The Achievement Consequences of the No Child Left Behind Act

Thomas S. Dee
Swarthmore College and NBER
dee@swarthmore.edu

Brian A. Jacob
University of Michigan and NBER
bajacob@umich.edu

Working Paper

**Please do not quote or cite without permission*

Paper presented at the *NCLB: Emerging Findings Research Conference* at the Urban Institute, Washington, D.C. on August 12, 2009. The conference was supported, in part, by the National Center for the Analysis of Longitudinal Data in Education Research (CALDER), funded by Grant R305A060018 to the Urban Institute from the Institute of Education Sciences, U.S. Department of Education and the National Center for Performance Incentives (NCPI) at Vanderbilt University. The authors would like to thank Rob Garlick and Elias Walsh for their research assistance. The views expressed in the paper are solely those of the authors and may not reflect those of the funders or supporting organizations. Any errors are attributable to the authors. An earlier version of this work was presented by Jacob at the David N. Kershaw Lecture, November 2008, at the Annual Fall meeting of the *Association of Public Policy and Management.*

Abstract

The No Child Left Behind (NCLB) Act has compelled states to design school accountability systems based on annual student assessments. The effect of this Federal legislation on the distribution of student achievement is a highly controversial but centrally important question. This study presents evidence on whether NCLB has influenced student achievement based on an analysis of state-level panel data on student test scores from the National Assessment of Educational Progress (NAEP). This study identifies the impact of NCLB by relying on comparisons of the test-score changes across states that already had school accountability policies in place prior to NCLB and those that did not. Results indicate that NCLB generated statistically significant increases in the average math performance of 4th graders (effect size = 0.22 by 2007) as well as improvements at the lower and top percentiles. However, the authors do not find consistent evidence that NCLB generated similarly broad improvements in reading achievement or achievement among 8th graders.

1. Introduction

The No Child Left Behind (NCLB) Act is arguably the most far-reaching education-policy initiative of the last four decades. The hallmark features of this Federal legislation, which was signed by President Bush in January of 2002, compelled states to conduct annual student assessments linked to state standards, to identify schools that are failing to make "adequate yearly progress" (AYP) towards achievement-based proficiency goals and to institute sanctions for chronically under-performing schools. A fundamental motivation for this reform is the notion that publicizing detailed information on school-specific performance and linking that "high-stakes" test performance to the possibility of meaningful sanctions (e.g., public school choice, staff replacement, and school restructuring) can improve the focus and productivity of public schools. However, several critics have charged that test-based school accountability has several unintended, negative consequences for the broad cognitive development of children (e.g., Nichols and Berliner 2007). Critics have also pointed to evidence that achievement trends and white-minority achievement gaps have not changed recently as evidence that "the law's sanctions don't work" (Ravitch 2009).

This study presents new evidence on whether NCLB influenced student achievement using state-level panel data on student test scores from the National Assessment of Educational Progress (NAEP). This study identifies the impact of NCLB by relying on comparisons of the test-score changes across states that already had school-accountability policies in place prior to NCLB and those that did not. Our results indicate that NCLB generated statistically significant increases in the math achievement of 4^{th} graders (effect size = 0.22 by 2007) and that these gains were concentrated among white and Hispanic students and among students at all levels of

performance. However, our evidence suggests that NCLB had more narrow (or non-existent) effects on reading achievement and achievement among 8th graders.

Section 2 briefly reviews the literature on school accountability and NCLB and situates the contributions of this study within that literature. Sections 3 and 4 discuss the methods and data used in this study. Section 5 summarizes the key results and robustness checks. Section 6 concludes.

2. Prior Literature on the Effects of School Accountability and NCLB

NCLB mandated that states implement several forms of school-focused accountability. For example, NCLB requires annual testing of public-school students in reading and mathematics in grades 3 through 8 (and at least once in grades 10-12) and that states rate schools, both as a whole and for key subgroups, with regard to whether they are making "adequate yearly progress" (AYP) towards their state's proficiency goals. Schools that fail to make AYP for two consecutive years are identified as needing improvement and can be subjected to increasingly severe sanctions that can include allowing students to enroll elsewhere and the closure or reconstitution of the school.

Several states protested the introduction of NCLB, arguing that these federally mandated reforms were likely to be both costly to implement and educationally unproductive. Interestingly, several states also argued that NCLB "needlessly duplicates" their previously developed school accountability systems (Dobbs 2005). A number of research studies have evaluated the achievement consequences of accountability policies by exploiting this variation in state policies prior to the introduction of NCLB.[1] For example, Carnoy and Loeb (2002) found that the within-

[1] Several studies have also focused on district or state-specific evaluations. See Figlio and Ladd (2008) for a review of this literature.

state growth in student performance on the math NAEP between 1996 and 2000 was larger in states with higher values on an accountability index, particularly for Black and Hispanic students in 8th grade.[2]

Similarly, Hanushek and Raymond (2005) evaluated the impact of state school-accountability policies on state-level NAEP math and reading achievement measured by the difference between the performance of a state's 8th graders and that of 4th graders in the same state four years earlier. This gain-score approach applied to the NAEP data implied that there were two cohorts of state-level observations in both math (1992-1996 and 1996-2000) and reading (1994-1998 and 1998-2002).

Hanushek and Raymond (2005) classified state accountability policies as either "report-card accountability" or "consequential accountability." Report-card states provided a public report of school-level test performance. States with consequential accountability both reported school-level performance and could attach consequences to that performance. The types of potential consequences states could implement were diverse. However, virtually all of the accountability systems in consequential-accountability states included key elements of the school-accountability provisions in NCLB (e.g., replacing a principal, allowing students to enroll elsewhere, and the takeover, closure, or reconstitution of a school). Hanushek and Raymond (2005) note that "all states are now effectively consequential accountability states (at least as soon as they phase in NCLB)."

Hanushek and Raymond (2005) find that the within-state timing of the introduction of consequential accountability implied statistically significant increases in the gain-score measures. The achievement gains implied by consequential accountability were particularly large

[2] The accountability index constructed by Carnoy and Loeb (2002) ranged from 1 to 5 and combined information on whether a state required student testing and performance reporting to the state, whether the state imposed sanctions or rewards and whether the state required students to pass an exit exam to graduate from high school.

for Hispanic students and, to a lesser extent, White students. However, the estimated effects of consequential accountability for the gains scores of Black students were statistically insignificant as were the estimated effects of report-card accountability. Hanushek and Raymond (2005) argue that these achievement results provide support for the controversial school-accountability provisions in NCLB because those provisions are so similar to the consequential-accountability policies that had been adopted in some states.

The broad interest in understanding whether NCLB has influenced the distribution of student achievement, both overall and for key subgroups, has motivated careful scrutiny of the most recent trend data. For example, in a report commissioned by the U.S. Department of Education's Institute of Education Sciences (IES), Stullich, Eisner, McCrary and Roney (2006) note that achievement trends on both state assessments and the NAEP are "positive overall and for key subgroups" through 2005. Similarly, using more recent data, a report by the Center on Education Policy (2008) concludes reading and math achievement measures based on state assessments have increased in most states since 2002 and that there have been smaller but similar patterns in NAEP scores. Both reports were careful to stress that these national gains are not necessarily attributable to the effects of NCLB. However, a press release from the U.S. Department of Education (2006) pointed to the improved NAEP scores, particularly for the earlier grades where NCLB was targeted, as evidence that NCLB is "working."

