0.5391	0.4985	0.5451	0.4980	0.5291	0.4992	0.0159***
Oferrs grade 4	0.5347	0.4988	0.5411	0.4983	0.5240	0.4994	0.0171***
Oferrs grade 5	0.5205	0.4996	0.5278	0.4992	0.5083	0.4999	0.0195***
Oferrs grade 6	0.3667	0.4819	0.3593	0.4798	0.3791	0.4852	-0.0198***
Oferrs grade 7	0.2791	0.4486	0.2751	0.4466	0.2859	0.4518	-0.0108***
Oferrs grade 8	0.2812	0.4496	0.2784	0.4482	0.2859	0.4518	-0.0074***
Oferrs grade 9	0.2278	0.4194	0.2259	0.4182	0.2308	0.4214	-0.0049***
Oferrs grade 10	0.2236	0.4166	0.2221	0.4156	0.2261	0.4183	-0.0040***
Offers grade 11	0.2208	0.4148	0.2184	0.4132	0.2248	0.4175	-0.0064***
Oferrs grade 12	0.2174	0.4124	0.2134	0.4097	0.2240	0.4169	-0.0106***
Unique schools	121,202		75,437		45,771		121,202
Unique schools	1,509,827		943,955		565,872		1,509,827

Notes: The unit of observation is the school year from 1995 to 2010. FRPL: Free/Reduced Price Lunch.
Source: Author's calculations using the NLSD.

Table 2: DiD Estimates of the Effect of NCLB on School Closure Rate

Variables	"Dee and Jacob" Classification		"Carnoy and Loeb" Classification		"Dee and Jacob" Classification		"Carnoy and Loeb" Classification	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Post x Treatment	0.0009	0.0004	0.0030	0.0022	0.0014	0.0010	0.0034*	0.0025
	(0.0020)	(0.0024)	(0.0021)	(0.0026)	(0.0018)	(0.0024)	(0.0018)	(0.0024)
Mean Rate	0.011		0.011		0.011		0.011	
Unique schools	101,170	97,075	73,241	70,155	101,170	97,075	73,241	70,155
Observations	1,132,957	904,160	812,576	648,436	1,132,957	904,160	812,576	648,436
School controls	No	Yes	No	Yes	No	Yes	No	Yes
Year, State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Post-treatment year	2003		2003		2001		2001	

Notes: The unit of observation is the school year from 1998 to 2010. The sample only includes Title I schools. Treated states are those where NCLB introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the control states implemented test-based accountability before NCLB. In the "Carnoy and Loeb" classification, the control states implemented test-based accountability with strong repercussion to low performance before NCLB. Standard errors in parenthesis are clustered by state. * Statistically significant at the 0.10 level, ** at the 0.05 level, *** at the 0.01 level. Source: Author's calculations using the NLSD.

Table 3: DiD Estimates of the Effect of NCLB on Low-Performing School Closure Rate

Variables	"Dee and Jacob" Classification (1)	"Carnoy and Loeb" Classification (2)	"Dee and Jacob" Classification (3)	"Carnoy and Loeb" Classification (4)	"Dee and Jacob" Classification (5)	"Carnoy and Loeb" Classification (6)	"Dee and Jacob" Classification (7)	"Carnoy and Loeb" Classification (8)
Post x Treatment	-0.0025	-0.0030	-0.0026	-0.0030	-0.0026	-0.0030	-0.0027	-0.0030
	(0.0017)	(0.0019)	(0.0017)	(0.0019)	(0.0017)	(0.0019)	(0.0017)	(0.0019)
Observations	553,522	388,196	553,522	388,196	553,522	388,196	553,522	388,196
School controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year, State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Low performers	Bottom 5%		Bottom 10%		Bottom 15%		Bottom 20%	

Notes: The unit of observation is the school year from 1998 to 2010. The sample only includes Title I schools. Treated states are those where NCLB introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the control states implemented test-based accountability before NCLB. In the "Carnoy and Loeb" classification, the control states implemented test-based accountability with strong repercussion to low performance before NCLB. Standard errors in parenthesis are clustered by state. * Statistically significant at the 0.10 level, ** at the 0.05 level, *** at the 0.01 level. Source: Author's calculations using the NLSD and the SLAD.

Table 4: DiD Estimates of the Heterogeneous Effect of NCLB on School Closure Rate by School and Student Characteristics

Variables	Elementary Schools		Middle Schools		High Schools	
	"Dee and Jacob" Classification	"Carnoy and Loeb" Classification	"Dee and Jacob" Classification	"Carnoy and Loeb" Classification	"Dee and Jacob" Classification	"Carnoy and Loeb" Classification
	(1)	(2)	(3)	(4)	(5)	(6)
Post x Treatment	0.0017	0.0042	-0.0009	0.0023	-0.0018	-0.0009
	(0.0031)	(0.0030)	(0.0022)	(0.0026)	(0.0017)	(0.0022)
Observations	641,389	456,879	214,705	154,149	194,192	141,043
School controls	Yes	Yes	Yes	Yes	Yes	Yes
Year, State Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Variables	High Share of Minority Students		Low Share of Minority Students	
	"Dee and Jacob" Classification	"Carnoy and Loeb" Classification	"Dee and Jacob" Classification	"Carnoy and Loeb" Classification
	(7)	(8)	(9)	(10)
Post x Treatment	0.0000	0.0043	0.0015	0.0027
	(0.0062)	(0.0058)	(0.0021)	(0.0023)
Observations	196,606	154,837	707,554	493,599
School controls	Yes	Yes	Yes	Yes
Year, State Fixed Effects	Yes	Yes	Yes	Yes

Variables	Small-Size Schools		Large-Size Schools	
	"Dee and Jacob" Classification	"Carnoy and Loeb" Classification	"Dee and Jacob" Classification	"Carnoy and Loeb" Classification
	(11)	(12)	(13)	(14)
Post x Treatment	0.0146	0.0341**	-0.0003	0.0006
	(0.0132)	(0.0142)	(0.0022)	(0.0023)
Observations	42,413	29,844	861,747	618,592
School controls	Yes	Yes	Yes	Yes
Year, State Fixed Effects	Yes	Yes	Yes	Yes

Notes: The unit of observation is the school year from 1998 to 2010. The sample only includes Title I schools. Treated states are those where NCLB introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the control states implemented test-based accountability before NCLB. In the "Carnoy and Loeb" classification, the control states implemented test-based accountability with strong repercussion to low performance before NCLB. High share of minority students: Share of black/Hispanic students is larger than 80 percent. Small-/Large-size Schools: less/more than 200 total students. Standard errors in parenthesis are clustered by state. * Statistically significant at the 0.10 level, ** at the 0.05 level, *** at the 0.01 level. Source: Author's calculations using the NLSD.