Other studies have taken a less sanguine view of these achievement gains. For example, Fuller, Wright, Gesicki, and Kang (2007) are sharply critical of relying on trends in state assessments, arguing that they are subject to spurious variation as states adjust their assessment systems over time. They also document a growing disparity between student performance on state assessments and the NAEP since the introduction of NCLB and conclude that "it is

important to focus on the historical patterns informed by the NAEP." Using NAEP data on fourth graders, they conclude that the *growth* in student achievement has actually become flatter since the introduction of NCLB. Similarly, an analysis of NAEP trends by Lee (2006) concludes that reading achievement is flat over the NCLB period while the gains in math performance simply tracked the trends that existed prior to NCLB.

Several more recent studies have directly assessed the achievement consequences of NCLB through analyses of student-level data. Most of these studies have focused on the distributional consequences of NCLB within particular cities and states and using data that are exclusively from the post-NCLB period. For example, Neal and Schanzenbach (in press) present evidence that, following the introduction of NCLB in Illinois, the performance of Chicago school students near the proficiency threshold (i.e., those in the middle of the distribution) improved while the performance of those at the bottom of the distribution of was the same or lower. Similarly, using data from the state of Washington, Krieg (1998) finds that the performance of students in the tails of the distribution is lower when their school faces the possibility of NCLB sanctions. However, in a study based on data from seven states over four years, Ballou and Springer (2008) conclude that NCLB generally increased performance on a low-stakes test, particularly for lower-performing students. Their research design leveraged the fact that the phased implementation of NCLB meant that some grade-year combinations mattered for calculating AYP while others did not.

The results presented in this study contribute to the existing literature in at least three critical ways. First, by using state-year NAEP data, this study relies on consistent measures of student achievement that are more nationally representative and that span the periods both before and after the implementation of NCLB. Second, by relying on the "low-stakes" NAEP data

rather than the "high-stakes" data from state assessments, this study's results are comparatively immune to concerns about whether policy-driven changes in achievement merely reflect "teaching to the test" rather than broader gains in cognitive performance. Third, this study adopts an alternative identification strategy based on comparing the achievement changes in states where NCLB catalyzed a new state-level school-accountability system relative to the corresponding changes in states consequential school-accountability policies had already existed.

It should be noted that this approach is broadly similar to one used in an earlier study by Lee (2006), which used hierarchical linear models (HLM) to compare the post-NCLB achievement trends across states with and without prior (i.e., "first-generation") accountability policies. Lee (2006) concluded that NCLB did not have any achievement effects. However, the study by Lee (2006) might be underpowered both because it could only use the NAEP data since 2005 and because HLM models may fail to exploit the precision gains associated with conditioning on state fixed effects.[3] Conditioning on state fixed effects may also be important because of changes over time in the composition of states participating in NAEP testing.

3. Methods

Many observers have pointed to national time trends in student achievement to gauge the impact of NCLB. Figures 1-4 present national trends on the Main NAEP from 1990 to 2007 for 4th grade math, 4th grade reading, 8th grade math and 8th grade reading respectively. The dashed horizontal line in 2002 visually identifies the point at which NCLB was implemented. These figures suggest that NCLB may have had some positive effects on 4th grade math achievement

[3] In fact, like our study, Lee (2006, Table C-7) finds evidence for a positive NCLB effect on math scores among 4th graders. Lee (2006, page 44) dismisses these results because they become statistically insignificant after conditioning on additional covariates. However, the estimated NCLB effect actually increases by roughly 20 percent after conditioning on these controls so the insignificance of this estimate reflects a substantial loss of precision in the saturated specification.

but, with a few exceptions, provide little evidence of impacts in the other three grade-subject combinations.[4] Figures 5-8 show similar trends for math and reading achievement on the Long-Term Trend NAEP for 9- and 13-year olds from the 1970s through 2004. These data tell a similar story.

Given the myriad of other social, economic and educational factors occurring over this time period, however, it is not clear that one should draw strong causal inferences from these data. For example, the nation was suffering from a recession around the time NCLB was implemented, which may have been expected to reduce student achievement in the absence of other forces. Conversely, there were a number of national education policies or programs that may have influenced student achievement at this time. For example, the National Council of Teachers of Mathematics (NCTM) adopted new standards in 2000, which likely shifted the content of math instruction in many elementary classrooms over this period (NCTM website). Similarly, the Reading Excellence Act of 1999 (the precursor to the Reading First program within NCLB) provided more than $750 million to states and LEAs to adopt scientifically-based instructional practices and professional development activities (Moss 2006).

3.1 Comparative Interrupted Time Series

To circumvent these concerns, we rely on a comparative interrupted time series (CITS) approach (also known as an interrupted time series with a non-equivalent comparison group). Specifically, we compare the deviation from prior achievement trends among a "treatment group" that was subject to NCLB with the analogous deviation for a "comparison group" that was arguably less affected by NCLB. The intuition is that the deviation from trend in the comparison group will reflect other hard-to-observe factors (e.g., the economy, other education

[4] One exception is a noticeable improvement in 8th grade math scores among African-Americans.

reforms) that may have influenced student achievement in the absence of NCLB. This strategy has a long tradition in education research (see, for example, the discussion in Bloom 1999 and Shadish et al. 2002), and has been used recently to evaluate reforms as diverse as Accelerated Schools (Bloom et al. 2001) and pre-NCLB accountability policies (Jacob 2005).

As discussed in more detail below, there are several important threats to causal inference in a CITS design. One such example involves the endogenous student mobility, as might occur if NCLB caused families to leave or return to the public schools. If this NCLB-induced mobility were random with respect to characteristics influencing achievement, it would not be a concern. On the other hand, if the most motivated parents pulled their children from public schools at the onset of NCLB, the resulting compositional change may have decreased student achievement in the absence of any changes to the schools themselves. A similar concern arises if NCLB induced states to selectively change the composition of students tested for the NAEP (e.g., increasing exclusion rates).

It is worth noting that all NCLB-induced changes do not necessarily invalidate our research design. For example, states may have responded to NCLB by increasing funding for schools, or instituting kindergarten testing for early identification of at-risk students. In this case, one could still interpret the estimates presented below as the causal "net" effect of NCLB, where funding and early identification are viewed as mechanisms through which the policy operated. Of course, if one wanted to ascertain the impact of specific components of NCLB (i.e., sanctioning schools, school choice provisions), one would need to adopt an alternative strategy.

The central challenge for any CITS design is to identify a plausible comparison group. In the case of NCLB, this is particularly difficult. As noted earlier, the policy was signed into law in January 2002 and implemented nationwide in the 2002-03 school year. It applied to all

schools receiving federal Title I fund, which in practice meant that all states and school districts were subject to the provisions of the law.

3.2 Catholic versus Public Schools

One potential comparison group is the set of Catholic schools in the U.S (Jacob 2008). While Catholic (and other private) schools do receive federal Title I funding and are thus entitled to participate in NCLB, a recent federal study indicates that few if any Catholic school students participate in the program (DOE 2007). One key reason is that very few students in these schools are eligible for free or reduced price lunch and are thus not affected by NCLB. Another reason is that Catholic schools have traditionally had little interaction with state DOEs or local LEAs, and thus were not well informed about the details of the legislation.

Figures 9 and 10 show achievement trends for public and Catholic school students from the national NAEP. In Figures 9 we see that students in Catholic schools outperformed their counterparts in public schools over the entire period 1990-2007. While both groups showed increasing achievement during the pre-NCLB period, public school students (particularly in 4^{th} grade) experienced a shift in achievement in 2003 and continued at roughly the same slope afterwards. Students in Catholic schools, by contrast, experienced no such shift and achievement trends appeared to flatten for this group after 2003. These comparisons reinforce the story told by the earlier figures – that is, a modest positive impact for 4^{th} grade math and a potential (smaller) effect for 8^{th} grade math. Figure 10 suggests a similar pattern for reading – potentially positive impacts in 4^{th} grade, but no evidence of impacts at 8^{th} grade.

As mentioned above NCLB-induced compositional changes in Catholic or public schools could comprise this design. To explore this, Figure 11a shows trends in public and Catholic

elementary school enrollment. To facilitate interpretation of the trends, the y-axis measures the natural logarithm of enrollment, demeaned by the initial year (1992) value so that both trends are zero in 1992 by construction. The trends thus reflect percent changes relative to 1992 in each sector. Catholic enrollment declined slightly prior to NCLB, but then dropped by nearly 10 percent between 2002 and 2004, and fell an additional 7 percent between 2004 and 2006. In contrast, public school enrollment increased steadily prior to NCLB, and leveled off following 2002. Figure 11b important differences across sectors in pupil-teacher ratios following 2002 (relative to prior trends). Pupil-teacher ratios in public schools appeared to increase modestly in absolute terms (relative to steady decline in prior years) while ratios in Catholic schools dropped relative to prior trends. Together, these figures are consistent with enrollment shifts from Catholic to public schools around the time of NCLB, possibly in response to the economic downturn.

While these figures raise important concerns, only non-random enrollment shifts related to student achievement (e.g., the most or least capable students switched from Catholic to public schools) will comprise the validity of the inferences above. Figures 12a-c show trends in the racial composition within public and Catholic schools over this time period. While there are no notable differences across sectors in terms of post-NCLB changes, it is still possible that the composition of each sector was changing in important ways that are not easily captured by race or other student demographics.

In summary, the comparison of Catholic versus public schools provides some suggestive evidence that NCLB increased math achievement in 4^{th} (and to a lesser extent in 8^{th}) grade, but the possibility of selective compositional changes limit the confidence one can place on the conclusions.

3.3 Early vs. Late Adopters of Accountability

A second approach is to compare trends in student achievement across states that had varying degrees of experience with consequential school accountability prior to NCLB. The intuition behind this approach is that NCLB represented less of a "treatment" in states that had adopted NCLB-like school accountability policies prior to 2002. To the extent that NCLB had positive or negative effects on measured student achievement, we would expect to observe those effects most distinctly in states that had not previously introduced similar policies.

Here we are assuming that the effect of pre-NCLB school accountability policies is comparable to the effect of NCLB – that is, the two types of accountability regimes are similar in the most relevant respects. To ensure that this is the case, we categorize states according to the features of the their own accountability policies that most closely resemble the key aspects of NCLB. For example, we do not consider states, which merely required districts to inform parents of school achievement through report cards to have adopted pre-NCLB accountability. Of course, it is possible that prior experience with school accountability may have prepared a state to respond even more effectively to NCLB. To the extent that this phenomenon dominates, our estimates will understate any positive effects of NCLB.[5]

Following the intuition of this comparative interrupted time series design, we estimate the following regression model:

$$
\begin{aligned}
Y_{st} = {} & \beta_0 + \beta_1 YEAR_t + \beta_2 NCLB_t + \beta_3 \left(YR\_SINCE\_NCLB_t\right) + \\
(1) \quad & \beta_4 \left(T_s \times YEAR_t\right) + \beta_5 \left(T_s \times NCLB_t\right) + \beta_6 \left(T_s \times YR\_SINCE\_NCLB_t\right) + \\
& \beta_7 X_{st} + \mu_s + \varepsilon_{st}
\end{aligned}
$$

where Y_{st} is NAEP-based measure of student achievement for state s in year t, $YEAR_t$ is a trend variable (defined as $YEAR_t$ – 1989 so that it starts with a value of 1 in 1990), and $NCLB_t$ is a

[5] More generally, this phenomenon would suggest the presence of heterogeneous treatment effects, which our model rules out.

dummy variable equal to one for observations from the NCLB era. For the the majority of our analysis, we assume the NCLB era begins in the academic year 2002-03, which was effectively the first year of full implementation since the legislation was signed into law in January 2002. In sensitivity analyses, we demonstrate that our results are robust to models that assume NCLB began in 2002 or even 2001.

$YR\_SINCE\_NCLB_t$ is defined as $YEAR_t - 2002$, so that this variable takes on a value of 1 for the 2002-03 year, which corresponds to the 2003 NAEP testing. T_s is a time-invariant variable that reflects the extent to which NCLB was a novel form of school accountability in state s and X_{st} represents covariates varying within states over time (e.g., per pupil expenditures, NAEP test exclusion rates, etc.). The variables, μ_s and ε_{st} represent state fixed effects and a mean-zero random error respectively.

The variable, T_s, can be thought of as simply identifying "treatment" states. For example, in our most basic application, T_s is a dummy variable that identifies whether a given state had *not* instituted consequential accountability prior to NCLB. This regression specification then allows for an NCLB effect that can be reflected in both a level shift in the outcome variable (i.e., β_5) as well as a shift in the achievement trend (i.e., β_6). Thus, the total estimated NCLB effect as of 2007 would be $\hat{\beta}_5 + 5 \times \hat{\beta}_6$.

This approach effectively compares the level and trend differences during the NCLB era across states that did and did not have a prior experience with school accountability. However, this simplistic definition of T_s could lead to somewhat attenuated estimates of NCLB's effects because it includes in the "control" group several states that had implemented school accountability only shortly before the onset of NCLB. That is, the "control" group includes some states for which the effects of prior state policies and NCLB are intertwined.

One approach to this concern is to simply omit states that adopted state accountability within several years of NCLB. However, this approach has two important disadvantages: (1) it reduces our statistical power and (2) it requires one to make largely arbitrary decisions about which states to omit from the analysis. As an alternative, we estimate a model in which we define T_s as a measure of NCLB's treatment intensity. To do so, we define T_s as the number of years during our panel period that a state did *not* have school accountability. As a practical matter, we show that all of these approaches generate quite similar results.

4. Data

4.1 The National Assessment of Educational Progress (NAEP)

This analysis uses data on math and reading achievement from the state-representative NAEP. Because our identification strategy depends on measuring achievement trends prior to NCLB, we limit our sample to states that administered the state NAEP at least two times prior to the implementation of NCLB. Because so few states administered the 8th grade math exam in 1990, when looking at math we focus on the pre-NCLB years of 1992, 1996 and 2000. For reading, we focus on 1994, 1998 and 2002. We chose to include 2002 as a pre-NCLB data point in our analysis because, given the timing of the passage and implementation of the law, it seems unlikely that Spring 2002 scores could have been substantially influenced by NCLB. All states administered NAEP in 2003, 2005 and 2007.

Our final sample includes 39 states (227 state x years) for 4th grade math, 38 states (220 state x years) for 8th grade math, 37 states (249 state x years) for 4th grade reading and 34 states (170 state x years) for 8th grade reading. A complete list of states in our sample can be found in

Appendix Table 1.[6] Since our estimates will rely on achievement changes across these states over time, it is worth exploring how representative these states are with respect to the nation. Table 1 presents some descriptive statistics that compare our analysis sample to the nation as a whole. With a few exceptions, our analysis sample closely resembles the nation in terms of student demographics and NAEP achievement.