Table 5: Robustness Check: DiD Estimates of the Effect of NCLB on School Closure Rate by Length of Exposure of State Accountability

Variables	"Dee and Jacob" Classification			
	(1)	(2)	(3)	(4)
Post x Treatment	0.0004	0.0011	0.0038**	0.0034*
	(0.0024)	(0.0021)	(0.0017)	(0.0017)
Observations	904,160	904,160	904,160	904,160
School controls	Yes	Yes	Yes	Yes
Year, State Fixed Effects	Yes	Yes	Yes	Yes
Controls are states that implemented accountability before:	2000	1999	1998	1997

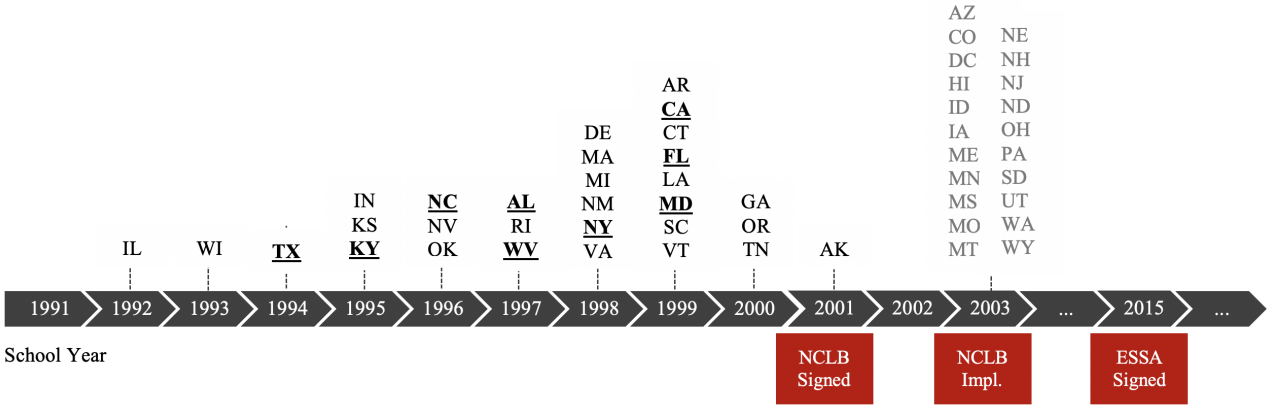
Notes: The unit of observation is the school year from 1998 to 2010. The sample only includes Title I schools. Treated states are those where NCLB introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the control states implemented test-based accountability before NCLB. Standard errors in parenthesis are clustered by state. * Statistically significant at the 0.10 level, ** at the 0.05 level, *** at the 0.01 level. Source: Author's calculations using the NLSD.

Table 6: Sanction Stage by Accountability and Closure Status

Sanction Stage	Schools in states with accountability before NCLB			Schools in states where NCLB introduced accountability		
	Eventually closed		% Closed	Eventually closed		% Closed
	No	Yes		No	Yes	
None	41,646	4,657	10.1	17,432	2,449	12.3
School Improvement	8,350	718	7.9	5,237	336	6.0
Corrective Action	3,783	232	5.8	1,477	98	6.2
Restructuring	7,630	501	6.2	3,004	144	4.6
Total	61,409	6,108	9.0	27,150	3,027	10.0

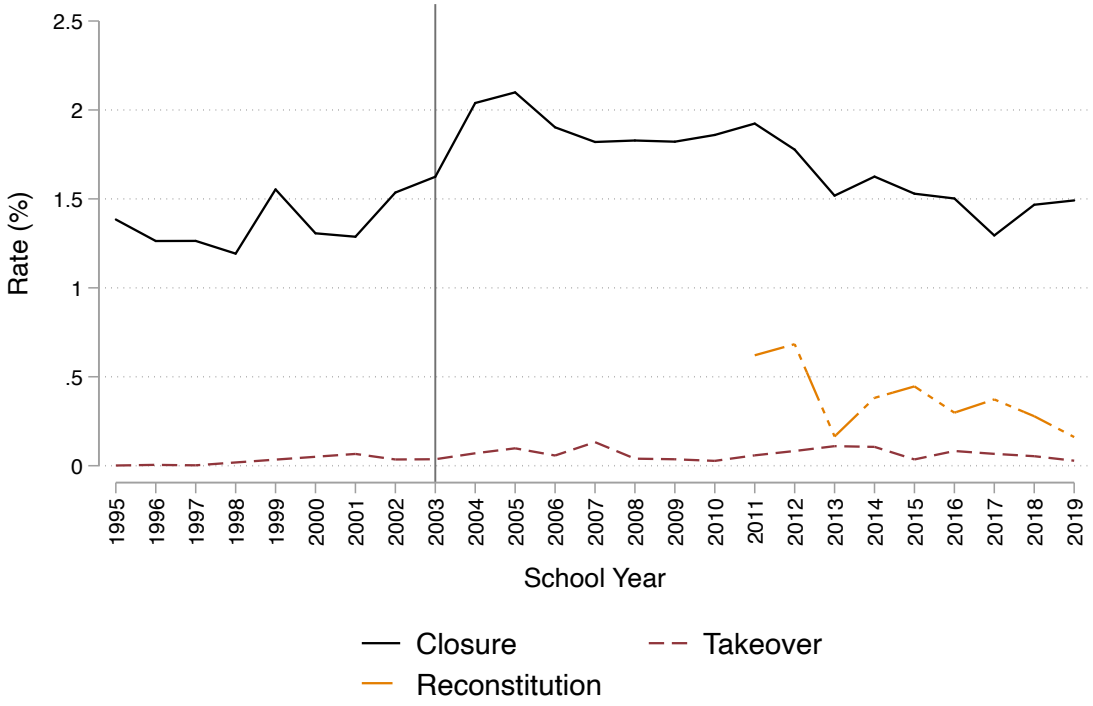
Notes: The unit of observation are schools that were ever open between 2004 and 2010. The sample only includes Title I schools. The sanction stage is the most severe NCLB-related sanction faced by the school between 2004 and 2010. Source: Author's calculations using the CSPR, the NAYPI, and the NLSD.

Figure 1: Timeline of the Implementation of Accountability Policies



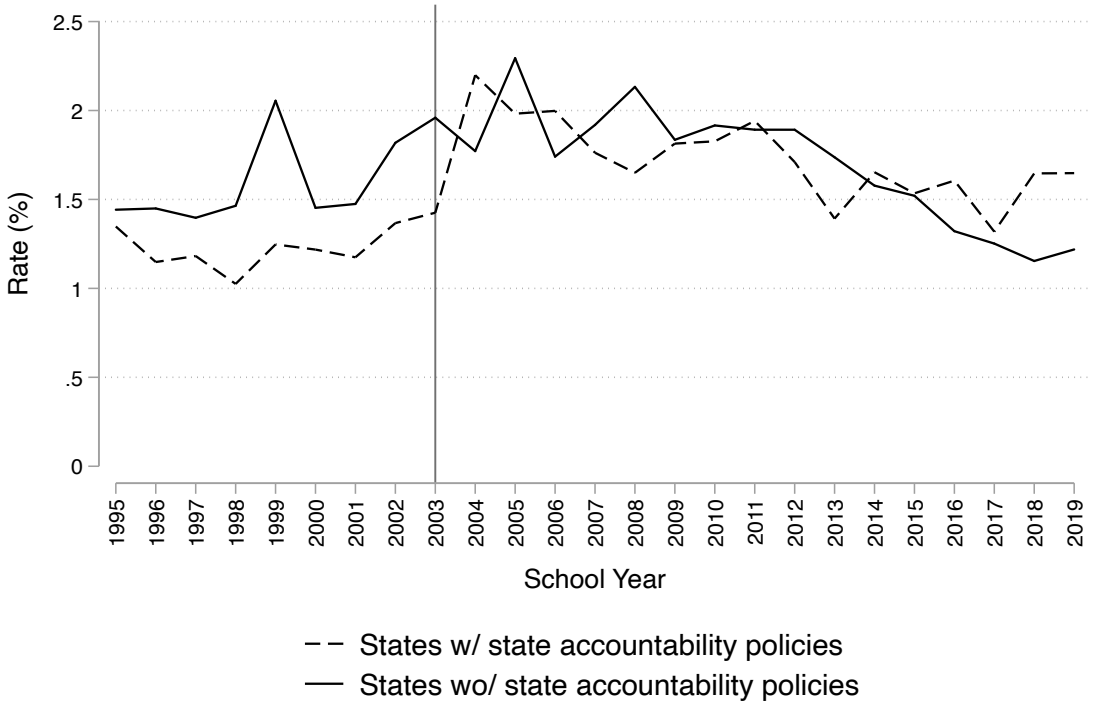
Notes: The 30 states in black font are the ones that implemented state accountability policies before NCLB (of those, the 9 underlined are the ones with stronger accountability policies). The 21 states in gray font are the ones where NCLB introduced test-based accountability for the first time. Source: Author's based on Carnoy and Loeb (2002) and Dee and Jacob (2011).

Figure 2: Closure, Takeover, and Reconstitution as Percentage of All Schools (%)



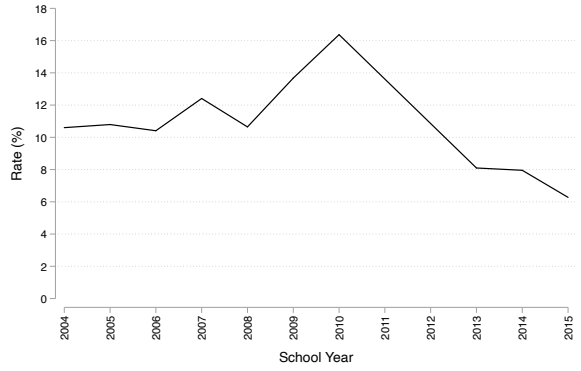
Notes: Closure (takeover/reconstitution) rate is the ratio between closed (taken over/reconstituted) schools and all schools. The vertical line identifies the year in which NCLB was implemented. Source: Author's calculations using the NLSD.

Figure 3: Closed Schools as a Share of All Schools (%), by Implementation of State Accountability

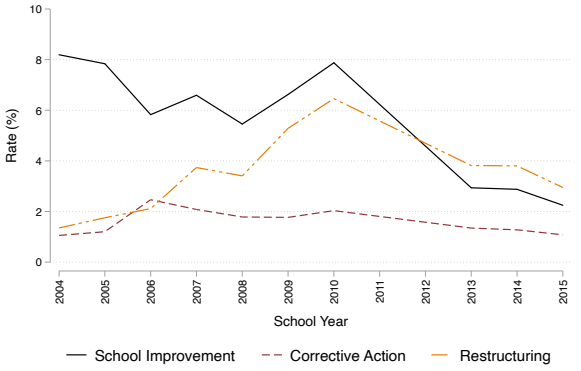


Notes: The vertical line identifies the year in which NCLB was implemented. Source: Author's calculations using the NLSD.

Figure 4: Schools Sanctioned Under NCLB's Improvement Process



(a) Sanctioned Schools as Percentage of All Schools (%)

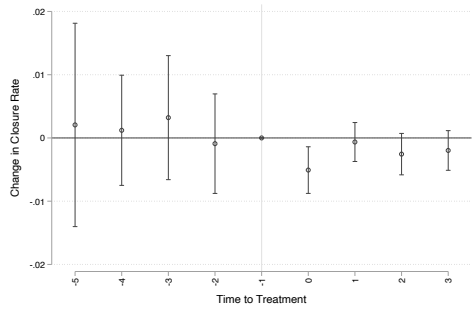


(b) Sanctioned Schools as Percentage of All Schools (%), by Sanction Level

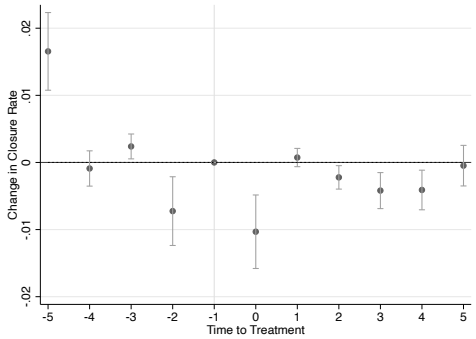
Notes: Sanctioned schools are those under any stage of NCLB's improvement process. The school improvement stage was meant for schools undergoing 2 to 3 years of non-improvement, while the corrective action stage was meant for schools undergoing 4 years of non-improvement, and the restructuring stage was meant for schools undergoing 5 or more years of non-improvement. Source: Author's calculations using the CSPR, the NAYPI, and the NLSD.

Figure 5: Event Study Estimates of the Effect of State Accountability Policies on School Closure Rate