4.2 School Accountability Policies Before NCLB

The research design used in this study relies on identifying states that had already implemented school-accountability policies similar to NCLB as well as the timing of those policies. To determine the pre-NCLB accountability policies of each state, we relied on a number of different sources including three recent studies of state accountability policies (Carnoy and Loeb, 2002; Lee and Wong, 2004; Hanushek and Raymond, 2005). The taxonomy developed by the more recent Hanushek and Raymond (2005) study is particularly salient in this context because it most closely tracked the key school-accountability features of NCLB. More specifically, Hanushek and Raymond (2005, Table 1) identified 25 states, which implemented "consequential accountability" prior to NCLB by coupling the public reporting of data on school performance to the possibility of meaningful sanctions based on that performance. We reviewed their coding with information from a variety of sources including the Quality Counts series put out by Education Week (1999), the state-specific "Accountability and Assessment Profiles" assembled by the Consortium for Policy Research in Education (Goertz and Duffy 2001), annual surveys on state assessment programs fielded by the Council of Chief State School Officers

[6] In order to ensure that we are accurately capturing the pre-NCLB trends, in addition to requiring that a state have at least two NAEP scores prior to 2003, we also require that states in our math sample participated in the 2000 NAEP and states in our reading sample participated in both the 1998 and 2002 NAEP. However, as shown in Table 4, our results are not particularly sensitive to this sample restriction.

(CCSSO), information from state Department of Education web sites, Lexis-Nexis searches of state and local newspapers, and conversations with academics and state officials in several states.

Our review generally confirmed their coding for the existence and timing of these state accountability policies. Furthermore, our review indicated that these pre-NCLB school-accountability systems closely resembled the state policies shaped by NCLB in both rating school performance and in attaching the possibility of invasive sanctions to those ratings (e.g., takeover, closure, reconstitution, replacing the principal and/or allowing student mobility). However, there are also a few notable distinctions between our classification of consequential-accountability states (Table 2) and the coding reported by Hanushek and Raymond (2005).

First, we reviewed a small number of states that were not included in the study by Hanushek and Raymond (2005) and identified two (i.e., Illinois and Alaska) that implemented consequential accountability in advance of NCLB (i.e., in 1992 and 2001, respectively). Second, our review suggested that the timing of consequential-accountability policies differed from that reported by Hanushek and Raymond (2005) in four states: Connecticut, New Mexico, North Carolina and Tennessee. We identified Connecticut as implementing consequential accountability in 1999 (i.e., with the adoption of Public Act 99-288) rather than in the early 1990s. While Connecticut reported on school performance in the early 1990s, it only rated schools that were receiving Title I schools and schools for which a district made a request during this period. We also identified New Mexico as implementing school accountability (i.e., rating school performance and providing financial rewards as well as the threat of possible sanctions) with the 1998 implementation of the Incentives for School Improvement Act rather than in 2003. We identified North Carolina as implementing school accountability in 1996 under the "ABCs of Public Education" rather than in 1993. We identified Tennessee as implementing consequential

school accountability in the fall of 2000 rather than in 1996. While Tennessee did begin reporting school performance in 1996, it did not rate schools, identify low performers or attach other school-level consequences until the State Board of Education approved a new accountability system in 2000.

Third, there are four additional states (Indiana, Kansas, Wisconsin and Virginia), which are identified as having consequential accountability in our baseline coding but could be viewed as marginal cases. Hanushek and Raymond (2005) identified both Wisconsin and Virginia as having consequential accountability prior to NCLB. However, in both Wisconsin and Virginia, the available state sanctions appear to have been clearly limited to school ratings. For example, Education Week (1999) notes that "Wisconsin law strictly limits the state's authority to intervene in or penalize failing schools." Similarly, Virginia began identifying low-performance schools through an accreditation system that became effective during the 1998-99 school year. However, because of limited state authority, the loss of accreditation was not clearly tied to the possibility of other explicit school sanctions (e.g., school closure).

Hanushek and Raymond (2005) also identify Indiana and Kansas as introducing report-card, rather than consequential, accountability prior to NCLB (i.e. in 1995). However, in addition to school-level performance reporting, Kansas had an accreditation process that rated schools and could culminate in several possible sanctions for low-performing schools (e.g., closure). Furthermore, Education Week (1999) indicated that, in addition to rating schools, Indiana rewarded high performing schools and state officials viewed vague state statutes as suggesting they could also close low-performing schools. In our baseline coding, we identify all four of these states as having consequential accountability prior to NCLB. However, we also report the results of a robustness check in which these designations are switched.

5. Results

5.1 Achievement Trends by Pre-NCLB Accountability Status

Before presenting formal estimates from equation (1), we show the trends in NAEP scores by pre-NCLB accountability (Figures 13-16). In each case, we present trends for three groups: states that adopted school accountability between 1994 and 1998; states that adopted school accountability between 1999 and 2001; and states that did not adopt school accountability prior to NCLB. The dots reflect the simple mean for each group x year, and the connecting lines show the predicted trends from the model described above.

Consider first Figure 13a, which shows trends in 4th grade math achievement. We see that in 1992, states that never adopted accountability scored roughly 5 scale points (.18 standard deviations) higher on average than other states. While all states made modest gains between 1992 and 2000, the states that adopted accountability policies prior to 2001 experienced more rapid improvement during this period. Indeed, this is the type of evidence underlying the conclusions in Carnoy and Loeb (2002) and Hanushek and Raymond (2005). Mean achievement in all three groups jumped noticeably in 2003, although relative to prior trends, this shift was largest among the "no prior accountability" group, which had the most modest prior trend. Interestingly, there was less noticeable change in the growth rates across period. In particular, for the two groups that had adopted prior accountability, the slope from 2003 through 2007 appears roughly identical to the slope from 1992 to 2000. The trends for percent of students meeting the basic standard, shown in Figure 13b, are similar. These figures suggest that NCLB had a positive impact on 4th grade math achievement.

The trends for 8th grade math (Figure 14) are similar to those for 4th grade math, but somewhat less clear in showing a positive achievement effect. In particular, the late adopters

(1999-2001 group) look quite similar to the never adopters in terms of prior trends and post-NCLB deviations.

The pattern for 4th grade reading in Figure 15 is much less clear. The pre-NCLB reading trends for all three groups are much noisier than the math trends, with all groups experiencing a decline in achievement in 1994, little change in 1998 (relative to 1992) and then very large gains in 2002. Both "early adopter" groups show little if any increase relative to trend. In contrast, the no accountability group saw a steeper growth rate post-NCLB, suggesting the possibility of a modest improvement relative to the other groups (and thus a modest positive impact of NCLB). It is worth noting, however, that if one focuses on the 8 years surrounding NCLB adoption (1998 to 2007), there is no evidence of any NCLB effect. The trends for 8th grade reading (Figure 16) show no evidence of any effects. (Note that the graph is scaled to accentuate what are really quite small absolute changes from year to year.)

5.2 Estimation Results

Table 3 shows our baseline estimates of equation (1). The outcome measure in all cases is the mean scale score. All models include linear and quadratic terms for the state-year exclusion rate as well as state fixed effects. Standard errors clustered at the state level are shown in parentheses. In Panel A, we define our treatment group to include only states that did not adopt school accountability prior to NCLB. Consistent with the earlier figures, we find that NCLB increased 4th grade math achievement by roughly 4.7 points by 2007 in states with no prior accountability relative to other states. Given a standard deviation of 31, this reflects an effect size of .15. We find no effect for 8th grade math or reading and a small but significant effect for 4th grade reading.