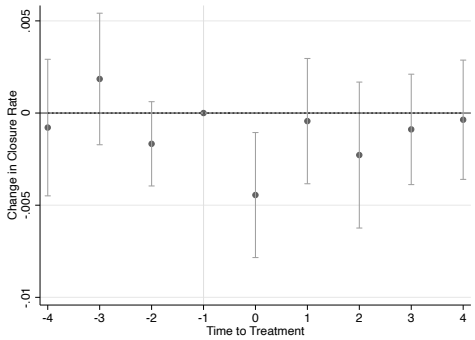
"Dee and Jacob" Classification



(a) TWFE

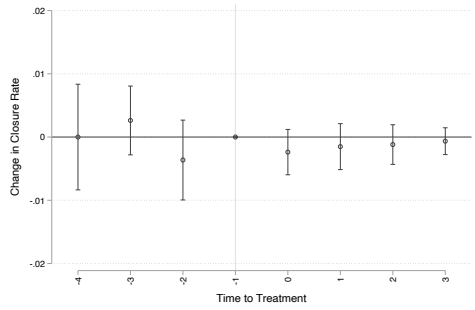


(b) Callaway and Sant'Anna

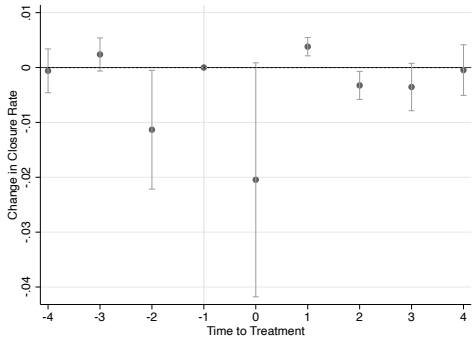


(c) Sun and Abraham

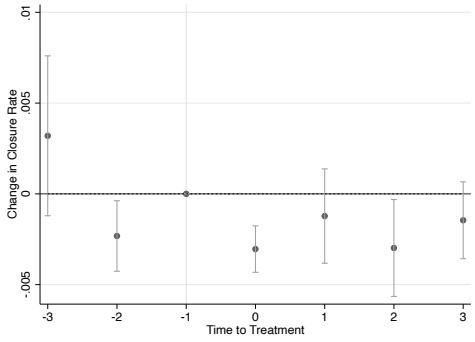
"Carnoy and Loeb" Classification



(d) TWFE



(e) Callaway and Sant'Anna

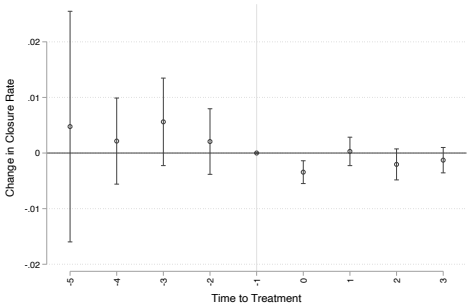


(f) Sun and Abraham

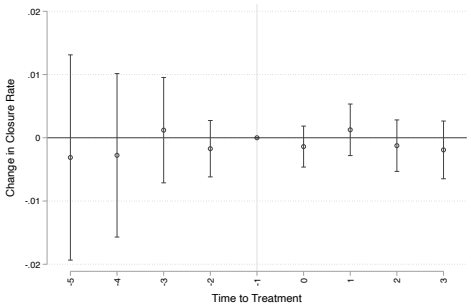
Notes: The unit of observation is the school year from 1995 to 2000. Treated states are those where state accountability introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the treated states implemented test-based accountability. In the "Carnoy and Loeb" classification, the treated states implemented test-based accountability with strong repercussion to low performance. Estimates expressed in relative time. 95% confidence intervals are reported based on standard errors clustered at the state level. Source: Author's calculations using the NLSD.

Figure 6: Event Study Estimates of the Effect of State Accountability Policies on School Closure Rate by School and Student Characteristics

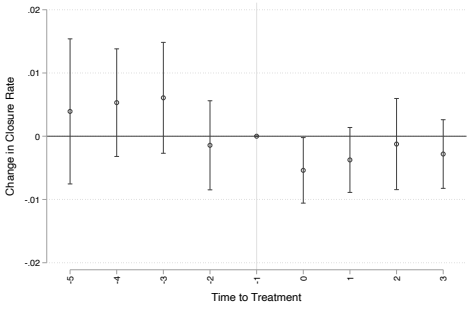
"Dee and Jacob" Classification



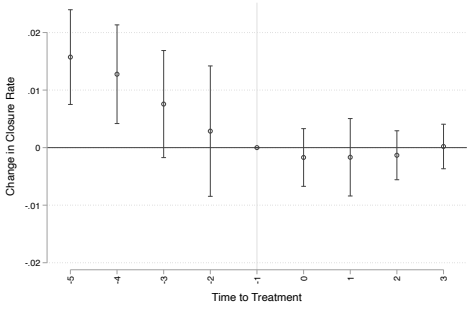
(a) Elementary Schools



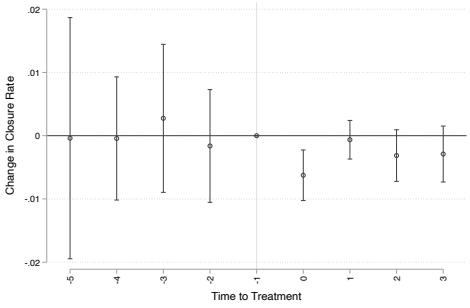
(b) Middle Schools



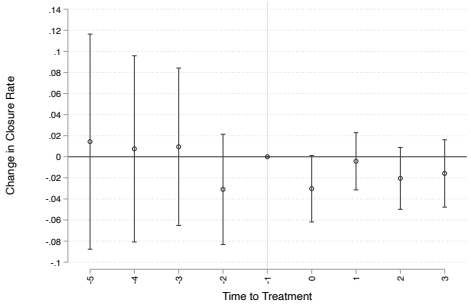
(c) High Schools



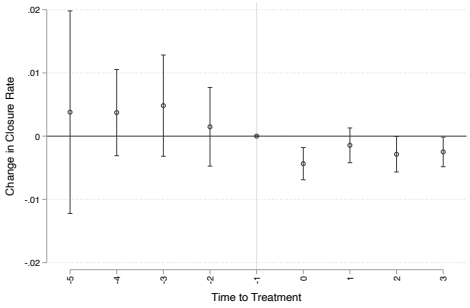
(d) High Share of Minority Students



(e) Low Share of Minority Students

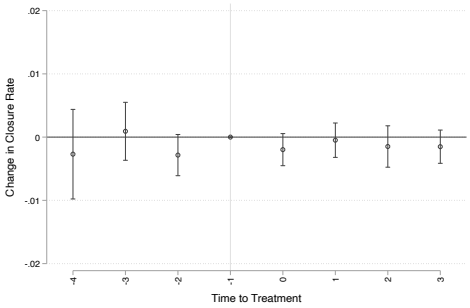


(f) Small-Size Schools

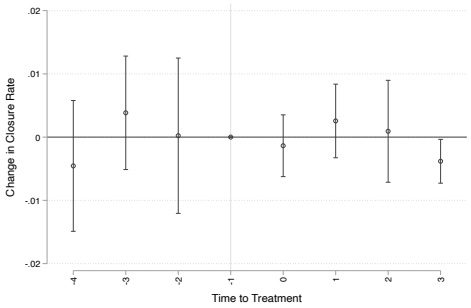


(g) Large-Size Schools

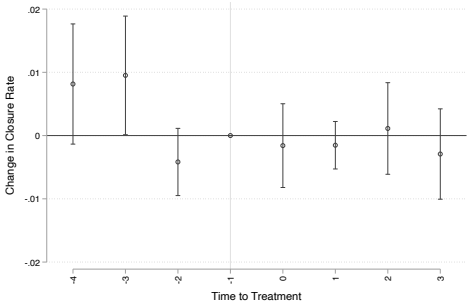
"Carnoy and Loeb" Classification



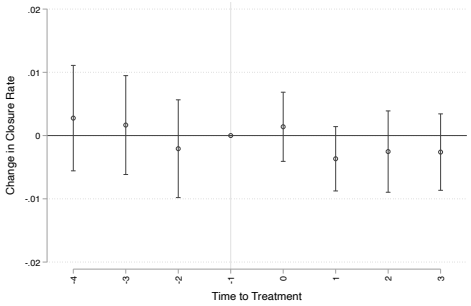
(h) Elementary Schools



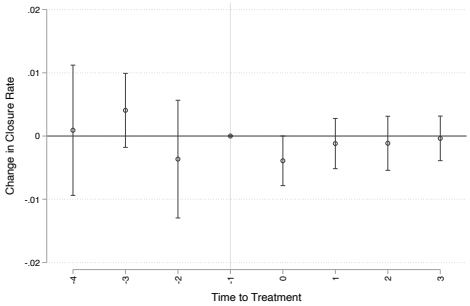
(i) Middle Schools



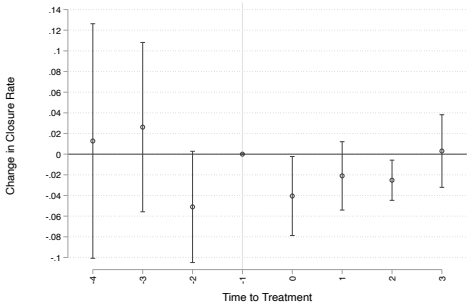
(j) High Schools



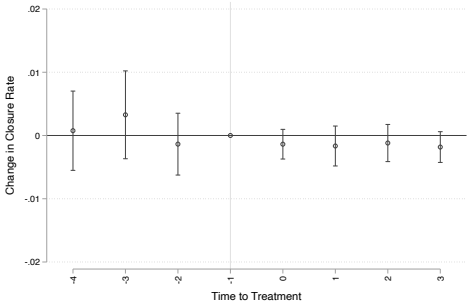
(k) High Share of Minority Students



(l) Low Share of Minority Students



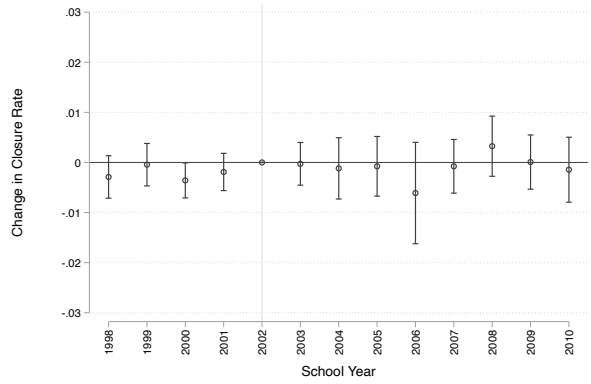
(m) Small-Size Schools



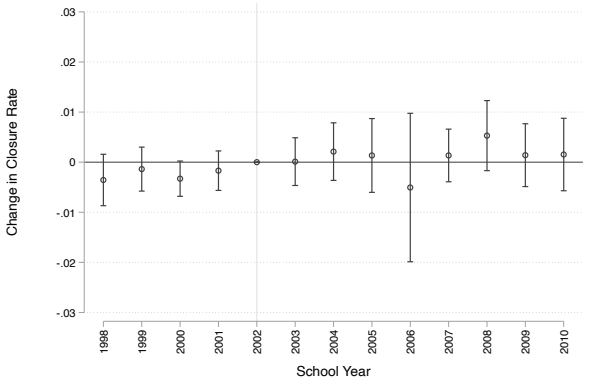
(n) Large-Size Schools

Notes: The unit of observation is the school year from 1995 to 2000. Treated states are those where state accountability introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the treated states implemented test-based accountability. In the "Carnoy and Loeb" classification, the treated states implemented test-based accountability with strong repercussion to low performance. Estimates expressed in relative time. TWFE estimates. 95% confidence intervals are reported based on standard errors clustered at the state level. High share of minority students: Share of black/Hispanic students is larger than 80 percent. Small-/Large-size Schools: less/more than 200 total students. Source: Author's calculations using the NLSD.

Figure 7: Event Study Estimates of the Effect of NCLB on School Closure Rate



(a) "Dee and Jacob" Classification



(b) "Carnoy and Loeb" Classification

Notes: The unit of observation is the school year from 1998 to 2010. The sample only includes Title I schools. Treated states are those where NCLB introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the control states implemented test-based accountability before NCLB. In the "Carnoy and Loeb" classification, the control states implemented test-based accountability with strong repercussion to low performance before NCLB. Estimates are relative to 2002, one year before the implementation of NCLB. 95% confidence intervals are reported based on standard errors clustered at the state level. Source: Author's calculations using the NLSD.

A Appendix

Table A.1: Pre-NCLB Test Score Information, by State

State	Number of schools with available information	Total number of schools	Test type	Grades	Subjects	Year
Alabama	1,182	1,323	PCT	5, 8, 11	Math, Reading	2000
Arizona	1,196	1,417	ASC	5, 8, 11	Math, Reading	2000
Arkansas	849	1,040	ASC	4, 8	Math, Reading	2000
California	7,285	8,358	ASC	5, 8, 11	Math, Reading	2000
Connecticut	772	802	RAW	4, 8	Math, Reading	2000
Delaware	101	151	ASC	4, 8, 10	Math, Writing	2000
Florida	2,559	2,776	ASC	5, 8, 10	Math, Reading	2000
Georgia	1,437	1,605	PCT	5, 8	Math, Reading	2000
Idaho	520	702	PCT	5, 8, 11	Math, Reading	2000
Illinois	3,574	3,847	ASC	3, 8, 10	Math, Reading	2000
Indiana	1,795	1,871	NCE	3, 8, 10	Math, Reading	2000
Kansas	1,374	1,392	RAW	R: 3, 7, 10	Math	1999
				M: 4, 7, 10	Reading	2000
Kentucky	476	1,230	ASC	6	Math, Reading	2000
Maine	571	596	ASC	4, 8, 11	Math, Reading	2000
Massachusetts	1,629	1,757	ASC	4, 8, 10	Math, ELA	2000
Minnesota	1,215	1,719	ASC	5, 8	Math, Reading	2000
Mississippi	575	859	PCT	5, 8	Math, Reading	2000
Montana	681	746	NCE	4, 8, 11	Math, Reading	1998
Nevada	380	382	PCT	4, 8, 10	Math, Reading	1999
New Hampshire	325	432	ASC	3, 10.	Math, Reading	2000
New Jersey	1,953	1,977	ASC	4, 8, 11	Math, Language	1999/2000
New Mexico	624	686	ASC	5, 8, 10	Math, Reading	1997
New York	3,178	3,248	ASC	4, 8	Math, Language	2000
North Carolina	1,513	1,868	ASC	5, 8	Math, Reading	1998
North Dakota	489	559	PCT	6, 8, 10	Math, Reading	2000
Pennsylvania	2,689	2,829	ASC	5, 8, 11	Math, Reading	2000
South Carolina	742	1,039	ASC	5, 8	Math, Reading	2000
Tennessee	1,071	1,311	ASC	5, 8	Math, Reading	1998
Utah	690	720	PCT	5, 8, 11	Math, Reading	2000
Virginia	1,415	1,839	ASC	5, 8	Math, Reading	1998
Wisconsin	1,946	2,015	PCT	4, 8, 10	Math, Reading	2000
Wyoming	299	305	ASC	4, 8, 11	Math, Reading	2000
Total	45,105	51,401				