As discussed earlier, the inclusion of late adopting states may understate any positive effects of NCLB. Hence, in panel B we estimate the same models but exclude states that adopted school accountability policies between 1999-2001. This nearly doubles the size of the 4th grade math effect, and leads to a 5.2 (.14 standard deviation) effect for 8th grade math. Unfortunately, this approach reduces the precision of our estimates and relies on a somewhat arbitrary decision of which states to exclude.

For this reason, our preferred specification, shown in panel C, relies on a continuous treatment measure. Here we define the treatment as the years *without* prior school accountability, starting in 1990-91. Hence, states with no prior accountability have a value of 11. Illinois, which adopted its policy in the 1992-03 school year, would have a value of 2. Texas would have a value of 4 since its policy started in 1994-95, and Vermont would have a value of 9 since its program started in 1999-2000. The total effect we report is the impact of NCLB in 2007 for states with no prior accountability relative to states that adopted school accountability in 1997 (the mean adoption year among states that adopt prior to NCLB). The results suggest moderate positive effects for 4th grade math and smaller effects for 8th grade math that are not statistically different than zero at conventional levels (p-value = .12). The 4th grade reading results are marginally significant and quite small (2.2 scale point, or .06 standard deviations). There is no effect for 8th grade reading.

Table 4 presents a series of sensitivity analyses using panel C from Table 3 as the baseline. Specifically, Table 4 reports the estimated NCLB effect for each grade-subject grouping across specifications that utilize weighted least squares (WLS) based on public-school enrollments and alternative coding for consequential accountability. Table 4 also reports the estimated NCLB effect for specifications that differ in terms of controls for state and year fixed,

state-specific trends and state-year covariates and the composition of the analytical sample with respect to the number of pre-NCLB observations. The key results from Table 3 are quite similar in these alternative specifications.

Table 5 re-estimates our baseline specification using alternative outcomes. Columns 1 and 2 show the "effect" of NCLB on per-pupil expenditures and pupil-teacher ratios, two potential mediating variables. For the math samples, we find that NCLB increases spending by roughly 7 percent by 2007, relative to states that adopted school accountability in 1997. Column 3 provides suggestive evidence that the introduction of NCLB may have increased test exclusion on NAEP. None of the estimates are significantly different than zero, but it is worth noting that the point estimates themselves are quite large given the baseline mean of 4-6 percent. Columns 5-7 examine whether NCLB was associated with student racial composition. These estimate are meant to provide a test of one key identifying assumptions of the model –namely, that the treatment did not influence the type of students enrolled in public schools. Column 4 focuses on state poverty rates in an effort to ascertain whether there may be some unobserved factors associated with both our treatment and student outcomes.

Table 6 shows the effect of NCLB on various measures of student achievement. As many have noted, the design of NCLB necessarily focused the attention of schools on helping students attain proficiency. Hence, one would expect NCLB to disproportionately influence achievement in the left tail of the NAEP distribution. We find results roughly consistent with this, although NCLB did seem to increase achievement at higher points on the achievement distribution than one might have expected. For example, in 4^{th} grade math, the impacts at the 75^{th} percentile were only 2 scale points lower than at the 10^{th} percentile. In particular, for 4^{th}

grade reading, the average impact appears to have come from increases at the top of the ability distribution.

Tables 7-10 show results separately by race for the four grade-subject combinations. In each table, we present OLS estimates as well as estimates weighted by student enrollment in the state-year. Several interesting findings emerge. First, the 4th grade math effects are somewhat larger for Black and Hispanic students relative to white students. Interestingly, in the case of Black students, weighting by enrollment substantially increases the magnitude of the effects. This suggests that NCLB had more positive effects on Black students in states with larger Black populations. Second, the 8th grade math results are driven almost entirely by Hispanic students, though the point estimates for Black students are large as well (but imprecise). Third, the 4th grade reading effects are driven entirely by white students. Finally, NCLB appeared to have a statistically significant and substantively important *negative* effect on 8th grade reading achievement among Black students.

6. Conclusions

NCLB is an extraordinarily influential and controversial policy that, over the last seven years, has brought test-based school accountability to scale at public schools across the United States. The implications of this Federally mandated reform for the patterns of student achievement is a question of central importance. This study presented evidence on this broad question using state-year panel data on multiple student-outcome measures from the NAEP and a research design that effectively relied on the changes over time in states that had no prior school-accountability system like those required by NCLB and those that did. Our results suggest that the achievement consequences of NCLB are decidedly mixed. Specifically, our results indicate

that NCLB generated large and broad gains in the math achievement of 4th graders. However, we do not find consistent evidence for similarly large and broad gains in reading achievement and achievement among 8th graders.

These mixed results suggest that NCLB has fallen short of its ambitious requirement of all students reaching proficiency in reading and mathematics (at least as defined in NAEP) by the 2013-14 school year. However, the targeted successes of NCLB documented here also suggest that school accountability can be an effective lever for improving student outcomes. Interestingly, the heterogeneous treatment effects documented here are similar to those reported by Hanushek and Raymond (2005) who found that the first-generation of state school-accountability policies were relatively effective for Hispanic and white students but not black students. Understanding the sources of this treatment heterogeneity is likely to be a particularly useful policy datum as the future of status of NCLB is considered.

REFERENCES

Angrist, Joshua D. and Jörn-Steffen Pischke. Mostly Harmless Econometrics: An Empiricist's Companion. Princeton University Press, 2009.

Ballou, Dale and Matthew G. Springer. "Achievement Trade-offs and No Child Left Behind," working paper, October 2008.

Bertrand, M., Duflo, E., Mullainathan, S. How Much Should We Trust Differences-in-Differences Estimates? Quarterly Journal of Economics 2004; 119(1), 249-75.

Bloom, Howard S. (1999). "Estimating Program Impacts on Student Achievement Using Short Interrupted Time Series" Manpower Demonstration Research Corporation, Working Paper.

Goertz, M.E., and M.E. Duffy. Assessment and Accountability Systems in the 50 States: 1999-2000. CPRE Research Report RR-046. Consortium for Policy Research in Education, Philadelphia PA, 2001.

Howard S. Bloom, Sandra Ham, Laura Melton, Julieanne O'Brien (2001). "Evaluating the Accelerated Schools Approach A Look at Early Implementation and Impacts on Student Achievement in Eight Elementary Schools." Manpower Demonstration Research Corporation.

Carnoy, Martin and Susanna Loeb. "Does External Accountability Affect Student Outcomes? A Cross-State Analysis," Educational Evaluation and Policy Analysis 24(4), Winter 2002, pages 305-331.

Center on Education Policy. Has Student Achievement Increased Since 2002: State Test Score Trends Through 2006-07. June 2008.

Dobbs, Michael. "Conn. Stands in Defiance on Enforcing 'No Child'", Washington Post, Sunday, May 8, 2005

DOE (2007). Department of Education, "Private School Participants in Federal Program Under the No Child Left Behind Act and the Individuals with Disabilities Education Act" (Department of Education. Washington, DC, 2007)

Figlio, D. N., & Ladd, H. (2008). School accountability and student achievement. In H. Ladd & E. Fiske (Eds.), *Handbook of research in education finance and policy* (pp. 166-182). New York and London: Routledge.

Fuller, Bruce, Joseph Wright, Kathryn Gesicki, and Erin Kang. "Gauging Growth: How to Judge No Child Left Behind?" Educational Researcher 36(5), 2007, pages 268-278.

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

Jacob, Brian A. (2008). Lecture for the David N. Kershaw Award, Annual Fall Meeting of the Association of Public Policy Analysis, November 2008, Los Angeles, CA.

Jacob, B. (2005). "Accountability, Incentives and Behavior: Evidence from School Reform in Chicago." *Journal of Public Economics*. 89(5-6): 761-796.