Notes: ASC: Average Scale Score, NCE: Normal Curve Equivalents, PCT: Percentile, and RAW: Raw score. Source: Based on the SLAD and Harris (2007).

Table A.2: Robustness Check: DiD Estimates of the Effect of NCLB on School Closure Rate

Variables	"Dee and Jacob" Classification		"Carnoy and Loeb" Classification	
	(1)	(2)	(3)	(4)
Panel (a) Dependent Variable: Closure/Takeover				
Post x Treatment	0.0005	0.0001	0.0025	0.0019
	(0.0020)	(0.0023)	(0.0020)	(0.0025)
Observations	1,132,957	904,160	812,576	648,436
School controls	No	Yes	No	Yes
Year, State Fixed Effects	Yes	Yes	Yes	Yes
Panel (b) Dependent Variable: Cumulative Number of Closures				
Post x Treatment	-5.4393		-8.8360	
	(5.0799)		(8.4953)	
Observations	663		442	
Year, State Fixed Effects	Yes		Yes	

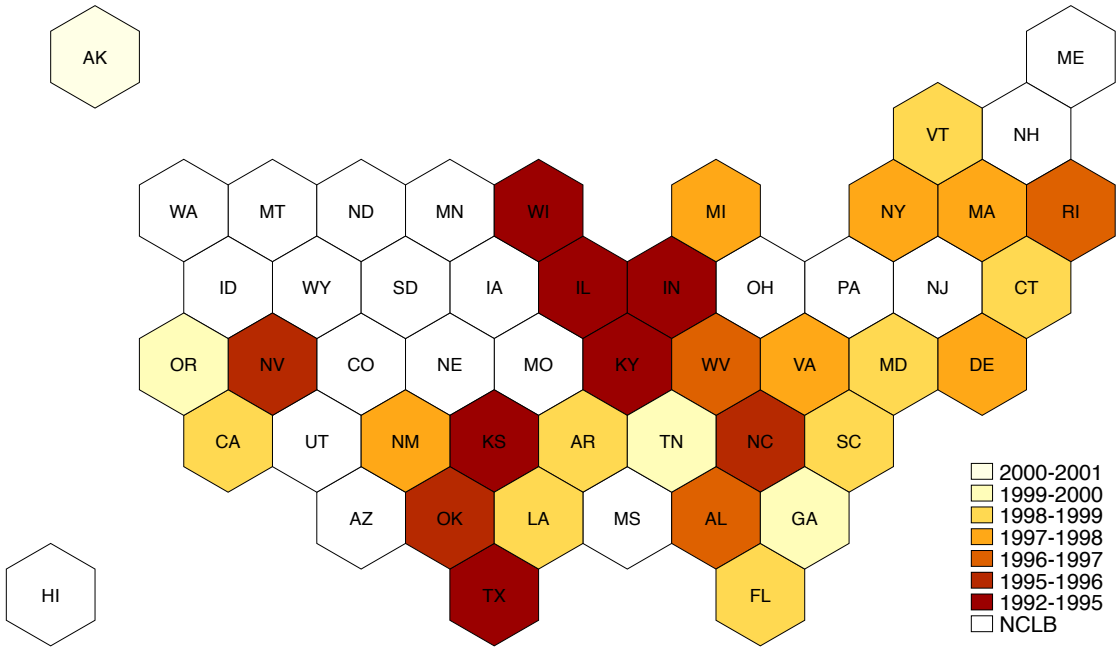
Notes: The unit of observation is the school year from 1998 to 2010. The sample only includes Title I schools. Treated states are those where NCLB introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the control states implemented test-based accountability before NCLB. In the "Carnoy and Loeb" classification, the control states implemented test-based accountability with strong repercussion to low performance before NCLB. Standard errors in parenthesis are clustered by state. * Statistically significant at the 0.10 level, ** at the 0.05 level, *** at the 0.01 level. Source: Author's calculations using the NLSD.

Table A.3: Robustness Check: DiD Estimates of the Effect of NCLB on School Size

Variables	"Dee and Jacob" Classification		"Carnoy and Loeb" Classification	
	(1)	(2)	(3)	(4)
Post x Treatment	1.3	3.5	12.4	15.7
	(8.2)	(11.8)	(11.8)	(17.2)
Observations	1,128,250	904,160	809,408	648,436
School controls	No	Yes	No	Yes
Year, State Fixed Effects	Yes	Yes	Yes	Yes

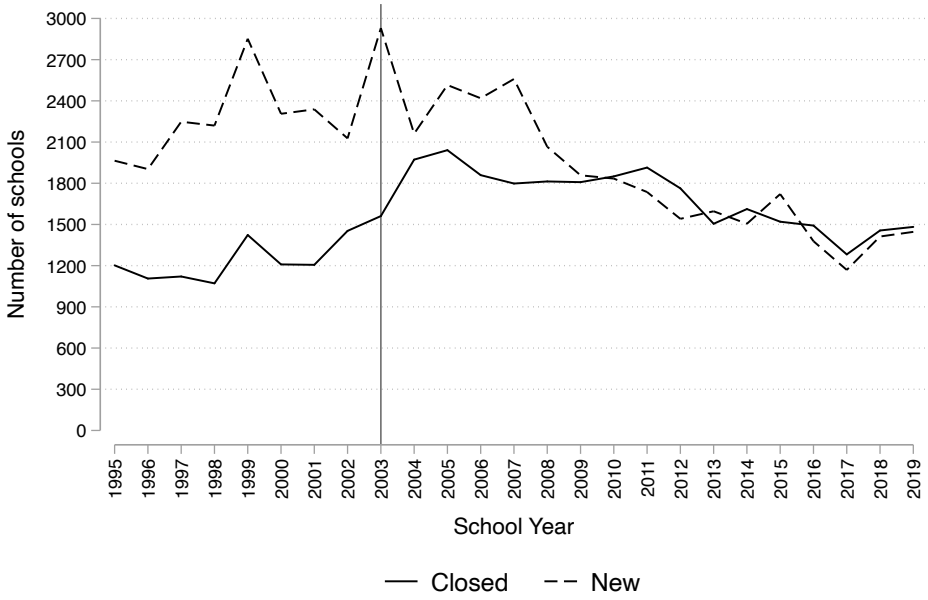
Notes: The unit of observation is the school year from 1998 to 2010. The sample only includes Title I schools. Treated states are those where NCLB introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the control states implemented test-based accountability before NCLB. In the "Carnoy and Loeb" classification, the control states implemented test-based accountability with strong repercussion to low performance before NCLB. Standard errors in parenthesis are clustered by state. * Statistically significant at the 0.10 level, ** at the 0.05 level, *** at the 0.01 level. Source: Author's calculations using the NLSD.

Figure A.1: Implementation of Accountability Policies



Notes: Marked in colors the 30 states that implemented state accountability policies before NCLB. Marked in white the 21 states where NCLB introduced test-based accountability for the first time. Source: Author's based on Carnoy and Loeb (2002) and Dee and Jacob (2011).

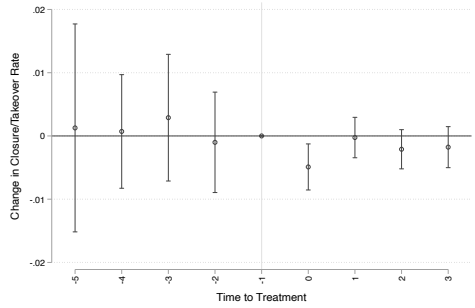
Figure A.2: Closed and New Schools



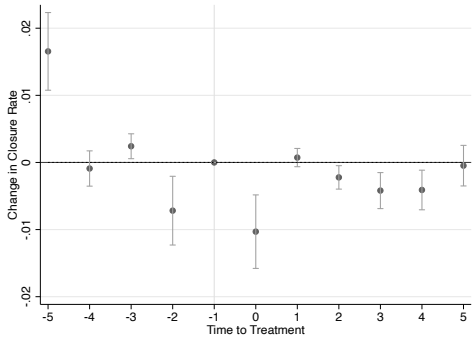
Notes: The vertical line identifies the year in which NCLB was implemented. Source: Author's calculations using the NLSD.

Figure A.3: Robustness Check: Event Study Estimates of the Effect of State Accountability Policies on School Closure/Takeover Rate

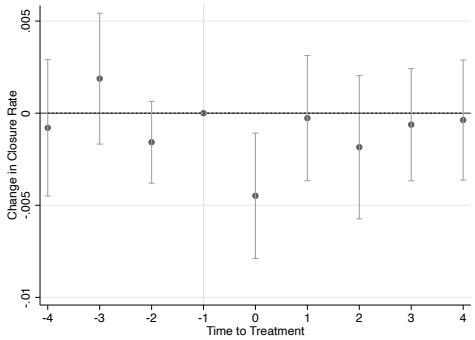
"Dee and Jacob" Classification



(a) TWFE

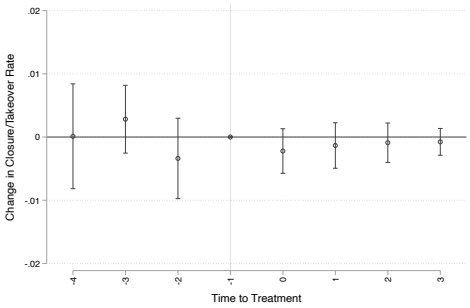


(b) Callaway and Sant'Anna

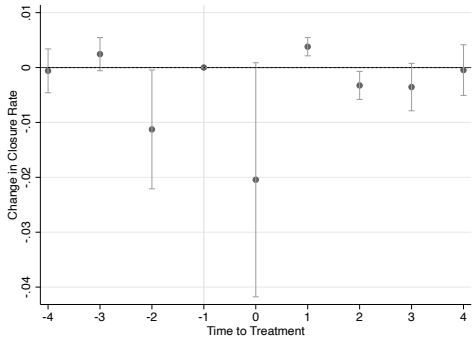


(c) Sun and Abraham

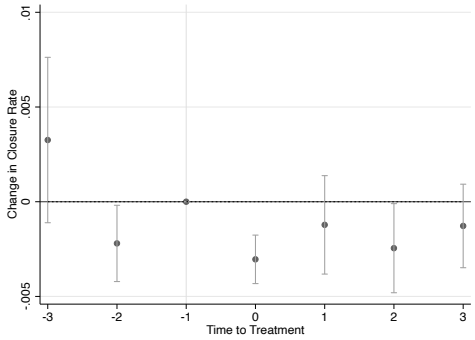
"Carnoy and Loeb" Classification



(d) TWFE



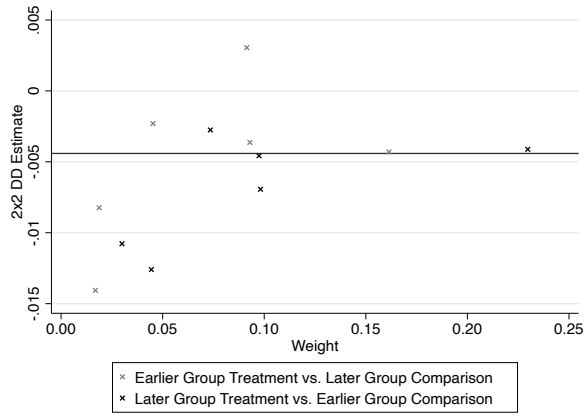
(e) Callaway and Sant'Anna



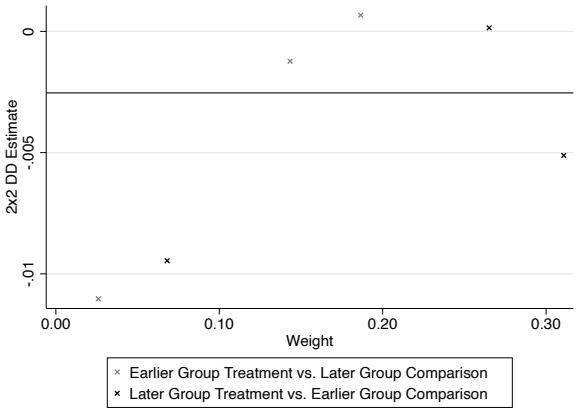
(f) Sun and Abraham

Notes: The unit of observation is the school year from 1995 to 2000. Treated states are those where state accountability introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the treated states implemented test-based accountability. In the "Carnoy and Loeb" classification, the treated states implemented test-based accountability with strong repercussion to low performance. Estimates expressed in relative time. 95% confidence intervals are reported based on standard errors clustered at the state level. Source: Author's calculations using the NLSD.