Krieg, John M. "Are Students Left Behind? The Distributional Effects of the No Child Left Behind Act," Education Finance and Policy 3(2), Spring 2008, pages 250–281.

Lee, Jaekyung. "Tracking Achievement Gaps and Assessing the Impact of NCLB on the Gaps: An In-depth Look into National and State Reading and Math Outcome Trends," The Civil Rights Project, Harvard University, June 2006.

Neal, Derek and Diane Whitmore Schanzenbach "Left Behind by Design: Proficiency Counts and Test-Based Accountability" *Review of Economics and Statistics*, forthcoming.

Nichols, Sharon L. and David C. Berliner. *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Harvard Education Press, 2007.

Ravitch, Diane "Time to Kill 'No Child Left Behind'" Education Week, June 10, 2009.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). Experimental and Quasi-Experimental Designs for Generalized Causal Inference. Boston: Houghton-Mifflin.

Springer, Matthew G. "The Influence of an NCLB Accountability Plan on the Distribution of Student Test Score Gains," Economics of Education Review 27, 2008, pages 556-563.

Stullich, Stephanie, Elizabeth Eisner, Joseph McCrary, and Collette Roney. *National Assessment of Title I Interim Report to Congress: Volume I: Implementation of Title I*, Washington, DC: U.S. Department of Education, Institute of Education Sciences, 2006.

U.S. Department of Education. "No Child Left Behind is Working," December, 2006, http://www.ed.gov/nclb/overview/importance/nclbworking.html, Accessed July 29, 2009.

Table 1 - Descriptive Statistics, National Data and State-Based Analysis Samples (1992-2007)

Variable	Nation	State-based Analysis Samples			
		4th Grade Math	8th grade Math	4th grade Reading	8th grade Reading
Pre-NCLB NAEP Performance					
4th grade math - 2000 average	224	224			
4th grade math -Percent change, 1992 to 2000	2.3%	3.4%			
8th grade math - 2000 average	272		271		
8th grade math - Percent change, 1992 to 2000	1.87%		2.65%		
4th grade reading - 2002 average	217			216	
4th grade reading -Percent change, 1994 to 2002	2.36%			3.35%	
8th grade reading - 2002 average	263				260
8th grade reading -Percent change, 1998 to 2002	0.77%				0.39%
Observed traits in 2000					
NAEP Exclusion rate, 4th Grade	4%	4.47%			
NAEP Exclusion rate, 8th Grade	4%		4.40%		
Poverty rate	11.30%	11.96%	11.97%		
Pupil teacher ratio	16.4	16.43	16.42095		
Current per pupil expenditures	$7,394	$7,286	$7,345		
Percent free lunch	26.92%	31.88%	31.86%		
Percent of students white	62.10%	60.40%	59.98%		
Percent of students black	17.20%	17.39%	17.85%		
Percent of students Hispanic	15.60%	16.74%	16.74%		
Percent of students other race	5.20%	5.49%	5.43%		
Observed traits in 2002					
NAEP Exclusion rate, 4th Grade	6%			4.40%	
NAEP Exclusion rate, 8th Grade	5%				5.91%
Poverty rate	12.10%			11.97%	13.10%
Pupil teacher ratio	16.2			16.4	16.6
Current per pupil expenditures	$8,259			$7,345	$7,960
Percent free lunch	28.81%			31.86%	34.39%
Percent of students white	60.30%			59.98%	54.51%
Percent of students black	17.20%			17.85%	18.37%
Percent of students Hispanic	17.10%			16.74%	20.86%
Percent of students other race	5.60%			5.43%	6.26%
Number of states		39	38	37	34
Sample size		227	220	249	170

Notes: State data are weighted by state-year public-school enrollment.

Table 2 - States with Consequential Accountability prior to NCLB

State	Implementation Year	Hanushek and Raymond (2005) Accountability Type (Year)	Carnoy and Loeb (2002) School Repercussions (1999-2000)	Lee and Wong (2004) Accountability Type (1995-2000)
IL	1992	n/a	Moderate	Strong
WI	1993	Consequential (1993)	Weak to Moderate	Moderate
TX	1994	Consequential (1994)	Strong	Strong
IN	1995	Report Card (1993)	Moderate	Strong
KS	1995	Report Card (1993)	Weak	Moderate
KY	1995	Consequential (1995)	Strong	Strong
NC	1996	Consequential (1993)	Strong	Strong
NV	1996	Consequential (1996)	Weak	Moderate
OK	1996	Consequential (1996)	Weak	Moderate
AL	1997	Consequential (1997)	Strong	Strong
RI	1997	Consequential (1997)	Weak implementation	Moderate
WV	1997	Consequential (1997)	Strong	Moderate
DE	1998	Consequential (1998)	None	Weak
MA	1998	Consequential (1998)	Implicit only	Weak
MI	1998	Consequential (1998)	Weak	Moderate
NM	1998	Consequential (2003)	Moderate to strong	Strong
NY	1998	Consequential (1998)	Strong	Strong
VA	1998	Consequential (1998)	Weak to Moderate	Moderate
AR	1999	Consequential (1999)	None	Weak
CA	1999	Consequential (1999)	Strong	Moderate
CT	1999	Consequential (1993)	Weak	Moderate
FL	1999	Consequential (1999)	Strong	Strong
LA	1999	Consequential (1999)	Moderate	Strong
MD	1999	Consequential (1999)	Strong	Strong
SC	1999	Consequential (1999)	Moderate	Moderate
VT	1999	Consequential (1999)	Weak	Moderate
GA	2000	Consequential (2000)	None	Moderate
OR	2000	Consequential (2000)	Weak to Moderate	Moderate
TN	2000	Consequential (1996)	Weak	Moderate
AK	2001	n/a	None	Weak

Additional sources: CPRE Assessment and Accountability Profiles, Education Week (1999), CCSSO annual surveys, state Department of Education websites and Lexis-Nexis searches of state and local newspaper archives.

Table 3 - The Estimated Effects of NCLB on Mean NAEP Scores

Independent variables	Grade 4 Math	Grade 8 Math	Grade 4 Read	Grade 8 Read
Panel A: T_s = no prior accountability, no sample exclusions				
$NCLB_t \times T_s$	1.538	0.177	1.053	0.104
	(1.209)	(1.350)	(0.869)	(1.035)
$NCLB_t \times T_s \times (Years\ since\ NCLB)_t$	0.649**	0.100	0.354	-0.217
	(0.266)	(0.268)	(0.222)	(0.394)
Total effect by 2007	4.782**	0.677	2.824**	-0.982
	(1.952)	(2.304)	(1.242)	(1.931)
Number of states	39	38	37	34
Sample size	227	220	249	170
Panel B: T_s = no prior accountability, excludes 1999-2001 adopters				
$NCLB_t \times T_s$	4.438**	2.602*	1.851	-0.287
	(1.261)	(1.346)	(1.205)	(1.260)
$NCLB_t \times T_s \times (Years\ since\ NCLB)_t$	0.755*	0.530	-0.086	-0.386
	(0.405)	(0.359)	(0.330)	(0.487)
Total effect by 2007	8.212**	5.253**	1.420	-2.219
	(2.318)	(2.457)	(1.531)	(2.404)
Number of states	24	23	21	19
Sample size	139	132	140	95
Panel C: T_s = Years without prior school accountability, no sample exclusions				
$NCLB_t \times T_s$	0.647**	0.273	0.307**	-0.074
	(0.212)	(0.194)	(0.148)	(0.215)
$NCLB_t \times T_s \times (Years\ since\ NCLB)_t$	0.112*	0.069	0.015	-0.055
	(0.058)	(0.060)	(0.046)	(0.074)
Total effect by 2007 relative to state with school accountability starting in 1997	6.684**	3.359	2.221*	-1.825
	(2.007)	(2.198)	(1.264)	(1.776)
Number of states	39	38	37	34
Sample size	227	220	249	170
Mean of Y before NCLB in states without prior accountability	224	272	216	261
Student-level standard deviation prior to NCLB	31	38	36	34

Notes: Each column within a panel is a separate regression. All specifications include state fixed effects and linear and quadratic exclusion rates. Standard errors are clustered at the state level. ***p<0.01, ** p<0.05, * p<0.1.