Figure A.4: DiD Decomposition of the Effect of State Accountability Policies on School Closure Rate



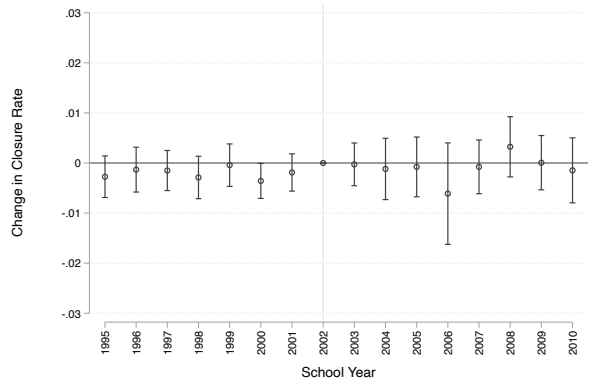
(a) "Dee and Jacob" Classification



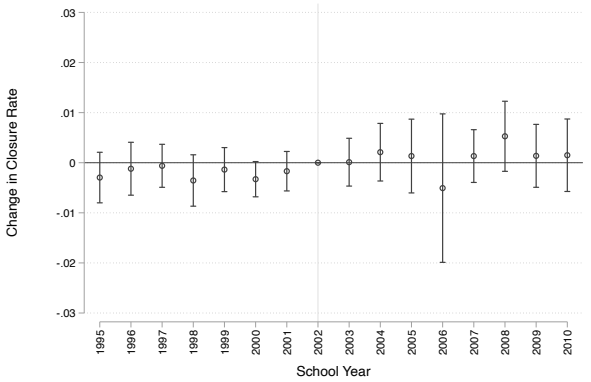
(b) "Carnoy and Loeb" Classification

Notes: The unit of observation is the school year from 1995 to 2000. Treated states are those where state accountability introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the treated states implemented test-based accountability. In the "Carnoy and Loeb" classification, the treated states implemented test-based accountability with strong repercussion to low performance. The figures present the average DiD estimate and total weight for all possible two-group/two-period DiD estimates. The DiD estimate based on the TWFE estimator is the weighted average of all those 2x2 DiD estimates. Source: Author's calculations using the NLSD.

Figure A.5: Robustness Check: Event Study Estimates of the Effect of NCLB on Schools Closure Rate, Longer Pre-Treatment Period



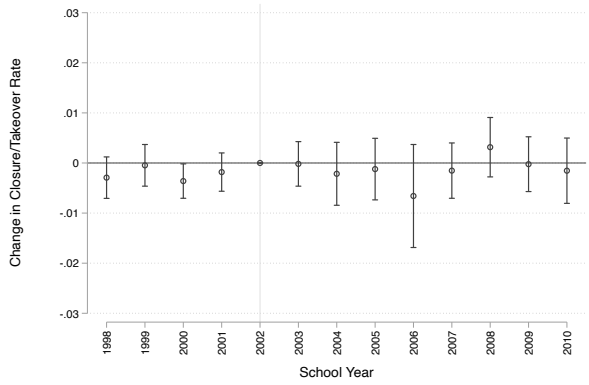
(a) "Dee and Jacob" Classification



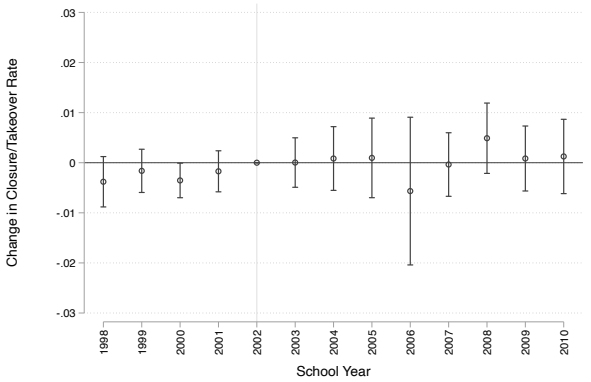
(b) "Carnoy and Loeb" Classification

Notes: The unit of observation is the school year from 1995 to 2010. The sample only includes Title I schools. Treated states are those where NCLB introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the control states implemented test-based accountability before NCLB. In the "Carnoy and Loeb" classification, the control states implemented test-based accountability with strong repercussion to low performance before NCLB. Estimates are relative to 2002, one year before the implementation of NCLB. 95% confidence intervals are reported based on standard errors clustered at the state level. Source: Author's calculations using the NLSD.

Figure A.6: Robustness Check: Event Study Estimates of the Effect of NCLB on Schools Closure/Takeover Rate



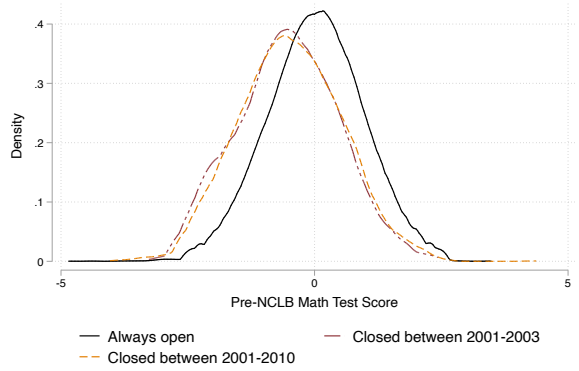
(a) "Dee and Jacob" Classification



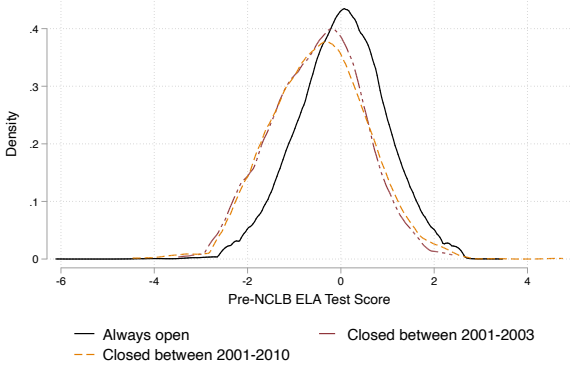
(b) "Carnoy and Loeb" Classification

Notes:The unit of observation is the school year from 1998 to 2010. The sample only includes Title I schools. Treated states are those where NCLB introduced test-based accountability for the first time. In the "Dee and Jacob" classification, the control states implemented test-based accountability before NCLB. In the "Carnoy and Loeb" classification, the control states implemented test-based accountability with strong repercussion to low performance before NCLB. Estimates are relative to 2002, one year before the implementation of NCLB. 95% confidence intervals are reported based on standard errors clustered at the state level. Source: Author's calculations using the NLSD.

Figure A.7: Distribution of the Pre-NCLB School Performance Measures Circa 2000, by Closure Status



(a) Pre-NCLB Math Test Score



(b) Pre-NCLB ELA Test Score

Notes: The unit of observation are schools that were ever open between 2001 and 2010 in 32 states. Z-scores. The sample only includes Title I schools. Source: Author's calculations using the CSPR, the NAYPI, the NLSD, and the SLAD.