Table 4 -The Estimated Effects of NCLB on Mean NAEP Scores, Sensitivity Analyses

Grade-Subject Sample	Baseline	WLS	Alternative coding for VA, WI, IN, KS	Full set of year fixed effects	State-specific trends	State-year covariates	No state fixed effects	States with 1+ Pre-NCLB Test Score	States with 3+ Pre-NCLB Test Scores
4th Grade Math									
Total effect by 2007	6.684**	6.683**	4.945**	6.692**	6.696**	6.422**	6.953**	5.398**	7.454**
	(2.007)	(2.666)	(2.268)	(2.016)	(2.274)	(2.140)	(2.388)	(1.842)	(2.196)
8th Grade Math									
Total effect by 2007	3.359	1.753	2.143	3.443	2.986	3.474	4.730	3.407*	2.507
	(2.198)	(3.841)	(2.267)	(2.210)	(2.667)	(2.340)	(3.019)	(1.898)	(2.481)
4th Grade Reading									
Total effect by 2007	2.221*	1.578	1.715	2.258*	1.698	3.036*	3.349*	1.962**	2.051*
	(1.264)	(1.266)	(1.220)	(1.192)	(1.426)	(1.677)	(1.960)	(0.991)	(1.246)
8th Grade Reading									
Total effect by 2007	-1.825	-1.569	-1.767	-1.805	-1.602	-1.492	0.619	-0.977	n/a
	(1.776)	(1.658)	(1.892)	(1.777)	(2.253)	(1.856)	(2.868)	(1.570)	

Notes: Each column within a panel is a separate regression as in Panel C of Table 3. The total NCLB effect by 2007 is relative to a state with school accountability starting in 1997. Specifications include state fixed effects and a quadratic in the exclusion rate, except where indicated otherwise. Standard errors are clustered at the state level. ***p<0.01, ** p<0.05, * p<0.1.

Table 5 - The Estimated Effects of NCLB on Mediating and Observed Variables

Grade-Subject Sample	ln(Per-Pupil Expenditures)	Pupil-Teacher Ratio	Exclusion Rate	Poverty Rate	%Black	%Hisp	%White
4th Grade Math (39 states, n=227)							
Total effect by 2007	0.073*	-0.332	0.680	-0.002	0.007*	-0.014	0.009
	(0.041)	(0.571)	(1.024)	(0.014)	(0.004)	(0.010)	(0.014)
8th Grade Math (38 states, n=220)							
Total effect by 2007	0.075**	-0.065	0.975	0.001	0.009**	-0.006*	0.002
	(0.037)	(0.558)	(1.093)	(0.013)	(0.004)	(0.004)	(0.006)
4th Grade Reading (37 states, n=249)							
Total effect by 2007	0.046	0.020	1.604	-0.005	0.014	-0.009	0.010
	(0.030)	(0.392)	(1.466)	(0.011)	(0.010)	(0.007)	(0.009)
8th Grade Reading (34 states, n=170)							
Total effect by 2007	-0.009	0.873**	1.872	0.006	0.018	0.001	0.002
	(0.051)	(0.409)	(1.771)	(0.014)	(0.019)	(0.002)	(0.004)

Notes: Each column is a separate regression as in Panel C of Table 3. The total NCLB effect by 2007 is relative to a state with school accountability starting in 1997. All specifications include state fixed effects. Standard errors are robust to clustering at the state level. ***p<0.01, ** p<0.05, * p<0.1.

Table 6 - The Estimated Effects of NCLB on Achievement Distributions by Grade and Subjects

Grade-Subject Sample	Mean	Percent Basic	Percent Proficient	Percent Advanced	10th percentile	25th percentile	50th percentile	75th percentile	90th percentile
4th Grade Math (39 states, n=227)									
Total effect by 2007	6.684**	9.296**	5.090**	0.424	8.472**	7.796**	7.318**	6.014**	4.617**
	(2.007)	(2.841)	(1.692)	(0.416)	(3.447)	(2.468)	(2.056)	(1.684)	(1.663)
Mean of Y before NCLB in states without prior accountability	224	64	21	2	186	205	225	244	259
8th Grade Math (38 states, n=220)									
Total effect by 2007	3.359	5.273**	1.038	-0.416	5.371*	4.652*	3.453*	3.833**	2.146
	(2.198)	(2.422)	(1.798)	(0.816)	(2.934)	(2.498)	(1.966)	(1.954)	(2.145)
Mean of Y before NCLB in states without prior accountability	272.4	64.2	24.4	3.5	227.6	251.4	274.5	295.8	314.4
4th Grade Reading (37 states, n=249)									
Total effect by 2007	2.221*	2.337*	2.444**	1.021**	3.308	2.218	2.259*	2.221**	2.038**
	(1.264)	(1.392)	(0.901)	(0.340)	(2.440)	(1.728)	(1.194)	(0.813)	(0.688)
Mean of Y before NCLB in states without prior accountability	215.9	61.4	28.6	5.7	170.6	194.2	218.1	239.7	257.9
8th Grade Reading (34, states, n=170)	-1.825	-3.295	1.734	0.013	-4.514*	-2.925	-0.825	1.094	1.065
Total effect by 2007	(1.776)	(2.177)	(2.021)	(0.609)	(2.691)	(2.284)	(1.862)	(2.007)	(2.610)
Mean of Y before NCLB in states without prior accountability	261	73	28	2	219	241	263	282	299

Notes: Each column is a separate regression as in Panel C of Table 3. The total NCLB effect by 2007 is relative to a state with school accountability starting in 1997. All specifications include state fixed effects and a quadratic in the exclusion rate. Standard errors are clustered at the state level. ***p<0.01, ** p<0.05, * p<0.1.

Table 7 - The Estimated Effects of NCLB on 4th Grade NAEP Math Scores by Race-Ethnicity

	OLS				WLS			
	Mean	% Basic	10th percentile	90th percentile	Mean	% Basic	10th percentile	90th percentile
White (39 states, n=227)								
Total effect by 2007	5.356**	8.212**	6.419**	3.600**	4.046**	6.836**	5.309*	3.119**
	(1.472)	(2.463)	(2.874)	(1.360)	(1.706)	(2.899)	(3.141)	(1.474)
Mean of Y before NCLB in states without prior accountability	232	76	197	265	233	77	198	265
Black (30 states, n=176)								
Total effect by 2007	4.550	7.192	3.315	4.179	12.144**	18.517**	13.370**	9.692**
	(4.909)	(5.994)	(7.342)	(5.581)	(3.109)	(5.263)	(4.793)	(2.801)
Mean of Y before NCLB in states without prior accountability	203	35	168	238	202	33	169	235
Hispanic (19 states, n=108)								
Total effect by 2007	10.664**	10.070*	19.751**	3.606	8.161**	20.747**	4.695*	8.429**
	(3.732)	(5.977)	(7.871)	(3.294)	(1.176)	(3.028)	(2.736)	(2.001)
Mean of Y before NCLB in states without prior accountability	204	40	164	242	204	36	168	240

Notes: Each column is a separate regression as in Panel C of Table 3. The total NCLB effect by 2007 is relative to a state with school accountability starting in 1997. All specifications include state fixed effects and a quadratic in the exclusion rate. Standard errors are clustered at the state level. ***p<0.01, ** p<0.05, * p<0.1.

Table 8 - The Estimated Effects of NCLB on 8th Grade NAEP Math Scores by Race-Ethnicity

	OLS				WLS			
	Mean	% Basic	10th percentile	90th percentile	Mean	% Basic	10th percentile	90th percentile
White (37 states, n=214)								
Total effect by 2007	2.400	3.997*	3.489	1.416	1.523	3.544	3.216	-0.786
	(2.266)	(2.398)	(2.527)	(2.714)	(3.066)	(2.611)	(2.889)	(3.668)
Mean of Y before NCLB in states without prior accountability	281	74	240	320	282	76	242	321
Black (27 states, n=158)								
Total effect by 2007	8.604	9.224	9.879	5.178	7.355	8.337	10.345	6.089
	(6.057)	(7.160)	(6.647)	(6.439)	(7.499)	(9.962)	(7.440)	(7.358)
Mean of Y before NCLB in states without prior accountability	241	28	198	284	242	28	200	283
Hispanic (16 states, n=90)								
Total effect by 2007	18.021**	19.700**	16.442**	18.594	6.850**	15.576**	2.053	7.691**
	(5.149)	(4.072)	(7.655)	(5.382)	(3.446)	(3.888)	(4.997)	(2.696)
Mean of Y before NCLB in states without prior accountability	246	36	200	291	247	36	204	292

Notes: Each column is a separate regression as in Panel C of Table 3. The total NCLB effect by 2007 is relative to a state with school accountability starting in 1997. All specifications include state fixed effects and a quadratic in the exclusion rate. Standard errors are clustered at the state level. ***p<0.01, ** p<0.05, * p<0.1.

Table 9 - The Estimated Effects of NCLB on 4th Grade NAEP Reading Scores by Race-Ethnicity

	OLS				WLS			
	Mean	% Basic	10th percentile	90th percentile	Mean	% Basic	10th percentile	90th percentile
White (37 states, n=249)								
Total effect by 2007	4.629**	4.295**	5.970**	3.426**	4.468**	4.732**	6.876**	2.928**
	(1.051)	(1.201)	(1.662)	(1.004)	(1.001)	(1.161)	(1.874)	(0.740)
Mean of Y before NCLB in states without prior accountability	226	73	184	265	225	72	183	264
Black (32 states, n=214)								
Total effect by 2007	-1.453	-1.880	-1.182	-0.676	-0.726	-3.555*	0.662	0.722
	(3.246)	(3.279)	(6.015)	(2.176)	(2.141)	(2.070)	(3.540)	(1.299)
Mean of Y before NCLB in states without prior accountability	200	43	154	244	195	36	151	238
Hispanic (22 states, n=140)								
Total effect by 2007	4.580	3.945	4.374	4.082	-0.537	-0.795	3.687	0.765
	(4.077)	(4.011)	(6.088)	(3.321)	(3.887)	(4.072)	(3.956)	(3.169)
Mean of Y before NCLB in states without prior accountability	199	43	154	244	193	37	144	241

Notes: Each column is a separate regression as in Panel C of Table 3. The total NCLB effect by 2007 is relative to a state with school accountability starting in 1997. All specifications include state fixed effects and a quadratic in the exclusion rate. Standard errors are clustered at the state level. ***p<0.01, ** p<0.05, * p<0.1.

Table 10 - The Estimated Effects of NCLB on 8th Grade NAEP Reading Scores by Race-Ethnicity

	OLS				WLS			
	Mean	% Basic	10th percentile	90th percentile	Mean	% Basic	10th percentile	90th percentile
White (33 states, n=165)								
Total effect by 2007	0.718	-1.971	-1.690	1.397	1.771	-0.239	0.771	1.815
	(1.968)	(2.043)	(3.183)	(2.820)	(1.785)	(1.723)	(2.770)	(3.337)
Mean of Y before NCLB in states without prior accountability	269	82	230	305	269	82	231	306
Black (27 states, n=135)								
Total effect by 2007	-12.440**	-14.330**	-17.700**	-7.309	-9.384**	-13.335**	-12.663**	-4.009
	(3.513)	(4.487)	(4.071)	(5.421)	(2.262)	(2.911)	(4.518)	(4.391)
Mean of Y before NCLB in states without prior accountability	245	54	205	282	244	53	205	280
Hispanic (20 states, n=100)								
Total effect by 2007	6.033	7.688	4.295	13.556	-1.526	-2.380	-8.032	6.230**
	(5.643)	(5.654)	(8.582)	(10.090)	(1.885)	(2.705)	(5.205)	(3.160)
Mean of Y before NCLB in states without prior accountability	243	53	196	285	243	51	196	285

Notes: Each column is a separate regression as in Panel C of Table 3. The total NCLB effect by 2007 is relative to a state with school accountability starting in 1997. All specifications include state fixed effects and a quadratic in the exclusion rate. Standard errors are clustered at the state level. ***p<0.01, ** p<0.05, * p<0.1.

Table A1 - States included in NAEP analysis samples

State	Subject-Grade			
	Grade 4 Math	Grade 8 Math	Grade 4 Read	Grade 8 Read
Alabama	1	1	1	1
Alaska	0	0	0	0
Arizona	1	1	1	1
Arkansas	1	1	1	1
California	1	1	1	1
Colorado	0	0	0	0
Connecticut	1	1	1	1
Delaware	0	0	1	1
District of Columbia	1	1	1	1
Florida	0	0	1	1
Georgia	1	1	1	1
Hawaii	1	1	1	1
Idaho	1	1	0	0
Illinois	0	1	0	0
Indiana	1	1	0	0
Iowa	1	0	1	0
Kansas	0	0	1	1
Kentucky	1	1	1	1
Louisiana	1	1	1	1
Maine	1	1	1	1
Maryland	1	1	1	1
Massachusetts	1	1	1	1
Michigan	1	1	1	0
Minnesota	1	1	1	0
Mississippi	1	1	1	1
Missouri	1	1	1	1
Montana	1	1	1	1
Nebraska	1	1	0	0
Nevada	1	0	1	1
New Hampshire	0	0	0	0
New Jersey	0	0	0	0
New Mexico	1	1	1	1
New York	1	1	1	1
North Carolina	1	1	1	1
North Dakota	1	1	0	0
Ohio	1	1	0	0

Oklahoma	1	1	1	1
Oregon	1	1	1	1
Pennsylvania	0	0	0	0
Rhode Island	1	1	1	1
South Carolina	1	1	1	1
South Dakota	0	0	0	0
Tennessee	1	1	1	1
Texas	1	1	1	1
Utah	1	1	1	1
Vermont	1	1	0	0
Virginia	1	1	1	1
Washington	0	0	1	1
West Virginia	1	1	1	1
Wisconsin	0	0	0	0
Wyoming	1	1	1	1
Total	39	38	37	34

Notes: Our analysis samples consist of states that have 1996 and 2000 NAEP scores in mathematics and 1998 and 2002 scores in reading. NAEP achievement data are not available for racial-ethnic subgroups within all participating state-year observations.


CALDER